IMA WORKSHOP: Integration of Sensing and Processing

Minneapolis

December 8, 2005

Wavelets in Biomedical Data Analysis: FDA, Γ-Minimax, and Scaling in Applications

Brani Vidakovic

Collaborators: Felix Abramovich, Claudia Angelini, Anestis Antoniadis, Theofanis Sapatinas

P L A N

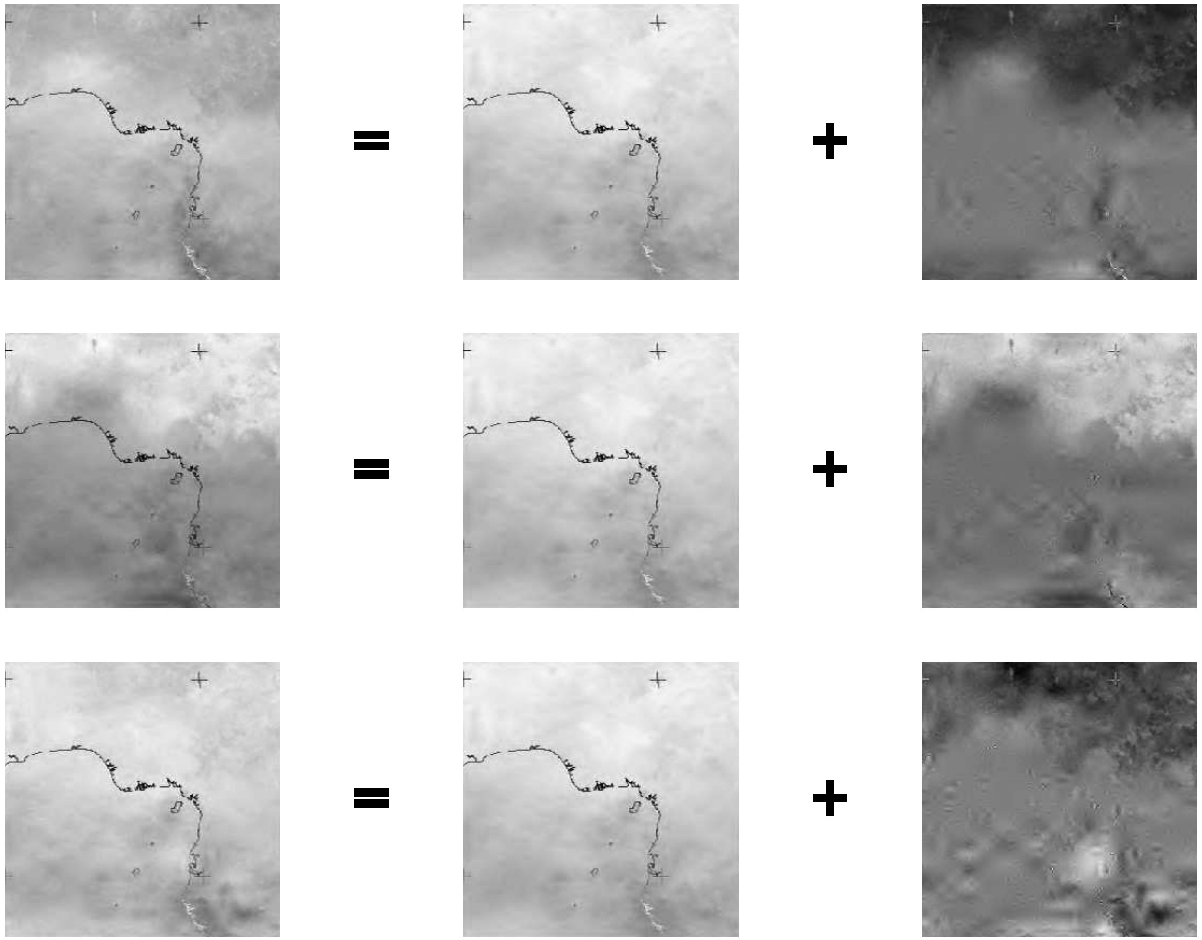
1. **Testing in Functional Analysis of Variance**

2. **Denoising: Gamma-Minimax Wavelet Shrinkage**

3. **Self Similarity: An Example**

Gulf of Guinea

Data set contains 36 IR satelite images of the Gulf of Guinea (West cost of Africa and South Atlantic Ocean) taken at 12 consecutive days (1/4/2001 - 1/15/2001). The images are divided into 3 groups according to time of their acquisition: (i) morning (6:00am) group, (ii) noon (12:00pm) group, and (iii) evening (18:00pm) group. These groups are treatment groups.



■ FDA, FANOVA, WANOVA,...

■ FDA: Predictors are curves, images, ..., s-dimensional objects – responses constants, vectors, curves, ..., s-dimensional objects.

■ *Analysis of Variance* (ANOVA) - one of the most widely used tools in applied statistics. Useful for handling low dimensional data, limitations in analyzing *functional* responses.

Functional Analysis of Variance (FANOVA) handles functional responses while still allowing a simple interpretation.

■ General: Ramsay and Coauthors, early to mid-90's, Ramsay & Silverman (1997, 2002). 2003 Gainesville Conference on FDA
`http://www.stat.ufl.edu/symposium/2003/fundat/`

■ Fitting and Estimation of GAM: Wahba *et al.*, 1995; Stone *et al.*, 1997; Huang, 1998; Lin, 2000; Gu, 2002.

• Huang, H.-C. and Cressie, N. 1997. • Raz and Turetsky, SPIE Proceedings 1999. • Rosner and Vidakovic 2000.

Wavelet-Based Functional Anova

$$y_{i\ell}(\mathbf{t}) = \mu(\mathbf{t}) + \alpha_i(\mathbf{t}) + \epsilon_{i\ell}(\mathbf{t}),\ \ i = 1, \ldots, r,$$

$$\mathbf{t} \in \mathcal{R}^d;\ \ \ell = 1, \ldots, n_i;\ \ \sum_{i=1}^{r} n_i = n, \quad \int \left| \sum_i n_i \alpha_i(\mathbf{t}) \right| d\mathbf{t} = 0.$$

Standard least square functional estimators for $\mu(\mathbf{t})$ and $\alpha_i(\mathbf{t})$:

$$\hat{\mu}(\mathbf{t}) = \bar{y}_{..}(\mathbf{t}) = \frac{1}{n} \sum_{i,\ell} y_{i\ell}(\mathbf{t}), \qquad \hat{\alpha}_i(\mathbf{t}) = \bar{y}_{i.}(\mathbf{t}) - \bar{y}_{..}(\mathbf{t}),$$

$\bar{y}_{i.}(\mathbf{t}) = \frac{1}{n_i} \sum_l y_{i\ell}(\mathbf{t})$. (Ramsay and Silverman, 1997, p. 141)

The fundamental ANOVA identity:

$$\boxed{\mathrm{SST}(\mathbf{t}) = \mathrm{SSTr}(\mathbf{t}) + \mathrm{SSE}(\mathbf{t}),}$$

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

For each fixed $\mathbf{t}$,

$$F(\mathbf{t}) = \frac{\mathrm{SSTr}(\mathbf{t})/(r-1)}{\mathrm{SSE}(\mathbf{t})/(n-r)} \sim \text{ noncentral } F_{r-1,n-r}\left(\frac{\sum_i n_i \alpha_i^2(\mathbf{t})}{\sigma^2}\right).$$

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

$$\begin{aligned} d_{i\ell}(j,\mathbf{k}) &= \theta_i(j,\mathbf{k}) + \epsilon'_{i\ell}(j,\mathbf{k}) \\ &= \theta(j,\mathbf{k}) + \tau_i(j,\mathbf{k}) + \epsilon'_{i\ell}(j,\mathbf{k}). \end{aligned}$$

- The discrete "times" $\mathbf{t}_m \longrightarrow$ double indexing $(j,\mathbf{k})$, j corresponds to a **scale level** and $\mathbf{k}$ corresponds to a **location position.**
- Identifiability condition: $\sum_{j,\mathbf{k}} |\sum_i n_i \tau_i(j,\mathbf{k})| = 0$.

Result: Let $\hat{\mu}$ and $\hat{\alpha}_i$ be the LS estimators of μ and α_i. If $\hat{\theta}$ and $\hat{\tau}_i$ are the LS estimators of θ and τ_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 $\hat{\alpha}_i \equiv \tilde{\alpha}_i$.

The *energy preservation* of orthogonal wavelet transformations also implies,

$$\begin{aligned}\sum_{\mathbf{t}\in\mathbf{T}} \mathrm{MSE}(\mathbf{t}) &= \sum_{j,\mathbf{k}\in\mathcal{I}} MSE(j,\mathbf{k})\\ \sum_{\mathbf{t}\in\mathbf{T}} \mathrm{MSTr}(\mathbf{t}) &= \sum_{j,\mathbf{k}\in\mathcal{I}} MSTr(j,\mathbf{k})\end{aligned}$$

where *MSE* and *MSTr* are the wavelet-domain counterparts of MSE and MSTr.

Benefits of carrying out an analysis in the wavelet domain, rather than in the time domain, come from

- **decorrelation**
- **regularization**
- **dimension reduction**

amenability of wavelet domains.

- How to regularize?
  - F-test shrinkage (with Claudia Angelini),

Bayesian version of FDR (Ongoing Project with Felix Abramovich, Tel Aviv).

Barbara, Lena & Cameraman in 2-D FANOVA

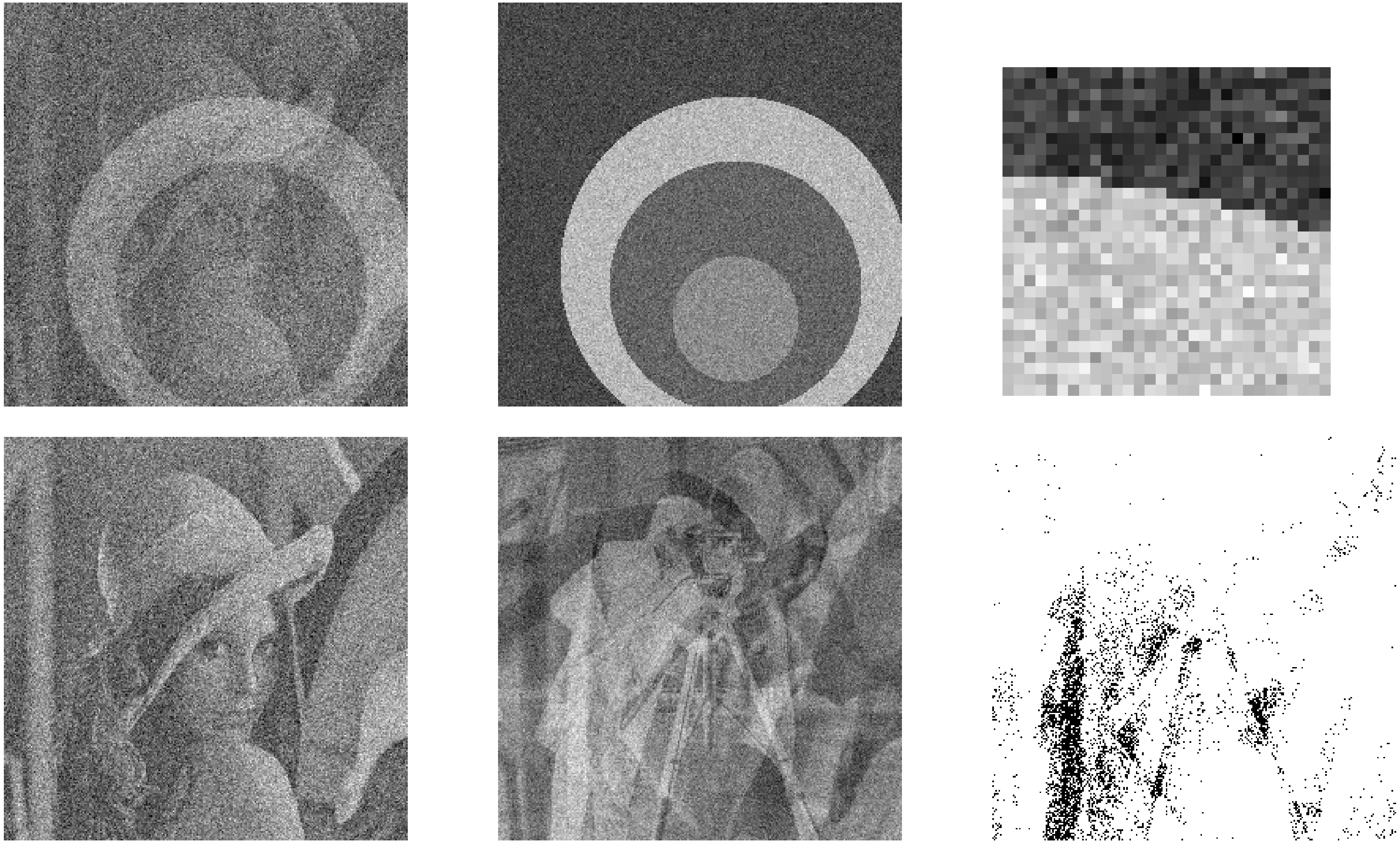
The FANOVA components:

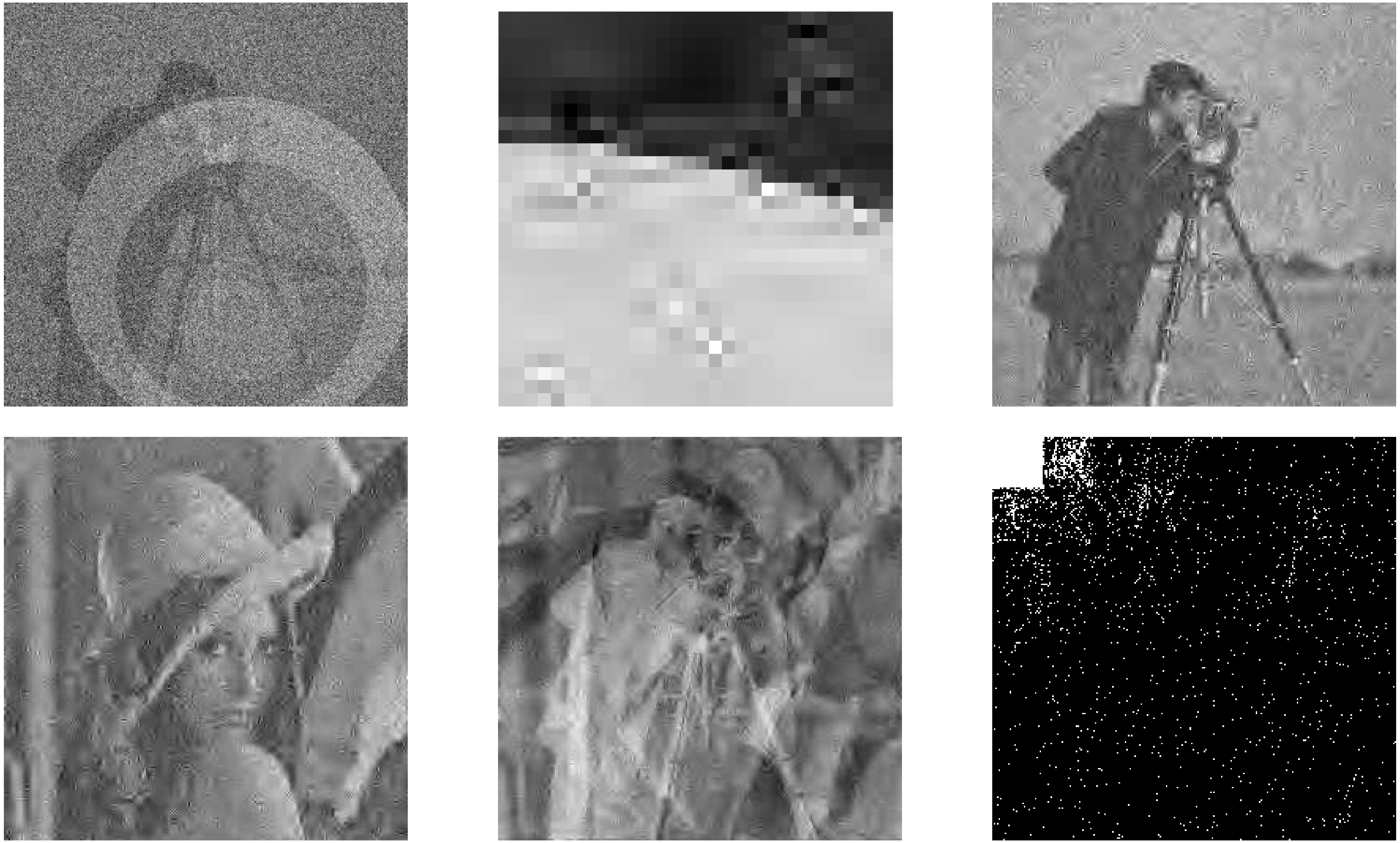
- Main effect - circle image $\mu(\mathbf{t})$.
- Treatment images $\alpha_1(\mathbf{t}), \alpha_2(\mathbf{t})$ and $\alpha_3(\mathbf{t})$ are **`lena`**, **`cameraman`**, and **`barbara`**, respectively, while the treatment (**`mish-mash`**) $\alpha_4(\mathbf{t}) = -\frac{1}{n_4}(n_1\alpha_1(\mathbf{t}) + n_2\alpha_2(\mathbf{t}) + n_3\alpha_3(\mathbf{t}))$ ensures identifiability.
- The number of observations: $n = 20$.
- Treatment sample sizes $n_1 = 5, n_2 = 5, n_3 = 5$, and $n_4 = 5$.

- Each observation ℓ from the treatment i is a 256×256 image, $\mu_i(\mathbf{t}) + \epsilon_{i\ell}(\mathbf{t}) = \mu(\mathbf{t}) + \alpha_i(\mathbf{t}) + \epsilon_{i\ell}(\mathbf{t})$ observed at equally spaced grid points $\mathbf{t}_m$

- Signal-to-noise ratio [SNR] of $\mu + \alpha_i$ (signal) to ϵ (noise) is 0.4 and the size of noise is 200 (i.e., the noise at each pixel is distributed as $N(0, 200^2)$).

- Wavelet: 6-tap Coiflet.



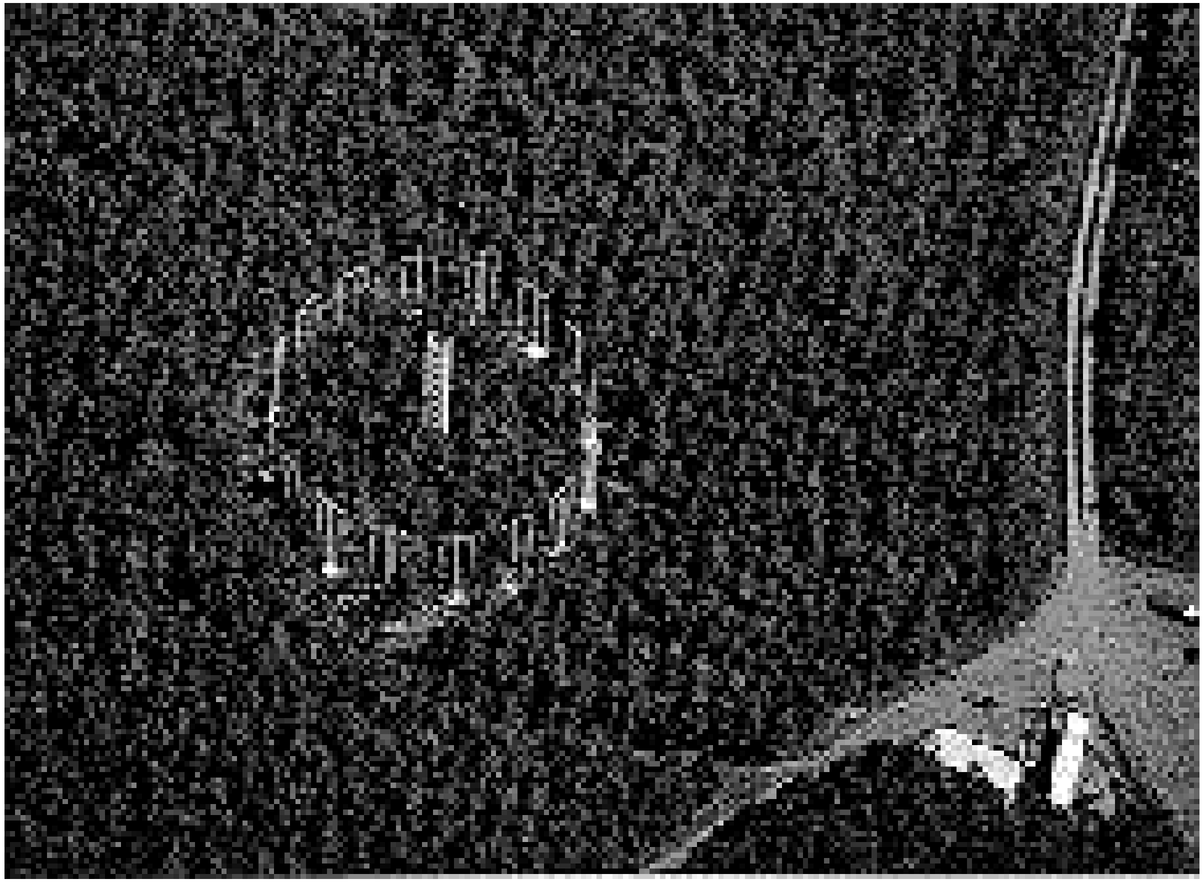


2-Way FANOVA

- Evaluation of functional treatment effects and interactions, on environmental measurements [H_2O fluxes] at Duke forest site [http://152.16.58.129/facts-1/facts.html]

$$y_{ijk}(t) = \mu(t) + \alpha_i(t) + \beta_j(t) + (\alpha\beta)_{ij}(t) + \epsilon_{ijk}(t),$$
$$1 \le i \le I,\ 1 \le j \le J,\ 1 \le k \le K_{ij}$$

- $y_{ijk}(t)$ – flux of water ground $\rightarrow$ tree
- $\alpha_i(t)$ – effect of CO_2 (2 levels)
- $\beta_j(t)$ – effect of fertilizer (2 levels)
- $(\alpha\beta)_{ij}(t)$ interaction term



Why it is difficult to test functional components?

▪ Leakage of power, due to large number of simultaneous tests!

- Example:

$$H_0 : \boldsymbol{\theta} = 0 \quad vs. \quad H_1 : \boldsymbol{\theta} \neq 0,$$

If $\mathbf{X} \sim \mathcal{MVN}_n(\boldsymbol{\theta}, \mathbf{I}_n)$ is observed, the ML ratio test based on sufficient statistics $||\mathbf{X}||^2$.

For the alternative $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta_1}$ α-level ML ratio test has appr. power

$$1 - \Phi\left(z_{1-\alpha} - ||\boldsymbol{\theta_1}||^2/\sqrt{2n}\right).$$

- If $||\boldsymbol{\theta_1}||^2 \to \infty$ as $o(\sqrt{n})$, the power tends to α !

Adaptive-Neyman truncation test

Fan (1996), Fan and Lin (1997)

• $\mathcal{I}_m$ – set of indices $(j, \mathbf{k})$ with m largest magnitudes of (treatment averaged) wavelet coefficients.

$$F^*_{\hat{m}} = \max_{1 \le m \le n} \frac{1}{\sqrt{2(r-1)m}} \left\{ \sum_{(j,\mathbf{k}) \in \mathcal{I}_m} \sum_{i=1}^{r} \left(\frac{\bar{d}_i(j,\mathbf{k}) - \bar{d}(j,\mathbf{k})}{\hat{\sigma}(j,\mathbf{k})/\sqrt{n_i}} \right)^2 - (r-1)m \right\}$$

$\bar{d}_i(j, \mathbf{k}) = 1/n_i \sum_{\ell=1}^{n_i} d_{i\ell}(j, \mathbf{k})$, $\quad \bar{d}(j, \mathbf{k}) = 1/n \sum_{i=1}^{r} n_i \bar{d}_i(j, \mathbf{k})$,

$\hat{\sigma}^2(j, \mathbf{k}) = WMSE(j, \mathbf{k})$.

The level α test rejects $H_0 : \alpha_1(t) = \alpha_2(t) = \ldots \alpha_r(t) = 0$ if

$$F^*_{\hat{m}} \ge \frac{1}{\sqrt{2(r-1)\hat{m}}} \left(\chi^2_{\hat{m}(r-1)}(1-\alpha) - (r-1)\hat{m} \right).$$

■ MODEL

■ Diffusion formulation of FANOVA. One observes a series of sample paths of a stochastic process driven by

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

where $\epsilon > 0$ is the diffusion coefficient, r and d are finite integers, m_i are (unknown) d-dimensional functional responses, and W_i are independent d-dimensional standard Wiener processes.

■ Results of Brown & Low (1996): Under general conditions, the corresponding discrete model with $N(0, \sigma^2)$ errors is asymptotically equivalent to the diffusion model with $\epsilon = \sigma/\sqrt{n}$.

■ MODEL

[Antoniadis, 1984; Abramovich, Antoniadis, Sapatinas, and Vidakovic, 2004]: Each of the r response functions in the FANOVA model admits the following unique decomposition

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

where m_0 is a constant function (the *grand mean*), $\mu(\mathbf{t})$ is either zero or a non-constant function of $\mathbf{t}$ (the *functional main effect* of $\mathbf{t}$), a_i is either zero or a non-constant function of i (the *treatment effect* of i) and $\gamma_i(\mathbf{t})$ is either zero or a non-zero function (functional *treatment* effect).

■ IDENTIFIABILITY CONSTRAINTS

- $$\int_{[0,1]^d} \mu(\mathbf{t})\, d\mathbf{t} = 0, \quad \sum_{i=1}^{r} a_i = 0,$$

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

■ Hypotheses to be Tested 1

Testing the significance of the main effects and the interactions

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

Identifiability constraints $\longrightarrow$

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

where $Y_i^* = \int_{[0,1]^d} dY_i(\mathbf{t})$ and ξ_i are independent $\mathcal{N}(0,1)$ random variables. This is the classical one-way fixed-effects ANOVA model.

■ Hypotheses to be Tested 2

We assume that m_i (and, hence, μ and γ_i as well) belong to a Besov ball of radius $C > 0$ on $[0,1]^d$, $B^s_{p,q}(C)$, where $s > 0$ and $1 \leq p, q \leq \infty$.

■ Interested in: **Rate** at which the distance between the null and alternative hypotheses decreases to zero, while still permitting consistent testing. Alternatives are separated away from the null by ρ in the $L^2([0,1]^d)$.

■ Alternatives are of the form

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

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

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

■ Minimax Optimality 1

Consider the general model

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

where W is a d-dimensional standard Wiener process.

We wish to test

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

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

■ For prescribed α and β, the rate of decay to zero of the "indifference threshold" $\rho = \rho(\epsilon)$, as $\epsilon \to 0$, can be viewed as a **measure of goodness of a test. It is natural to seek the test with the optimal (fastest) rate.**

■ Minimax Optimality 2

Definition $\rho(\epsilon)$ is the minimax rate of testing if $\rho(\epsilon) \to 0$ as $\epsilon \to 0$ and the following two conditions hold

(i) for any $\rho'(\epsilon)$ satisfying $\rho'(\epsilon)/\rho(\epsilon) = o_\epsilon(1)$, one has

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

where $o_\epsilon(1) \to 0$ 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

$$\alpha(\phi_\epsilon^*) \le \alpha + o_\epsilon(1), \quad \beta(\phi_\epsilon^*, c\rho(\epsilon)) \le \beta + o_\epsilon(1).$$

ϕ_ϵ^* is called an asymptotically optimal (minimax) test.

■ Minimax Optimality 3

■ Ingster (1993) and Lepski & Spokoiny (1999) showed that for $sp > d$ the asymptotically optimal (minimax) rate is

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

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

■ The proposed asymptotically optimal (minimax) tests were consistent but *non-adaptive* [involve the smoothness parameters s and p of the corresponding Besov ball].

■ Spokoiny (1996) and Horowitz & Spokoiny (2001): Problem of *adaptive* minimax testing where s and p are unknown. **No adaptive test can achieve the exact optimal rate uniformly over all s and p** (in some given range).

■ Minimax Optimality 4

■ Price for Adaptivity: If one allows increase of $\rho(\epsilon)$ by an additional log-log factor $t_\epsilon = (\ln \ln \epsilon^{-2})^{1/4}$, i.e, considers $\rho(\epsilon t_\epsilon)$ instead of $\rho(\epsilon)$, then [Horowitz & Spokoiny (2001)] the optimal rate of adaptive testing is

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

■ The "price" factor t_ϵ is unavoidable and cannot be reduced.

■ Wavelet Bases

We assume $d = 1$ and work with periodic o.n. wavelet bases in $L^2([0,1])$ generated by shifts of a compactly supported scaling function ϕ, i.e.

$$\phi^{\mathrm{p}}(t) = \sum_{\ell\in\mathbb{Z}} \phi(t-\ell), \; \psi^{\mathrm{p}}_{jk}(t) = \sum_{\ell\in\mathbb{Z}} \psi_{jk}(t-\ell), \; j \geq 0, k = 0, \ldots, 2^j-1$$

where

$\psi_{jk}(t) = 2^{j/2}\psi(2^j t - k)$. $\{\phi^{\mathrm{p}};\ \psi^{\mathrm{p}}_{jk},\ j \geq 0,\ k = 0, 1, \ldots, 2^j - 1\}$ generates an o.n. basis in $L^2([0,1])$.

If the MRA is of regularity $r > 0$, the corresponding wavelet basis is **unconditional** for Besov spaces $B^s_{p,q}([0,1])$ for $0 < s < r$, $1 \leq p, q \leq \infty$. Such bases characterize Besov balls in terms of wavelet coefficients.

■ Testing in FANOVA

Averaging over r paths + identifiability conditions:

$$\begin{aligned} d\overline{Y}(t) &= (m_0 + \mu(t))\, dt + \epsilon\, d\overline{W}(t),\ t \in [0,1] \\ d(Y_i - \overline{Y})(t) &= (a_i + \gamma_i(t))\, dt + \epsilon\, d(W_i - \overline{W})(t),\ i = 1, \dots, r. \end{aligned}$$

$\{W_i - \overline{W};\ i = 1, \dots, r\}$ are Wiener processes with the same covariance kernel $C(s,t) = \frac{r-1}{r} \min(s,t)$ [but no independent].

$$dZ(t) = f(t)\, dt + \eta\, dW(t),\ t \in [0,1],$$

- $Z(t) = \overline{Y}(t),\ f(t) = m_0 + \mu(t),\ \eta = \epsilon/\sqrt{r}$
- $Z(t) = (Y_i - \overline{Y})(t),\ f(t) = a_i + \gamma_i(t),\ \eta = \epsilon\sqrt{(r-1)/r}$

■ NON-ADAPTIVE TEST 1

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

$$H_0 : \ f(t) \equiv 0 \quad \text{vs} \quad H_1 : \ f(t) \in \mathcal{F}(\rho),$$

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

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

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

■ NON-ADAPTIVE TEST 2

■ ϕ^* is based on the sum of squares of the thresholded empirical wavelet coefficients Y_{jk} with properly chosen level-dependent thresholds. The null hypothesis is rejected when this sum of squares exceeds some critical value.

■ j_η the largest integer: $j_\eta \leq \log_2 \eta^{-2}$.

■ $j(s)$ resolution level given by

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

■ Levels split as:

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

■ NON-ADAPTIVE TEST 3

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

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

■ For each $j \in \mathcal{J}_+$ and for given threshold $\lambda > 0$, define

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

where $b(\lambda) = \mathbb{E}\left[\xi^2 \mathbf{1}(|\xi| > \lambda)\right]$ and $\xi \sim \mathcal{N}(0,1)$.

■ NONADAPTIVE TEST 4

■ Define

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

and

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

where $\lambda_j = 4\sqrt{(j - j(s) + 8)\ln 2}$.

■ Under H_0,

$v_0^2(j(s)) = 2\eta^4 2^{j(s)}$ and $w_0^2(j(s)) = \eta^4 \sum_{j=j(s)}^{j_\eta - 1} 2^j d(\lambda_j)$, are the variances of $T(j(s))$ and $Q(j(s))$, respectively, where $d(\lambda_j) = \mathbb{E}\left[\xi^4 \mathbf{1}(|\xi| > \lambda_j)\right]$.

■ NON-ADAPTIVE TEST 5

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

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

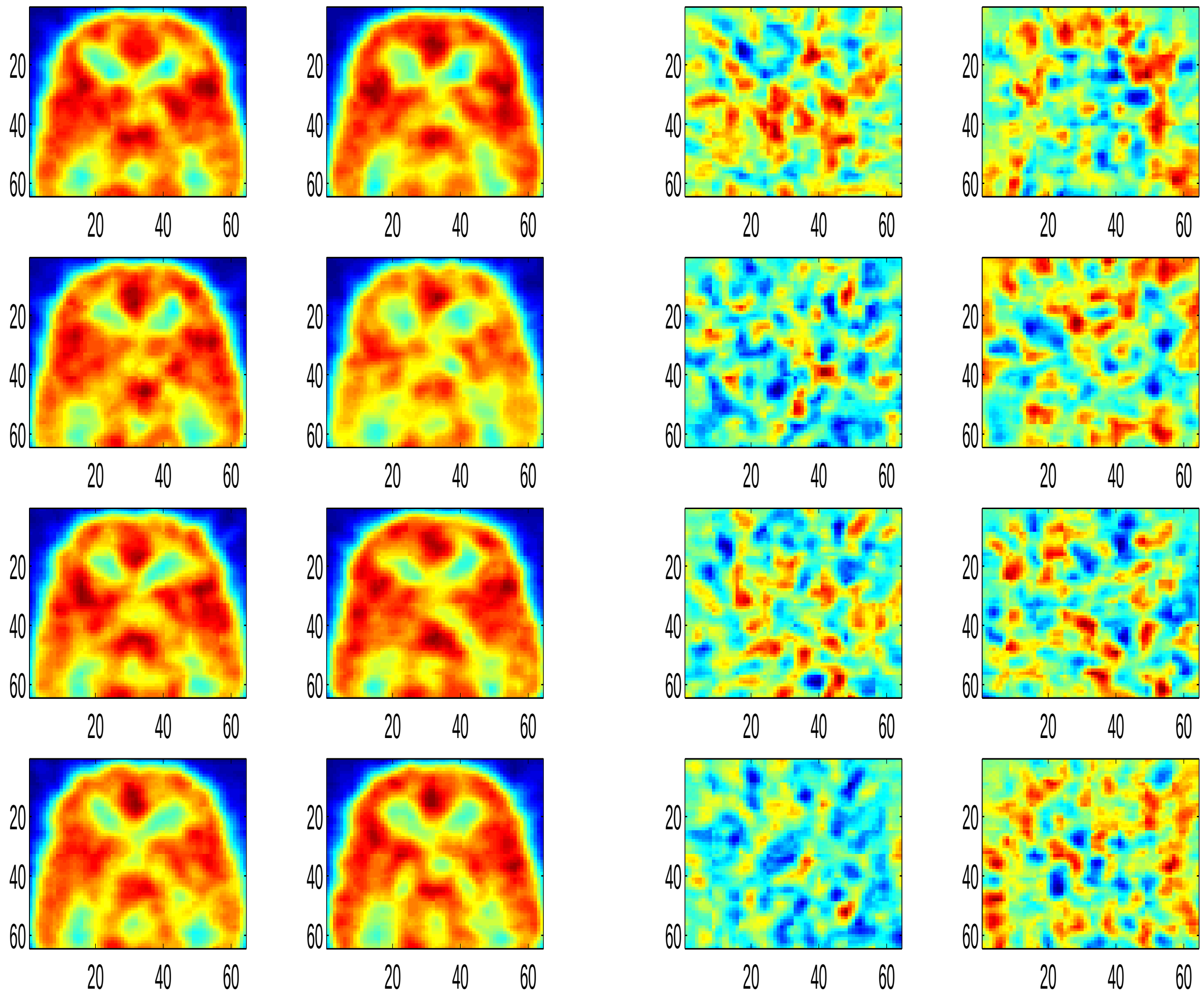
fMRI Example

- PET-fMRI data from a single subject experiment (Kinahan & Noll, 1999; Nichols & Holmes, 2001).
- The subject used left hand to perform finger opposition task
- DATA: 8 ANALYZE format $64 \times 64 \times 26$ scans, cut at $z = 8$. Odd scans (1,3,5,7) activation; even scans (2,4,6,8) baseline. Details:

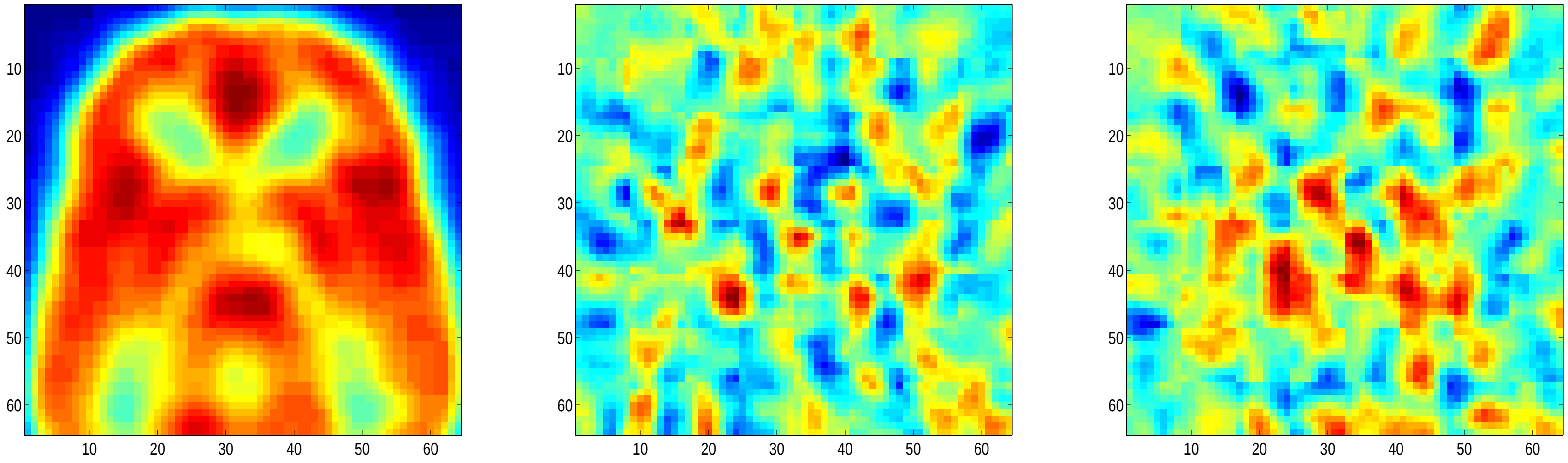
`http://www.fil.ion.ucl.ac.uk/spm/data/motor.html`

MODEL: $dY_i(\mathbf{t}) = m_i(\mathbf{t})\ d\mathbf{t} + \sigma dW_i(\mathbf{t}),$

$$m_i(\mathbf{t}) = m_0 + \mu(\mathbf{t}) + a_i + \gamma_i(\mathbf{t}),\ i = 1, \ldots, 8,\ \mathbf{t} \in [0,1]^2$$



Row data and the treatment effects $\hat{\gamma}_i(\mathbf{t}),\ i = 1, \ldots, 8.$



(left) The estimator of the grand mean, $\hat{\mu}(\mathbf{t})$;

(middle) Estimator $\hat{\gamma}_2(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$; hypothesis $H_0 : \gamma_2(\mathbf{t}) - \gamma_8(\mathbf{t}) = 0$ tested not significant at 5% level;

(right) Estimator $\hat{\gamma}_1(\mathbf{t}) - \hat{\gamma}_8(\mathbf{t})$, hypothesis $H_0 : \gamma_1(\mathbf{t}) - \gamma_8(\mathbf{t}) = 0$ tested significant at 5% level.

Γ-Minimax Wavelet Shrinkage

- Model induced wavelet shrinkage.
- Dimensionality of the model (Do not worry – wavelets decorrelate) $\longrightarrow$ Independence Models (coefficient-by-coefficient).
- Accounting for dependence (neighbors, parent-children), Blocking strategies (classical), Many Bayes solutions (MCMC, hidden MC's).
- Model complexity/efficiency compromise.
- Simple models/Fast shrinkage $\circ$ Realistic? Complex models $\circ$ Useful?

- What kind of prior information about signal can be incorporated? Smoothness, Scaling, Periodicity, Energy,...
- Joint work with Claudia Angelini, from CNR-Naples, Italy.
- Approach: *Gamma*-Minimax. Rule δ^* is Γ-minimax if $\rho(\delta^*, \pi^*) = \inf_{\delta \in \mathcal{D}} \ \sup_{\pi \in \Gamma} \ \rho(\delta, \pi)$

  ρ – cost function (Bayes risk, posterior expected loss; posterior regret; ...) ; $\mathcal{D}$ – class of decision rules; Γ – class of priors.
- Standard priors in the wavelet-domain have the form: $\pi(\theta) = \epsilon\delta_0 + (1-\epsilon)q(\theta)$. We assume that q-part belongs to a class of compactly supported, symmetric and unimodal densities.

Γ-Shrink

Theorem. Under the model

$$\begin{cases} d|\theta & \sim \quad \mathcal{N}(\theta, 1) \\ \theta & \sim \quad \pi(\theta) \in \Gamma \\ \mathcal{L}(\theta, \delta) & = \quad (\theta - \delta)^2 \ \text{ Squared Error Loss} \end{cases}$$

where $\Gamma = \{\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0) q(\theta), q(\theta) \in \Gamma_{SU[-m,m]}\}$, holds

$$\inf_{\delta \in \mathcal{D}} \sup_{\pi \in \Gamma} r(\pi, \delta) = \sup_{\pi \in \Gamma} \inf_{\delta \in \mathcal{D}} r(\pi, \delta),$$

where $r(\pi, \delta) = E^{\theta} E^{d|\theta} (\theta - \delta)^2$ is Bayes risk. The associated Γ-minimax rule is Bayes with respect to the least favorable distribution π in Γ.

The least favorable distribution π in Γ is of the form

$$\pi(\theta) = (\epsilon_0 + (1-\epsilon_0)\alpha_0)\delta_0 + (1-\epsilon_0)\sum_{k=1}^{p} \alpha_k \mathcal{U}[-m_k, m_k]$$

$$\alpha_k = \alpha_k(\epsilon_0) \geq 0; \quad \sum_{k=0}^{p} \alpha_k = 1;$$

$$m_k = m_k(\epsilon_0) \ \text{ s.t. } \ 0 < m_1 < m_2 < \ldots < m_p = m,$$

and the corresponding Bayes rule is

$$\delta_\pi(d) = d - \frac{(\epsilon_0 + (1-\epsilon_0)\alpha_0) d\phi(d) - (1-\epsilon_0)\sum_{k=1}^{p} \frac{\alpha_k}{2m_k}\left(\phi(d+m_k) - \phi(d-m_k)\right)}{(\epsilon_0 + (1-\epsilon_0)\alpha_0)\phi(d) + (1-\epsilon_0)\sum_{k=1}^{p} \frac{\alpha_k}{2m_k}\left(\Phi(d+m_k) - \Phi(d-m_k)\right)},$$

where ϕ and Φ are the density and the cdf of the standard normal random variable and $\mathcal{U}$ denotes the uniform distribution.

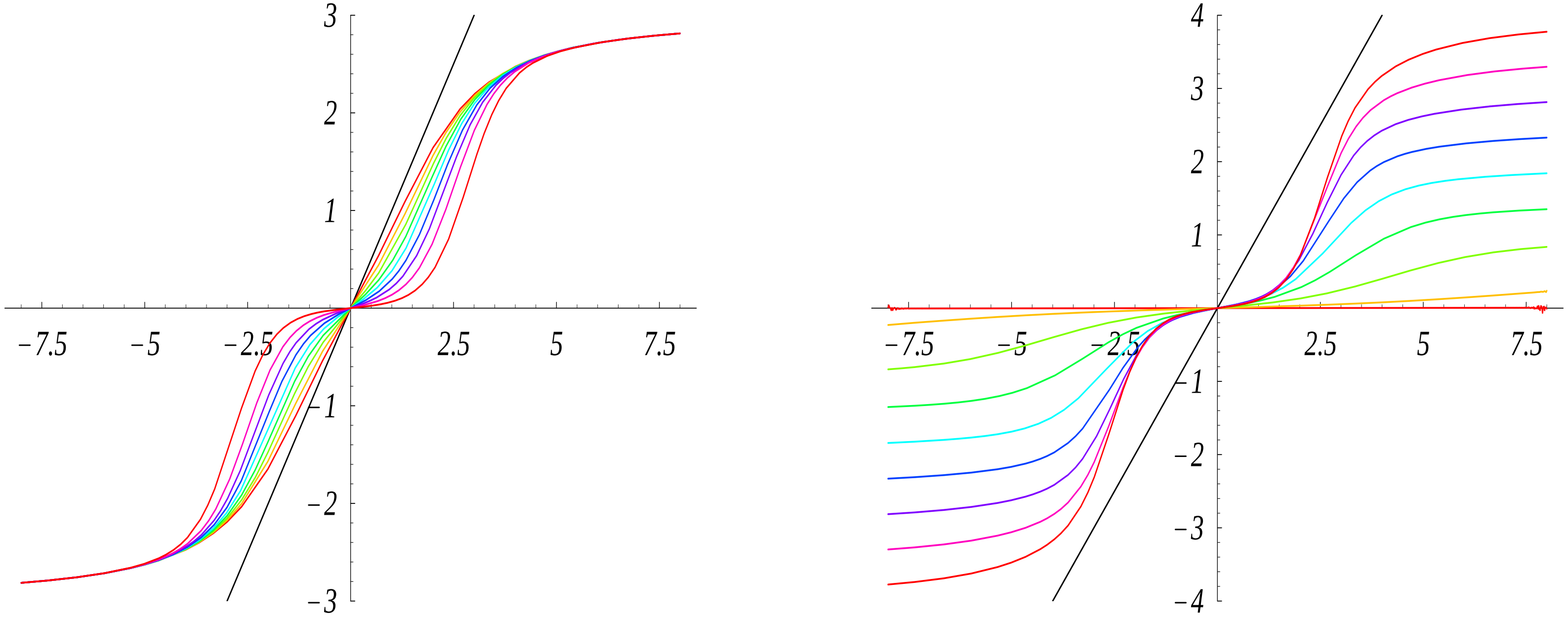
For any ϵ_0 there exists $m^* = m^*(\epsilon_0)$ such that, for any $m \leq m^*$, the least favorable prior is

$$\pi(\theta) = \epsilon_0 \delta_0 + (1 - \epsilon_0)\mathcal{U}[-m, m]$$

and the Γ-minimax rule has the form

$$\delta_\pi(d) = d - \frac{\epsilon_0 d \phi(d) - \frac{1-\epsilon_0}{2m}\left(\phi(d+m) - \phi(d-m)\right)}{\epsilon_0 \phi(d) + \frac{1-\epsilon_0}{2m}\left(\Phi(d+m) - \Phi(d-m)\right)}.$$

Hyperparameters m and ϵ_0 are elicited by signal-to-noise ratio (SNR) and smoothness considerations.

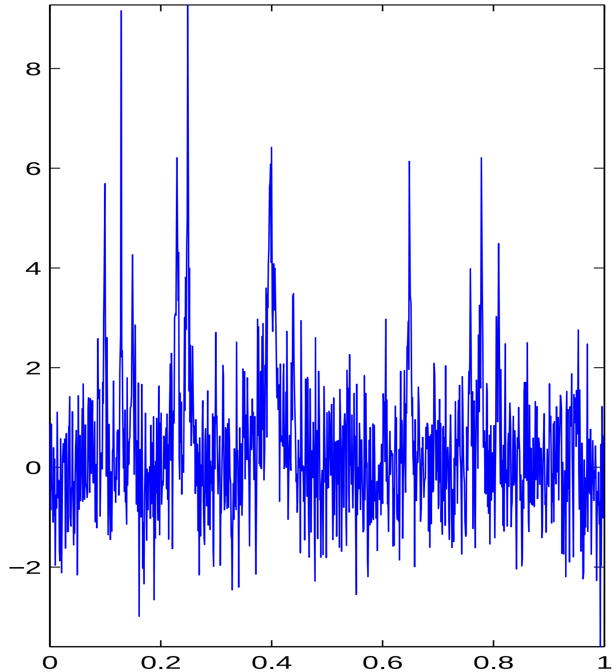


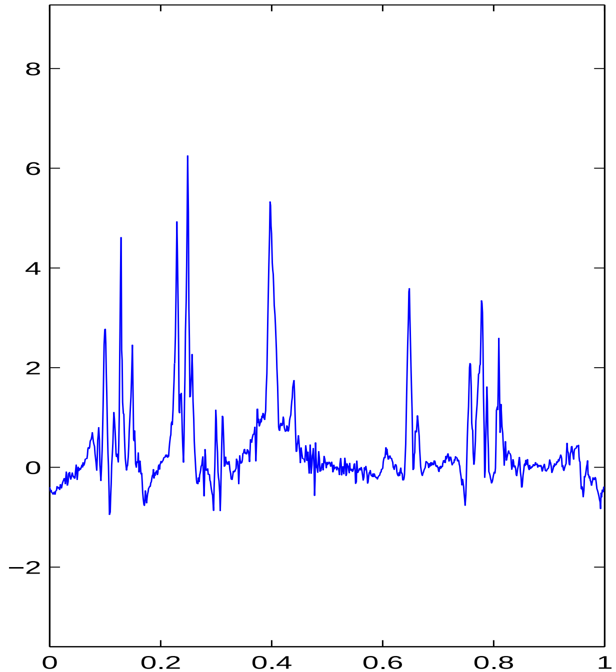
Left: Γ-minimax rules for $m = 3$ and ϵ_0 ranging from 0.1 (upper envelope function) to 0.9 (lower envelope function); Right: Γ-minimax rules for $\epsilon_0 = 0.8$ and m ranging from 0.5 to 4.

	BLOCKS	BUMPS
Γ-SHRINK	0.1094 (0.0573+0.0521)	0.1919 (0.1002+0.0917)
VISUSHRINK	0.2772 (0.2605+0.0167)	0.5326 (0.5075+0.0251)
SURESHRINK	0.2337 (0.0270+0.0207)	0.2390 (0.0771+0.1619)
VISUSHRINK (5)	0.1870 (0.1531+0.0339)	0.4041 (0.3626+0.0415)
SURESHRINK (5)	0.1738 (0.0832+0.0906)	0.2381 (0.0726+0.1655)

	HEAVISINE	DOPPLER
Γ-SHRINK	0.0168 (0.0081+0.0087)	0.0670 (0.0330+0.0340)
VISUSHRINK	0.0178 (0.0097+0.0081)	0.1890 (0.1735+0.0155)
SURESHRINK	0.0355 (0.0088+0.0267)	0.1427 (0.0221+0.1216)
VISUSHRINK (5)	0.0348 (0.0031+0.0317)	0.1244 (0.0898+0.0346)
SURESHRINK (5)	0.0383 (0.0030+0.0353)	0.1420 (0.0193+0.1227)

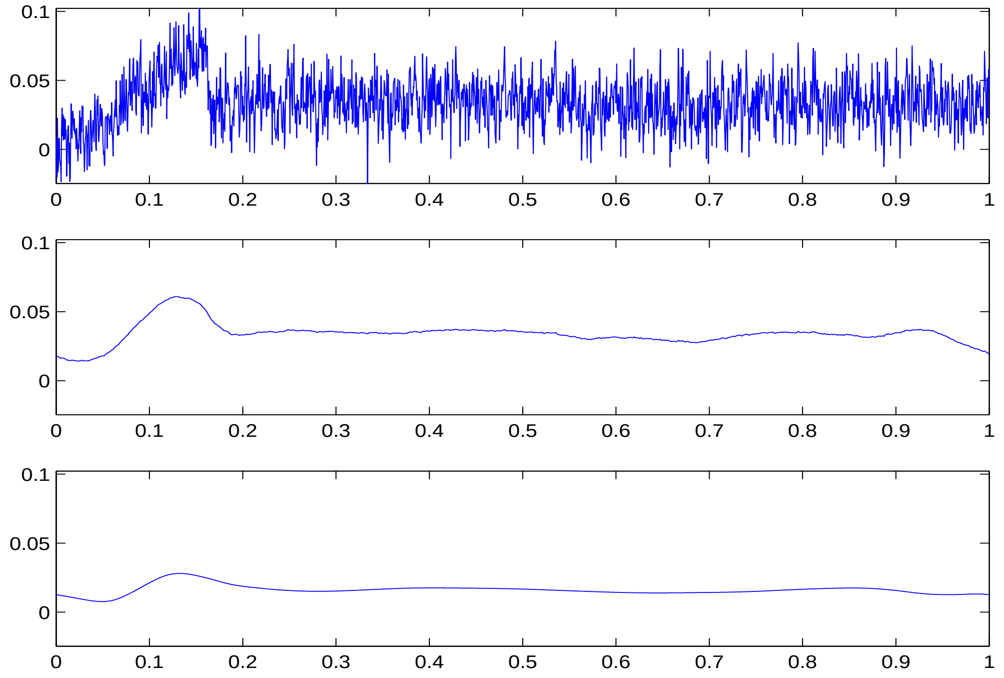
$SNR = 1$, Symmlet 8 for `doppler`, `heavisine`, Haar for `blocks`, and Daub3 (6 taps) for `bumps`. Sample size $n = 1024$.





Left: noisy **bumps** [SNR=1, $n = 1024$, $\sigma^2 = 1$]; Right: The reconstructed signal using the Γ-minimax rule.

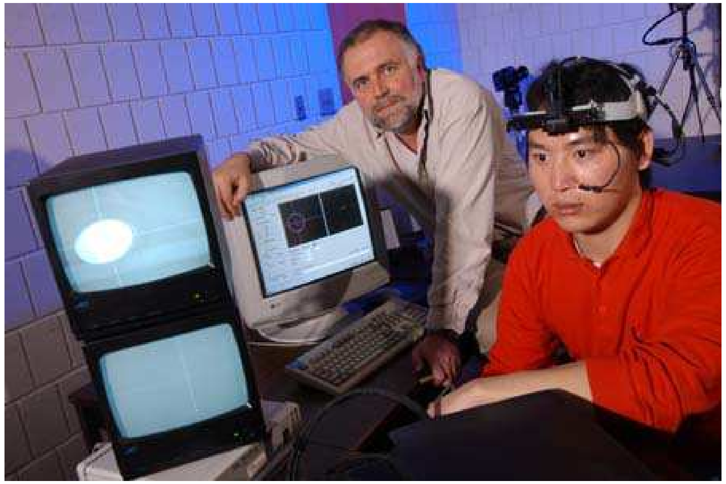
Atomic Force Microscopy Data: The adhesion measurements between carbohydrate and the cell adhesion molecule (CAM) E-Selectin, (Bryan Marshal, Dpt of BME at GaTech).

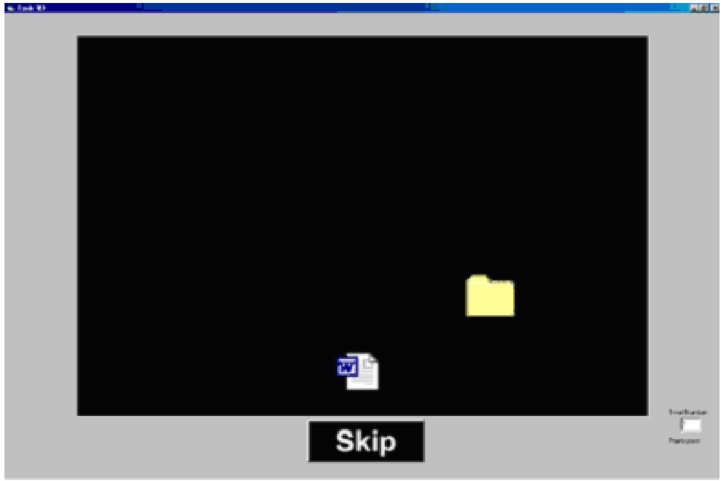


Top: Original AFM measurements; Middle: Γ-Shrink; Bottom: SureShrink.

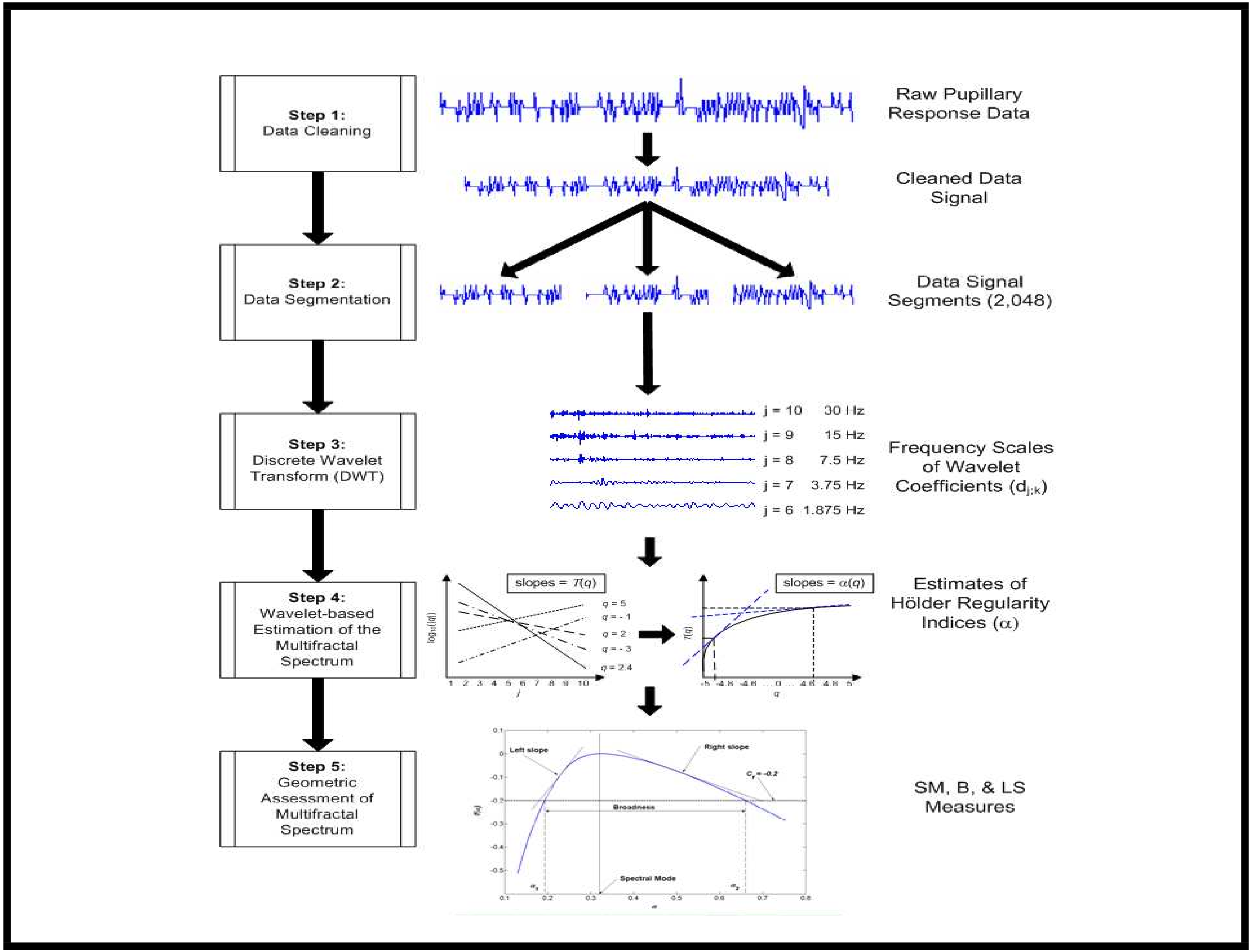
Wavelets in Assessing AMD

- AMD ≡ Age-Related Macular Degeneration
- Joint work with Moloney, Shi, Leonard, Jacko, and Sainfort, Health Care Informatics, BME GaTech
- Pupillary Response Behavior (PRB): *pupil diameter dynamics* affected by *age, AMD status, cognitive workload*
- Participants: 3 groups, $n = 28$ (average age 76), $n_1 = 14$ control [20/20-20/35], $n_2 = 8$ AMD visual acuity [20/35-20/80], $n_3 = 6$ AMD visual acuity [20/80-20/200]
- Computer task Drag-and-drop, 105 trials
- 287 Data Sets: Pupil diameter at 60 Hz





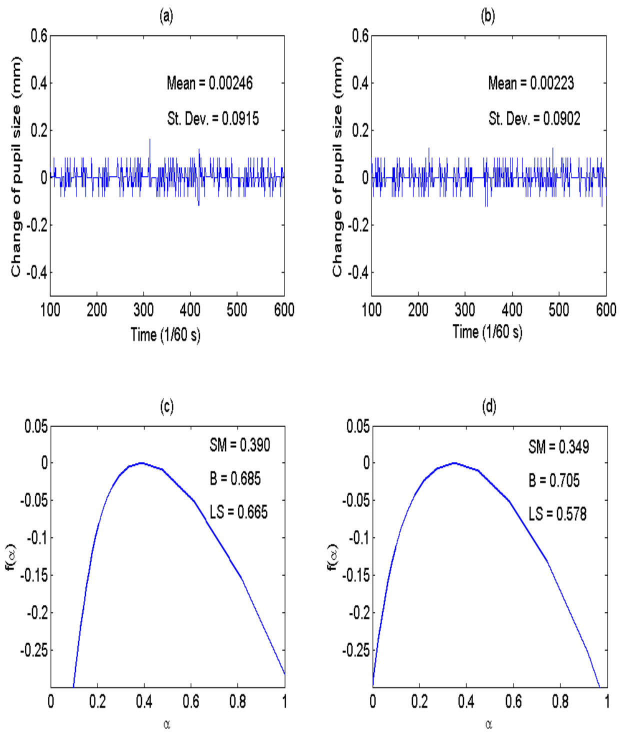
- PRB highly irregular for older adults with ocular disease.
- Use PRB to examine intrinsic user differences between older adults with/without AMD to explore underlying visual differences that affect human-computer interaction.

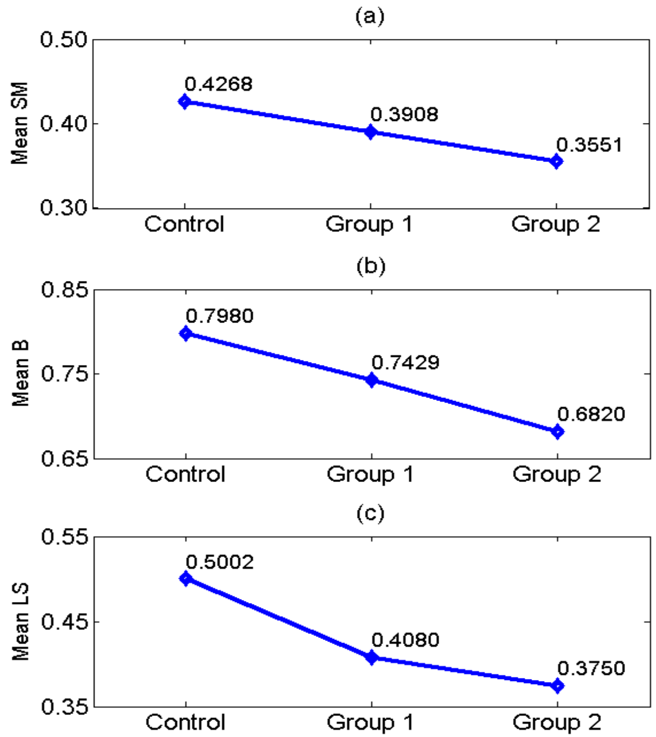

Step 1:
Data Cleaning
Step 2:
Data Segmentation
Step 3:
Discrete Wavelet Transform (DWT)
Step 4:
Wavelet-based Estimation of the Multifractal Spectrum
Step 5:
Geometric Assessment of Multifractal Spectrum
Raw Pupillary Response Data
Cleaned Data Signal
Data Signal Segments (2,048)
j = 10 30 Hz
j = 9 15 Hz
j = 8 7.5 Hz
j = 7 3.75 Hz
j = 6 1.875 Hz
Frequency Scales of Wavelet Coefficients ($d_{j;k}$)
slopes = $\tau(q)$
slopes = $\alpha(q)$
Estimates of Hölder Regularity Indices (α)
Left slope
Right slope
Broadness
Spectral Mode
SM, B, & LS Measures

- Monotonic trend Control Group $<$ Group 1 $<$ Group 2

Visual impairment ↓ = *LS* ↑ (*Multifractality* ↑)

- ANOVA: Group effect on LS ($F = 32.238, p < 0.01$) Group 2 $<$ Group 1 $<$ Control group ($p < 0.05$)





■ CONCLUSIONS

■ Formal statistical inference with nonstandard (from the statistical point of view) observations. Dimension reduction and whitening properties of transformed domains critical.

■ Prior information robustness: Being reasonably conservative and subjective. Incorporating prior information about SNR.

■ Generating descriptors as inputs for classifiers using scaling and multifractal properties of the signal/image.

■ Data, Manuscripts, Software at Jacket's Wavelets:
`http://www2.isye.gatech.edu/ ˜brani/wavelet.html`

■ SOME ADDITIONAL MATERIAL (DETAILS AND EXAMPLES)

■ One Slide Intro to Besov Spaces

■ Function f belongs to Besov $B^s_{p,q}$ space if there exist $f_0, g_0, g_1, \ldots$ from the Sobolev space $W^m_p = \{f| \quad \|(1+|\omega|^2)^{m/2}\hat{f}(\omega)\|_{L_p} < \infty\}$ and a sequence $\{\epsilon_0, \epsilon_1, \epsilon_2, \ldots\} \in \ell^q$ such that

- $f = f_0 + g_0 + g_1 + \cdots \in L_p,$
- $\|g_j\|_{L_p} < \epsilon_j 2^{-sj}, \; j = 0, 1, 2, \ldots$
- $\|g_j^{(m)}\|_{L_p} \leq C\epsilon_j 2^{(m-s)j}, \; j = 0, 1, 2, \ldots$

where $C > 0$ is a constant and m is an integer such that $m < s$.

■ Functions from Besov space $B^s_{p,q}$ for which Besov norm $< C$ form the Besov ball $B^s_{p,q}(C)$.

■ APPLICATIONS

■ SIMULATION STUDY 1

■ Synthetic data from the battery of standard test functions of Donoho & Johnstone (1995): BLOCKS, BUMPS, DOPPLER and HEAVISINE. Additional test function MISHMASH, defined as

$$\text{MISHMASH} = -(\text{ BLOCKS} + \text{ BUMPS} + \text{DOPPLER} + \text{HEAVISINE}),$$

added because of the identifiability constraints.

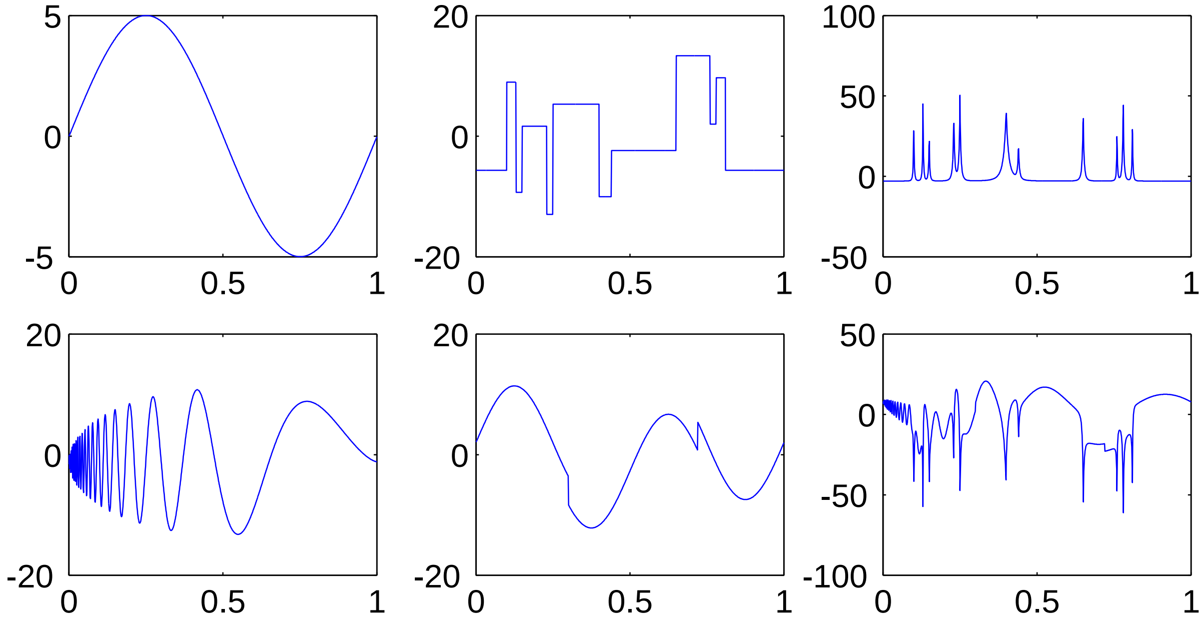


Figure 1: The mean function $\mu(t) = 5\sin(2\pi t)$ and the centered treatment effect functions $\gamma_i(t)$, $i = 1, \ldots, 5$ (i.e., centered BLOCKS, BUMPS, DOPPLER, HEAVISINE, and MISHMASH), sampled at $n = 1024$ data points.

■ SIMULATION STUDY 3

■ $m_0 = 1,\ \mu(t) = 5\sin(2\pi t)$

■ Five simulated observations (one for each test function shown; length ($n = 1024$), two SNRs (SNR = 3 and 7).

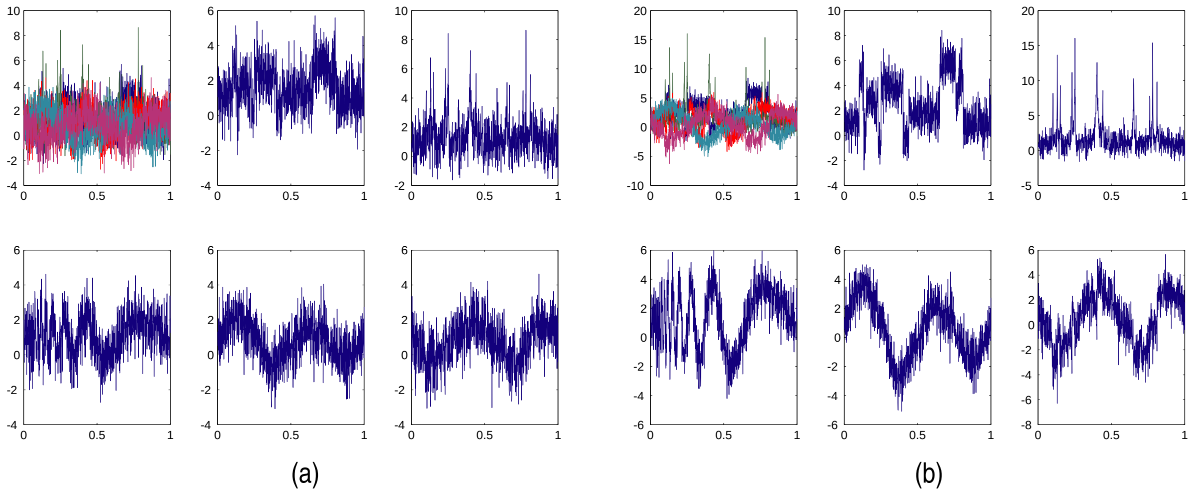


Figure 2: Five simulated observations (one for each test function shown in Figure 1) sampled at $n = 1024$ data points are shown superimposed (first plot) and separately (remaining five plots) for (a) SNR = 3 and (b) SNR = 7.

■ SIMULATION STUDY 4

■ To test the hypothesis $H_0 : \mu(t) = 0$, nonadaptive test, $p \geq 2$. ■ Symmlet 8-tap

■ $j(s) = 3$

■ SNR=3: $T(3) = 15.28$ critical value 1.5949

■ SNR=7: $T(3) = 97.52$ critical value 1.6316.

■ $H_0 : \gamma_i(t) = 0$ $(i = 1, \ldots, 5)$, non-adaptive test, $1 \leq p < 2$. ■ Daubechies 6-tap

■ $j(s) = 3$ ■ $j_\eta = 7$.

■ SNR=3, $T(3) + Q(3) = 275.3326$ critical value 154.6294

■ SNR=7, $T(3) + Q(3) = 5941.099$ critical value 156.4943

■ SIMULATION STUDY 5

■ Extensive power analysis for the above tests against the composite alternatives

$$H_1 : \mu \in \mathcal{F}(\rho) \quad \text{and} \quad H_1 : \frac{1}{5}\sum_{i=1}^{5} \gamma_i \in \mathcal{F}(\rho). \qquad (1)$$

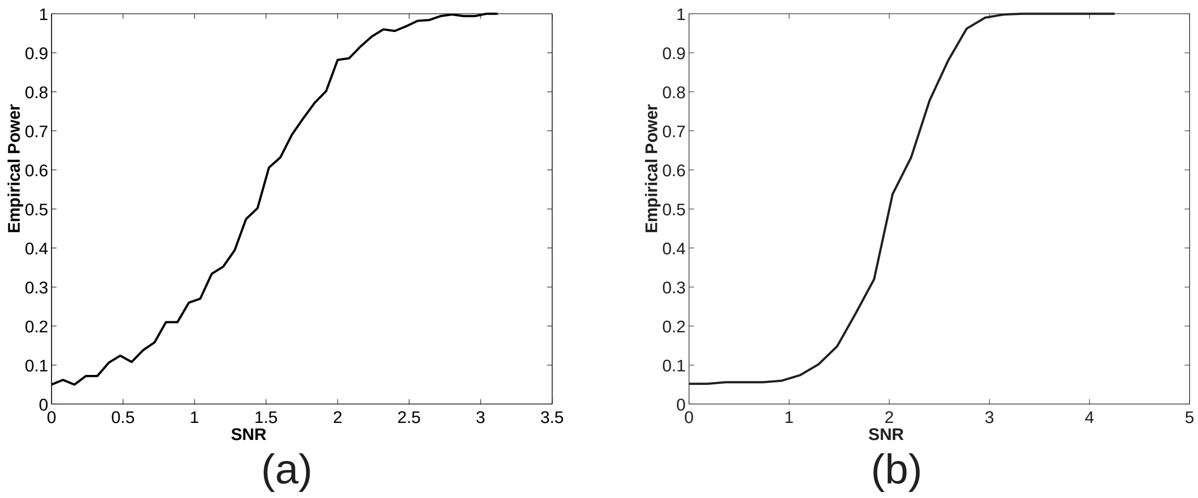


Figure 3: Empirical power functions for testing (a) $H_0 : \mu(t) = 0$ versus $H_1 : ||\mu||_2 = \rho$ and (b) $H_0 : \gamma_i(t) = 0$ $(i = 1, \ldots, 5)$ versus $H_1 : ||\sum_i \gamma_i/5||_2 = \rho$. In both panels, the sample size was $n = 512$ and the number of trials at a fixed discretized SNR was 500.

■ ORTHOSIS DATA ANALYSIS 1

■ Interesting data on human movement.

■ Data: Amarantini David and Martin Luc, Laboratoire Sport et Performance Motrice, Grenoble University

■ Underlying movement under various levels of an externally applied force to the knee.

■ Seven young male volunteers wore a spring-loaded orthosis of adjustable stiffness under 4 experimental conditions:

- Control condition (without orthosis),
- Orthosis condition,
- Two conditions (Spring1, Spring2) stepping in place was perturbed by fitting a spring-loaded orthosis onto the right knee.

■ ORTHOSIS DATA ANALYSIS 2

What is Orthosis? Another name for an orthosis is a brace. It is prescribed by a physician to provide correction, support, or protection to a part of the body.

Figure 4: Knee Orthoses

■ ORTHOSIS DATA ANALYSIS 3

■ The data set consists in 280 separate runs and involves the seven subjects over four described experimental conditions, replicated ten times for each subject.

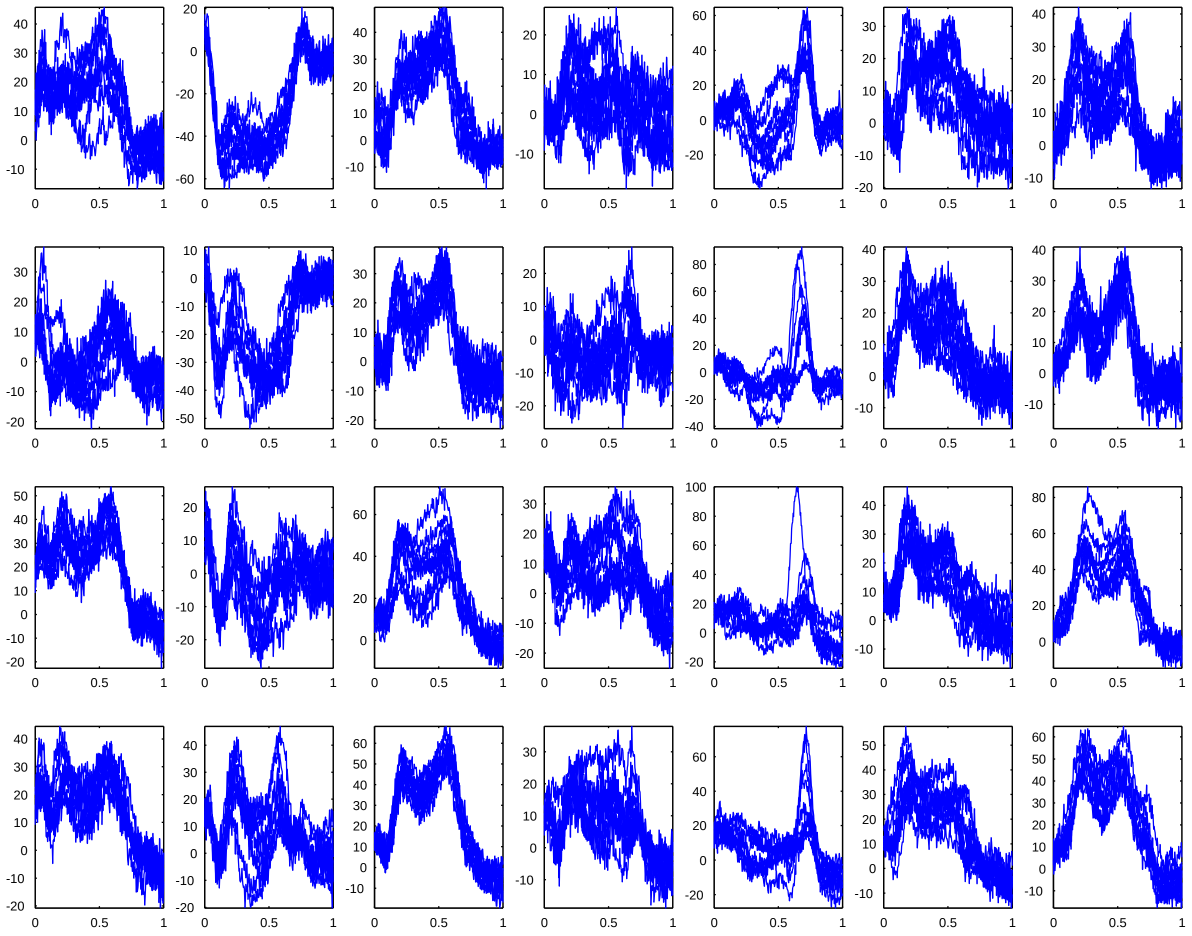


Figure 5: Orthosis data set: panels in rows correspond to *Treatments* while the panels in columns correspond to *Subjects*.

■ ORTHOSIS DATA ANALYSIS 4: MODEL

■ Model

$$\begin{aligned} dY_{ijk}(t) &= m_{ij}(t)\,dt + \epsilon\,dW_{ijk}(t), \\ &\quad i = 1,\ldots,I;\ j = 1,\ldots,J;\ k = 1,\ldots,K;\ t \in [0,1], \end{aligned}$$

with

$$\begin{aligned} m_{ij}(t) &= m_0 + \mu(t) + \alpha_i + \gamma_i(t) + \beta_j + \delta_j(t), \\ &\quad i = 1,\ldots,I;\ j = 1,\ldots,J;\ t \in [0,1], \end{aligned}$$

where i is the condition index, j is the subject index, k is the replication index, and t is the time.

■ Subjects in the above model are naturally considered as **block effects**; subjects obviously differ but the researchers are not interested in their differences.

■ ORTHOSIS DATA ANALYSIS 5: MODEL

$$d\bar{Y}_{i..}(t) = m_i(t)\,dt + \eta\,dW_{i..}(t),\ i = 1,\dots,I;\ t \in [0,1],$$

with

$$m_i(t) = m_0 + \mu(t) + \alpha_i + \gamma_i(t),\ i = 1,\dots,I;\ t \in [0,1],$$

where $\eta = \epsilon/\sqrt{JK}$.

■ *Coiflet 18-tap filter*

■ $j(s) = 4$ and $j_\eta = 6$.

■ ORTHOSIS DATA ANALYSIS 6

- Tests $H_0 : \mu(t) = 0$ and $H_0 : \gamma_i = 0$ were both significant.
- The researchers were interested in contrasts:
  - Control and Orthosis functional treatment effects are equal $(H_0 : \gamma_1(t) = \gamma_2(t))$. Not significant, p-value 0.157
  - Spring 1 and Spring 2 functional treatment effects are equal $(H_0 : \gamma_3(t) = \gamma_4(t))$. Not significant, p-value 0.198.
  - Contrast $(\gamma_1(t) + \gamma_2(t)) - (\gamma_3(t) + \gamma_4(t))$. Significant, p-value is 0.0103.

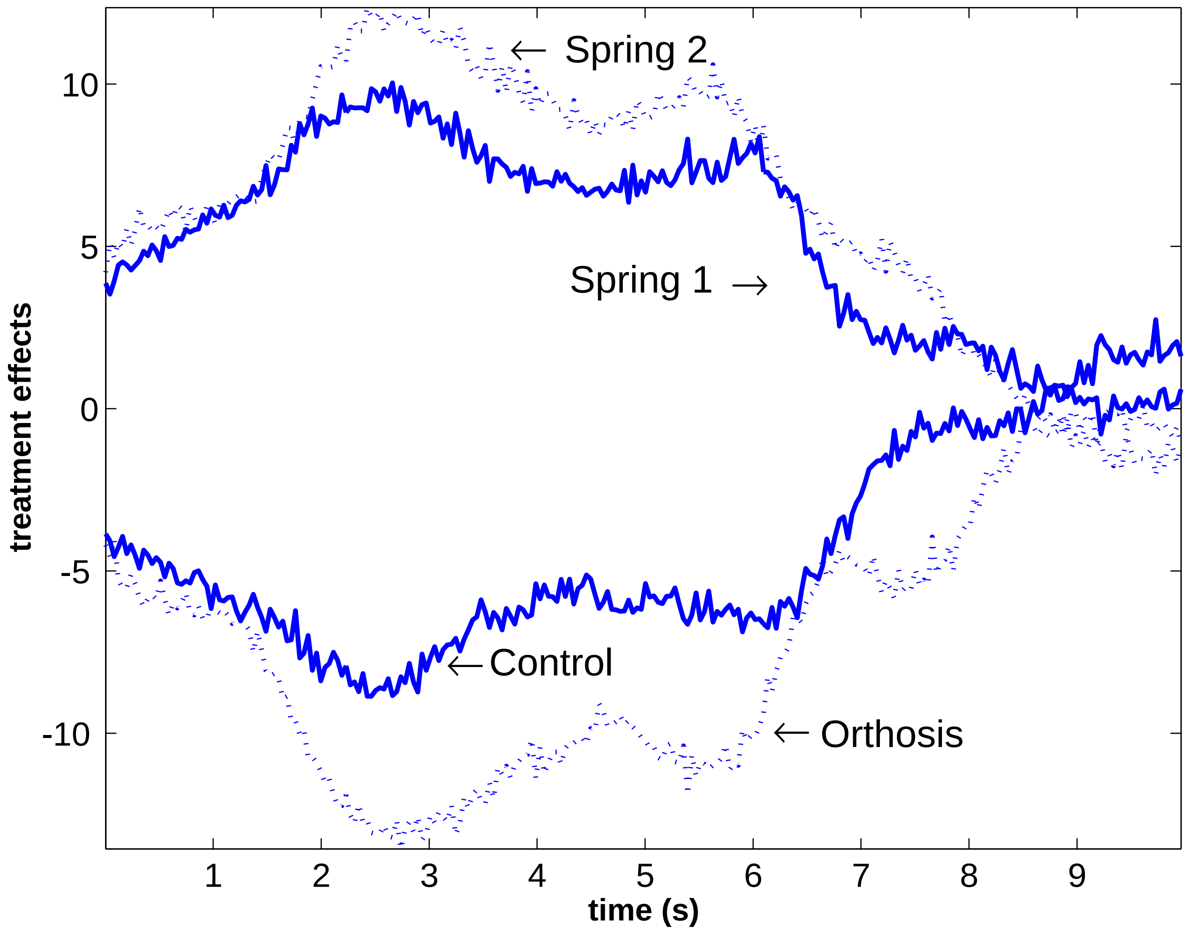


Figure 6: Empirical estimators of the treatment effects of interest. Constant and functional components α_i and $\gamma_i(t)$ $(i = 1, \ldots, 4)$ are not separated.

■ ADAPTIVE TEST 1

■ The parameters s, p, q and the radius C of the corresponding Besov ball $B^s_{p,q}(C)$ are unknown. Assume that $0 < s \le s_{\max}$, $1 \le p, q \le \infty$, $sp > 1$, $s - \frac{1}{2p} + \frac{1}{4} > 0$ and $0 < C \le C_{max}$.

■ Let $t_\eta = (\ln\ln \eta^{-2})^{1/4}$ and $j_{\min} = \frac{2}{4s_{\max}+1} \log_2 \eta^{-2}$.

■ Regularity of MRA: $r > s_{\max}$.

■ The idea: Consider the range of $j(s) = j_{\min}, \ldots, j_\eta - 1$ and reject H_0 if it is rejected at least for one selected level $j(s)$.

■ ADAPTIVE TEST 2

Since $\mathrm{card}(\{j_{\min}, \ldots, j_\eta - 1\}) = O(\ln \eta^{-2})$, Bonferroni type testing leads to the asymptotically *adaptive* test

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

■ ADAPTIVE TEST 3

Spokoiny (1996) showed that the test ϕ_η^* is an adaptive optimal test, i.e.

$$\alpha(\phi_\eta^*) = o_\eta(1)$$

and

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

where $\rho(\eta t_\eta) = (\eta t_\eta)^{4s''/(4s''+1)}$, $o_\eta(1) \to 0$ as $\eta \to 0$, and c is a constant.

If it is known that $p \geq 2$ then the adaptive test can be simplified to

$$\phi_\eta^* = \mathbf{1}\left[\max_{j_{\min} \leq j(s) \leq j_\eta - 1} \left\{ \frac{T(j(s))}{\sqrt{v_0^2(j(s))}} \right\} > \sqrt{2 \ln \ln \eta^{-2}}\right].$$

■ A COMMENT

The test ϕ_η^* is similar in spirit to that in Fan (1996) and Fan & Lin (2000), though they apply a *global* threshold.