Wavelet-Based Nonparametric Modeling of Hierarchical Functions in Colon Carcinogenesis

Jeffrey S. Morris

University of Texas, MD Anderson Cancer Center

Joint work with Marina Vannucci,

Philip J. Brown, and Raymond J. Carroll

Outline

- Introduction
  - Colon Carcinogenesis Studies
  - Hierarchical Functional Data
  - Wavelets and Wavelet Regression
- Overview of Method
- Choice of Smoothing Parameters
- Application
- Conclusions

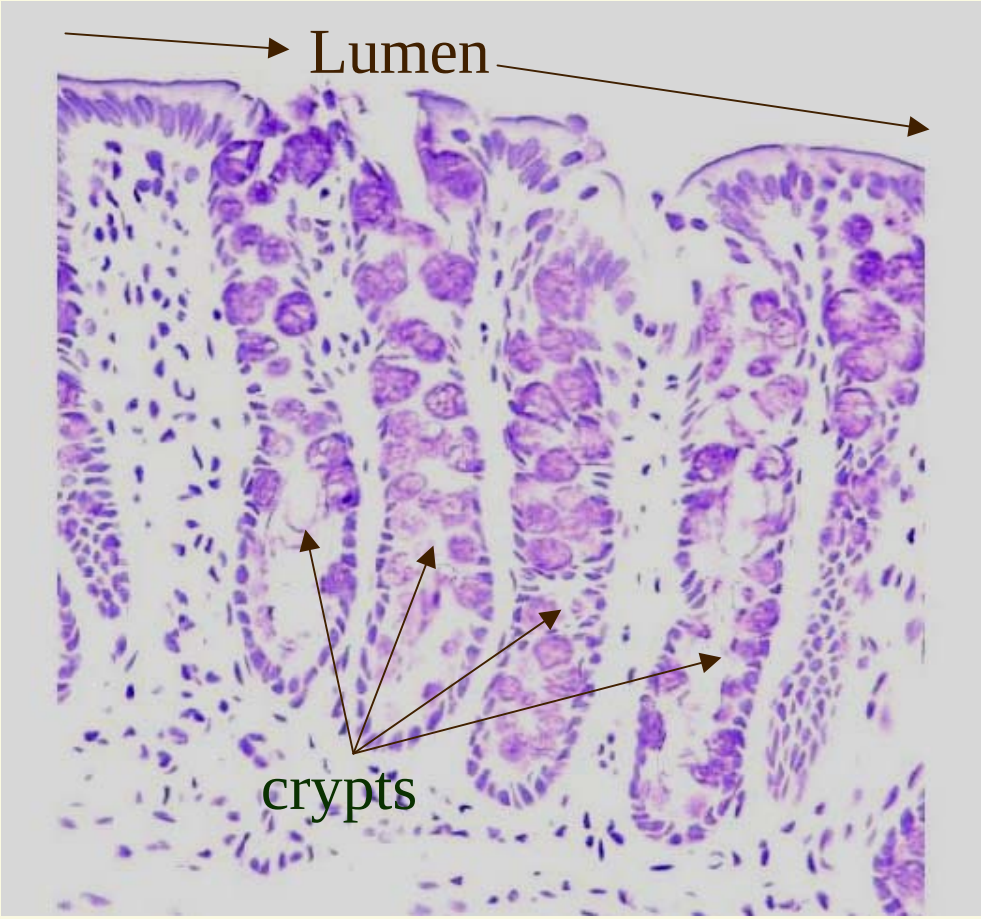
Colon Carcinogenesis Studies

- Colon cancer asymptomatic until advanced
  - Limited treatment options for advanced disease
- Preventive approaches crucial
  - Better understand carcinogenic mechanisms
  - Identify & understand risk factors
- Dietary factors key in colon cancer risk

Colon Carcinogenesis Studies

- "Carcinogen-induced colon cancer in rats" model used to study biological mechanisms *in vivo*
  - Rats given treatments of interest (diet) over time
  - Exposed to carcinogen
  - Animals euthanized, colon removed and examined.
- IHC staining used to quantify responses
  - Slides treated with chemical
  - Staining intensity indicates level of response.

Architecture of Colon



- ***Stem Cells***:
  Mother cells near bottom
- Depth in crypt ~ age of cells
- ***Relative Cell Position*** (Depth):
  $t \in (0,1)$

Colon Carcinogenesis Studies

- Data structure:
  - Multiple treatment groups (diets)
  - Multiple rats for each treatment
  - Multiple crypts sampled for each rat
  - Response quantified over fixed grid of *t* along side of each selected crypt
- Hierarchical functional data
  - Multi-level random effects model with functional observations

Hierarchical Functional Data

- **2-level HF model:**

$$\boxed{\begin{aligned} \mathbf{Y}_{abc} &= g_{abc}(t) + \mathbf{e}_{abc}, \\ g_{abc}(t) &= g_{ab}(t) + \eta_{abc}(t) \\ g_{ab}(t) &= g_{a}(t) + \xi_{ab}(t) \end{aligned}}$$

where $\mathbf{e}_{abc} \sim MVN(\mathbf{0}, \sigma_e^2 \mathbf{I})$,

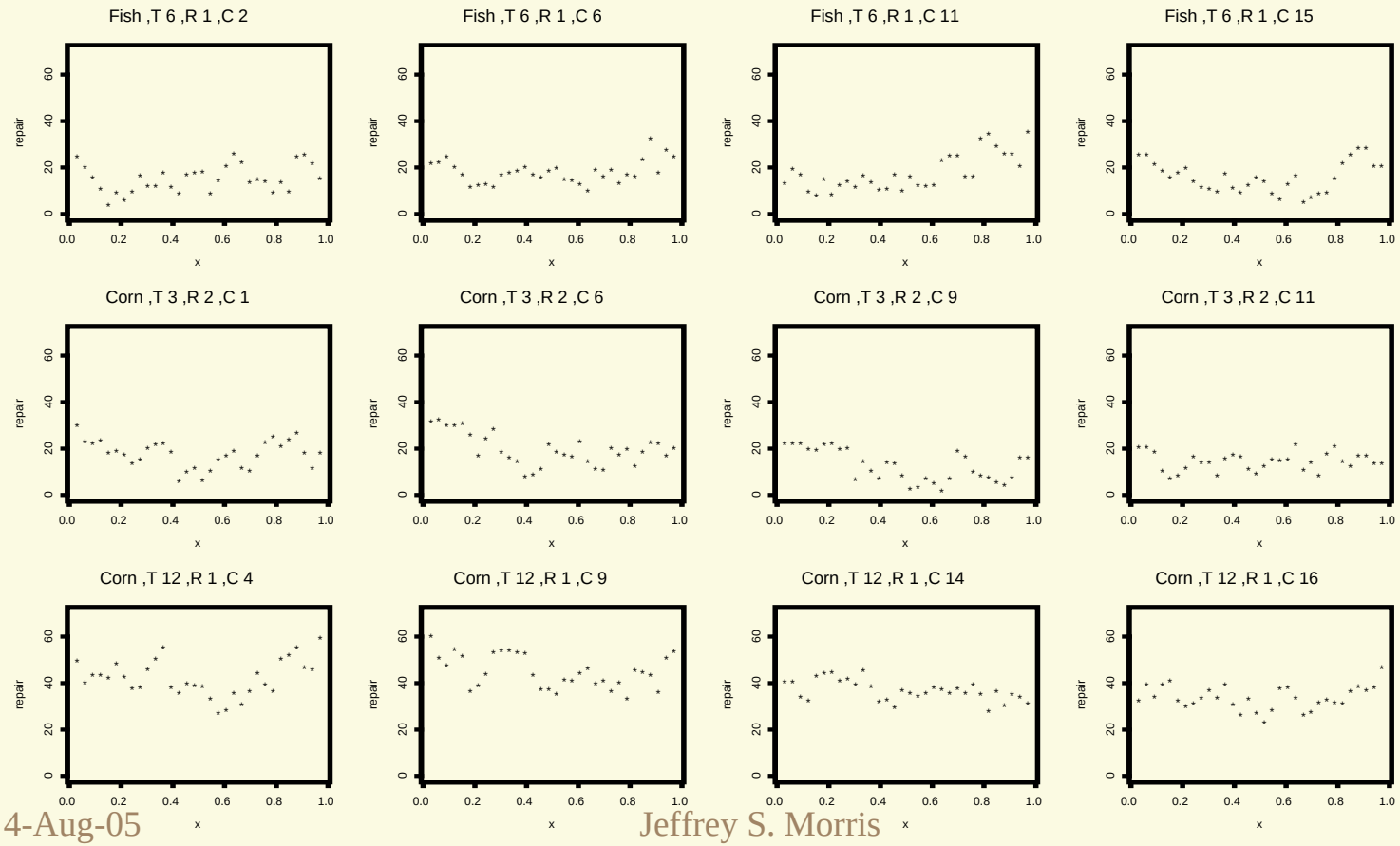
$\eta_{abc}(\bullet)$ and $\xi_{\mathrm{ab}}(\bullet)$: mean 0 with covariance matrices $\Sigma_1(t_1, t_2)$ and $\Sigma_2(t_1, t_2)$.

Hierarchical Functional Data

- Parametric model (linear):
  - $g(\mathbf{t}) = X(\mathbf{t})\,\beta$ at each level of hierarchy:
    - e.g. line or parabola at each level.
  - i.e. $g_{abc}(\mathbf{t}) = X(\mathbf{t})\,\beta_{abc}$, $\mathrm{Cov}(\beta_{abc}|\beta_{ab}) = \Sigma_{\beta 1}$
    $g_{ab}(\mathbf{t}) = X(\mathbf{t})\,\beta_{ab}$, $\mathrm{Cov}(\beta_{ab}|\beta_{a}) = \Sigma_{\beta 2}$
    $g_{a}(\mathbf{t}) = X(\mathbf{t})\,\beta_{a}$
- Can be fit using standard mixed models software.
- Efficient and easy IF you can find a parametric form for the g(•) functions.

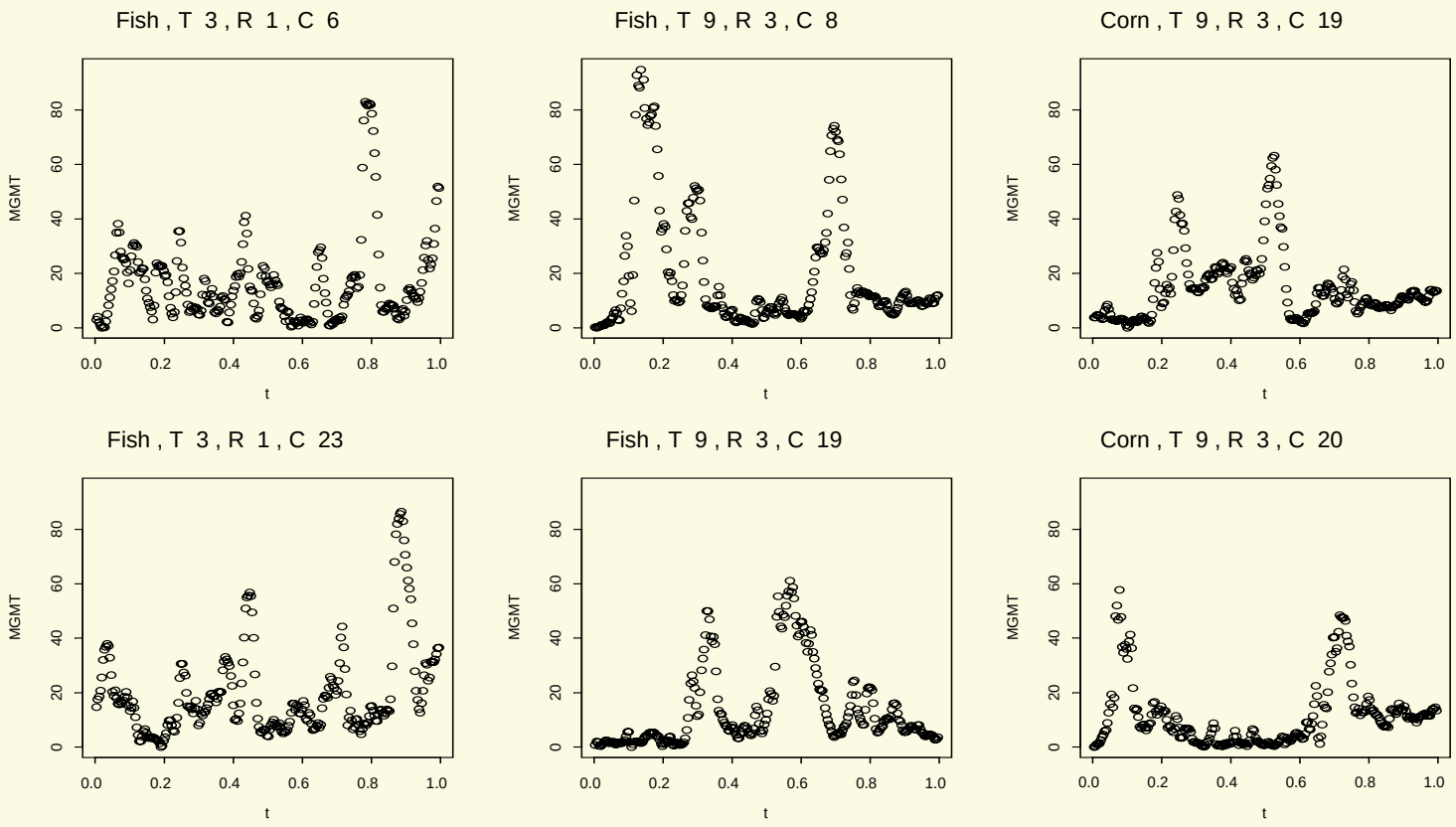
Hierarchical Functional Data

DNA Adduct Level (Damage)



Hierarchical Functional Data

MGMT Level (DNA Repair Enzyme)



Goals

- Develop method to analyze H.F.D.
  - Est. population-level functions $g_a(t)$ (nonparametrically)
  - Perform Bayesian inference
  - Uncover relative variability at each hierarchical level.
  - Work for hierarchical models with 1 or 2 levels
  - Adjust for possibly unequal sample sizes
  - Model spike-like data at lowest hierarchical level
- Our method first to do all these
- Apply to colon carcinogenesis data set
- Use of wavelets fundamental to our approach

Wavelets & Wavelet Regression

- Wavelets: families of orthonormal basis functions.
- Wavelet series for $f(t)$: $f(t) = \sum_{j,k \in \Im} d^{j,k} \psi^{j,k}(t)$
- Wavelet **basis functions**: $\psi^{j,k}(t) = 2^{j/2} \psi(2^j t - k)$
  - **Mother** wavelet: $\psi(t)$
- Wavelet **coefficients**: $d^{j,k} = \int f(t) \psi^{j,k}(t)\, dt$

Wavelets & Wavelet Regression

- Wavelet coefficients calculated using Discrete Wavelet Transform (DWT)
- Efficient -- $O(n)$
- n data points $\boldsymbol{y} \Rightarrow n$ wavelet coefficients $\boldsymbol{d}$
  - Indexed by: resolution level j *(scale)*, wavelet number k *(location)*.
- Can be written: $\boldsymbol{d} = \mathrm{W}\,\boldsymbol{y}$ (W orthogonal)
- Inverse Transform (IDWT): $\boldsymbol{y} = \mathrm{W}^t\,\boldsymbol{d}$

Wavelets & Wavelet Regression

- Key Properties:
  - Compact support -- model local features in time (spikes)
  - Whitening property -- less correlation in wavelet space, allowing parsimonious modeling
  - Sparsity -- most wavelet coefficients 0 for given function
- Can be used to perform Nonparametric Regression (Donoho & Johnstone, 1994, many others)

Wavelets & Wavelet Regression

- Data space model: $\boldsymbol{y} = f(\boldsymbol{t}) + \boldsymbol{e}$
  - $\boldsymbol{t}$ = equally spaced grid, length $n=2^J$, on (0,1)
  - $\boldsymbol{e} \sim \text{MVN}(\mathbf{0},\sigma^2 \text{ I})$
- In wavelet space: $\boldsymbol{d} = \boldsymbol{W}\boldsymbol{y} = \boldsymbol{\theta} + \boldsymbol{e}^*$
- $\boldsymbol{d}$ = 'empirical' wavelet coefficients
  $\boldsymbol{\theta}$ = 'true' wavelet coefficients
- By orthogonality, $\boldsymbol{e}^* \sim \text{MVN}(\mathbf{0},\sigma^2 \text{ I})$

Wavelets & Wavelet Regression

- Wavelet Regression procedure:
  - (1) Transform to wavelet space
  - (2) Threshold/shrink coefficients $\rightarrow 0$
  - (3) Transform back to data space (IDWT)
- Thresholding/shrinkage filters out noise, performs regularization.
- Shrinkage can be done by placing a certain prior structure on the θ.

Method: Overview

1. Convert data $\boldsymbol{y_{abc}}$ to wavelet space $\boldsymbol{d_{abc}}$
   - Involves 1 DWT for each crypt

2. Fit hierarchical model in wavelet space to obtain
   * Posterior distribution of 'true' wavelet coefficients $\boldsymbol{\theta_a}$ corresponding to $g_a(t)$
   * Variance component estimates to assess relative variability

3. Use IDWT to obtain posterior distributions of $g_a(t)$ for estimation and inference

Method: Model in Wavelet Space

- $\mathbf{d}_{abc} = \{ d_{abc}^{j,k} \} = \mathrm{W}\ \mathbf{y}_{abc}$

$$\boxed{\begin{aligned} d_{abc}^{j,k} &\sim iid\ N(\theta_{abc}^{j,k}, \sigma_e^2) \\ \theta_{abc}^{j,k} &\sim iid\ N(\theta_{ab}^{j,k}, \sigma_{1,j}^2) \\ \theta_{ab}^{j,k} &\sim iid\ N(\theta_{a}^{j,k}, \sigma_{2,j}^2) \end{aligned}}$$

  - σ_e^2 assumed known, in practice estimated by $\mathrm{Var}(d_{abc}^{1,k})$
  - IG priors on $\sigma_{1,j}^2$ and $\sigma_{2,j}^2$ when unknown

Method; "Shrinkage" Prior

- Prior on $\theta_a^{j,k}$ $\Rightarrow$ thresholding/shrinkage
- Mixture of point mass at 0 and Normal

$$\theta_a^{j,k} \sim N(0, \gamma_a^{j,k} \tau_j^2)$$

$$\gamma_a^{j,k} \sim \text{Bernoulli}(p_j)$$

- p_j and τ_{j}^2 : smoothing parameters

Method: Estimating $g_a(t)$

- If VC known $\Rightarrow$ closed form posterior

$$(\theta_a^{j,k} \mid \mathbf{d}, \sigma_{1,j}^2, \sigma_{2,j}^2) \sim \text{Normal}(\mu, \sigma^2)$$

$$\mu = \hat{\Theta}_{a,NS}^{j,k} \times h\left(\tau_j^2, p_j, \hat{\Theta}_{a,NS}^{j,k}\right)$$

$$\sigma^2 = \text{Var}\left(\hat{\Theta}_{a,NS}^{j,k}\right) \times h\left(\tau_j^2, p_j, \hat{\Theta}_{a,NS}^{j,k}\right)$$

$\hat{\Theta}_{a,NS}^{j,k} = MLE$ of $\theta_a^{j,k}$ assuming **no shrinkage**

$h(.)$ = "shrinkage" function

- Using inverse wavelet transform, we get entire posterior distribution of $g_a(t)$.

Method: Estimating $\theta_a^{j,k}$

- VC Unknown -- no closed form posterior
- Metropolis-within-Gibbs Sampler used on marginalized model.
- Sampling scheme: Let $\Omega^j = \{\sigma_{1,j}^2, \sigma_{2,j}^2\}$

  and $\mathbf{\Theta}_a^j = \left\{\Theta_a^{j,k}, k = 1, \ldots, K_j\right\}$

  (1) Sample from $\left(\Theta_a^{j,k} \mid \mathbf{d}, \Omega^j\right)$ for each j, k

  (2) Sample from $\left(\Omega^j \mid \mathbf{d}, \mathbf{\Theta}_a^j\right)$ for each j

Method: Estimating Variability

- Relative variability at different hierarchical levels can help guide design of future experiments
- Assessed using the trace: $\mathrm{tr}\{\Sigma_1\}$, $\mathrm{tr}\{\Sigma_2\}$, and $\mathrm{tr}\{\sigma_e^2 I\}$
- Orthogonality $\Rightarrow$ Compute trace in wavelet space

$$V_0 = \mathrm{tr}\{\sigma_e^2 I\} = n\sigma_e^2$$

$$V_1 = \mathrm{tr}\{\Sigma_1\} = \sum_{\mathrm{j}} K_j \sigma_{1,j}^2$$

$$V_2 = \mathrm{tr}\{\Sigma_2\} = \sum_{\mathrm{j}} K_j \sigma_{2,j}^2$$

Relative variability at level i:

$$V_i / (V_0 + V_1 + V_2)$$

Choice of Smoothing Parameters

Shrinkage Function:

$$h(Z, T_j^2, p_j) = \underbrace{\left(\frac{T_j^2}{T_j^2+1}\right)}_{\text{Linear Shrinkage}} \underbrace{\Pr\left\{\gamma_a^{j,k} = 1 \mid \mathbf{d}_a^{j,k}\right\}}_{\text{Nonlinear Shrinkage}}$$

$$\Pr\left(\gamma_a^{j,k} = 1 \mid \mathbf{d}_a^{j,k}\right) = \frac{BF}{BF+1}, \qquad BF = \text{Bayes Factor}$$

$$\underbrace{BF}_{\text{Bayes Factor}} = \underbrace{\left(\frac{p_j}{1-p_j}\right)}_{\text{Prior Odds}} \underbrace{\left(1+T_j^2\right)^{-1} \exp\left\{\frac{Z^2}{2}\left(\frac{T_j^2}{T_j^2+1}\right)\right\}}_{\text{Likelihood Ratio}}$$

$$Z = \hat{\theta}_{a,NS}^{j,k} / \sqrt{\mathrm{Var}(\hat{\theta}_{a,NS}^{j,k})} \qquad T_j^2 = \tau_j^2 / \mathrm{Var}(\hat{\theta}_{a,NS}^{j,k})$$

Choice of Smoothing Parameters

- p_j : Expected proportion of non-negligible wavelet coefficients at resolution j
  - $p_j \downarrow \Rightarrow$ Belief that many coeffs. are essentially noise.
  - Results in more ‘smoothing’ of features at scale j
- Generally -- expect p to decrease in j
  - Higher frequency behavior mostly noise

Choice of Smoothing Parameters

- τ_j^2 : Variation of nonzero wavelet coeffs.

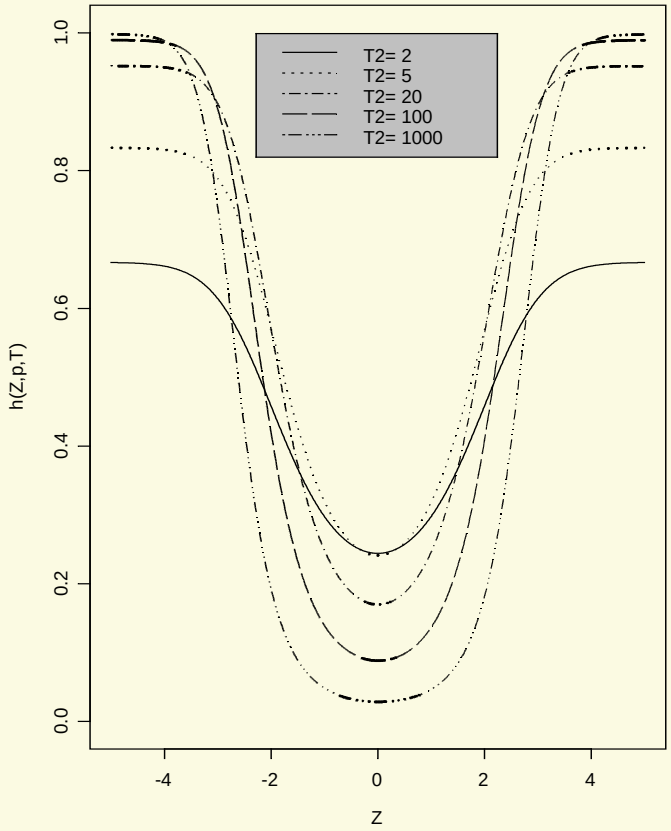
$$T_j^2 = \tau_j^2 / \mathrm{Var}\left(\hat{\Theta}_{a,NS}^{j,k}\right)$$

- Must take care in choosing
  - Too small $\Rightarrow$ too much linear shrinkage
  - Too large $\Rightarrow$ *Lindley's paradox*
    - Too much nonlinear shrinkage for some Z

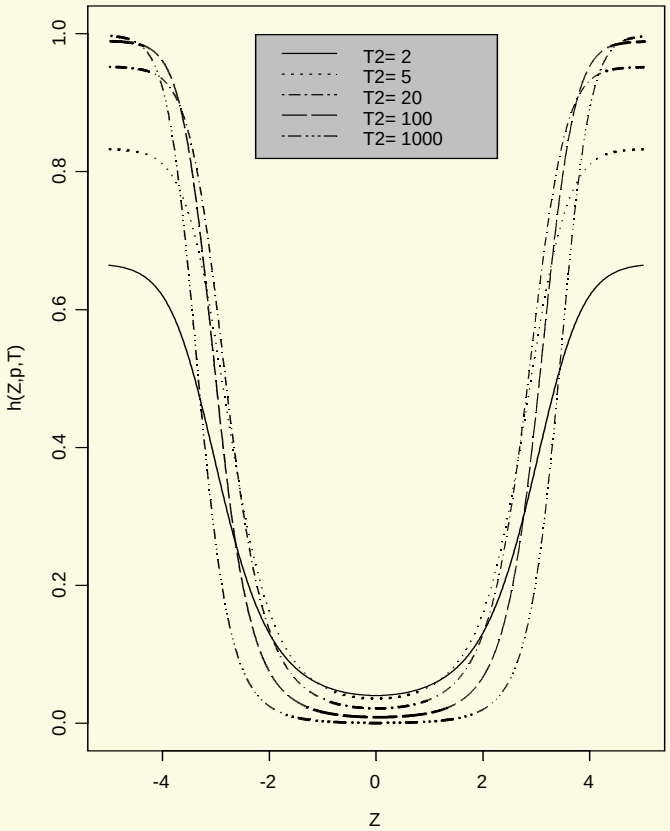
$T_j^2 \in (10,100)$ seem to work quite well.

Choice of Smoothing Parameters

(a) Shrinkage functions for p= 0.5



(b) Shrinkage functions for p= 0.1

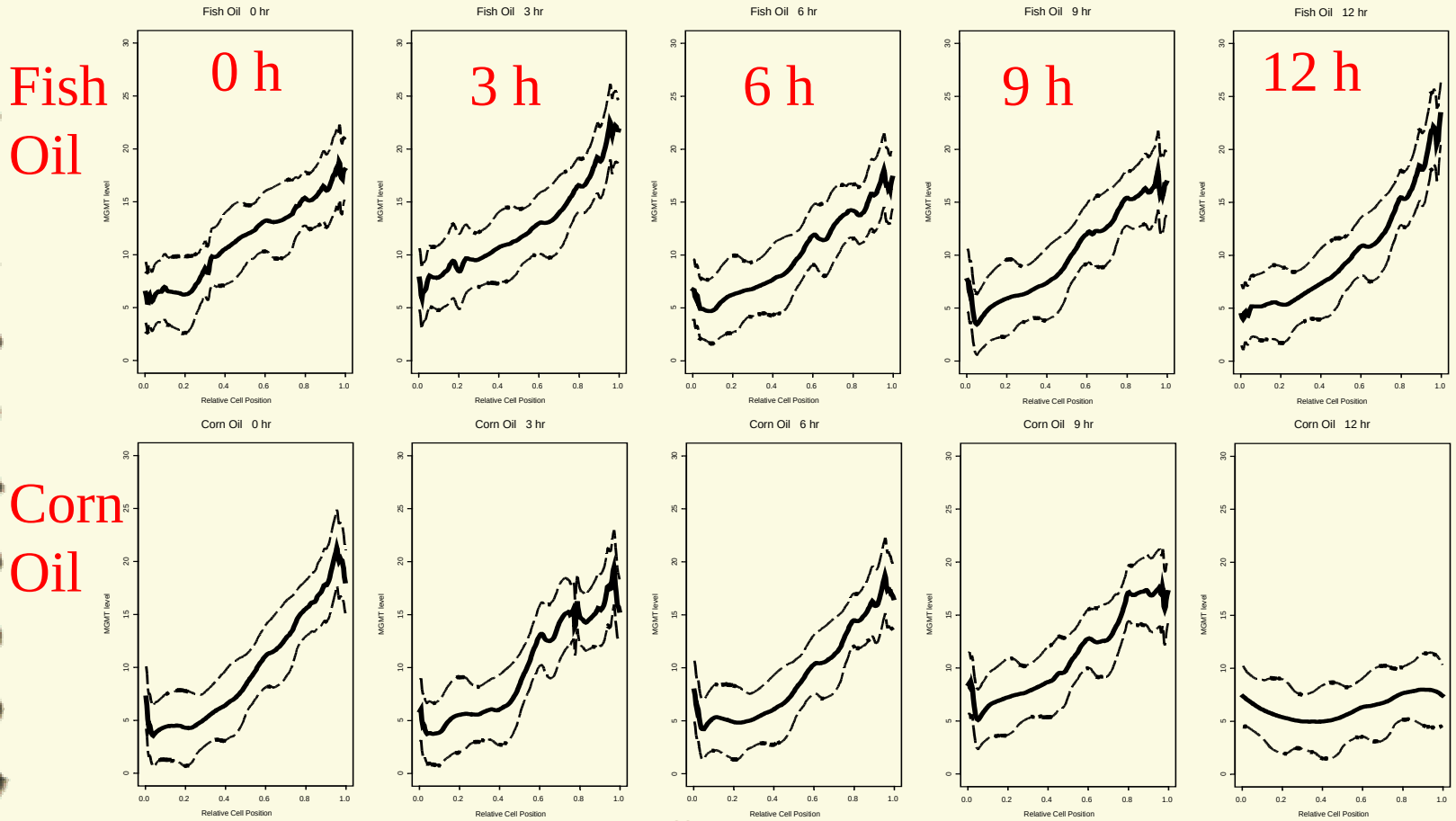


Example: Data

- MGMT expression (DNA repair enzyme)
  - 2 diets (fish/corn) x 5 times (0h,3h,6h,9h,12h)
  - 3 rats per diet x time combination
  - ~ 25 crypts per rat,
  - function sampled at $256=2^8$ equally spaced t
- Daubechies-8 wavelet chosen
  - $\boldsymbol{p}=\{1,.20,.07,.03,.015,.01,0.005\}$, $T_j^2 = 20 \ \forall j$
  - MCMC 5000 iterations; Burn-in=5000

Example: Results

Estimates & 90% posterior bounds by diet/time



Conclusion: Biological Results

- More MGMT at top of crypts
- Diet difference at 12 h near luminal surface ($t \sim 1$)
  - Fish oil-fed rats -- MGMT ↓ from 9 h to 12 h
  - Corn oil-fed rats -- MGMT ↑ from 9h to 12 h
  - Fish MGMT > Corn MGMT
- Crypt-level dominates variability
  - 90.0% at crypt level, 9.6% at rat level, 0.4% residual
  - Need lots of crypts!!
- MGMT operates on cell-by-cell basis

Conclusion: Summary

- Method to fit hierarchical functional data
  - nonparametrically estimate mean functions w/ bounds
  - evaluate relative variability at hierarchical levels
  - perform Bayesian inference
  - can handle 1 or 2 hierarchical levels & unbalanced designs
  - appropriate when functions at base level irregular
- Fitting done in wavelet space
  - Whitening property allows parsimonious modeling of Σ_i
  - Can represent spatially heterogeneous functions well
- Motivated by colon carc., but applicable elsewhere
- Possible extensions of method