Wavelet-Based Modeling of Hierarchical Functions in Colon Carcinogenesis

Jeffrey S. Morris

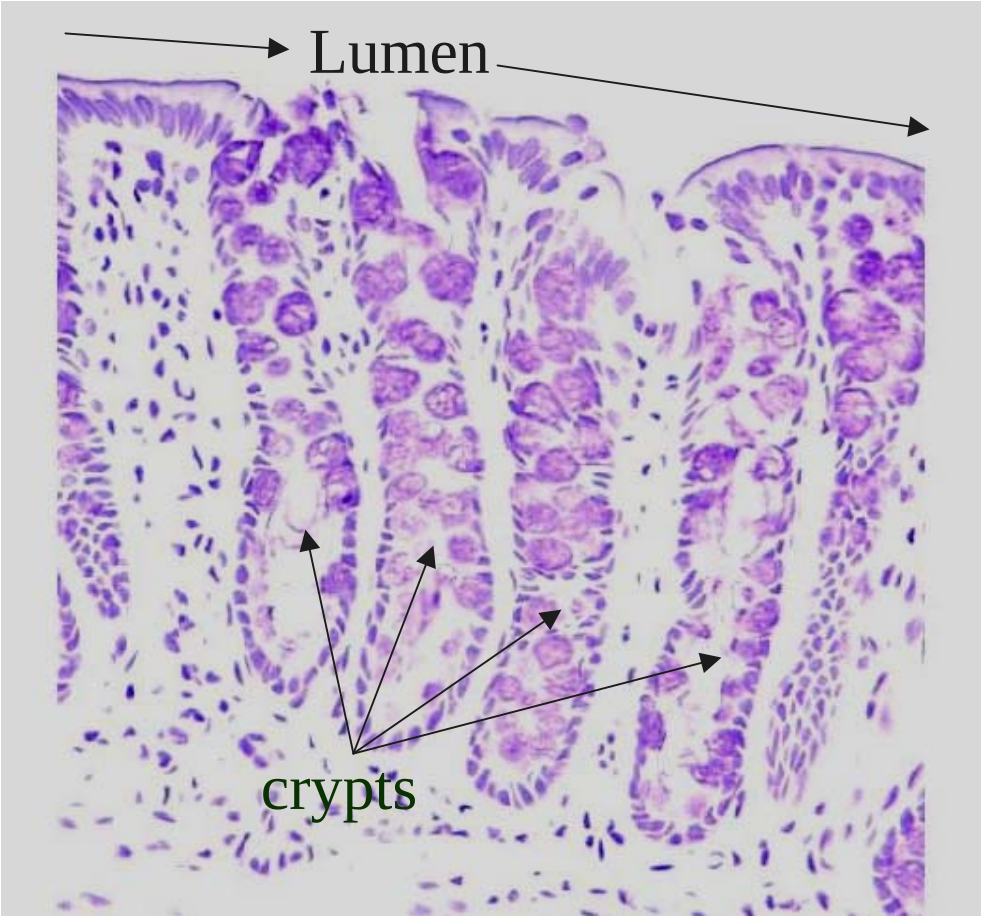
MD Anderson Cancer Center

Joint work with Marina Vannucci, Philip J. Brown, and Raymond J. Carroll,

Outline

- Introduction
  - Scientific Background/Case Study
  - Hierarchical Functional Model
- Our Method
  - Wavelet Space Modeling
  - Model Fitting and Inference
- Case Study
- Conclusions

Architecture of Colon

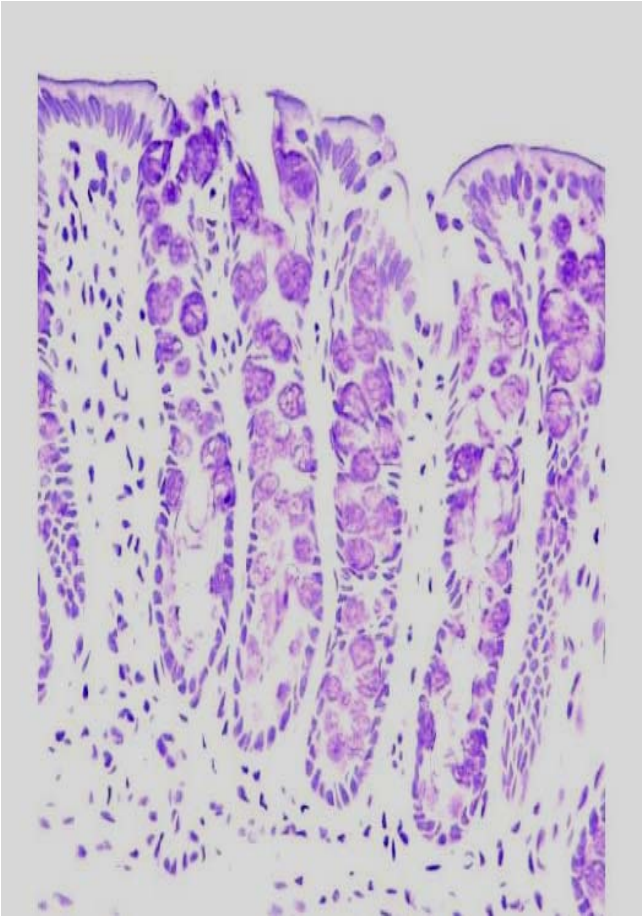


- ***Stem Cells***: Mother cells near crypt base
- Depth in crypt ~ age of cells
- ***Relative Cell Position***

  (Depth w/in crypt):

  $\boldsymbol{t} \in (0,1)$

Colon Carcinogenesis Studies



- Rats fed different diets, exposed to carcinogen, then euthanized.
- DNA adducts, DNA repair, apoptosis quantified as staining intensities along crypt wall
- Data has hierarchical structure
  - Treatment groups - rats - crypts - cells
  - Hierarchical Functional Data

Colon Carcinogenesis Study

- Fish oil diet protective compared with corn oil diet
- Study:
  - 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 on fine grid – interpolated to 256-grid
- Response: DNA Repair Enzyme (MGMT)
- Questions of interest:
  - At what depths within crypts is MGMT expressed?
  - Diet effect on MGMT? Depend on time/depth?
  - Does MGMT vary more from crypt-crypt or rat-rat?

Hierarchical Functional Model

- **Hierarchical Functional Model:**

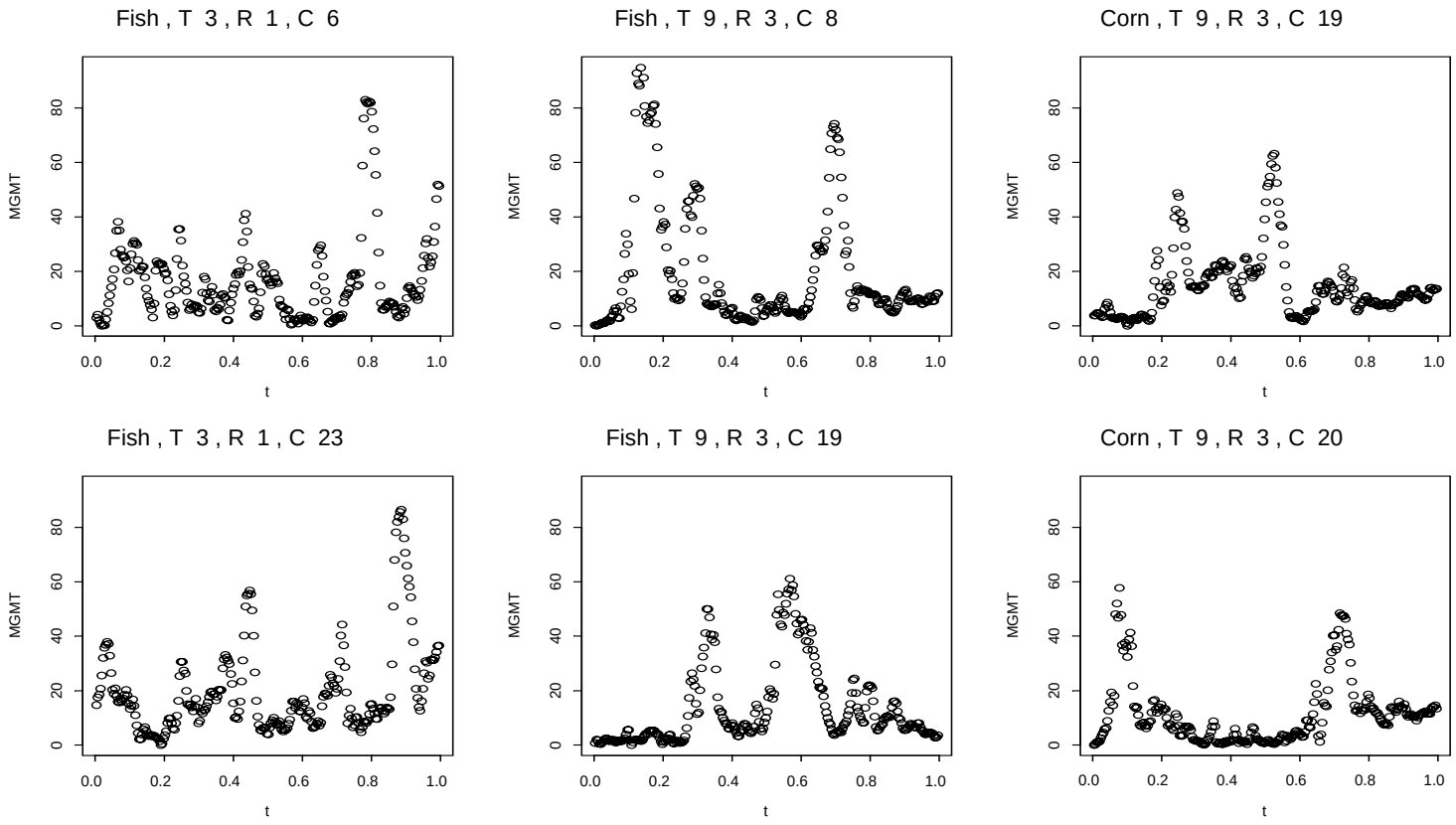
$$\boxed{\begin{aligned} \mathbf{Y}_{abc} &= g_{abc}(t) + \mathbf{e}_{abc}, \\ g_{abc}(t) &= g_{ab}(t) + \eta_{abc}(t) \\ g_{ab}(t) &= g_{a}(t) + \xi_{ab}(t) \end{aligned}}$$

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

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

DNA Repair Data

DNA Repair Enzyme for Selected Crypts



Functional Mixed Models

- Methods for nonparametrically modeling nested functional data:
- Brumback and Rice, 1998
  - Model nested and crossed functional data using splines.
- Wang, 1998; Guo, 2002
  - Functional mixed models using splines
- Splines effective for spiky data like ours?

Introduction to Wavelets

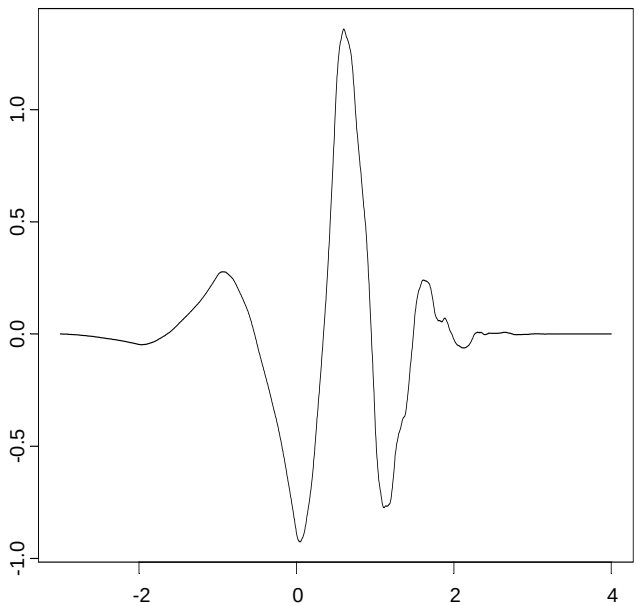
- **Wavelets:** families of orthonormal basis functions

$$g(x) = \sum_{j,k \in \Im} d^{j,k} \psi^{j,k}$$

$$\psi^{j,k} = 2^{-j/2} \psi(2^{-j/2} x - k)$$

$$d^{j,k} = \int g(x) \psi^{j,k}(x) dx$$

Daubechies Basis Function



Wavelet Space Model

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

Discrete Wavelet Transform

$$d_{abc}^{j,k} \sim N(\theta_{abc}^{j,k}, \sigma_e^2)$$

$$\theta_{abc}^{j,k} \sim N(\theta_{ab}^{j,k}, \sigma_{1,j}^2)$$

$$\theta_{ab}^{j,k} \sim N(\theta_{a}^{j,k}, \sigma_{2,j}^2)$$

- IDWT can be used to convert wavelet space quantities to data space

Shrinkage Prior

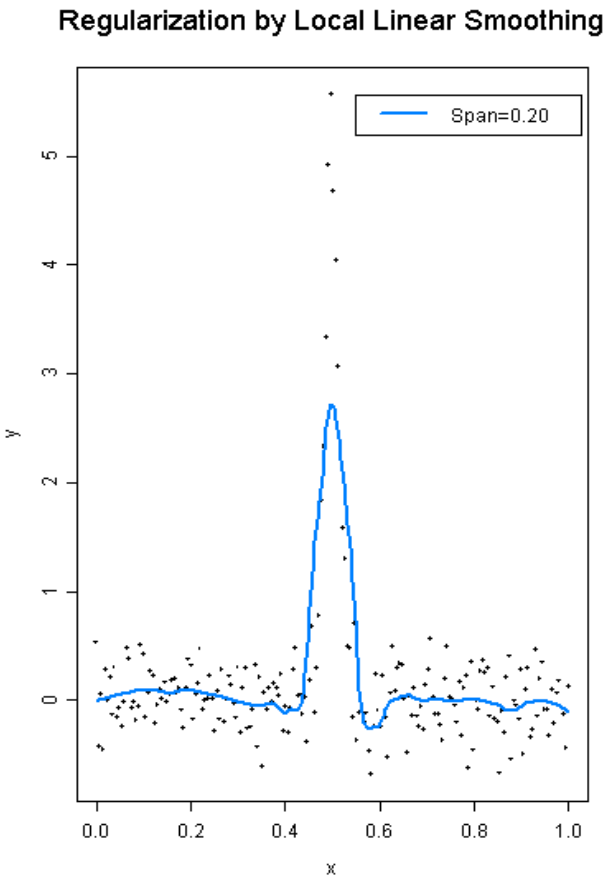
- Prior on $\theta_a^{j,k} \Rightarrow$ Normal-0 mixture

$$\theta_a^{j,k} = \gamma_a^{j,k} N(0, \tau_j^2) + (1 - \gamma_a^{j,k}) \mathrm{I}_0$$

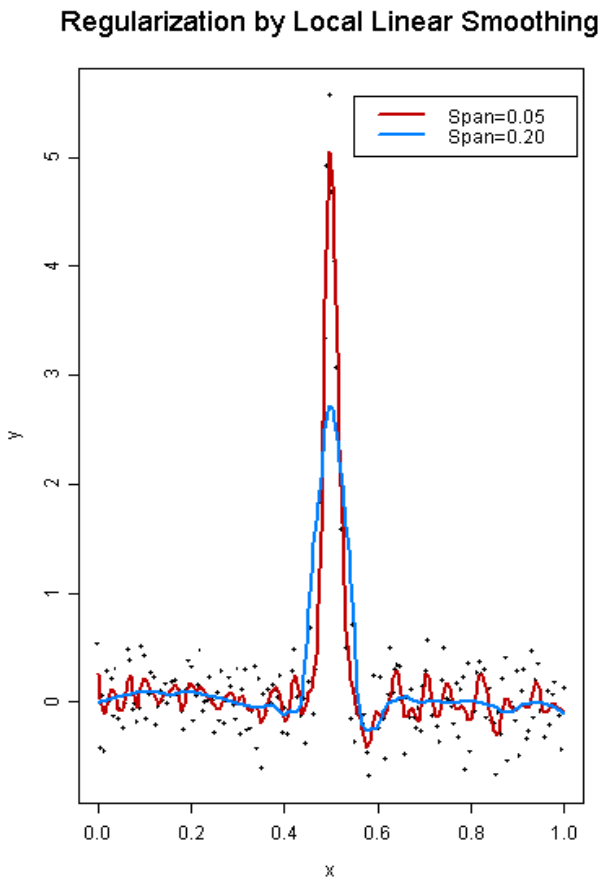
$$\gamma_a^{j,k} \sim \text{Bernoulli}(p_j)$$

- Nonlinear shrinkage – adaptive regularization
- p_j and τ_j^2 - regularization parameters (scale dependent regularization)
- Smaller p_j : More smoothing at scale 2^{j-1}

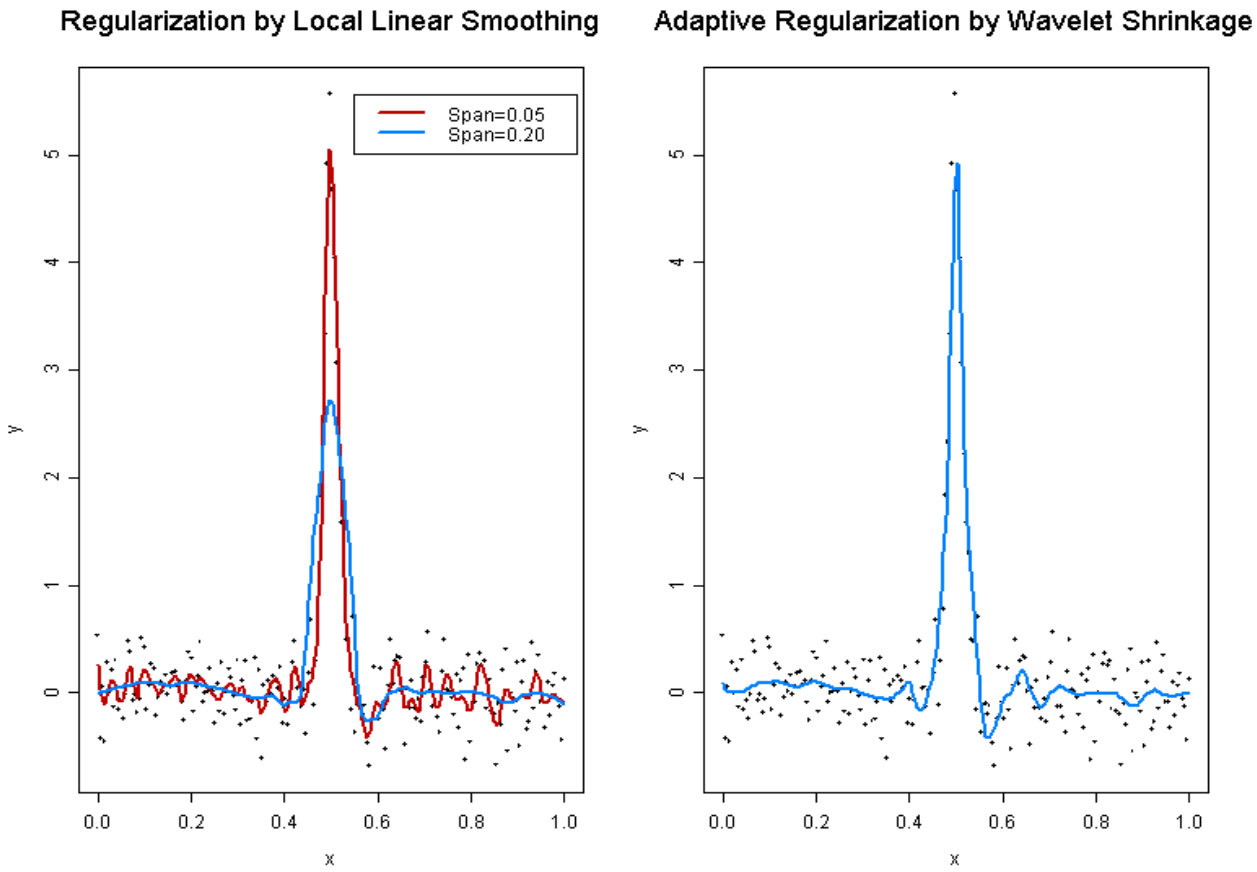
Adaptive Regularization



Adaptive Regularization



Adaptive Regularization



Fitting the Model

- Hierarchical model fit using MCMC
  - Marginalized model used
  - Steps: (Ω^j=VCs at wavelet level j)

1. Sample $\theta_a^{j,k}$ from $(\theta_a^{j,k} \mid \mathbf{d}_a^{j,k}, \Omega^j)$ for each j, k, a
   - Mixture of point mass at 0 and Normal
2. Sample Ω^j from $(\Omega^j \mid \mathbf{d}^{j,k}, \theta_a^{j,k})$ for each j
   - Metropolis step used

- IDWT applied to samples of $\{\theta_a^{j,k}\}$ yield posterior samples of $g_a(t)$.

Fitting the Model

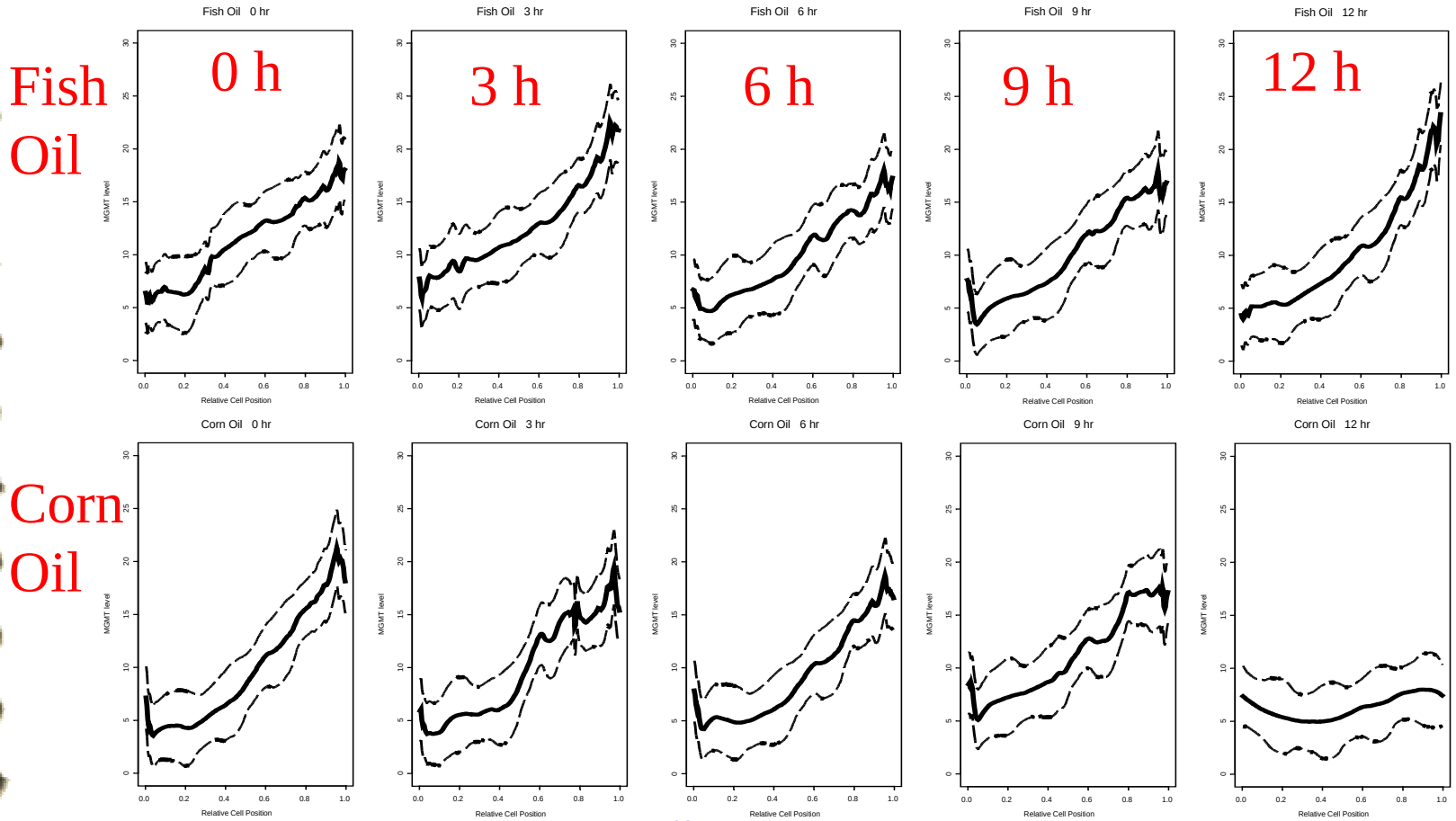
- Relative variability at hierarchical levels, measured by trace of covariance matrices, can be computed directly from wavelet space variance components.
- Can also obtain posterior samples for rat- and crypt-level curves.

$$1.\ \text{Sample from } (\theta_{ab}^{j,k} \mid \theta_{a}^{j,k}, \Omega^{j}, \mathbf{d}_{ab}^{j,k}) \text{ for each } a, b, j, k$$

$$2.\ \text{Sample from } (\theta_{abc}^{j,k} \mid \theta_{ab}^{j,k}, \Omega^{j}, \mathbf{d}_{abc}^{j,k}) \text{ for each } a, b, c, j, k$$

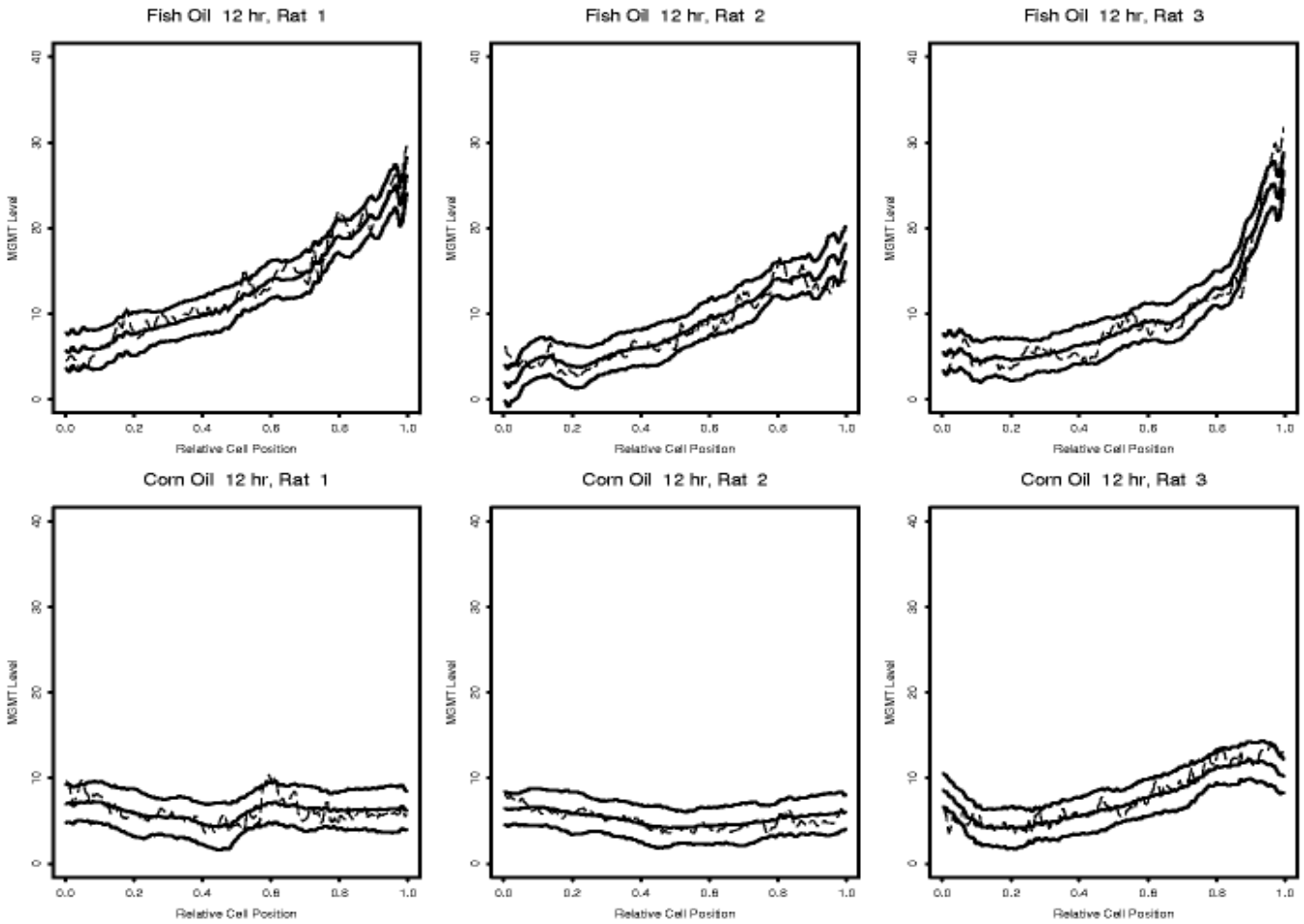
Results: DNA Repair Enzyme

Estimates & 90% posterior bounds by diet/time



Results: DNA Repair Enzyme

Estimates of individual rat profiles for T12 rats



Results: DNA Repair Enzyme

- More repair enzyme at base of crypt
- Possible diet effect at 12 hr time point
- **Relative variability:**
  - 79% between crypts
    20% between rats
    1% within crypts
  - Important to sample lots of crypts per rat

Conclusions

- Bayesian method for hierarchical functional data
  - Nonparametrically estimate curves at mean, individual and subsampling levels using wavelet shrinkage.
  - Covariance parameter estimates
- Posterior samples from MCMC allow inference to be done
- All done with unified, ‘nonparametric’ model that appropriately adjusts for imbalance & correlation.
- Wavelets allow use of simpler covariance structures and work well for irregular functions
- Applicable to other functional data sets

Mixture Prior and Nonlinear Shrinkage

Shrinkage Function:

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

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

$$\underbrace{O^{j,k}}_{\textbf{Posterior Odds}} = \underbrace{\left(\frac{p_j}{1-p_j}\right)}_{\textbf{Prior Odds}} \underbrace{\left(1+T_j^2\right)^{-1} \exp\left\{\frac{(Z^{j,k})^2}{2}\left(\frac{T_j^2}{T_j^2+1}\right)\right\}}_{\textbf{Bayes Factor}}$$

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

Shrinkage Curves