Analysis of DNA Damage and Repair in Colonic Crypts

Jeffrey S. Morris

University of Texas, MD Anderson Cancer Center

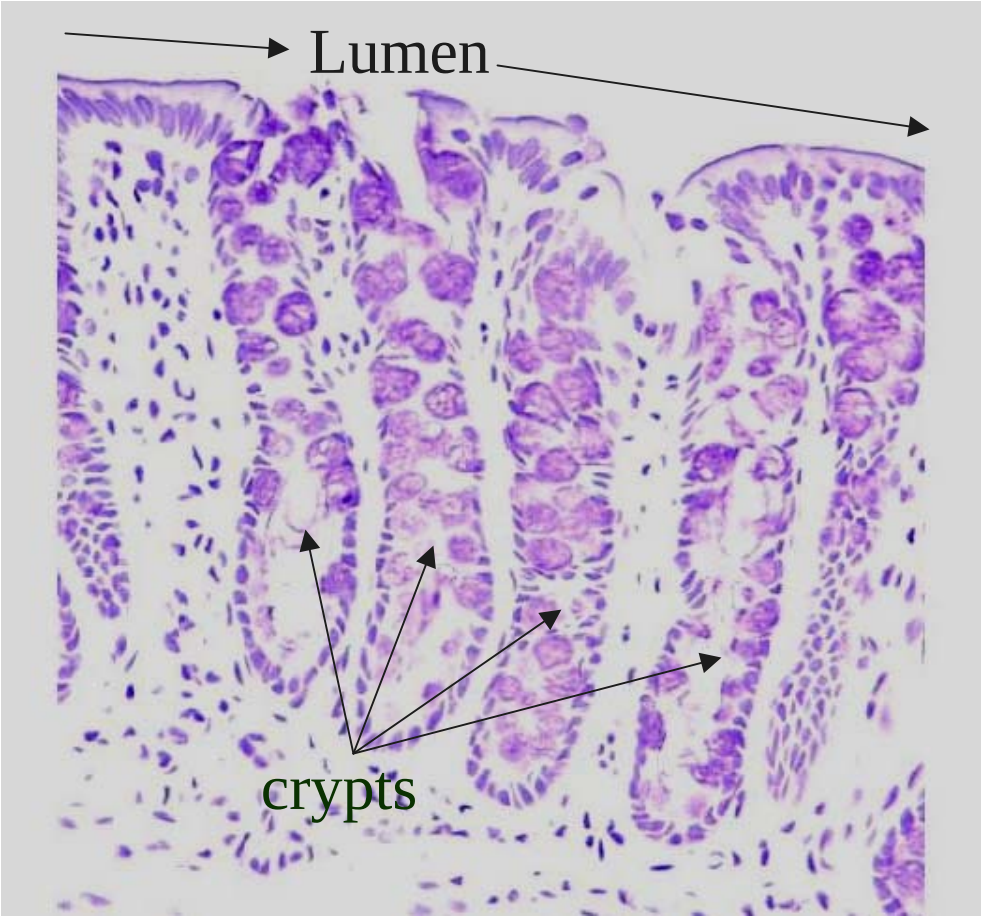
Marina Vannucci and Raymond J. Carroll,

Texas A&M University

Outline

- Introduction
  - Colon Architecture and Carcinogenesis Studies
  - Hierarchical Functional Model
- Wavelet-based Estimation of Hierarchical Functions
- Kernel-based Estimation of Correlations
- Bayesian Modeling of Crypt Cell Architecture
- Conclusion

Architecture of Colon

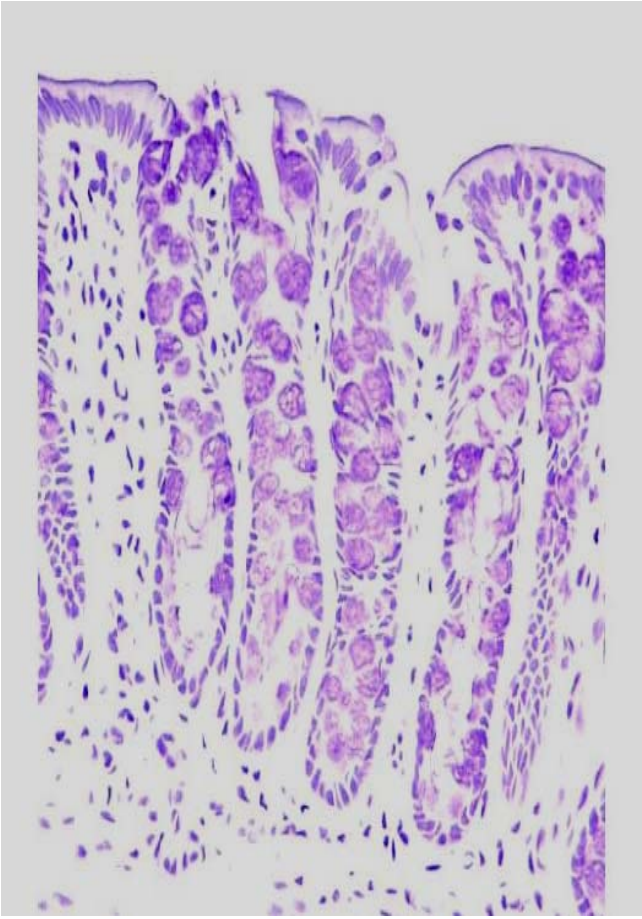


- ***Stem Cells***: Mother cells near crypt base
- Depth in crypt ~ age of cells
- ***Relative Cell Position***

  (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
- Data has hierarchical structure
  - Treatment groups - rats - crypts - cells/pixels
  - Hierarchical Longitudinal Data

Hierarchical Functional Model

- **2-level HF model:**

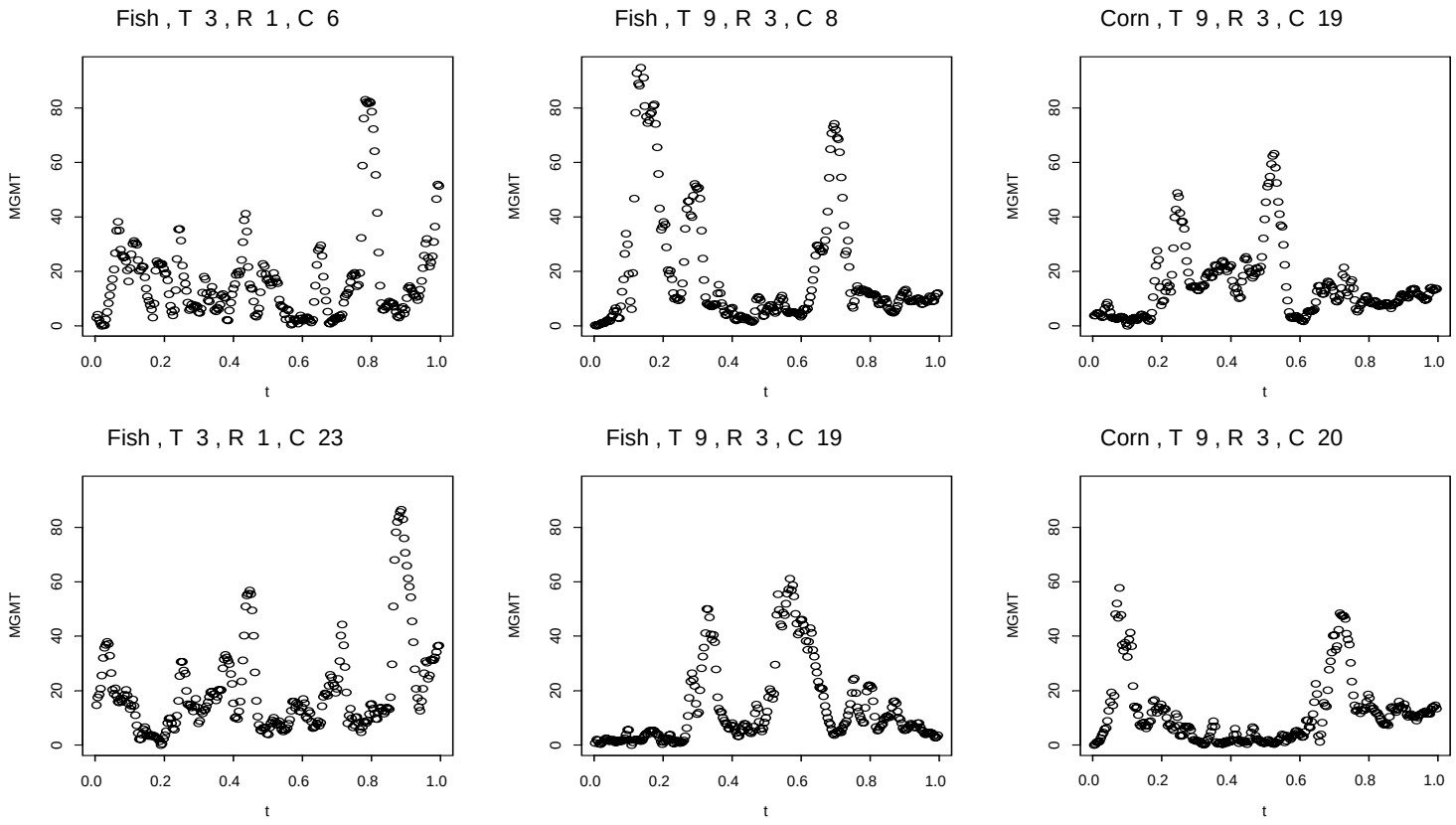
$$\boxed{\begin{aligned} \mathbf{Y}_{abc} &= g_{abc}(t) + \mathbf{e}_{abc}, \\ g_{abc}(t) &= g_{ab}(t) + \eta_{abc}(t) \\ g_{ab}(t) &= g_{a}(t) + \xi_{ab}(t) \end{aligned}}$$

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 $\Sigma_2(t_1, t_2)$.

DNA Repair Data

DNA Repair Enzyme for Selected Crypts



Wavelet Space Model

- **Wavelets:** families of orthonormal basis functions
- $\mathbf{d}_{abc} = \{d_{abc}^{j,k}\} = \mathrm{W}\,\mathbf{y}_{abc}$

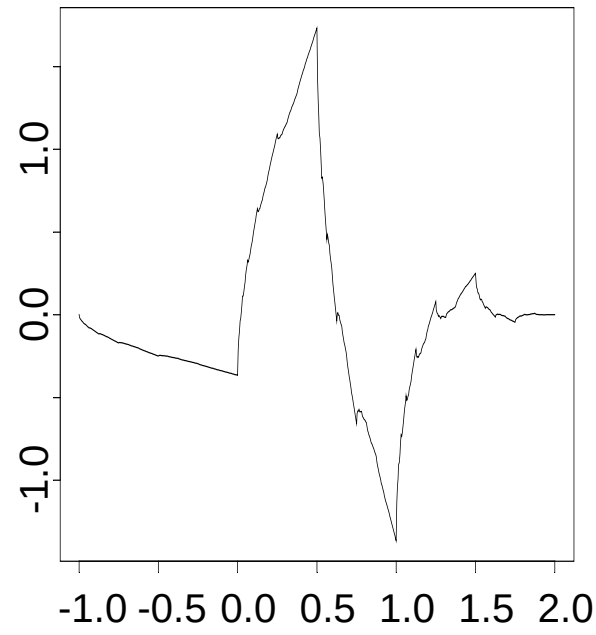
Discrete Wavelet Transform

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

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

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

Daubechies Basis Function



"Shrinkage" Prior

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

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

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

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

Model Fitting

- Hierarchical model fit using MCMC
  - Marginalization done to get stationary distribution of treatment-level coefficients.
  - IDWT done on set of coefficient estimates to obtain posterior distributions of $g_a(t)$.
- Relative variability at hierarchical levels, measured by trace/determinant, given directly by wavelet space results.
- Possible to also obtain individual and subsample-level curve predictors.

Results: DNA Repair

Estimates & 90% posterior bounds by diet/time

	0 h	3 h	6 h	9 h	12 h
Fish Oil					
Corn Oil					

Summary

- Method to fit hierarchical longitudinal data
- Nonparametrically estimate mean profiles for:
  - 1. Treatments
  - 2. Individuals
  - 3. Subsampling units
- Do Bayesian inference
- Handles spatially hetero. data, unbalanced designs
- Estimate of relative variability at hier. levels
- Paper under review -- tech. report available.

Estimating Correlations

- Assess interrelationships between 2 responses as function of cell position
  - e.g. DNA adducts and DNA repair
- Modeled as correlation: $\rho(t_1,t_2)$

  $\rho(t)$

Model

Response 1, e.g.
DNA Adducts

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

Response 2, e.g.
DNA Repair

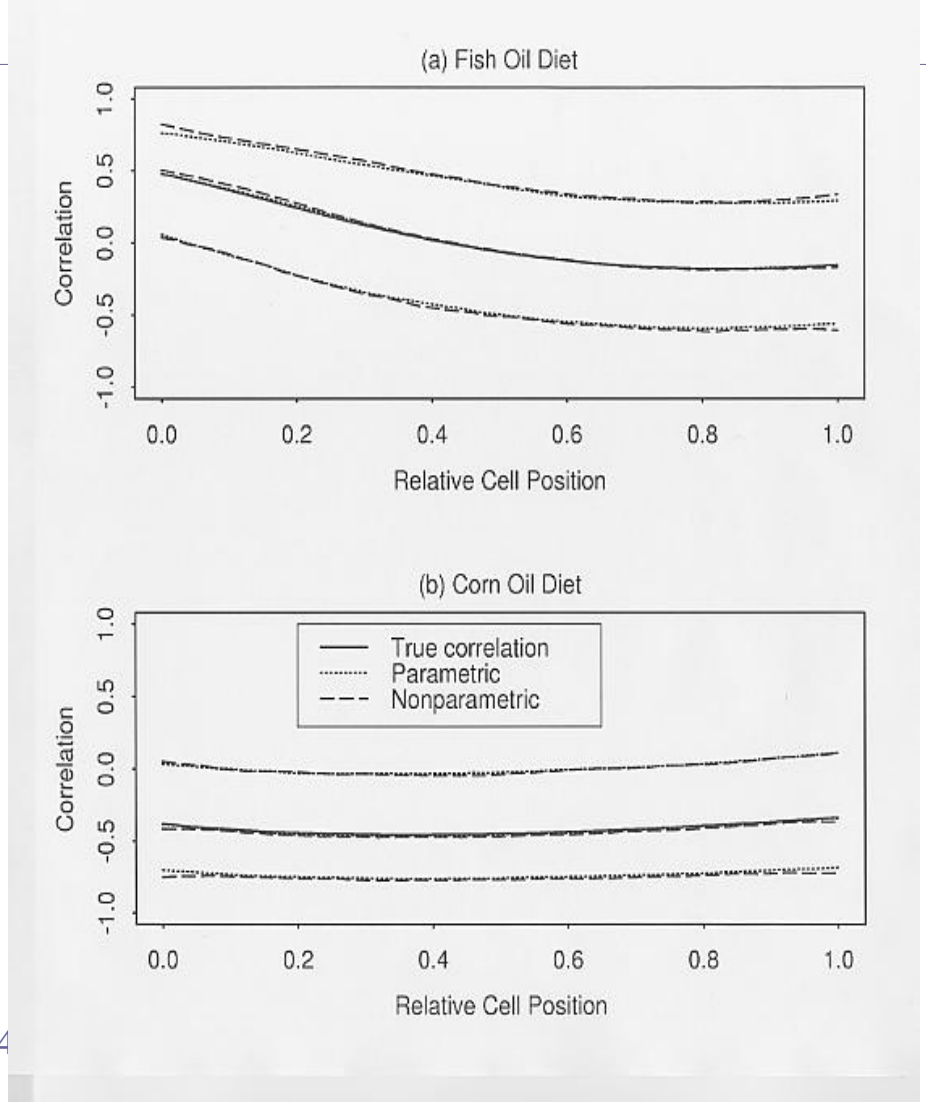
$$\mathbf{Z}_{abc} = f_{abc}(t) + \mathbf{u}_{abc},$$
$$f_{abc}(t) = f_{ab}(t) + \theta_{abc}(t)$$
$$f_{ab}(t) = f_{a}(t) + \psi_{ab}(t)$$

$$\rho(t_1, t_2) = \mathrm{Corr}\{\xi_{ab}(t_1), \psi_{ab}(t_2)\}$$

Nonparametric Estimation

- Nonparametric estimation method:
  - Estimate crypt-level functions $g_{abc}(t)$, $f_{abc}(t)$ on grid
  - Fit mixed model at each grid point
  - Correlation estimated from variance components
- Undersmooth at crypt level
- Virtually no loss of efficiency from nonparametric estimation for our data
  - Asymptotics: Same dominating term as estimator with $g_{abc}(t)$, $f_{abc}(t)$ known

Parametric vs. Nonparametric Methods



- Simulation study done to compare parametric and noparametric methods
- Study confirmed efficiency of nonparametric method relative to parametric method under certain conditions

Summary

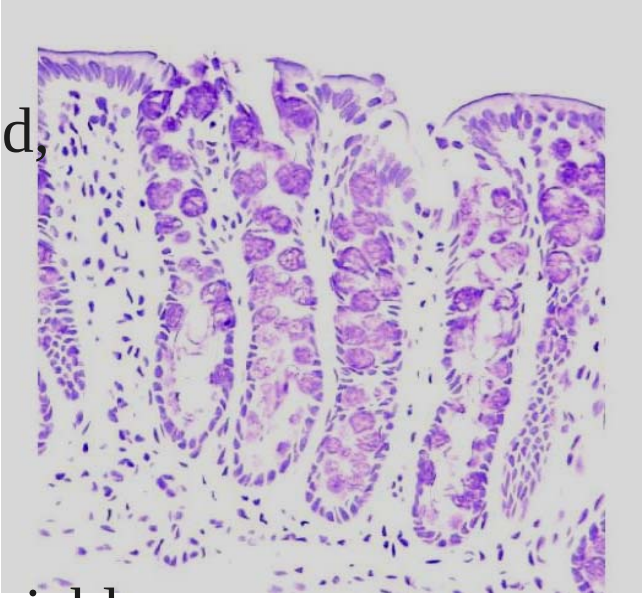
- Method to assess linear relationships between responses that depend on longitudinal variable
- Example where little cost in doing nonparametric estimation

- For details -- see paper in 11/2001 *JASA*

Modeling Crypt Cell Architecture

- When measurements taken ‘by cell’,

$$\mathbf{t} = \frac{\text{distance from each cell's midpoint to crypt base}}{\text{length of crypt}}$$



  - Relative cell positions not observed, since tedious to measure
  - SURROGATE OBSERVED:

$$\mathbf{u} = \frac{\#\text{ of cells below each cell}}{\#\text{ of cells in crypt}}$$

- Results in missing longitudinal variable

Modeling Crypt Cell Architecture

- Model for relative cell positions **t**
  - **t** ~ o.s. of Beta, or mixture of Betas
- Given validation data set (**t**,**u**):
  - Can efficiently generate missing **t** using EM or within MCMC step
  - Characterize crypt architecture in colonic regions
- *Biostatistics* paper -- to appear

Conclusion

- Cell-based colon carcinogenesis studies
  - Hierarchical Longitudinal data
  - Rich in information -- challenging to extract
- New methods developed
  - Wavelet regression methods for long. data
  - Kernel methods for longitudinal correlations
  - Method for missing data with order constraints
- Email: `jeffmo@mdanderson.org`