Hierarchical Functional Data With Correlated Functions

Raymond J. Carroll

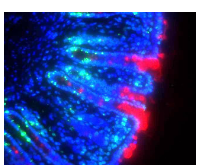
Department of Statistics

Center for Statistical Bioinformatics

Institute for Applied Mathematics and Computational Science

Texas A&M University

http://stat.tamu.edu/~carroll



STATISTICS
TEXAS A&M UNIVERSITY

BCS History

- 2006 season: Florida win BCS beating Ohio State, 41-14

- 2006: Carroll gives Challis Lectures, University of Florida

BCS History

- 2007 season: Florida goes 9-4
- 2007: Carroll does not visit the University of Florida

BCS History

- 2008 season: Florida win BCS beating Oklahoma, 24-14
- 2008: Carroll invited to speak at Winter Workshop, University of Florida

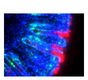
BCS History

- 2009 season: Florida -----???

- 2009: Carroll speaks at the University of Florida Winter Workshop

Outline

- **Problem**: Hierarchical functional data where the functions at the deepest level of the hierarchy are correlated
  - Functions might be spatially correlated
- Biological background and motivating example
- Fixed effects methods
- Random Effects methods

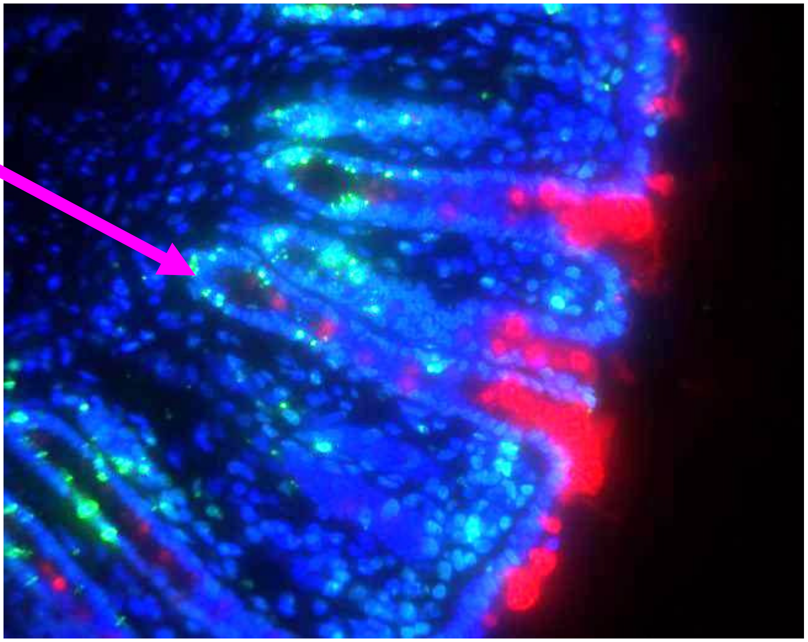


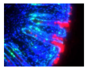
Basic Background

- **Apoptosis**: Programmed cell death
- **Cell Proliferation**: Effectively the opposite
- **p27**: Differences in this marker are thought to stimulate and be predictive of apoptosis and cell proliferation
- **Our experiment**: understand some of the structure of p27 in the colon when animals are exposed to a carcinogen

Data Collection

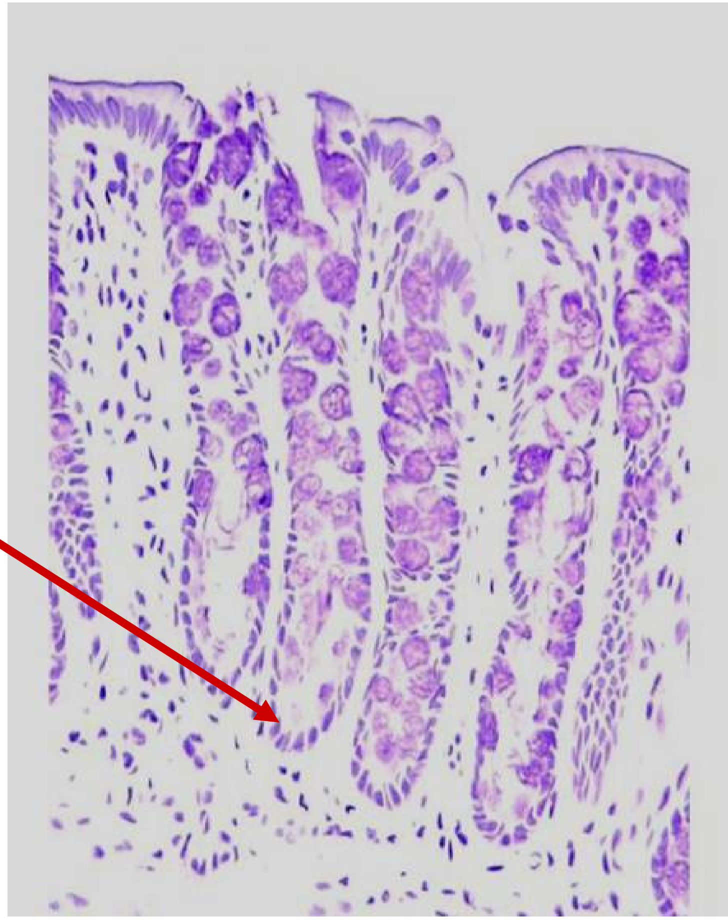
- Structure of Colon
- Note the finger-like projections
- These are **colonic crypts**
- We measure expression of cells within colonic crypts





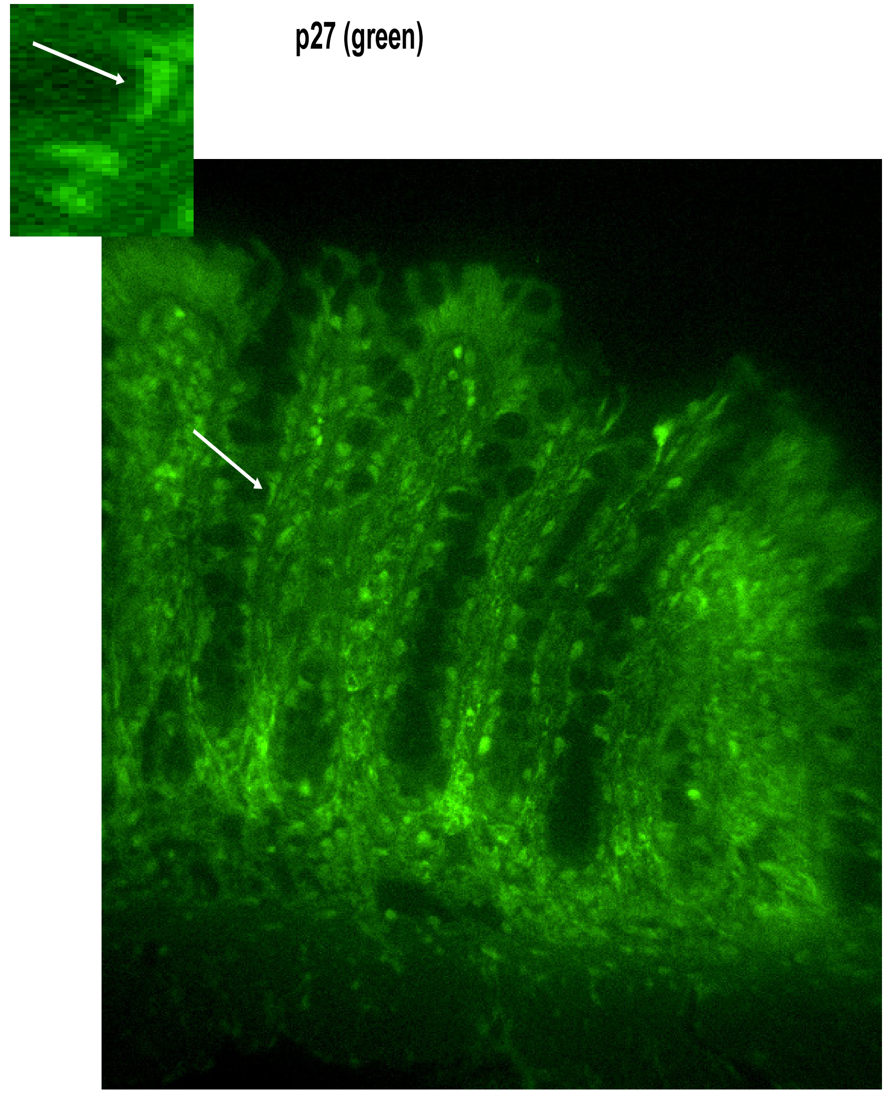
Another View

- Structure of Colon
- Note the finger-like projections
- These are **colonic crypts**
- We measure expression of cells within colonic crypts



Another View

- **p27 expression**: Measured by staining techniques
- Brighter intensity = higher expression
- Done on a cell by cell basis within selected colonic crypts
- Very time intensive



p27 (green)

Spatial Layout of Crypts

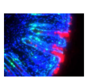
Top View of the colon.

White dots are crypts

Sampling is done in a very small part of the colon

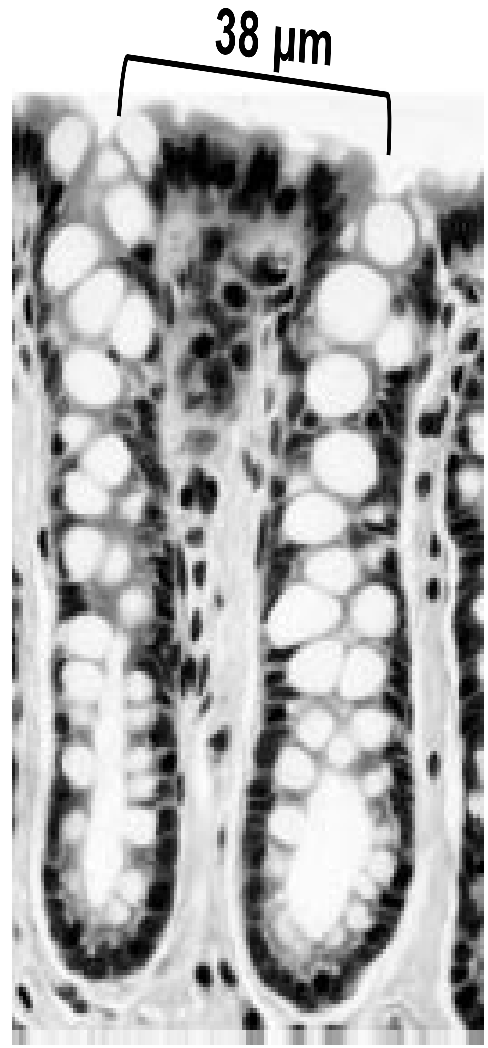
Data Collection

- **Animals sacrificed at 4 times**: 0 = control, 12hr, 24hr and 48hr after exposure
- **Rats**: 12 at each time period, split into 4 diets
- **Crypts**: 20 are selected
- **Cells**: all cells collected, about 30 per crypt
- **p27**: measured on each cell, with logarithmic transformation



Nominal Cell Position

- **X = nominal cell position**
- **Differentiated cells: at top, X = 1.0**
- **Proliferating cells: in middle, X=0.5**
- **Stem cells: at bottom, X=0**



Standard Model

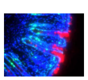
- **Hierarchical structure**: cells within crypts within rats within times

$$Y_{trc}(x) = \mu_t(x) + Z_{tr}(x) + Q_{trc}(x) + \varepsilon_{trc}(x)$$

$$\mu_t(x) + Z_{tr}(x) = \text{rat-level function}$$

$$Q_{trc}(x) = \text{crypt-level functions,}$$

typically assumed independent



Standard Model

- **<u>Hierarchical structure</u>**: cell locations within crypts within rats within times/diets

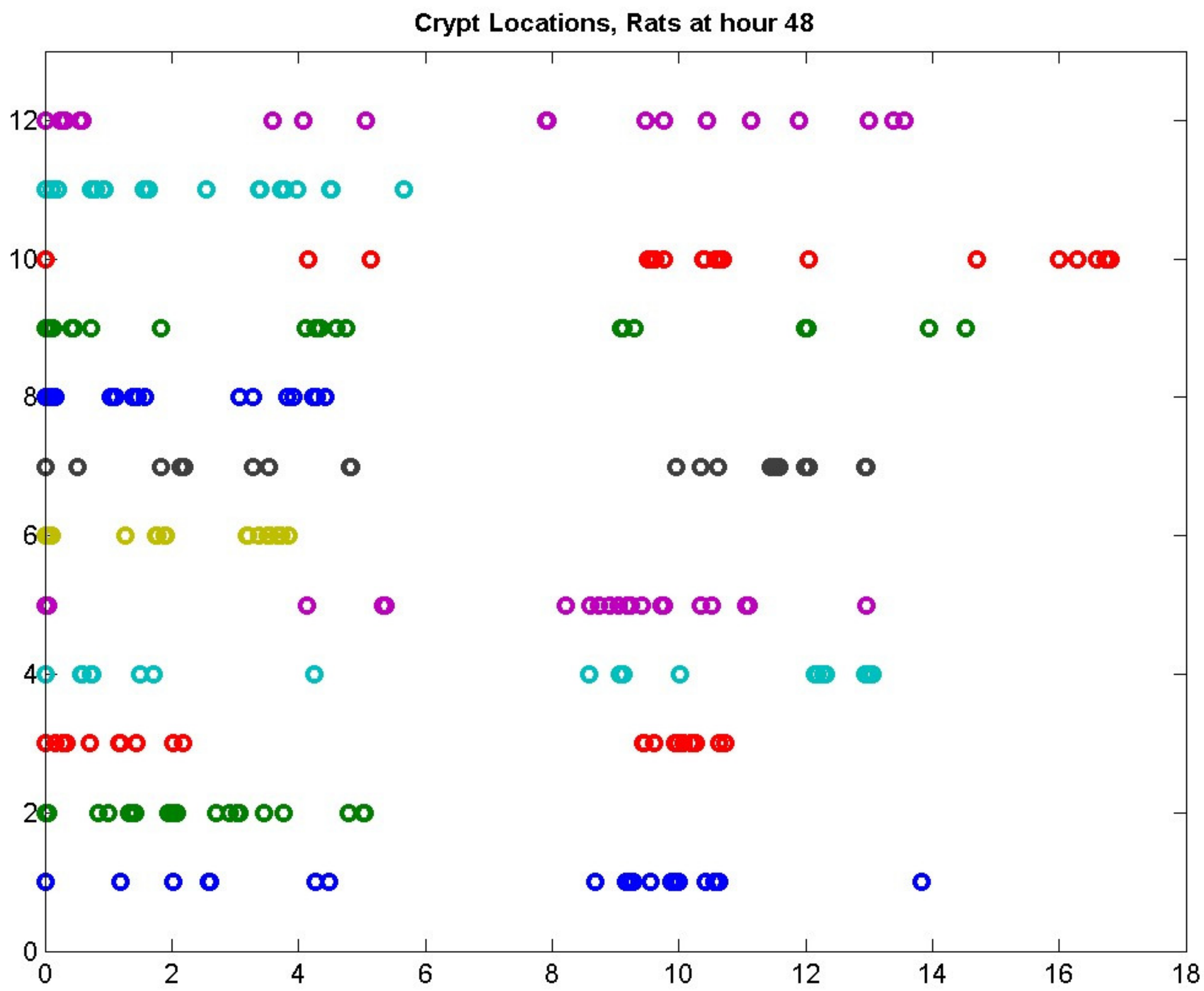
- In our experiment, the residuals from fits at the crypt level are essentially white noise

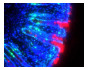
- However, we also measured the **<u>location</u>** of the colonic crypts

Crypt Distances to a nominal zero

<u>Scale</u>: 1000's of microns

<u>Our interest</u>: relationships at between 25-200 microns



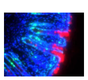


Standard Model

- **Hypothesis**: it is biologically plausible that the nearer the crypts to one another, the greater the relationship of overall p27 expression.

- **Expectation**: The effect of the carcinogen might well alter the relationship over diet

- **Technically**: **What is different** is that this is functional data where the functions are themselves correlated



Fixed Effect Methods

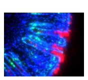
- **Fixed Effects**: Treat the rat-level functions as fixed effects
- **Residualize**: to get at the crypt level structure

$$Y_{trc}(x) = \mu_t(x) + Z_{tr}(x) + Q_{trc}(x) + \varepsilon_{trc}(x)$$

$$\mu_t(x) + Z_{tr}(x) = \text{rat-level function}$$

$$Q_{trc}(x) = \text{crypt-level functions,}$$

typically assumed independent



STATISTICS
TEXAS A&M UNIVERSITY

Fixed Effect Methods

- **<u>Nonparametrically</u>**: (with Yehua Li and Naisyin Wang) We developed kernel-based methods

- These methods assume that there are lots of data to estimate each rat-level function

- In our case, we have 600 observations per rat

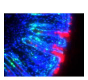
Yehua Li as a student

Naisyin Wang, Marcia Ory and Raymond Carroll in Taiwan, January 1, 2008

Nonparametric Fits

- **Define:** $\boxed{V(x_1,x_2,\Delta)}$ = covariance between crypt-level functions that are Δ apart, one at cell depth x_1 and the other at cell depth x_2.
  - Assumed not to depend on the rat, of course
- Often convenient to assume separable covariance structure as well

$$\boxed{V(x_1,x_2,\Delta)=G(x_1,x_2)\,\rho(\Delta)}$$



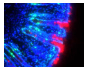
Nonparametric Fits

- Define the rat-level deviations at cell depth x and crypt spatial location δ for crypt c as

$$\boxed{\mathbf{R}_{\mathbf{tr}}(\mathbf{x},\delta)=\mathbf{Q}_{\mathbf{trc}}(\mathbf{x})+\varepsilon_{\mathbf{trc}}(\mathbf{x})}$$

- Then when $\boxed{|\delta_1-\delta_2|=\Delta}$ our function is just

$$\boxed{\mathbf{V}(\mathbf{x}_1,\mathbf{x}_2,\Delta)=\mathbf{E}\left\{\mathbf{R}_{\mathbf{tr}}(\mathbf{x}_1,\delta_1)\mathbf{R}_{\mathbf{tr}}(\mathbf{x}_2,\delta_2)\,\middle|\,|\delta_1-\delta_2|=\Delta\right\}}$$



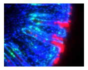
Nonparametric Fits

- Note what we want:

$$\boxed{E\left\{R_{tr}(x_1,\delta_1)R_{tr}(x_2,\delta_2) \;\middle|\; |\delta_1-\delta_2|=\Delta\right\}}$$

- This target is just **a regression function** on the distances among crypts within a subject, given cells at x_1 for one crypt and at x_2 for the other crypt.

$$\boxed{V(x_1,x_2,\Delta)=G(x_1,x_2)\;\rho(\Delta)}$$

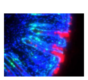


Nonparametric Fits

- Note what we want:

$$\boxed{E\left\{R_{tr}(x_1,\delta_1)R_{tr}(x_2,\delta_2) \mid |\delta_1-\delta_2|=\Delta\right\}}$$

- **Nonparametric methods** (kernels for theory, splines, etc.) are then simple to construct
- **For kernels**, one takes all crypts that are Δ plus or minus a target bandwidth apart
- **Crossvalidation** to estimate the bandwidth



Nonparametric Fits

- **Discrete Version**: Pretend Δ, x_1 and x_2 take on a small discrete set of values (we actually use a kernel-version of this idea)
- Form the sample covariance matrix per rat at Δ, x_1 and x_2 , then average across rats.
- Call this estimate

$$\boxed{\hat{V}(x_1, x_2, \Delta)}$$

Nonparametric Fits

- **Separability**: Now use the separability to get a rough estimate of the correlation surface.

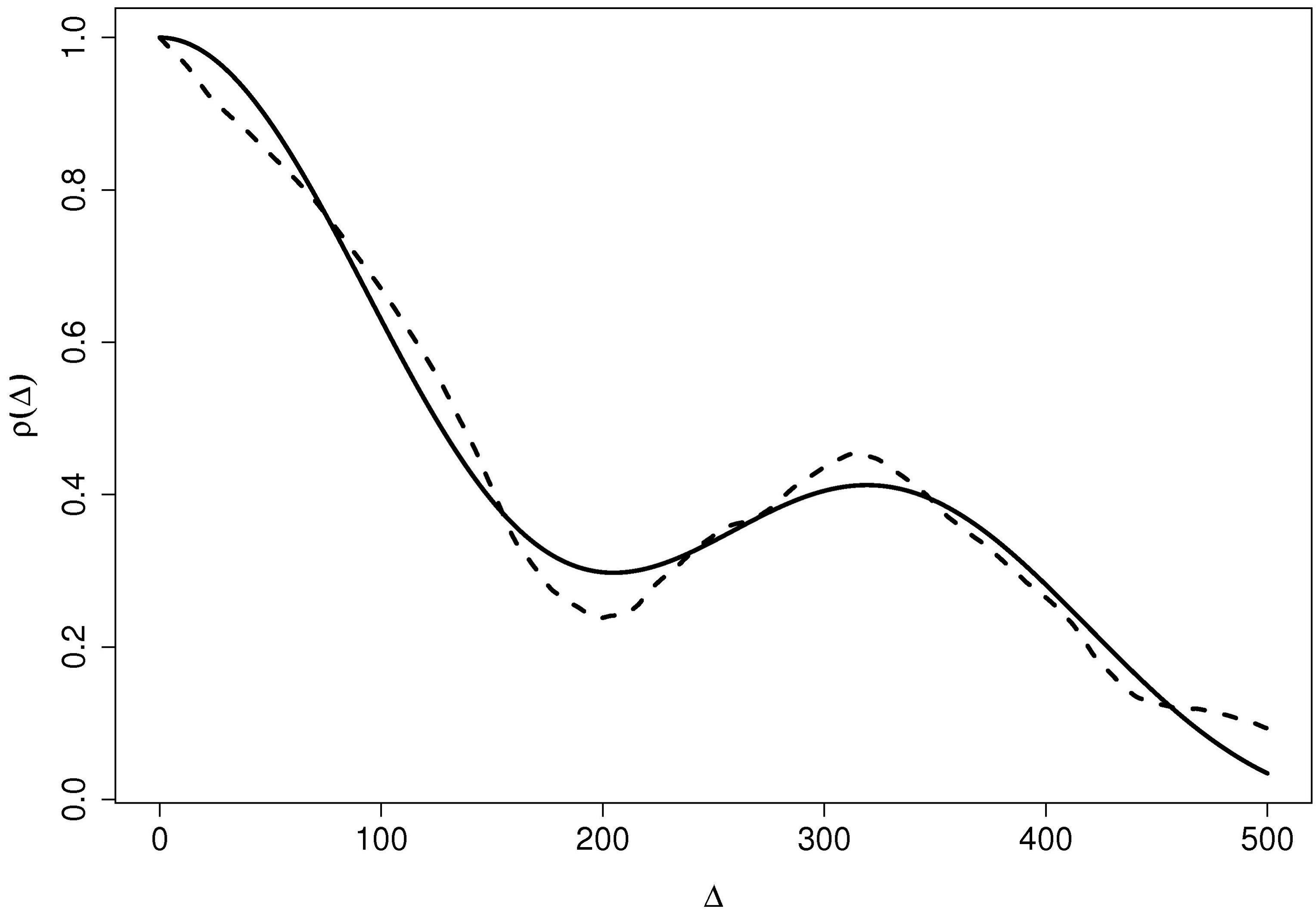
$$V(x_1,x_2,\Delta) = G(x_1,x_2)\ \rho(\Delta)$$

$$\tilde{\rho}(\Delta) = \frac{\sum_{x_1,x_2} \hat{V}(x_1,x_2,\Delta)}{\sum_{x_1,x_2} \hat{V}(x_1,x_2,0)}$$

Nonparametric Fits

- The estimate $\boxed{\tilde{\rho}(\Delta)}$ is not a proper correlation function
- We fixed it up using a trick due to Peter Hall (1994, Annals), thus forming $\boxed{\hat{\rho}(\Delta)}$, a real correlation function
- **<u>Basic idea</u>** is to do a Fourier transform, force it to be non-negative, then invert
- Actually improves the look of the correlation function and lowers MSE
- Asymptotic theory worked out

Nonparametric Fits, 24 hours



Fixed Effect Methods

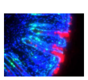
- **<u>Some conclusions</u>**:
  - Up to 100 microns, the estimate correlations are all above 0.4
  - The estimated correlation is non-monotone, quite odd
  - We have generated data with a non-monotone shape in the correlation function, and the method captures it

Fixed Effect Methods

- **Many methods**: There are also many parametric ways to get at the crypt-level structure after residualizing

$$\boxed{Y_{trc}(x) - \mu_t(x) - Z_{tr}(x) = Q_{trc}(x) + \varepsilon_{trc}(x)}$$

- We have done more or less clever things such as spline structure on the crypt functions with separable Matern correlations of the coefficients

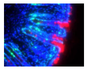


Fixed Effect Methods

- **Scaling**: The operative feature though of fixed effects methods is that they **require enough data per rat to estimate the marginal rat-level functions**

- This works for our example, maybe not for others

- Plus we lose the "borrow strength" aspects of hierarchical models



Random Effect Methods

- **Random Effects**: We have developed a variety of random effect methods that deal with the entire structure of the data

- One method is completely Bayesian (with Veera Baladandayuthapani and Bani Mallick)

- All functions are treated as regression splines, with fixed or random coefficients

Bani Mallick

Veera Baladandayuthapani as a student

Bayesian Model

- **Crypt-Level**: A regression spline, with few knots, in a **parametric mixed-model formulation**

$$Q_{trc}(x) = C(x)\beta_{trc}$$

$$C(x) = \text{spline basis functions}$$

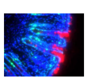
$$\mathrm{cov}(\beta_{trc}) = \Sigma_S$$

Bayesian Model

- **Crypt-Level**: regression spline, few knots
- Separable covariance structure with a parametric (Matern) correlation structure

$$\mathbf{Q}_{trc}(\mathbf{x}) = \mathbf{C}(\mathbf{x})\beta_{trc}$$

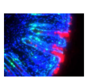
$$\mathbf{cov}(\beta_{tri}, \beta_{trj}) = \begin{pmatrix} \mathbf{1} & \rho(\Delta_{ij}) \\ \rho(\Delta_{ij}) & \mathbf{1} \end{pmatrix} \otimes \Sigma_S$$



Bayesian Model

- **Correlation**: The correlation is directly interpretable and at same cell positions, identical

$$\boxed{\begin{aligned} &\mathbf{Q}_{\mathbf{trc}}(\mathbf{x}) = \mathbf{C}(\mathbf{x})\beta_{\mathbf{trc}} \\[2em] &\mathbf{corr}\{\mathbf{C}(\mathbf{x})\beta_{\mathbf{tri}}, \mathbf{C}(\mathbf{x})\beta_{\mathbf{trj}}\} = \rho(\Delta) \end{aligned}}$$



Bayesian Model

- **Correlation**: However, the correlation is not the same across arbitrary cell locations

$$Q_{trc}(\mathbf{x}) = C(\mathbf{x})\beta_{trc}$$

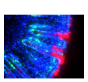
$$\mathrm{corr}\{C(\mathbf{x})\beta_{tri}, C(\mathbf{t})\beta_{trj}\}$$
$$= \rho(\Delta)\frac{C(\mathbf{x})\Sigma_S C^T(\mathbf{t})}{\sqrt{C(\mathbf{x})\Sigma_S C^T(\mathbf{x}) \bullet C(\mathbf{t})\Sigma_S C^T(\mathbf{t})}}$$

Bayesian Model

- **Matlab Code**: There is Matlab code for this methodology available from Veera

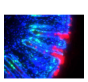
- The method works well in simulations and gives answers that fit with the nonparametric method where the two can be compared

- Seamless Bayesian inference for important questions such as the effects of diets, variability of the correlation estimates, etc.



Bayesian Model

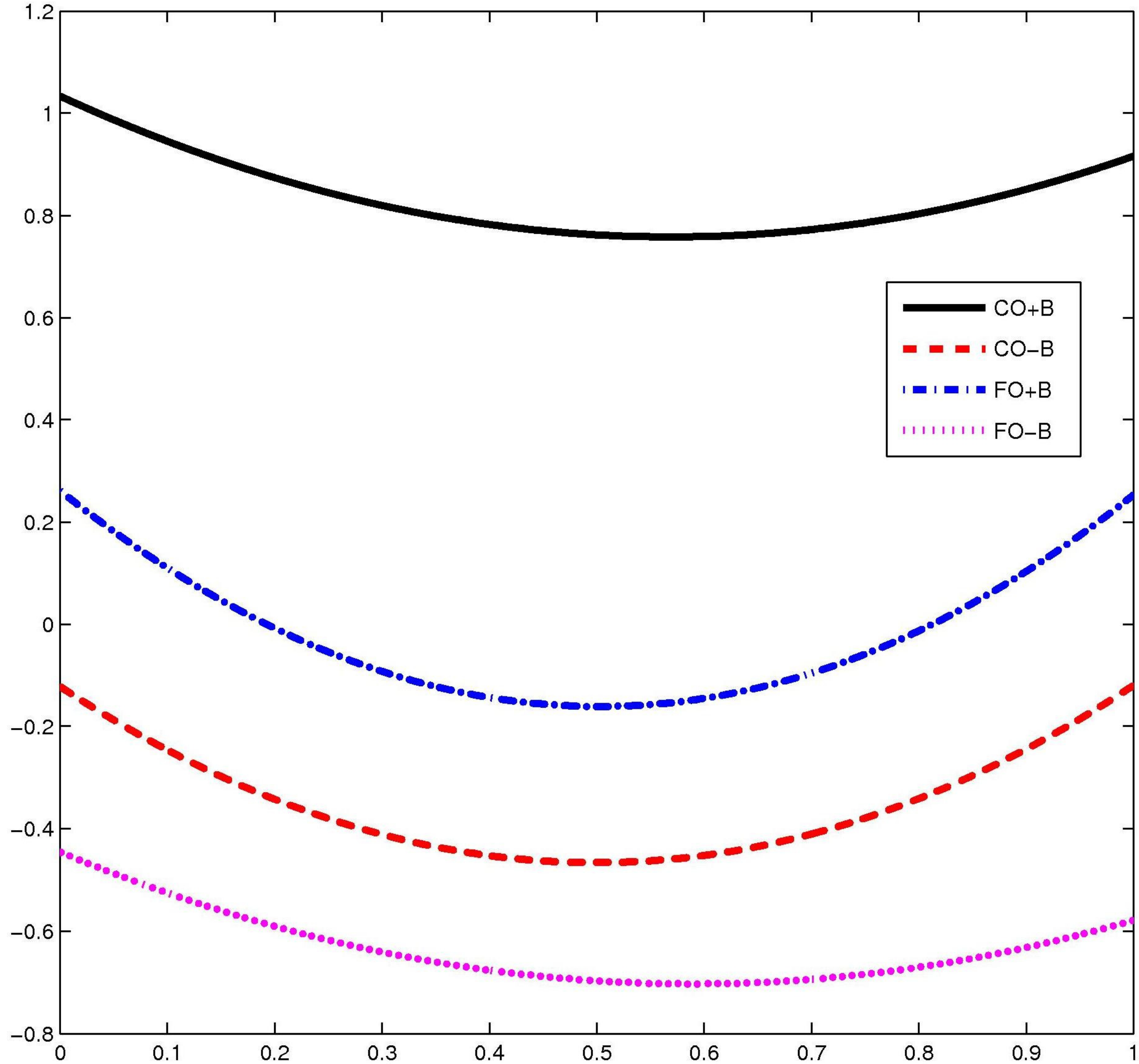
- **<u>Our Implementation</u>** can handle small numbers of observations per subject, unlike the fixed effect methods



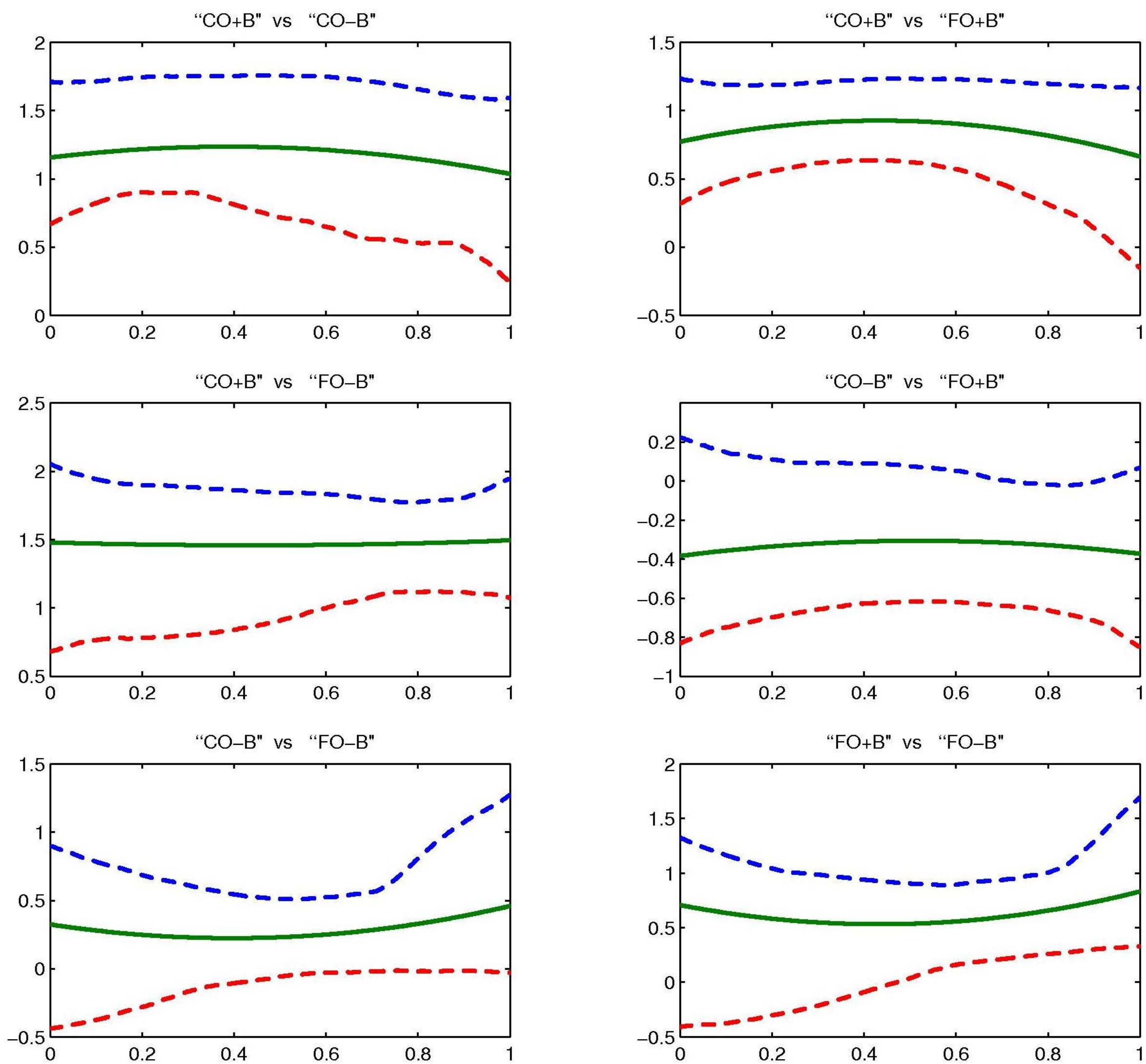
Bayesian Model

- **<u>Our Implementation</u>** is slow
- It is not clear how well it scales up to having many subjects
- To handle many knots it requires an ad hoc dimension reduction
- Need multiple processors to see if one animal drives the results (leave one out, etc.)

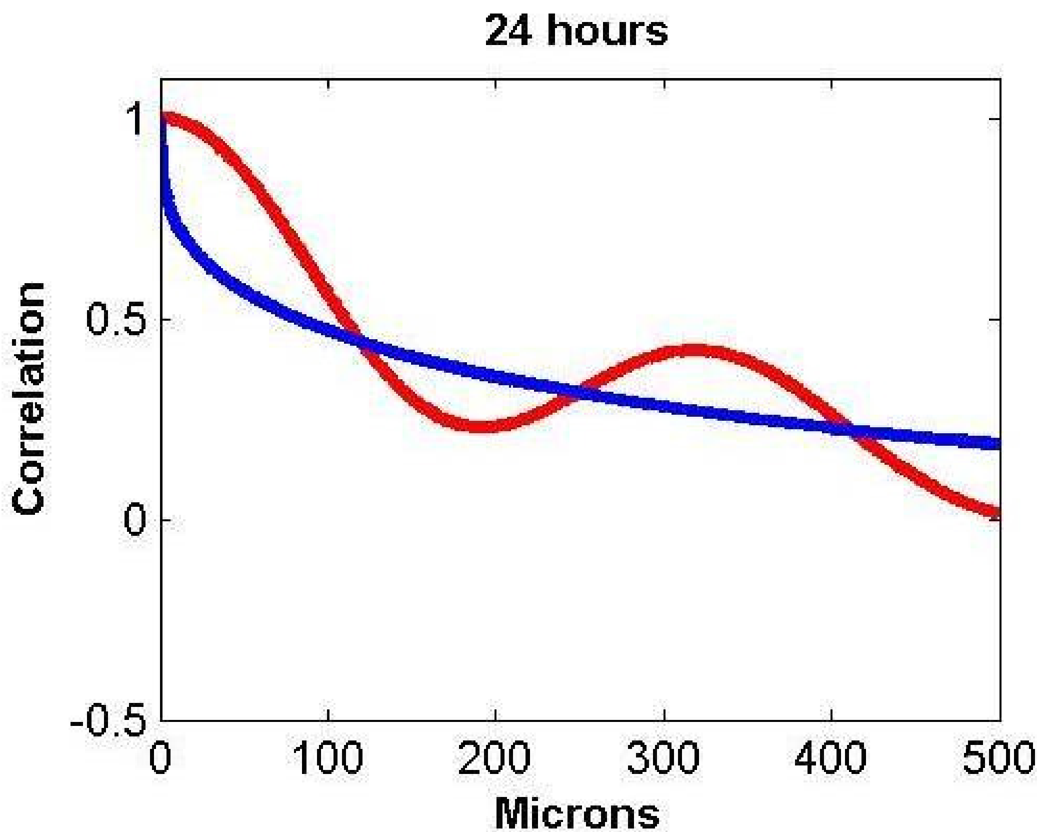
Parametric Mean Fits

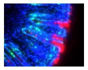


Parametric Mean Fits



Parametric and Nonparametric Fits





STATISTICS
TEXAS A&M UNIVERSITY

Other Hierarchical Methods

- For computational reasons then, we have worked out principal component approaches to the problem
- The methods are flexibly parametric with some nonparametric flavor
- Parametric bootstrap for inference, although technical issues remain, see e.g., N. Wang's talk

Other Hierarchical Methods

- The major issue with frequentist inference in PC methods is the model selection inherent in them

- Model selection methods cannot be analyzed by the bootstrap, because they are not asymptotically normally distributed at contiguous alternatives

Other Hierarchical Methods

- We hope soon to report on Bayesian methods that account for the model selection in the PC methods

- I will next talk about one such PC method

Basis Functions

- The essential issue with basis functions is dimensionality

$$Y_{trc}(x) = \eta_t(x) + Z_{tr}(x) + Q_{trc}(x) + \varepsilon_{trc}(x)$$

$$Z_{tr}(x) = C(x)\gamma_{tr}$$

$$Q_{trc}(x) = C(x)\theta_{trc}$$

- What distributions are assumed for the random effects, while accounting for spatial correlation?

Basis Functions

- In the usual mixed model formulation, massive dimension reduction is made. "**Effectively**",

$$Q_{trc}(x) = C(x)\theta_{trc}$$
$$\text{cov}(\theta_{trc}) = \Sigma_S = \sigma_\theta^2 K$$
$$K = \text{known}$$

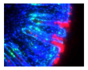
- There is no real reason to assume this is true. If there are 10 basis functions, 55 free parameters become 1 free parameter. **Convenient**!

Basis Functions

- In Veera B., et al., we allowed a general covariance matrix Σ_S but only a few knots
- EM implementations have the same issue: number of parameters is about the square of the number of knots
- Ruppert shows that 20 knots with regression splines solve all problems, but that is a lot of parameters!

Basis Functions

- Dimension reduction of covariance matrices has to be done (**or I think it does!**)
- This means assumptions of one brand or another, none perfect
- We have two approaches, and I will outline one that is still massive dimension reduction, but relies on nothing more than the method of moments



Ana-Maria Staicu

Ciprian Crainiceanu

Simple Model

- Remember

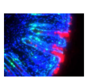
$$\boxed{Y_{trc}(x) = \mu_t(x) + Z_{tr}(x) + Q_{trc}(x) + \varepsilon_{trc}(x)}$$

- Force spatial correlation at locations δ_{trc} as

$$Q_{trc}(x) = W_{trc}(x) + U_{tr}(\delta_{trc})$$

$$W_{trc}(x) = \text{independent across crypts}$$

$$U_{tr}(\delta_{trx}) = \text{isotropic spatial process}$$



Simple Model

- In the spline approach, the spatial correlation is the **correlation** of $\mathbf{Q}_{\mathbf{trc}}(\mathbf{x})$ and $\mathbf{Q}_{\mathbf{trj}}(\mathbf{x})$ at same cell locations

- In the new simple model, the spatial feature is the **covariance** of $\mathbf{Q}_{\mathbf{trc}}(\mathbf{x})$ and $\mathbf{Q}_{\mathbf{trj}}(\mathbf{x})$ independent of cell location

Simple Model

- Now use a functional PCA approach to reduce dimension, i.e.,

$$Z_{tr}(x) = \sum_{k=1}^{K_z} \phi_{kZ}(x)\gamma_{trk}$$

$$\phi_{kZ}(\bullet) \;=\; \text{orthogonal}$$

$$\gamma_{trk} = \text{Normal}(0, \sigma_{kZ}^2) \;\text{ and independent}$$

$$\text{cov}\{Z_{tr}(x), Z_{tr}(s)\} = \sum_{k=1}^{K_z} \sigma_{kZ}^2 \phi_{kZ}(x)\phi_{kZ}(s)$$

Simple Model

- Similarly

$$W_{trc}(x) = \sum_{k=1}^{K_W} \phi_{kW}(x)\beta_{trck}$$

$$\phi_{kW}(\bullet) \;=\; \text{orthogonal}$$

$$\beta_{trck} = \text{Normal}(0, \sigma^2_{kW}) \text{ and independent}$$

$$\text{cov}\{W_{trc}(x), W_{trc}(s)\} = \sum_{k=1}^{K_W} \sigma^2_{kW}\phi_{kW}(x)\phi_{kW}(s)$$

Summary of the Simple Model

- With independence, etc.,

$$
\begin{aligned}
\mathbf{Y}_{trc}(\mathbf{x}) = \mu_t(\mathbf{x}) & \\
& + \sum_{k=1}^{K_z} \phi_{kZ}(\mathbf{x})\gamma_{trk} \\
& + \sum_{k=1}^{K_W} \phi_{kW}(\mathbf{x})\beta_{trck} \\
& + U_{tr}(\delta_{trc}) \\
& + \varepsilon_{trc}(\mathbf{x})
\end{aligned}
$$

Method of Moments

- Everything can be pushed through if we can estimate

$$\mathbf{K}_{\text{within}}(\mathbf{x},\mathbf{s},|\delta_i-\delta_j|=\Delta)$$
$$=\mathbf{cov}\left\{\mathbf{W}_{\text{tri}}(\mathbf{x})-\mathbf{W}_{\text{trj}}(\mathbf{x}),\mathbf{W}_{\text{tri}}(\mathbf{s})-\mathbf{W}_{\text{trj}}(\mathbf{s})\,\middle|\,|\delta_i-\delta_j|=\Delta\right\}$$

- Like Li, et al, this is nonparametric regression, although we use **KNN** averaging rather than **kernels**

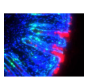
Method of Moments

- We have developed a series of method of moments based calculations to fit this model

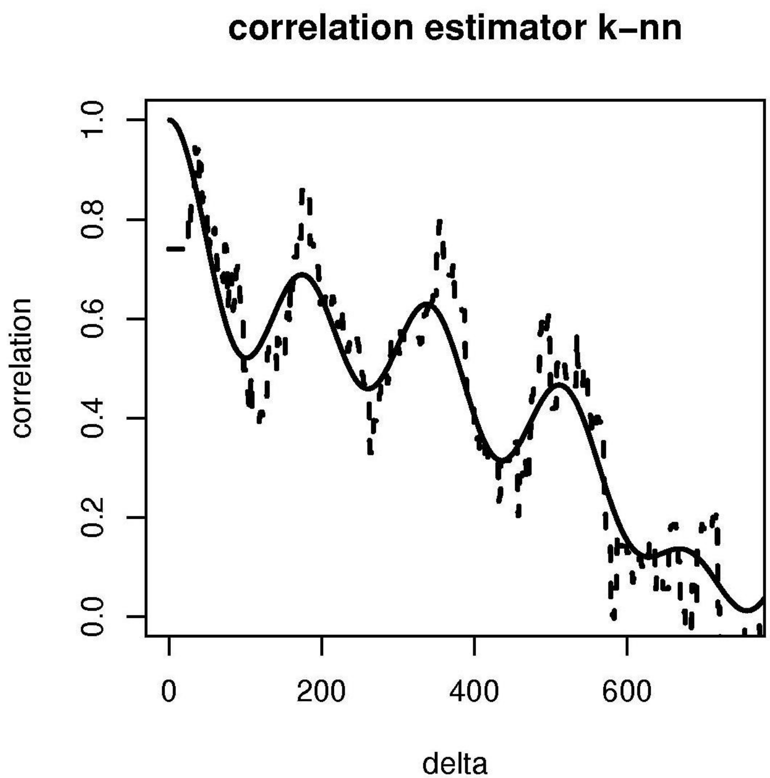
- There are some large covariance matrices that need to be inverted (BLUP) to compute estimates of the random effects, but we have developed dimension-reduction techniques to get around this

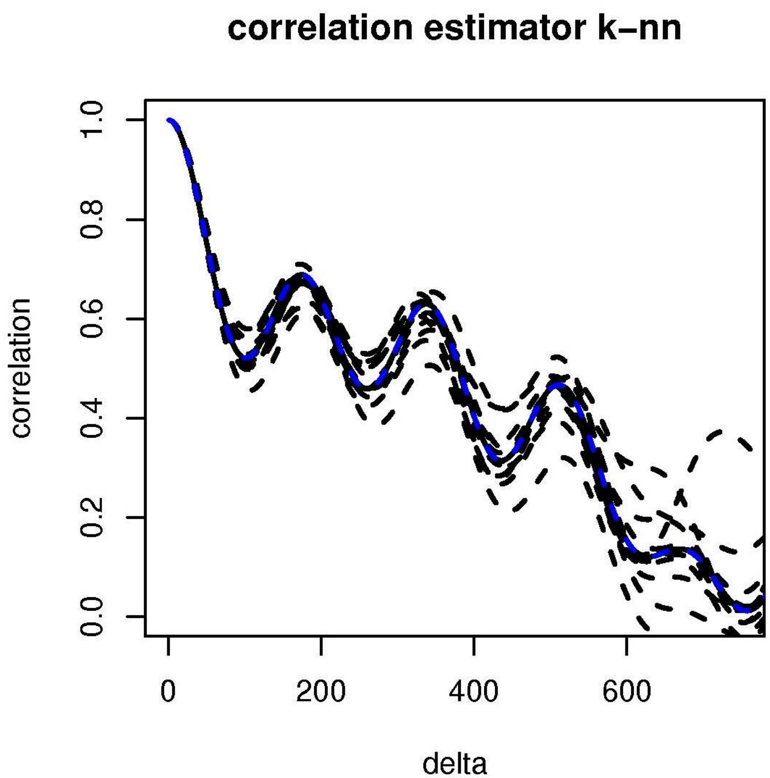
Method of Moments

- The method is fast
  - On our data, the Bayesian method takes about 5 hours on a very fast processor
  - Ours takes 12 seconds, including estimation of the number of principal components
- The speed allows us to do leave-one-subject out analyses, e.g., to see the sensitivity to individual subjects

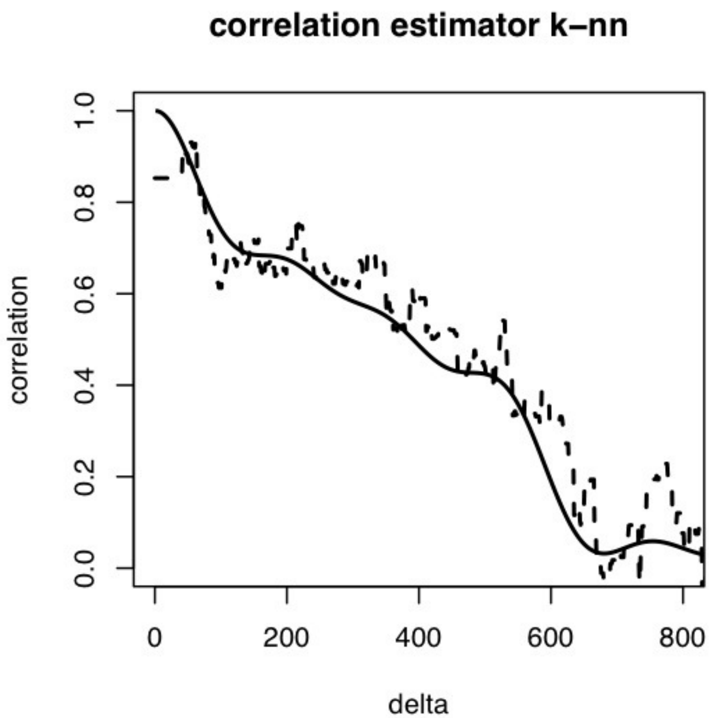


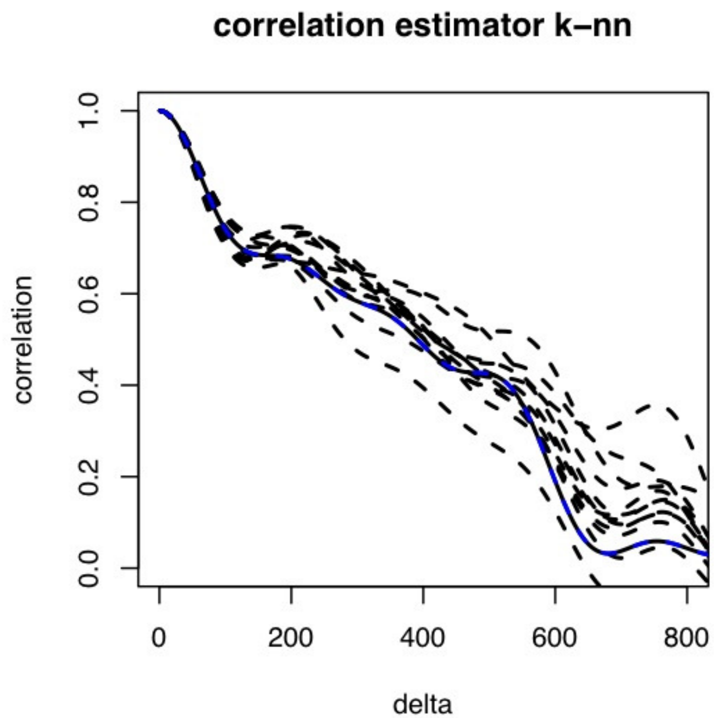
Method of Moments, 40NN

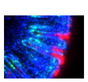




Method of Moments, 80NN

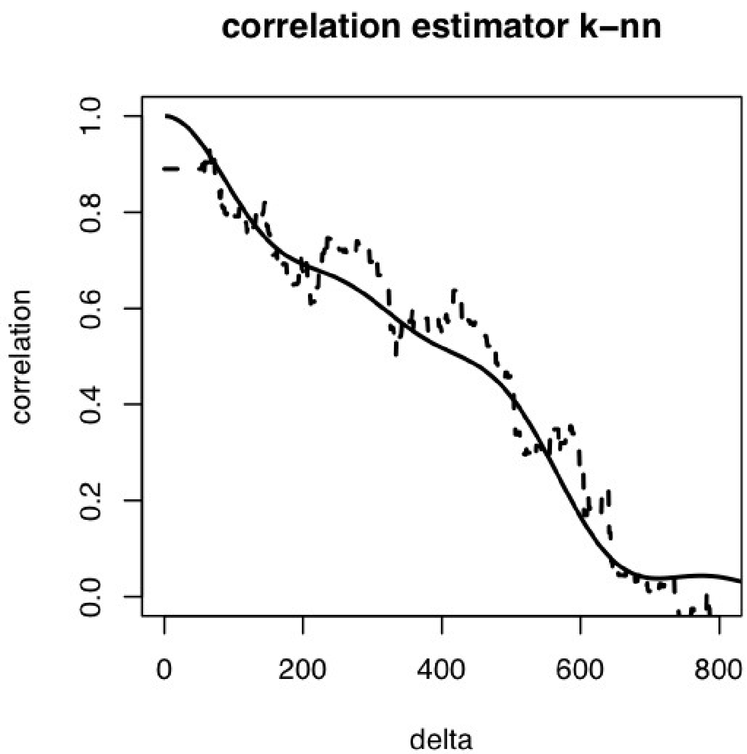


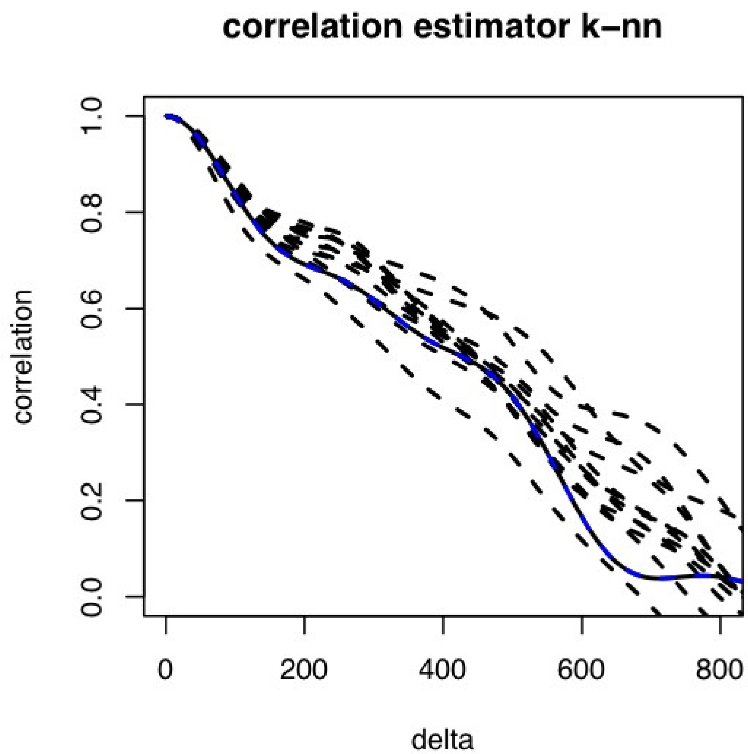


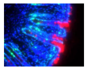


STATISTICS
TEXAS A&M UNIVERSITY

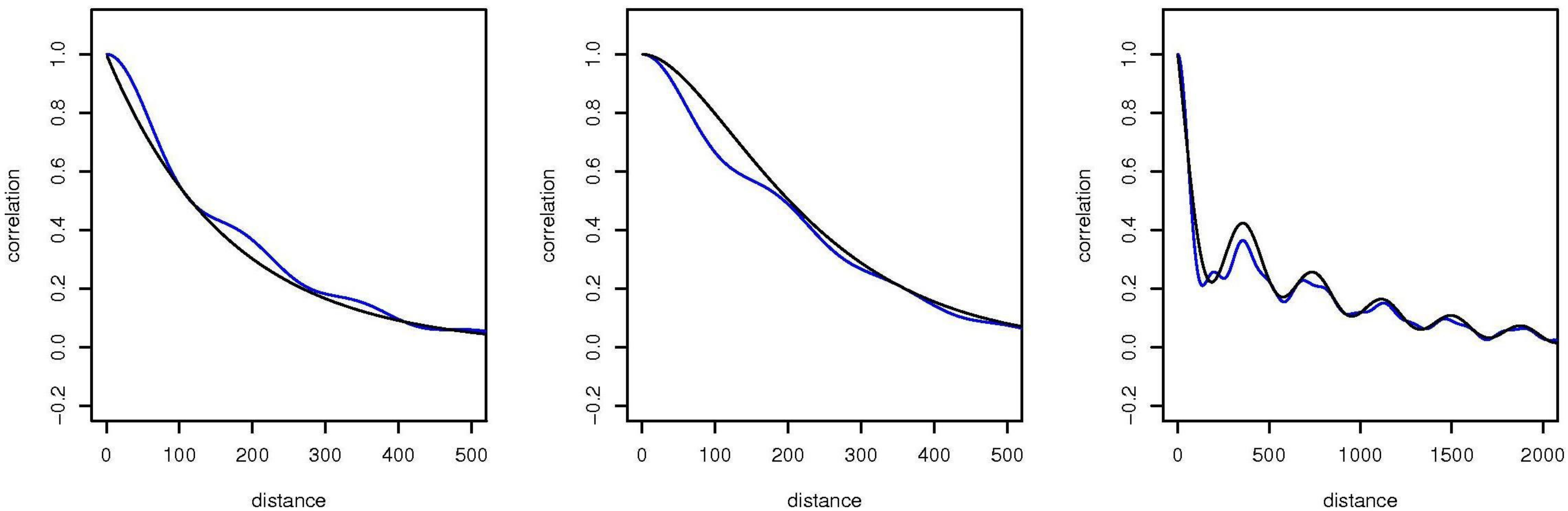
Method of Moments, 100NN







Method of Moments



Simulations: Mean fits for spatial structure as in the data:
Black = true,
Blue = estimated

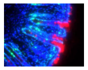
Other PC Approaches

- There are at least two other ways to use a PC approach that has a structure like the previous approaches
- Old Method

$$Q_{trc}(x) = W_{trc}(x) + U_{tr}(\delta_{trc})$$

$$W_{trc}(x) = \text{independent across crypts}$$

$$U_{tr}(\delta_{trx}) = \text{isotropic spatial process}$$



Other PC Approaches

- Also

$$W_{trc}(x) = \sum_{k=1}^{K_W} \phi_{kW}(x)\beta_{trck}$$

$$\phi_{kW}(\bullet) \;=\; \text{orthogonal}$$

$$\beta_{trck} = \text{Normal}(0, \sigma_{kW}^2) \;\text{ and independent}$$

$$\text{cov}\{W_{trc}(x), W_{trc}(s)\} = \sum_{k=1}^{K_W} \sigma_{kW}^2 \phi_{kW}(x)\phi_{kW}(s)$$

Other PC Approaches

- New Method

$$Q_{trc}(x) = W_{trc}(x)$$

$$W_{trc}(x) = \text{NOT independent across crypts}$$

Other PC Approaches

- Also

$$W_{trc}(x) = \sum_{k=1}^{K_W} \phi_{kW}(x)\beta_{trck}$$

$$\phi_{kW}(\bullet) \ = \ \text{orthogonal}$$

$$\beta_{trck} = \text{Normal}(0, \sigma^2_{kW}) \ \text{and independent}$$

$$\text{cov}\{W_{trc}(x), W_{trc}(s)\} = \sum_{k=1}^{K_W} \sigma^2_{kW}\phi_{kW}(x)\phi_{kW}(s)$$

Other PC Approaches

- However,

$$W_{trc}(x) = \sum_{k=1}^{K_W} \phi_{kW}(x)\beta_{trck}$$

$$\text{cov}(\beta_{trck}, \beta_{trjk}) = \rho_k(\Lambda)\sigma^2_{kW}$$

- Not necessarily separable

Other PC Approaches

- There are technical difficulties with this due to the construction of the principal component functions

$$\boxed{\phi_{kW}(\mathbf{x})}$$

- We are developing an alternative approach, more like Bsplines but with a PC flavor, that avoids this construction

Summary

- We have studied the problem of crypt-signaling in colon carcinogenesis experiments
- Technically, this is a problem of hierarchical functional data where the functions are not independent in the standard manner
- We developed constructive semiparametric and nonparametric methods
- The correlations we see in the functions are surprisingly large.