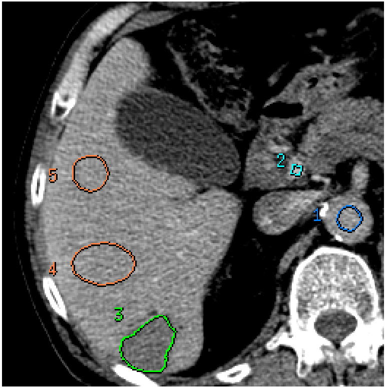
Analysis of Spatially Correlated Curves

Jianhua Hu
Department of Biostatistics
The University of Texas MD Anderson Cancer Center

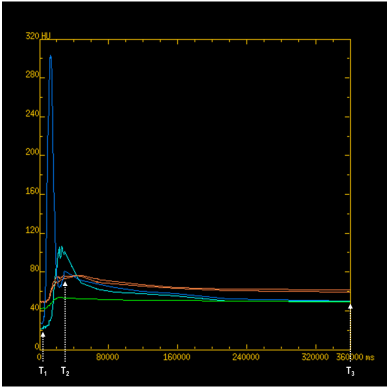
Banff, February 2016

Motivating Example

CT scan of liver



Time-attenuation curves



Functional Data Model

- Mathematical model

$$y_{ijk} = \mu(t_{ijk}) + x_{ij}(t_{ijk}) + \epsilon_{ijk}$$

  - y_{ijk}: observation of subject i and unit j taken at time t_{ijk}
  - $\mu(t)$: a smooth mean function
  - $x_{ij}(t)$: a random function for subject i unit j. It is assumed to be mean zero and correlated within the same subject with the cross-covariance function.

$$\text{cov}(x_{ij}(t), x_{ij'}(t')) = G(t, t'; \Delta), \quad (1)$$

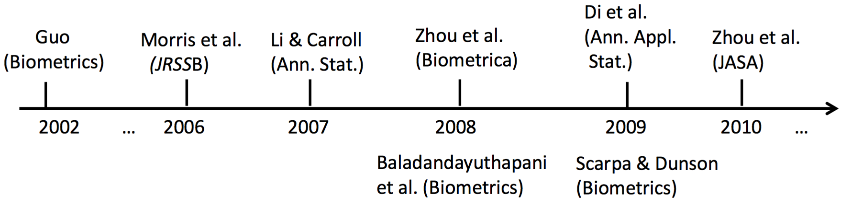
    where Δ denotes the distance between two units.
  - ϵ_{ijk}: random measurement noise$\sim$ i.i.d. $N(0, \sigma^2)$.

Modeling of Spatial Correlation

- functional mixed model
- multilevel functional model

$$x_{ij}(t) = \eta_i(t) + \xi_{ij}(t)$$

  - $\eta_i(t)$: random curve at subject level
  - $\xi_{ij}(t)$: random curve at unit level



The general covariance structure

The general covariance (1) imposes no restriction on the spatial and temporal correlation structure.

- For the hierarchical model (e.g., Baladandayuthapani et al., 2008; Zhou et al., 2010), it corresponds to the summation of the covariance at the subject level and the unit level.
- For the additive model (e.g. Staicu et al., 2010), it corresponds to the summation of the spatial covariance and temporal covariance.

Problems of interest are:

- estimate mean function $\mu(t)$
- estimate covariance function $G(t, t'; \Delta)$
- predict individual curves

Problems of interest are:

- estimate mean function $\mu(t)$
- estimate covariance function $G(t, t'; \Delta)$
- predict individual curves

Kernal Smoothing (et al., Yao et al., 2005): sparse observations, irregular time points

Kernel Smoothing Estimate for the Mean Function

The kernel smoothing estimate $\hat{\mu}_0(t)$ is defined as

$$\hat{\mu}_0(t) = \arg\min_{\beta_0} \sum_{i,j,k} \kappa(\frac{t_{ijk} - t}{h}) \left(y_{ijk} - f(t_{ijk} - t)\right)^2$$

- $f(t_{ijk} - t) = \beta_0 + \beta_1(t_{ijk} - t)$
- $\kappa(\cdot)$: kernel function
- h: bandwidth that controls the size of local window

Yao et al. (2005)

The kernel smoothing estimate $\hat{\mu}_0$ is uniformly consistent, i.e.,

$$\sup_{t \in \mathcal{T}} |\hat{\mu}_0(t) - \mu(t)| = O_p(\frac{1}{\sqrt{N}h}).$$

Kernel Smoothing Estimate for Covariance

We estimate covariance between measurements taken at time t_1 and t_2 by minimizing

$$\sum_{i,j,k_1 \neq k_2} w_{ijk_1k_2}(\hat{G}_i(t_{ijk_1}, t_{ijk_2}) - g(t_{ijk_1} - t_1, t_{ijk_2} - t_2))^2,$$

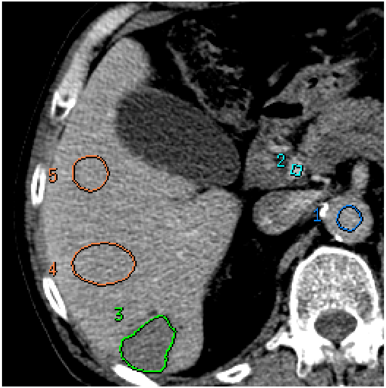
where $\hat{G}_i(t_{ijk_1}, t_{ijk_2})$ is the sample covariance from subject i, g is a bivariate local linear function, $w_{ijk_1k_2}$ is the local weight given by the kernel function and bandwidth h_G.

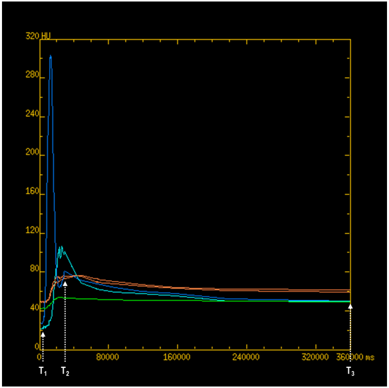
Yao et al. (2005)

The kernel smoothing estimate $\hat{G}_0$ is uniformly consistent, i.e.,

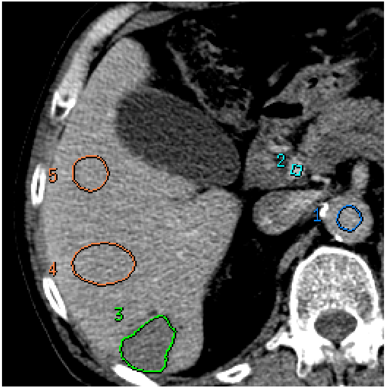
$$\sup_{t \in \mathcal{T}} |\hat{G}_0(t) - G(t)| = O_p(\frac{1}{\sqrt{N} h_G^2}).$$

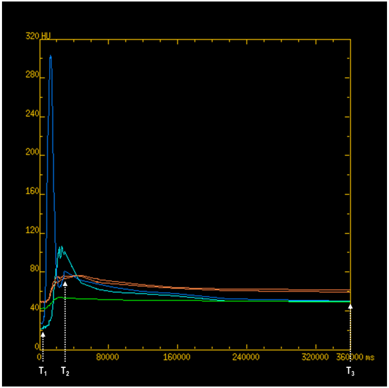
Problem of Interest





Problem of Interest





How can we incorporate the spatial correlation to improve mean estimation?

A New Mean Estimation

Let $\boldsymbol{y}_{i\cdot k} = [y_{i1k}, \ldots, y_{in_ik}]^T$ be the stacked observation from subject i at time t_{ik}. The simple kernel smoothing estimate in Yao (2005) equivalently minimizes

$$\sum_{i=1}^{N}\sum_{k=1}^{m_i}\kappa(\frac{t_{ik}-t}{h})(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})^T(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})$$

A New Mean Estimation

Let $\boldsymbol{y}_{i\cdot k} = [y_{i1k}, \ldots, y_{in_ik}]^T$ be the stacked observation from subject i at time t_{ik}. The simple kernel smoothing estimate in Yao (2005) equivalently minimizes

$$\sum_{i=1}^{N}\sum_{k=1}^{m_i} \kappa(\frac{t_{ik}-t}{h})(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})^T(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})$$

We propose a weighted kernel smoothing estimate minimizes:

$$\sum_{i=1}^{N}\sum_{k=1}^{m_i} \kappa(\frac{t_{ik}-t}{h})(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})^T\boldsymbol{W}_{ik}(\boldsymbol{y}_{i\cdot k} - f(t_{ik}-t)\mathbf{1}_{n_i})$$

- Special case: $\boldsymbol{W}_{ik} \equiv \boldsymbol{I}_n$.
- In general, weight is a function of time and space
$$\boldsymbol{W}_{ik} = \boldsymbol{w}(t_{ik}; \boldsymbol{s}_i).$$
- Requirement: all units within the same subject are sampled at the same time.

Asymptotic Results

Theorems

Under certain conditions, we demonstrate the following results:

1. Uniform consistency:
$$\sup_{t \in \mathcal{T}} |\hat{\mu}(t) - \mu(t)| = O_p(\frac{1}{\sqrt{Nh}})$$
2. Asymptotic distribution:
$$\sqrt{Nh}(\hat{\mu}(t) - \mu(t)) \xrightarrow{d} N\left(0, V(\boldsymbol{w})\right)$$
3. Asymptotic efficiency:
$$V(\boldsymbol{w}) \geq V(\boldsymbol{w}_0)$$

   Here $\boldsymbol{w}_0$ is the inverse of the conditional covariance of $\boldsymbol{y}$ given time and space.

Seperable Covariance Structure

- We specify a separable structure for the covariance:

$$G(t, t'; \Delta) = \psi(\Delta) G_0(t, t')$$

  where Δ is the distance between two units.

- The weight function is

$$\boldsymbol{W}_{ik} = \{\hat{G}_0(t_{ik}, t_{ik}) \hat{\boldsymbol{\Psi}}_i + \hat{\sigma}^2 \boldsymbol{I}_{n_i}\}^{-1},$$

  where $\hat{\boldsymbol{\Psi}}_i$ is an $n_i \times n_i$ matrix with elements $(\hat{\boldsymbol{\Psi}}_i)_{jj'} = \psi(\Delta_{jj'})$.

- In our studies, we focus on the compound symmetry structure.

Individual Curve Prediction

Idea: using discrete observations to predict each individual curve.

- Functional principal component analysis (PCA) approximation

$$\hat{y}_{ij}(t) = \hat{\mu}(t) + \sum_{l=1}^{L} \xi^*_{ijl} \hat{\phi}_l(t)$$

- Independent prediction

$$\xi^*_{ijl} = E[\xi_{ijl} | \boldsymbol{y}_{ij}]$$

- Simultaneous prediction

$$\xi^*_{ijl} = E[\xi_{ijl} | \boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in}]$$

Simulation

- A data generation scenario to mimick the pCT data.

$$y_{ij}(t) = \mu(t) + \xi_{ij1}(s_{ij})\phi_1(t) + \xi_{ij2}(s_{ij})\phi_2(t) + \epsilon(t)$$

- In total of 20 subjects, 4 units each.
- Separable covariance
- Covariance estimation
  - $\widehat{G}$: bivariate kernel smoothing
  - $\psi(\Delta)$: maximizes the likelihood given $\widehat{G}$.
- Cross-validation bandwidth selection

Simulation Results

(1) Sparse case: 3 observations for each curve

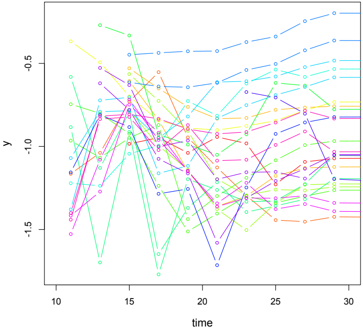
	Estimation		Prediction	
inter-region correlation	KS	WKS	Indep	Joint
0	0.111	0.101	1.306	1.298
0.4	0.333	0.316	1.446	1.405
0.8	0.502	0.487	1.638	1.466

(2) Non-sparse case: 7 observations for each curve

	Estimation		Prediction	
inter-region correlation	KS	WKS	Indep	Joint
0	0.055	0.052	0.572	0.571
0.4	0.213	0.206	0.612	0.594
0.8	0.308	0.303	0.672	0.601

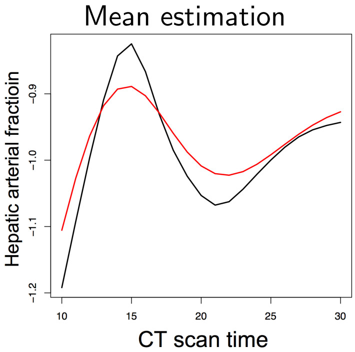
The CT Perfusion Study

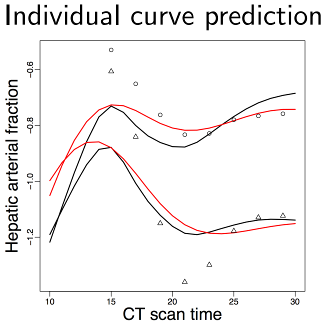
- 16 patients with 25 path verified neuroendocrine liver metastases, with CTp undertaken within the designated region
- CTp variables from neighboring regions are strongly correlated due to common vasculature.



Hepatic Arterial Fraction (Between-region correlation:0.83)

Results





Prediction error	
Indep	Joint
0.325	0.218

Summary

- Incorporating the correlation between curves improves the mean estimation efficiency and individual curve prediction accuracy
- The proposed weighted mean estimate is uniformly consistently and asymptotically most efficient
- The proposed method offers to improve our ability to characterize biomarkers that are acquired from a small number of scans by leveraging the information at neighboring sites

The multivariate functional data model

Denote $\boldsymbol{y}_{ij}(t) = [y_{ij1}(t), \ldots, y_{ijp}(t)]^T$ the $p \times 1$ measurement function for the jth unit of subject i

$$\boldsymbol{y}_{ij}(t) = \boldsymbol{\mu}(t) + \boldsymbol{x}_{ij}(t) + \boldsymbol{\epsilon}_{ij}(t),$$

- $\boldsymbol{\mu}(t)$ is the $p \times 1$ smooth mean curve that depicts the average temporal trajectory of the p features in the study
- $\boldsymbol{x}_{ij}(t)$ is the $p \times 1$ smooth random function that characterizes the random deviation from the mean curve for unit j of patient i
- $\boldsymbol{\epsilon}_{ij}(t)$ is a $p \times 1$ random function assumed to be independent white noise process with variance $\boldsymbol{\sigma}^2$

The three-way correlation structure

- Correlation between time and variables

$$\mathrm{cov}(\boldsymbol{x}_{ij}(t), \boldsymbol{x}_{ij}(t)) = \boldsymbol{G}(t,t') = \begin{bmatrix} G_{11}(t,t') & \dots & G_{1p}(t,t') \\ \vdots & \ddots & \vdots \\ G_{p1}(t,t') & \dots & G_{pp}(t,t') \end{bmatrix}.$$

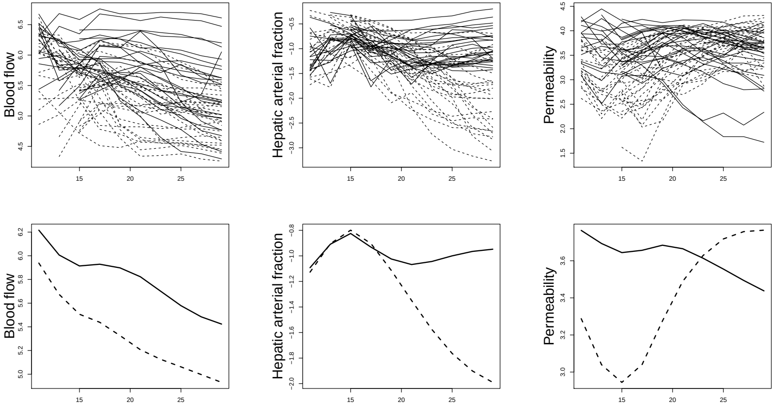
- Correlation between neighboring regions

$$cov(\boldsymbol{x}_{ij}(t), \boldsymbol{x}_{ij'}(t')) = \boldsymbol{G}(s_{ij}, s_{ij'}; t, t').$$

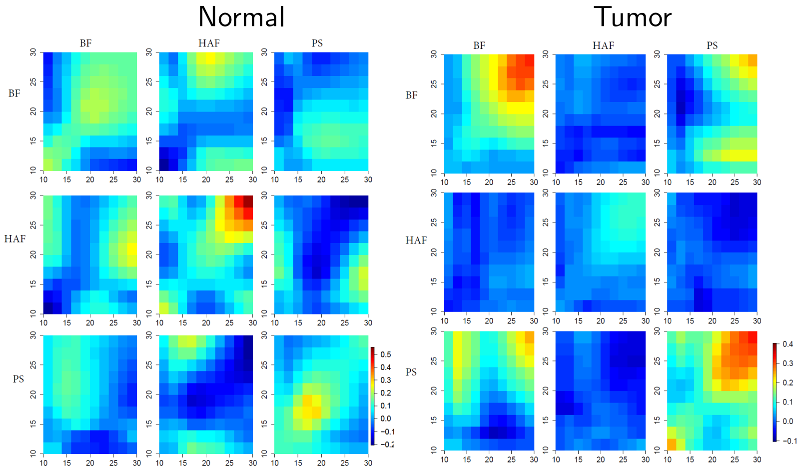
A separable structure

$$\boldsymbol{G}(s_{ij}, s_{ij'}; t, t') = \rho(s_{ij}, s_{ij'})\boldsymbol{G}(t,t'),$$

Raw Data and Mean Estimates



Cross-Covariance for Each Class of Tissue



Classification

Suppose the new subject has n_0 regions of interest. Let $\boldsymbol{z}_0 = (z_{01}, \ldots, z_{0n_0})$ be the class assignment for each region and $\boldsymbol{y}_{0\cdot\cdot} = [\boldsymbol{y}_{01\cdot}, \ldots, \boldsymbol{y}_{0n_0\cdot}]$ the measurements collected on all the regions.

- When $n_0 = 1$, we assign $z_{01} = 0$ if

$$\frac{f(\boldsymbol{y}_{01\cdot}|z_{01} = 0, \boldsymbol{\Theta}_0)p_0}{f(\boldsymbol{y}_{01\cdot}|z_{01} = 1, \boldsymbol{\Theta}_1)p_1} \geq 1$$

  and $z_{01} = 1$ otherwise. p_0 and p_1 are the prior probabilities for normal and tumor tissue, respectively.

- When $n_0 > 1$,

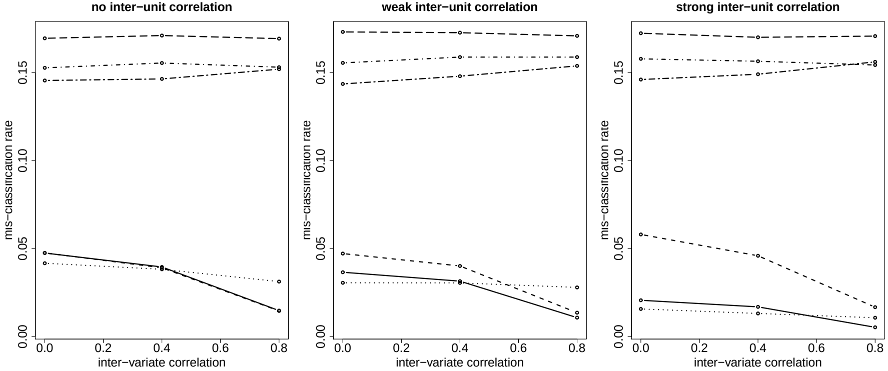
$$\widehat{\boldsymbol{z}}_0 = \arg\max_{\boldsymbol{z} \in \mathcal{D}} f(\boldsymbol{y}_{0\cdot\cdot}|\boldsymbol{z}, \boldsymbol{\Theta}_0, \boldsymbol{\Theta}_1) \prod_{j=1}^{n_0} p_{z_0 j}$$

Case Study

Method		MCR	TPR	FPR
three-way NB		0.135	0.800	0.074
two-way NB(no spatial corr)		0.327	0.480	0.148
two-way NB(no variable corr)		0.154	0.80	0.111
functional depth		0.173	0.880	0.222
GLM	BF	0.308	0.640	0.259
	HAF	0.231	0.800	0.259
	PS	0.269	0.720	0.259
KNN	BF	0.577	0.280	0.444
	HAF	0.173	0.960	0.296
	PS	0.346	0.640	0.333

Table : Leave-one-subject-out classification for the CT perfusion study.

Simulation Study



classification approaches: the proposed (solid), no inter-unit correlation(dashed), no inter-variate correlation (dotted), GLM (dotdash) , KNN (longdash), multivariate functional depth (twodash)