LECTURES ON FUNCTIONAL DATA ANALYSIS

Hans-Georg Müller
UC Davis

ERCIM

Oviedo, November 2012

OVERVIEW

- PART I: BASICS: MODELING OF RANDOM TRAJECTORES AND LONGITUDINAL DATA
  - Functional Principal Component Analysis
  - Sparse and Dense Functional Data
  - Derivatives
  - Empirical Dynamics
- PART II: FUNCTIONAL REGRESSION MODELS
  - Functional Linear Models
  - Diagnostics
  - Functional Dose-Response Models
  - Functional Additive Regression
  - Functional Quadratic Regression
  - Functional Gradients
- PART III: TIME WARPING AND NONLINEAR REPRESENTATIONS
  - Lecture on Saturday

INTRODUCTION

What characterizes functional data?

Per subject or experimental unit, one samples one or several functions $X(t)$, $t \in T$

High-dimensional (infinite-dimensional) data with a topology characterized by order, neighborhood and smoothness – in contrast to MDA (Multivariate Data Analysis).

Commonly adopted model: Data correspond to independent realizations of a stochastic process with smooth trajectories

LONGITUDINAL STUDIES AND DYNAMICS

Data: Longitudinal studies, e.g. Baltimore Longitudinal on Aging; e-Bay online auction data

Model: Sample of irregularly measured realizations of an underlying stochastic process, assumed to be smooth

Goals: Estimating derivatives for irregularly sampled random trajectories

Learning the underlying dynamics – empirical differential equation

Methods: Functional principal component analysis; Smoothing and differentiation (local least squares); Representations of stochastic processes

STOCHASTIC PROCESS PERSPECTIVE

Assume observed data are generated by underlying stochastic process $X \in L^2(T)$ with finite second moments:

$\mu(t) = E(X(t))$ mean function
$G(s,t) = \mathrm{cov}\{X(s), X(t)\}$ covariance function.

Define auto-covariance operator $(A_G f)(t) = \int f(s)G(s,t)\, ds$ with orthonormal eigenfunctions ϕ_k and ordered eigenvalues $\lambda_1 \geq \lambda_2 \geq \ldots$,

$$(A_G \phi_k)(t) = \lambda_k \, \phi_k(t)$$

Background Material

- Books
  - Ramsay, J.O. & Silverman, B.W. (2002) Applied Functional Data Analysis. Springer
  - Ferraty, F. & Vieu, P. (2006) Nonparametric Functional Data Analysis. Springer
  - Horvath, L. & Kokoszka, P. (2012) Inference for Functional Data with Applications. Springer
- Software
  - Ramsay's fda package (Matlab and R versions)
  - PACE 2.16: http://anson.ucdavis.edu/∼mueller/data/pace.html (Matlab)
  - Various R packages

FUNCTIONAL PRINCIPAL COMPONENTS (FPC)

KARHUNEN-LOÈVE REPRESENTATION USING FPCs

$$X(t) = \mu(t) + \sum_{k=1}^{\infty} A_k \phi_k(t),$$

where $A_k = \int_0^T \{X(t) - \mu(t)\}\phi_k(t)dt$, are uncorrelated r.v. with $EA_k = 0,\ EA_k^2 = \lambda_k$, the functional principal components.

Some key papers:

- Grenander 1950: Basic ideas (following up on Karhunen 1949)
- C.R. Rao 1958: Preliminary version for growth curves
- Castro, Lawton & Sylvestre 1987: Modes of Variation in industrial applications
- Rice & Silverman 1991, Rice & C. Wu 2001: B-splines and systematic study

- Book: Ramsay & Silverman 2005: Presmoothing (usually inefficient)
- Bali, Boente, Tyler & J.L. Wang 2012: Systematic study of robust FPCA

Why Functional Principal Components?

- Parsimonious description of longitudinal/functional data as it is the unique linear representation which explains the highest fraction of variance in the data with a given number of components.
- Main attraction is equivalence $X \equiv \{A_1, A_2, \ldots\}$ so that X can be expressed in terms of mean function μ and the countable sequence of eigenfunctions and uncorrelated FPC scores A_k.
- For modeling functional regression: Functions $f(X)$ have an equivalent function $g(A_1, A_2, \ldots)$ so that

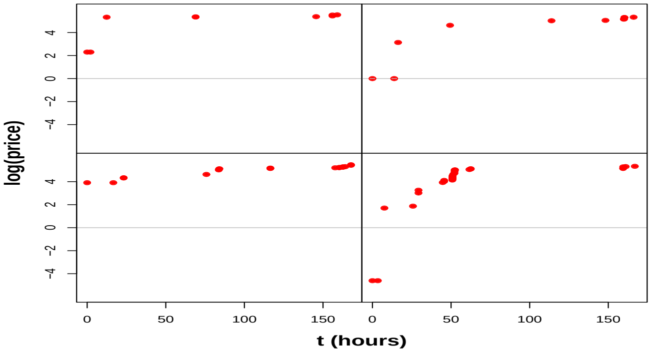
$$f(X) \equiv g(A_1, A_2, \ldots)$$

FUNCTIONAL DATA DESIGNS

- Fully observed functions without noise at arbitrarily dense grid
  Measurements $Y_{it} = X_i(t)$ available for all $t \in T$,
  $i = 1, \ldots, n$:
  Often unrealistic but mathematically convenient

- Dense design with noisy measurements
  Measurements $Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, where T_{ij} are recorded on a regular grid, $T_{i1}, \ldots, T_{iN_i}$, and $N_i \to \infty$:
  Applies to typical functional data

- Sparse design with noisy measurements = Longitudinal data
  Measurements $Y_{ij} = X_i(T_{ij}) + \varepsilon_{ij}$, where T_{ij} are random times and their number N_i per subject is random and finite.

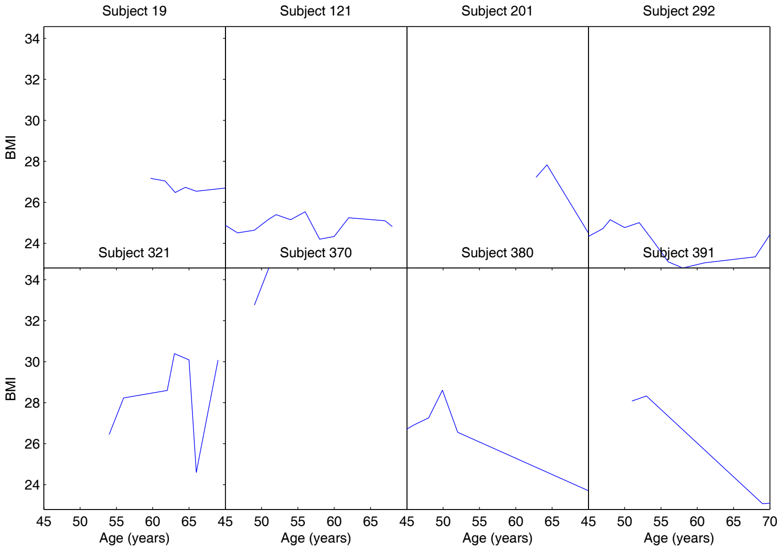


Four eBay auctions: willing-to-pay prices (log-transformed) recorded against time (in hours). Selected from 156 same-item auctions – data from W. Jank

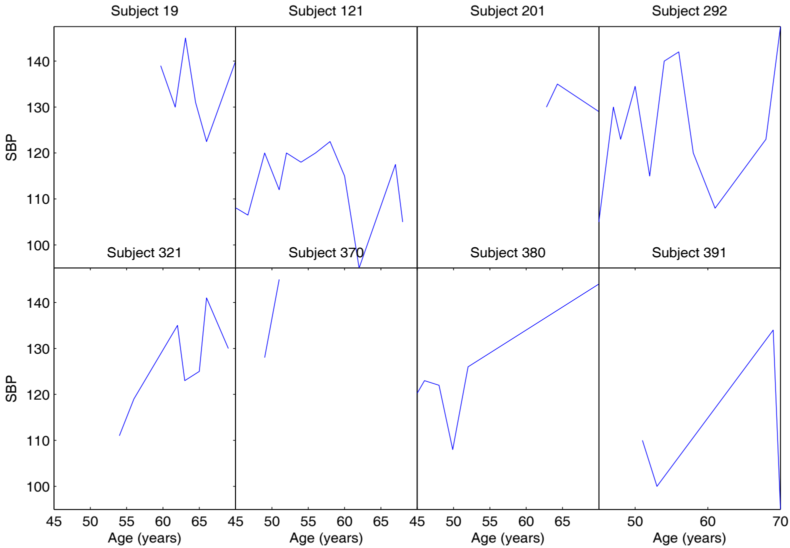
BALTIMORE LONGITUDINAL STUDY ON AGING

- Subset of $n = 507$ males whose Body Mass Index (BMI) and Systolic Blood Pressure (SBP) were measured at least twice between ages 45 and 70 and who survived beyond age 70.

- Measurements are both noisy and spaced irregularly, with both the measurement times and the number of available measurements varying from subject to subject.



Observations of BMI for eight randomly selected subjects



Observations of SBP for eight randomly selected subjects

PACE

Principal Analysis by Conditional Expectation (Yao, M, Wang 2005ab, Liu & M 2009) to obtain components of the functional principal component representation for all of these designs.

Idea: Borrowing strength from entire sample for estimation of individual trajectories

Implementation steps:

- Mean function: Smoothing across all pooled observations
- Covariance surface: Pooling products for pairs of observations from the same subject, then smoothing – denoising is achieved by separating out the diagonal (Staniswalis & Lee 1998)

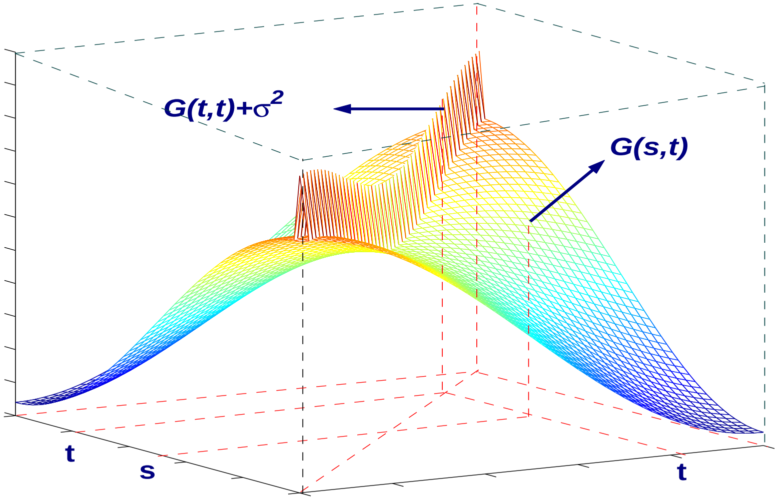
Local Linear Smoothing Estimators

- Mean function is given by $\hat{\mu}(t) = \hat{a}_0$, where

$$(\hat{a}_0, \hat{a}_1) = \arg\min \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{[Y_{ij} - a_0 - a_1(t_{ij} - t)]^2 \times K_h(t_{ij} - t)\}.$$

- Covariance function is given by $\hat{G}(t_1, t_2) = \hat{a}_0$, where

$$\begin{aligned}(\hat{a}_0, \hat{a}_1, \hat{a}_2) = \arg\min \sum_{i=1}^{n} \sum_{j \neq l} \{[&Y_{ij}^c Y_{il}^c - a_0 - a_1(t_{ij} - t_1) \\ &- a_2(t_{il} - t_2)]^2 \times K_b(t_{ij} - t_1) K_b(t_{il} - t_2)\}.\end{aligned}$$



Relationship between the covariance surface and variances on the diagonal: Decomposing diagonal into error and covariance components.

IMPLEMENTATION ISSUES

- Obtain eigenvalues/eigenfunctions:

  For k-th eigenvalue/eigenfunction pair (λ_k, ϕ_k) use discretized versions of eigenequations,

  $$\int_0^T \mathrm{cov}(X(s), X(t))\phi_k(s)ds = \lambda_k \phi_k(t),$$

  s.t. $\int_0^T \phi_k(t)^2 dt = 1, \quad \int_0^T \phi_k(t)\phi_m(t)dt = 0, \quad m \neq k,$
  substituting smoothed estimates for the covariance surface.

- Project initial smoothed covariance estimates on space of non-negative definite covariance matrices: (Hall, M, Yao 2008)

  $$\widehat{\mathrm{cov}}(X(s), X(t)) = \sum_{k=1, \hat{\lambda}_k > 0}^{K} \hat{\lambda}_k \hat{\phi}_k(s) \hat{\phi}_k(t).$$

- Obtain Functional principal components (the random effects):
  - Conditioning $E(A_k|U_i)$, where U_i is the vector of available data for the i-th subject (random dimension)
  - Best linear predictor for conditional expectation (best predictor under Gaussian assumptions)
  - Substitute estimates for eigenvalues, eigenfunctions, covariances
  - Regularization for inverses of cova matrices at random locations
  - Choice of regularization parameters (number of included components, smoothing parameters: GCV, FVE, BIC,...)
- Implementation of FPCA and functional regression models: PACE 2.14 at: http://anson.ucdavis.edu/∼mueller/data/programs.html

ESTIMATING DERIVATIVES FROM SPARSE DATA

Differentiating Karhunen-Loève representation:

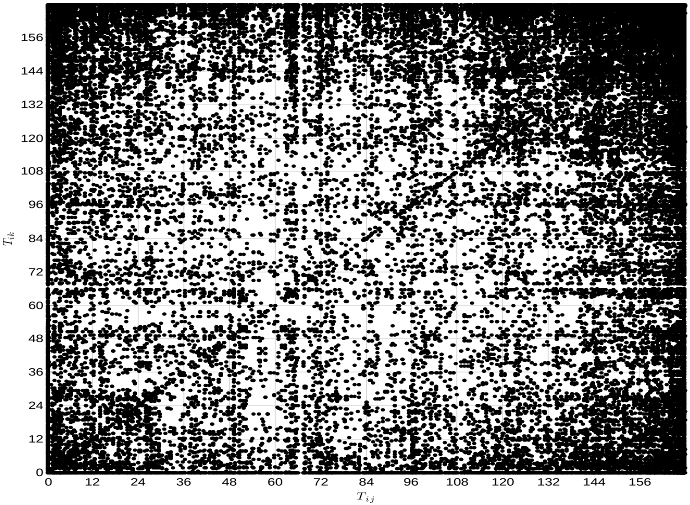
$$X_i^{(\nu)}(t) = \mu^{(\nu)}(t) + \sum_{k=1}^{\infty} A_{ik}\phi_k^{(\nu)}(t), \quad \nu = 0, 1, \ldots.$$

- Obtain estimated random effects A_{ik} by conditioning as before
- Estimate $\mu^{(\nu)}(t)$ by known nonparametric 1-d differentiation, applied to pooled scatterplots.
- How to obtain $\phi_k^{(\nu)}$? Observe

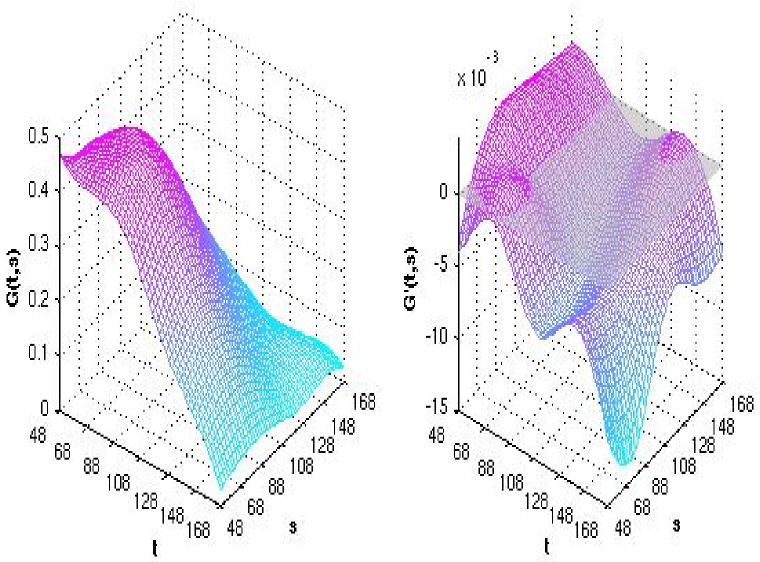
$$\frac{d^\nu}{dt^\nu}\int_{\mathcal{T}} G(t,s)\phi_k(s)ds = \lambda_k \frac{d^\nu}{dt^\nu}\phi_k(t),$$

  implying

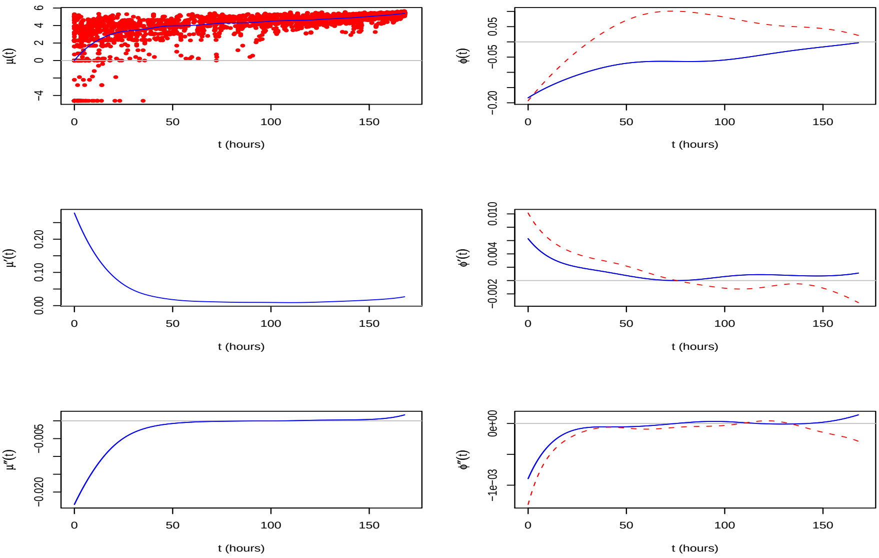
$$\phi_k^{(\nu)}(t) = \frac{1}{\lambda_k}\int_{\mathcal{T}} \frac{\partial^\nu}{\partial t^\nu} G(t,s)\phi_k(s)ds.$$



Locations of all pairs of points where bids are recorded for auction data.



Estimated covariance surface from all pairs and estimated partial derivative surface for auction data.



Estimates of mean and first two eigenfunctions and their first two derivatives for auction data.

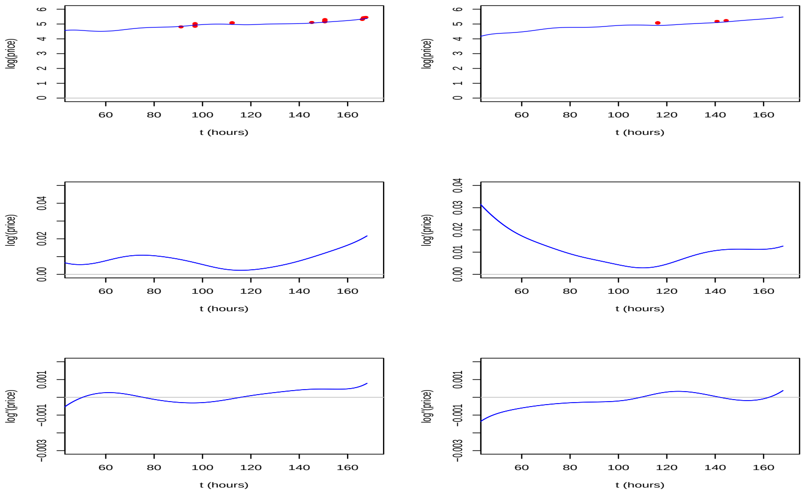
DERIVATIVES OF TRAJECTORIES

- Obtain

$$\hat{X}_{i,K}^{(\nu)}(t) = \hat{\mu}^{(\nu)}(t) + \sum_{k=1}^{K} \hat{A}_{ik}\hat{\phi}_k^{(\nu)}(t).$$

for the derivatives of the random trajectories X_i.

- Choosing the number of included components K: e.g. by Fraction of variance explained
- Asymptotic convergence results and confidence intervals for the case of a Gaussian process
- In simulations, this differentiation method works much better than single curve derivative estimation (splines, kernels, ...)



Fitted price trajectories and their first two derivatives for two auctions.

DYNAMICS OF GAUSSIAN PROCESSES

From the Karhunen-Loève representation of processes X, obtain for the covariance function for derivatives

$$\mathrm{cov}\{X^{(\nu_1)}(t), X^{(\nu_2)}(s)\} = \sum_{k=1}^{\infty} \lambda_k \phi_k^{(\nu_1)}(t)\phi_k^{(\nu_2)}(s),\ \nu_1, \nu_2 \in \{0, 1\},\ s, t \in \mathcal{T}$$

Assuming Gaussianity of X,

$$\begin{pmatrix} X^{(1)}(t) - \mu^{(1)}(t) \\ X(t) - \mu(t) \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{\infty} A_k \phi_k^{(1)}(t) \\ \sum_{k=1}^{\infty} A_k \phi_k(t) \end{pmatrix}$$

$$\sim N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)^2 & \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t) \\ \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t) & \sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2 \end{pmatrix}\right)$$

EMPIRICAL DIFFERENTIAL EQUATION

Population level: $E\{X^{(1)}(t) - \mu^{(1)}(t) \mid X(t)\} = \beta(t)\{X(t) - \mu(t)\}$

Subject level:

$$X^{(1)}(t) - \mu^{(1)}(t) = \beta(t)\{X(t) - \mu(t)\} + Z(t),\ \ t \in \mathcal{T},$$

with varying coefficient function

$$\begin{aligned} \beta(t) &= \frac{\operatorname{cov}\{X^{(1)}(t), X(t)\}}{\operatorname{var}\{X(t)\}} = \frac{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t)}{\sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2} \\ &= \frac{1}{2}\frac{d}{dt}\log[\operatorname{var}\{X(t)\}],\ \ t \in \mathcal{T}, \end{aligned}$$

and Gaussian drift process Z.

DRIFT PROCESS

Gaussian drift process is such that
(i) $Z(t)$, $X(t)$ are independent at each $t \in \mathcal{T}$; (ii) $E\{Z(t)\} = 0$;
(iii) Z has the representation

$$Z(t) = \sum_{k=1}^{\infty} \sqrt{\frac{\lambda_k}{2T^3}}(2k-1)\pi \int_0^T \sin\{\frac{(2k-1)\pi}{2T}u\} \times\{\phi_k^{(1)}(t) - \beta(t)\phi(t)\}\, dW(u)$$

Integral equation version

$$\begin{aligned} X(t) &= X(s) + \{\mu(t) - \mu(s)\} \\ &+ \int_s^t \beta(u)\{X(u) - \mu(u)\}\, du + \int_s^t Z(u)\, du, \end{aligned}$$

for any $s, t \in \mathcal{T}$, $s < t$.

LEARNING GAUSSIAN DYNAMICS

- For varying coefficient function β use plug-in estimates

$$\hat{\beta}(t) = \frac{\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^{(1)}(t) \hat{\phi}_k(t)}{\sum_{k=1}^{K} \hat{\lambda}_k \hat{\phi}_k^2(t)}.$$

- dynamic regression to the mean (negative β)

- dynamic exponential growth (positive β)

- Interpretation within population model
$E\{X^{(1)}(t) - \mu^{(1)}(t) \mid X(t)\} = \beta(t)\{X(t) - \mu(t)\}$

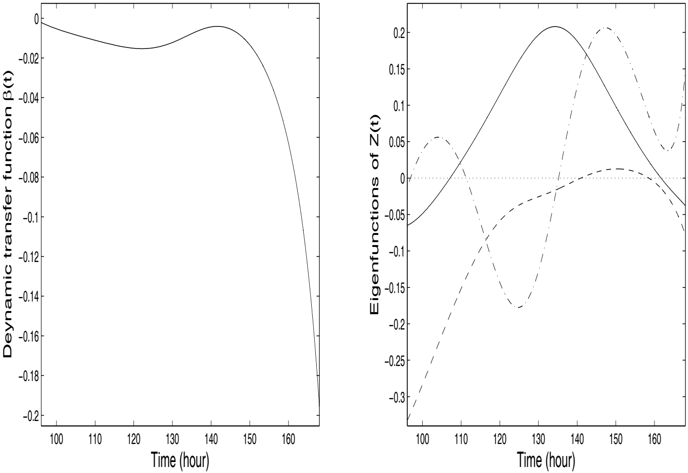
For drift process Z

$\mathrm{var}(Z(t)) =$
$\left(\sum_k \lambda_k(\phi_k^{(1)}(t))^2 \sum_k \lambda_k \phi_k^2(t) - \{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t)\}^2\right) / \sum_k \lambda_k \phi_k^2(t)$
and

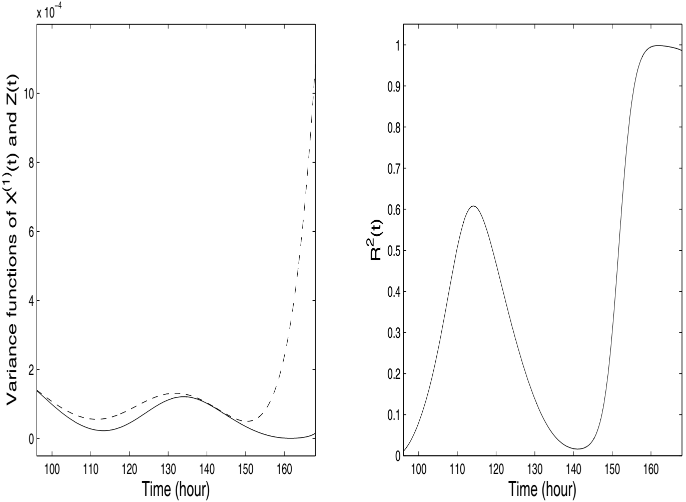
$$\mathrm{var}\{X^{(1)}(t)\} = \beta(t)^2 \mathrm{var}\{X(t)\} + \mathrm{var}\{Z(t)\}.$$

Then the fraction of the variance of $X^{(1)}(t)$ explained by the deterministic part of the differential equation is given by:

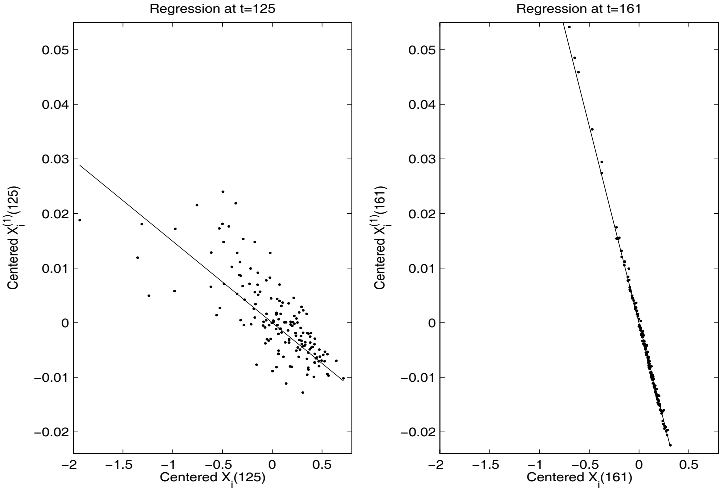
$$R^2(t) = \frac{\mathrm{var}\{\beta(t)X(t)\}}{\mathrm{var}\{X^{(1)}(t)\}} = \frac{\{\sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)\phi_k(t)\}^2}{\sum_{k=1}^{\infty} \lambda_k \phi_k(t)^2 \sum_{k=1}^{\infty} \lambda_k \phi_k^{(1)}(t)^2}.$$



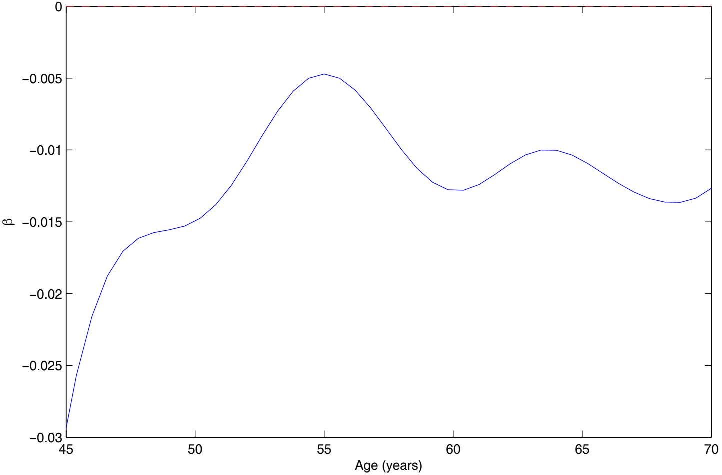
Left: Smooth estimate of the dynamic varying coefficient function β for auction data. Right: Smooth estimates of the first (solid), second (dashed) and third (dash-dotted) eigenfunction of drift process Z.



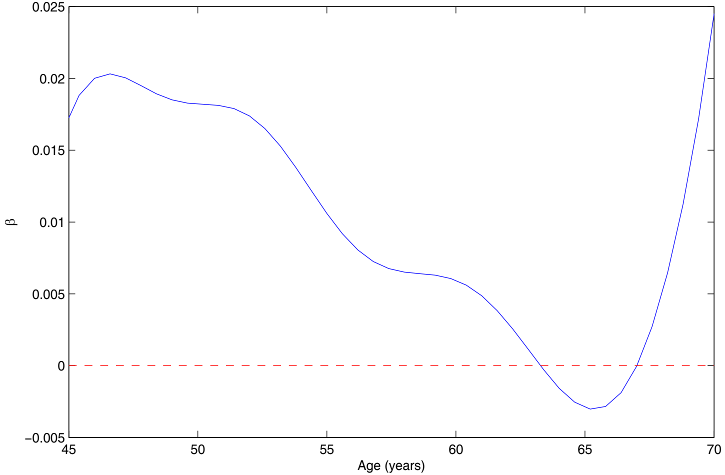
Left: Smooth estimates of the variance functions of $X^{(1)}(t)$ (dashed) and $Z(t)$ (solid). Right: Smooth estimate of $R^2(t)$, the variance explained by the deterministic part of the dynamic equation at time t.



Regression of $X_i^{(1)}(t)$ on $X_i(t)$ (both centered) at $t = 125$ hours (left panel) and $t = 161$ hours (right panel), respectively, with regression slopes $\beta(125) = -.015$ and coefficient of determination $R^2(125) = 0.28$, respectively, $\beta(161) = -.072$ and $R^2(161) = 0.99$.



Smooth estimate of the dynamic varying coefficient function β for Body Mass Index (BLSA).



Smooth estimate of the dynamic varying coefficient function β for Systolic Blood Pressure (BLSA).

LEARNING DYNAMICS – NON-GAUSSIAN CASE

- Data Model. For n realizations X_i of an underlying process X, have N_i measurements Y_{ij} ($i = 1, \ldots, n$, $j = 1, \ldots, N_i$),

$$Y_{ij} = Y_i(t_{ij}) \quad = \quad X_i(t_{ij}) + \epsilon_{ij},$$

with iid zero mean finite variance measurement errors ϵ_{ij}.

- Linear Gaussian Dynamics. As before, with varying coefficient function β,

$$X'(t) \quad = \quad \mu_{X'}(t) + \beta(t)\{X(t) - \mu_X(t)\} + Z_2(t),$$

where Z_2 is a zero mean drift process with $\mathrm{cov}\{Z_2(t), X(t)\} = 0$.

- General Dynamics. There always exists a function f with

$E\{X'(t) \mid X(t)\} = f\{t, X(t)\}, \quad X'(t) = f\{t, X(t)\} + Z(t)$,

with $E\{Z(t) \mid X(t)\} = 0$ almost surely and where f is unknown. Learning dynamics corresponds to inferring f.

- Special Case: Autonomous Dynamics.

$$E\{X'(t) \mid X(t)\} = f_1(X(t)), \quad f_1 \text{ unknown}$$

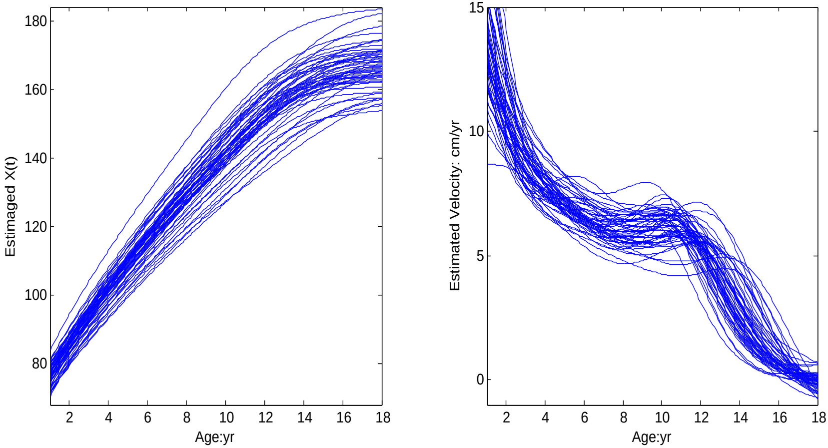
- Parametric Dynamics. Parametric differential equations

$$X_i'(t) = g\{t, X_i(t), \theta_i\}$$

require extensive knowledge of underlying system – often incorrect and hard to fit. Not much known for incorporating random effects θ_i.

BERKELEY LONGITUDINAL GROWTH STUDY

- Dynamics of Human Growth of Interest
- Nonlinear Parametric Models: Preece-Baines, Triple-Logistic
  Subject-by-subject fitting, limited efficiency
- Berkeley Growth Study – 54 girls with 31 height measurements for ages 1 to 18, recorded at different time intervals, ranging from three months (from 1 to 2 years old), six months (from 8 to 18 years old), to one year (from 3 to 8 years old).
- Learning dynamics:
  – Gain a better understanding of the growth process.
  – Distinguish between normal and pathological patterns of development.



Left panel: Estimated growth curves for 54 girls. Right panel: Estimated growth velocity trajectories for 54 girls.

ESTIMATING THE DRIVING FUNCTION f

Adopt a two-step kernel smoothing approach to obtain an estimator for f in $E\{X'(t) \mid X(t)\} = f\{t, X(t)\}$:

- Step 1: Obtaining estimates for $X(t)$ and $X'(t)$:

$$\widehat{X}_i(t) = \frac{1}{h_X} \sum_{j=1}^{N_i} \int_{s_{j-1}}^{s_j} Y_{ij} K\left(\frac{u-t}{h_X}\right) du,$$

$$\widehat{X'}_i(t) = \frac{1}{h_{X'}^2} \sum_{j=1}^{N_i} \int_{s_{j-1}}^{s_j} Y_{ij} K_2\left(\frac{u-t}{h_{X'}}\right) du,$$

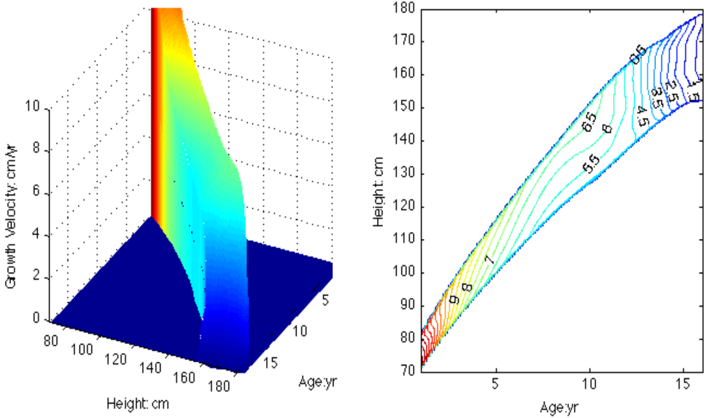
  where $s_j = (t_{ij} + t_{i,j+1})/2$ and $h_X > 0$ and $h_{X'} > 0$ are smoothing bandwidths.

- Step 2: Trajectory estimates $\widehat{X}(t)$ and $\widehat{X'}(t)$ from Step 1 are combined to obtain a Nadaraya–Watson kernel estimator for f,

$$\widehat{f}(t,x) = \frac{\sum_{i=1}^{n} K\{\frac{\widehat{X}_i(t)-x}{b_X}\}\widehat{X'}_i(t)}{\sum_{i=1}^{n} K\{\frac{\widehat{X}_i(t)-x}{b_X}\}} .$$

  utilizing bandwidths $b_X > 0$.

- Under regularity conditions, this gives consistent estimators.



Left panel: Estimated surface $\hat{f}(t, x)$ on a curved domain, characterizing the deterministic part of the nonlinear dynamic model. Right panel: Contour plot of the surface $\hat{f}(t, x)$.

DECOMPOSING VARIANCE

- Since $\mathrm{var}\{X'(t)\} = \mathrm{var}[f\{t, X(t)\}] + \mathrm{var}\{Z(t)\}$, on subdomains where the variance of the drift process $\mathrm{var}\{Z(t)\}$ is small, the deterministic approximation

$$X'(t) = f\{t, X(t)\} \quad (t \in \mathcal{T}),$$

  is reasonable. Then future changes of individual trajectories are easily predictable.

- Fraction of the variance of $X'(t)$ that is explained by the deterministic part

$$R^2(t) = \frac{\mathrm{var}[f\{t, X(t)\}]}{\mathrm{var}\{X'(t)\}} = 1 - \frac{\mathrm{var}\{Z(t)\}}{\mathrm{var}\{X'(t)\}} \ .$$

- Quantify predictability by

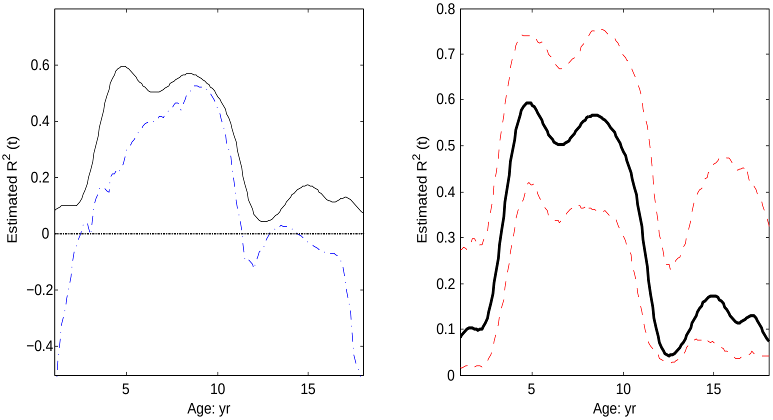
$$S(t,x) = \frac{f^2(t,x)}{E\{X'^2(t) \mid X(t)=x\}} = \frac{f^2(t,x)}{f^2(t,x) + \mathrm{var}\{Z(t) \mid X(t)=x\}}$$

  When $S(t,x)$ is close to one, then $f^2(t,x)$ is large compared to $\mathrm{var}\{Z(t) \mid X(t)=x\}$ and the process is well predictable when $X(t)=x$.

- Diagnostics for linearity. For the coefficient of determination for the linear dynamic model

$$R_L^2(t) = \frac{\mathrm{var}\,\{\beta(t)X(t)\}}{\mathrm{var}\{X'(t)\}}$$

  one expects that $R^2(t) \geq R_L^2(t)$ On subdomains of $\mathcal{T}$ where $R(t)$ is close to $R_L(t)$, one may infer that the data-driven differential equation is reasonably linear.



Left panel: Estimated coefficients of determination $\widehat{R}^2(t)$, corresponding to the fraction of variance explained by the deterministic part of the nonlinear dynamic model (solid), in comparison with the corresponding fractions of variance $\widehat{R}_L^2(t)$ explained by linear dynamics (dot-dashed). Right panel: 95% bootstrap confidence interval for $R^2(t)$.

- Linear concurrent model. Relating two stochastic processes $X(t)$ and $U(t)$ at each time $t \in \mathcal{T}$, the linear concurrent model captures a linear relationship between X and U through a deterministic function $\beta(t)$,

$$U(t) = \mu_U(t) + \beta(t)\{X(t) - \mu_X(t)\} + Z_2(t),$$

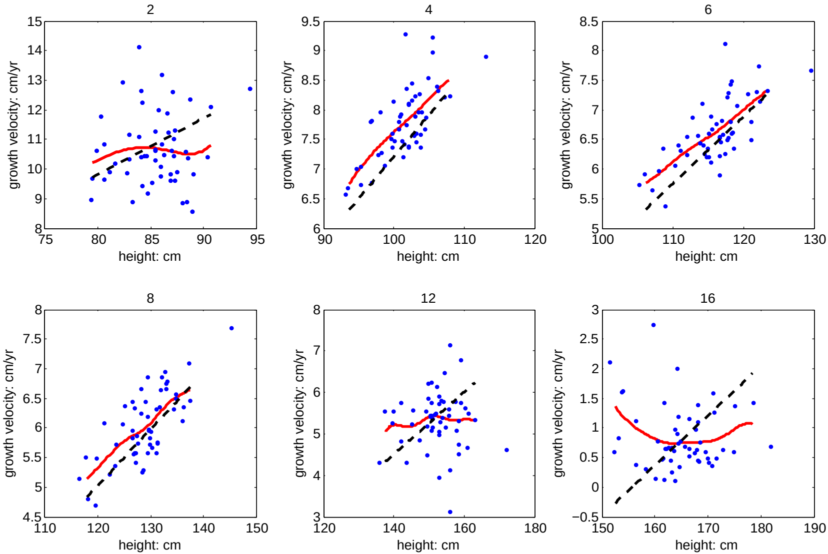
where $Z_2(t)$ is a zero mean drift process with $\mathrm{cov}\{Z_2(t), X(t)\} = 0$.

- Nonlinear concurrent model. Proposed methodology covers the case where the link between $U(t)$ and $X(t)$ is nonlinear,

$$U(t) = f\{t, X(t)\} + Z(t) \ ,$$

with $E\{Z(t) \mid X(t)\} = 0$ almost surely and $f\{t, X(t)\} = E\{U(t) \mid X(t)\}$. Can establish consistency and rates of convergence for two-step estimators.

- Learning Gaussian dynamics works for sparse data, learning non-Gaussian dynamics is viable only for dense data



Each of the panels, arranged for ages $t = 2, 4, 6, 8, 12$, from left to right and top to bottom, respectively, illustrates estimates $\hat{f}(t, \cdot)$ of the deterministic part of the nonlinear dynamic model (solid), the linear estimates (dashed) and the scatterplot of observed data pairs $(x(t), x^{(1)}(t))$.

PART II

FUNCTIONAL REGRESSION

FUNCTIONAL REGRESSION MODELS

X	$\mapsto$	Y	
$\mathbb{R}^d$		$\mathbb{R}$	Multiple Regression, GLM
$\mathbb{R}^{d_1}$		$\mathbb{R}^{d_2}$	Multivariate Regression
L^2		$\mathbb{R}$	"Functional Predictor Models"
$\mathbb{R}^d$		L^2	"Functional Response Models"
L^2		L^2	"Function to Function Regression"

MODELING FUNCTIONAL PREDICTORS

1. Functional Linear Regression

Idea: Extending the multivariate linear regression model $E(Y|X) = BX$ to functional data $(X(t), Y)$ or $(X(t), Y(t))$:

$$E(Y|X) \;=\; \mu_Y + \int (X(s) - \mu_X(s))\beta(s)\, ds,$$

the functional linear regression model with regression parameter function β and scalar responses (also generalized version by including link function (GFLM));

$$E(Y(t)|X) \;=\; \mu_Y(t) + \int (X(s) - \mu_X(s))\beta(s,t)\, ds,$$

model with functional responses (Ramsay & Dalzell 1991; Grenander 1950)

2. Functional Nonparametric Regression

$$E(Y|X) = \mu_Y + g(X)$$

for "smooth" function g, in analogy to nonparametric regression (Ferraty & Vieu 2006)

Problem: Curse of dimensionality, as predictor is infinite-dimensional. The infinite-dimensional curse can be quantified by using results on small ball probabilities for stochastic processes (Hall, M, Yao 2009).

$\Rightarrow$ Require new models that fall between these extremes

PRINCIPAL COMPONENT REPRESENTATION OF FUNCTIONAL LINEAR REGRESSION

With predictor representations

$$X(s) = \mu_X(s) + \sum_{k=1}^{\infty} A_k \phi_k(s)$$

obtain from normal equations for the functional linear model (FLM) $E(Y|X) = \mu_Y + \int \beta(s)(X(s) - \mu_X(s))ds$:

$$\beta(s) = \sum_{k=1}^{\infty} \frac{E(A_k Y)}{E(A_k^2)} \phi_k(s) = \sum_{k=1}^{\infty} \beta_k \phi_k(s),$$

implying

$$E(Y|X) = \sum_k \beta_k A_k$$

- Estimation: Can directly apply PACE to obtain all needed estimates. Alternative: Representation of regression parameter function β by B-splines or other basis expansions (Cardot et al. 1999; James et al. 2001)
- Special features of functional linear regression with PACE: Perturbation theory directly applicable for asymptotics Functional regression diagnostics: Based on decomposition of Linear Functional Regression into series of simple linear regressions on FPCs; e.g., Functional Cook's distance and Functional hat matrix (Chiou & M 2007)
- Choice of included predictor components: Nested sequence, can use AIC-type criteria

FUNCTIONAL RESPONSE MODELS

- Response process with mean function μ_Y, eigenfunctions ψ_m and functional principal component (FPC) scores B_m:

$$Y(t) = \mu_Y(t) + \sum_{m=1}^{\infty} B_m \psi_m(t)$$

- Given covariates $\boldsymbol{Z} \in \mathcal{R}^p$, this suggests conditioning approach

$$\begin{aligned} E\{Y(t)|\boldsymbol{Z}\} &= \mu_Y(t) + \sum_{m=1}^{\infty} E(B_m|\boldsymbol{Z})\psi_m(t) \\ &\leftarrow \hat{\mu}_Y(t) + \sum_{m=1}^{M} \hat{\eta}_m(\boldsymbol{Z})\hat{\psi}_m(t) \end{aligned}$$

  with nonparametric or semiparametric (e.g., single index) regressions

$$\hat{\eta}_m(\boldsymbol{Z}) = \hat{E}(B_m|\boldsymbol{Z}).$$

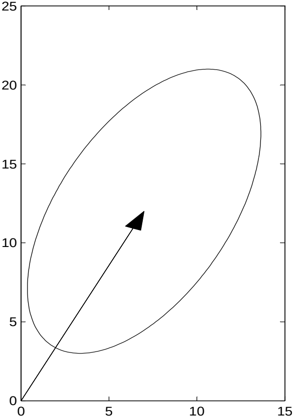
- Mean response models:

$$E(Y(t)|\boldsymbol{Z} = z) = \mu(t, z)$$

- Product Model:

$$\mu(t,z) = \mu_Y(t)\theta(z), \quad E\{Y(t)\} = \mu_Y(t),\ E\{\theta(Z)\} = 1,$$

  product form is motivated empirically (Chiou et al. 2004)
- Alternative: Functional ANOVA (Brumback & Rice 1998).



0
5
10
15
20
25
0
5
10
15
0
5
10
15
20
25
0
5
10
15

Least squares solution:

$$\theta(z) = \arg\min_{\theta}\{\int_T [E(Y(t)|\boldsymbol{Z} = z) - \mu_0(t)\theta]^2\, dt\},$$

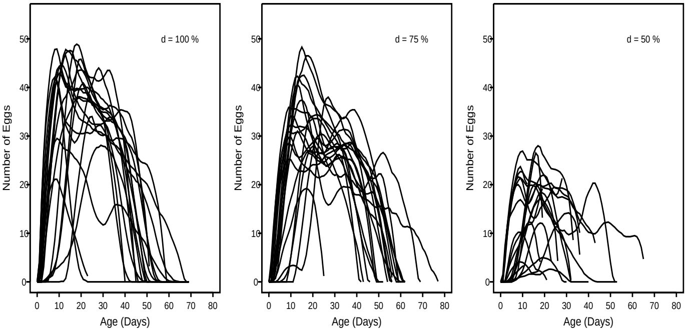
implies

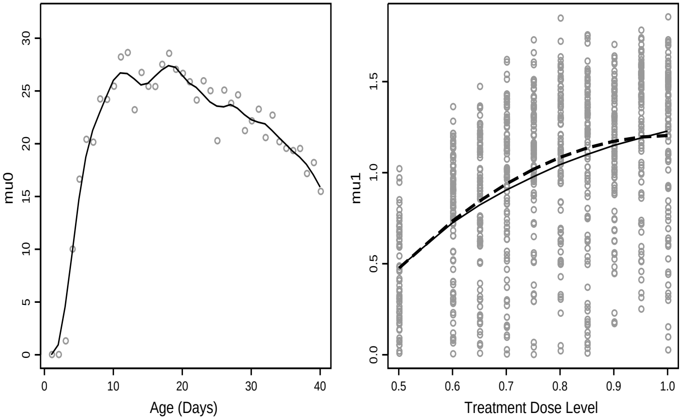
$$\theta(z) = \frac{\int \mu_0(t) E(Y(t)|\boldsymbol{Z} = z)\, dt}{\int \mu_0^2(t)\, dt}.$$

Add single index assumption: $\theta(z) = \mu_1(\gamma' z)$ for a smooth function μ_1 and a vector γ, $|\gamma| = 1$.
Consequence of above: $E\{\mu_1(\gamma' \boldsymbol{Z})\} = 1$.

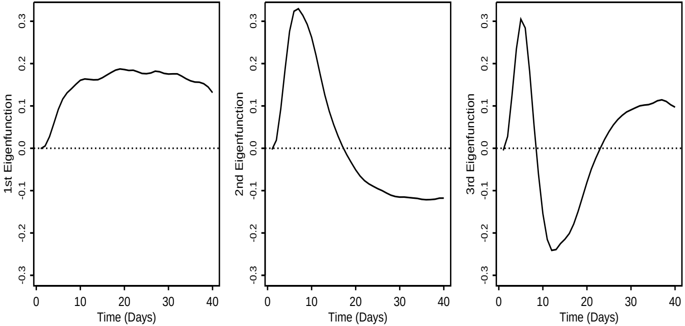
FUNCTIONAL DOSE-RESPONSE

Medfly (*Ceratitis capitata*) experiments on reproductive behavior in response to nutrition amount (Carey laboratory at UC Davis). Predictor is amount of protein in diet, between 30 and 100%, response is daily egg-laying profile ($n = 874$, $m = 10$ dose levels)


d = 100 %
d = 75 %
d = 50 %
Number of Eggs
Age (Days)
0
10
20
30
40
50
60
70
80



Function estimates of the mean function and multiplicative components, with overall mean function $\hat{\mu}_{\mathbf{0}}(t)$ (left) and multiplicative effect function $\hat{\mu}_{\mathbf{1}}(z)$ (right).



The first three estimated eigenfunctions $\{\hat{\psi}_{\boldsymbol{k}}\}_{\boldsymbol{k}=\mathbf{1},\ldots,\mathbf{3}}$. The first eigenfunction explains 35.31%, the second additional 16.84%, and the third additional 8.82% of total variation.

FLM FOR FUNCTIONAL PREDICTORS AND RESPONSES

Extending the multivariate linear regression model $E(Y|X) = BX$ to functional data $(X(t), Y(t))$:

$$E(Y(t)|X) = \mu(t) + \int X(s)\beta(s,t)\,ds.$$

Estimation of the parameter function $\beta(\cdot,\cdot)$ is an inverse problem.

- Idea: Extending the least squares normal equation $\mathrm{cov}(X, Y) = \mathrm{cov}(X)B$.
- "Functional Normal Equation" (He et al. 2000,2003) For auto-covariance operator A_G of predictors X and

$$r_{XY}(s,t) = \mathrm{cov}\left[X(s), Y(t)\right] : \qquad r_{XY} = A_G \beta.$$

- Since A_G is a compact operator in L^2, equation is not invertible. Require functional generalized inverse: Well-defined under regularity conditions and obtained by regularization – truncation of included components or penalty (Cai & Hall 2006, Hall & Horwitz 2007).

Solution of the functional normal equation:

$$\beta^*(s,t) = \sum_{j,k=1}^{\infty} \frac{\mathrm{cov}(\xi_j, \zeta_k)}{\mathrm{var}(\xi_j)} \varphi_j(s)\psi_k(t).$$

Practical solution: By discretization.

Existence of solution in image space of A_G.

REPRESENTATIONS OF FLR

With predictor and response representations

$$X(s) = \mu_X(s) + \sum_{k=1}^{\infty} A_k \phi_k(s), \; Y(t) = \mu_Y(t) + \sum_{m=1}^{\infty} B_m \psi_m(t)$$

obtain from normal equations for the model
$E(Y(t)|X) = \mu_Y(t) + \int \beta(s,t)(X(s) - \mu_X(s))ds$
the representation

$$\beta(s,t) = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} \frac{E(A_k B_m)}{E(A_k^2)} \phi_k(s) \psi_m(t) = \sum_{m=1}^{\infty} \sum_{k=1}^{\infty} \beta_{mk} \phi_k(s) \psi_m(t)$$

which implies $E(B_m|X) = \sum \beta_{mk} A_k$ and (as A_k are uncorrelated)

$$E(B_m|A_k) = E[E(B_m|A_1, A_2, \ldots)|A_k] = E[E(B_m|X)|A_k] = \beta_{mk} A_k.$$

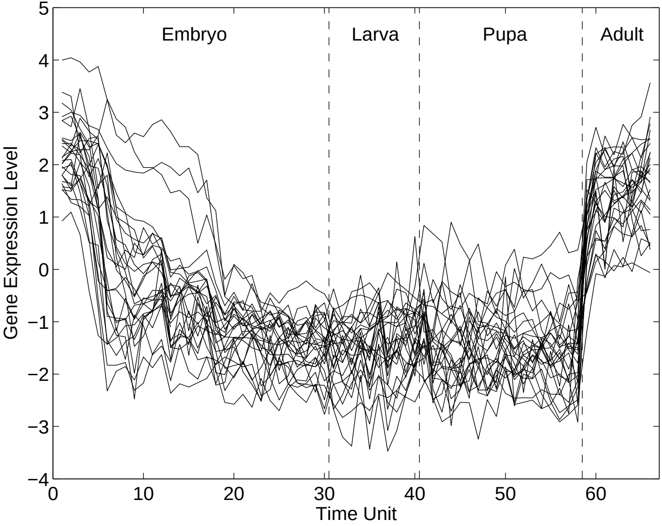
NOTES

- Other basis representations (wavelets, B-splines) have been considered, eigen-representation has advantages due to uncorrelatedness of scores (independence in Gaussian case) and (relative) sparseness of representation (often only few base functions needed, especially for prediction purposes)
- Obtaining estimated FPC scores $\hat{A}_k, \hat{B}_m$ through the PACE method (Yao et al 2005), then $\hat{\beta}_{mk} = \widehat{\text{cov}}(\hat{A}_k, \hat{B}_m)/\hat{\lambda}_k$, ie, all it takes is a series of simple linear regressions through the origin.
- Inference: Simultaneously sample predictor and response data for randomly resampled subjects, then recalculate functional regression and obtain bootstrap confidence regions.
- Alternatively, separately resample predictor and response data and obtain bootstrap distribution of suitable statistic under null hypothesis of no functional relationship.

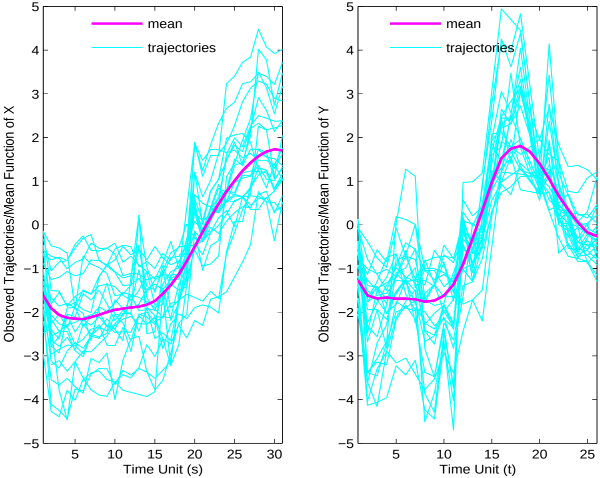
DROSOPHILA LIFE CYCLE GENE EXPRESSION

Consider gene time course data, where gene expression is repeatedly measured for:

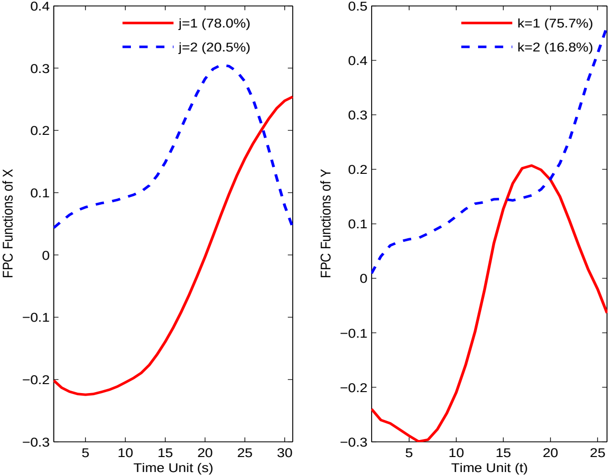
- 23 “muscle specific” genes: tissue-specific, muscle development
- 22 “skeleto-neural” genes



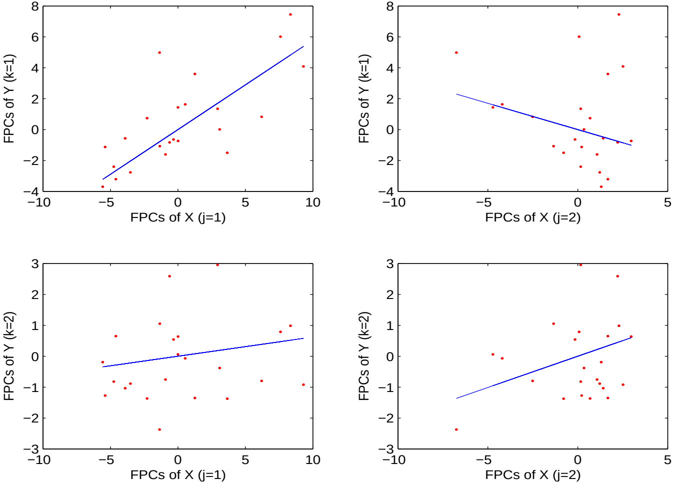
A subset of observed gene expression profiles (strict maternal genes). Each profile (or curve) is composed of expression levels of one gene at different time points.



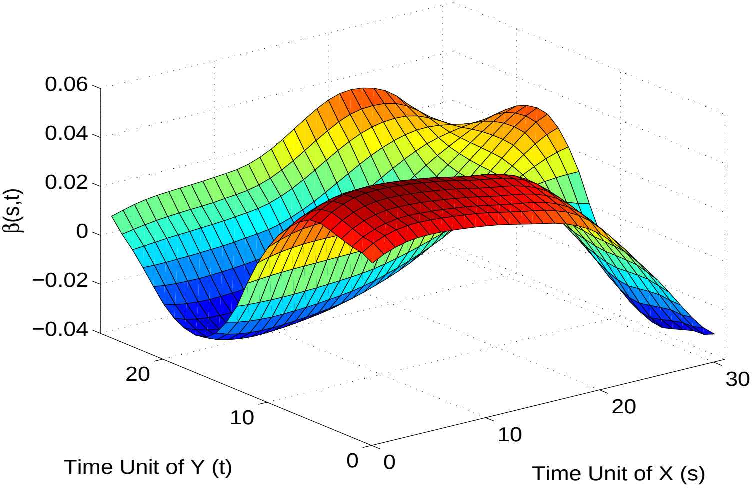
Observed trajectories and estimated mean function for muscle-specific genes for predictor profiles X (corresponding to gene expression profiles in embryo phase, left panel) and for response profiles Y (profiles for pupa-adult phase, right panel)



First two estimated eigenfunctions for temporal gene expression trajectories for the muscle-specific genes in embryo phase (predictors X, left panel) and pupa-adult phase (responses Y, right panel).



Scatterplots of functional principal component scores ζ_k of response trajectories versus ξ_j of predictor trajectories, for $j, k = 1, 2$, for muscle-specific genes



Estimated regression parameter function $\hat{\beta}(s, t)$ for muscle-specific genes with embryo phase as predictor $X(s)$ (plotted towards the right) and pupa-adult phase as response $Y(t)$ (plotted towards the left)

FUNCTIONAL COEFFICIENT OF DETERMINATION AND DIAGNOSTICS

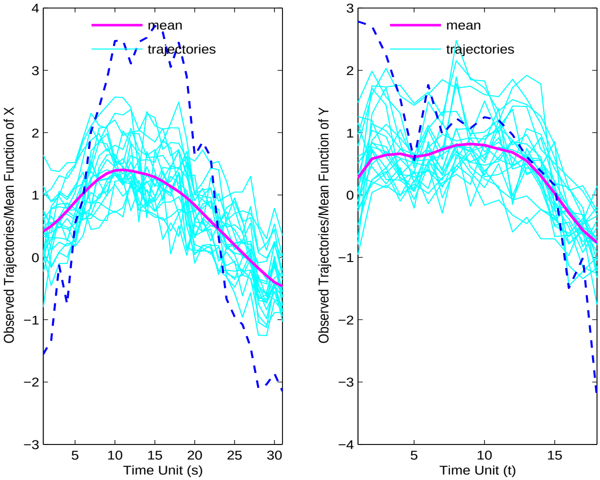
Extension from the multiple linear regression case:

$$R^2 = \frac{\int_{\mathcal{T}} \mathrm{var}(E[Y(t)|X])dt}{\int_{\mathcal{T}} \mathrm{var}(Y(t))dt} = \sum_{j=1}^{\infty} \frac{\sum_{k=1}^{\infty} R_{kj}^2 \tau_k}{\sum_{k=1}^{\infty} \tau_k},$$

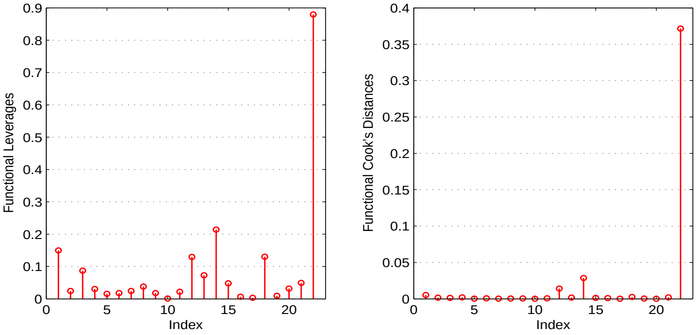
where

$$R_{kj}^2 = \frac{[\mathrm{cov}(\xi_j, \zeta_k)]^2}{\lambda_j \tau_k}$$

are the coefficients of determination for the simple linear regressions of ζ_k on ξ_j. Obtain estimate $R^2 = 0.85$ for muscle-specific genes ($p = 0.0010$ from bootstrap test)
Functional diagnostics can be obtained by a similar weighting scheme: Functional hat matrix, functional Cook's distance, etc.



Observed trajectories and estimated mean function for cytoskeleton/neural genes in embryo phase (for predictor X, left panel) and pupa phase (for response Y, right panel), respectively. Trajectories of gene CG2198 are dashed.



Functional leverages obtained as diagonal elements of functional hat matrix H (left panel) and functional Cook's distances (right panel) for the functional regression of cytoskeleton/neural genes.

FUNCTIONAL LINEAR MODEL FOR LONGITUDINAL DATA

Regress processes $Y(\cdot)$ on processes $X(\cdot)$ under sparse data situation. Notation:

$X_i(s)$ on $[0, \mathcal{S}]$: smooth predictor curve

U_{il}: measurements of $X_i(\cdot)$ at S_{il}, $1 \leq i \leq n, 1 \leq l \leq L_i$

$Y_i(t)$ on $[0, \mathcal{T}]$: smooth response curve

V_{ij}: measurements of $Y_i(\cdot)$ at T_{ij}, $1 \leq j \leq N_i$

Functional Regression Model

$$E[Y(t)|X(\cdot)] = \mu_Y(t) + \int_0^{\mathcal{S}} \beta(s,t)X(s)ds.$$

$\beta(s,t)$: smooth regression function, $\int_0^{\mathcal{T}} \int_0^{\mathcal{S}} \beta^2(s,t)dsdt < \infty$.

Modelling Predictor and Response Curves:

$$U_{il} = X_i(S_{il}) + e_{il} = \mu_X(S_{il}) + \sum_{m=1}^{\infty} A_{im}\phi_m(S_{il}) + e_{il},$$

$$V_{ij} = Y_i(T_{ij}) + \epsilon_{ij} = \mu_Y(T_{ij}) + \sum_{k=1}^{\infty} B_{ik}\psi_k(T_{ij}) + \epsilon_{ij}.$$

BASIS REPRESENTATION

$$\beta(s,t) = \sum_{k,m=1}^{\infty} \frac{E[A_m B_k]}{E[A_m^2]} \phi_m(s)\psi_k(t)$$

Estimating $E[A_m B_k]$:

$$\widehat{E}[A_m B_k] = \int_0^{\mathcal{T}} \int_0^{\mathcal{S}} \hat{\phi}_m(s)\widehat{\Gamma}_{XY}(s,t)\hat{\psi}_k(t)dsdt,$$

where $\widehat{\Gamma}_{XY}(s,t)$ is local linear smoothing estimate of the covariance surface $\Gamma_{XY}(s,t) = \mathrm{cov}(X(s), Y(t))$.

CONDITIONAL METHOD

Objective: Predict trajectory Y^* of a new subject, given observations $U^* = (U_1^*, \cdots, U_{L^*}^*)^T$ of $X^*(\cdot)$.

$$
\begin{aligned}
E[Y^*(t)|X^*(\cdot)] &= \mu_Y(t) + \int_0^{\mathcal{S}} \beta(s,t) X^*(s) ds \\
&= \mu_Y(t) + \sum_{k,m=1}^{\infty} \frac{E[A_m B_k]}{E[A_m^2]} A_m^* \psi_k(t)
\end{aligned}
$$

Constraint: $\mu_Y(t) = \int_0^{\mathcal{S}} \beta(s,t) \mu_X(s) ds$.

PREDICTION OF $Y^*(t)$

$$\widehat{Y}^*_{KM}(t) = \hat{\mu}_Y(t) + \sum_{m=1}^{M} \sum_{k=1}^{K} \frac{\widehat{E}[A_m B_k]}{\widehat{E}[A_m^2]} \widehat{E}[A_m^* | U^*] \hat{\psi}_k(t),$$

where $\widehat{E}[A_m^*|U^*]$ is estimated by the conditional method, given observations $U^* = (U_1^*, \cdots, U_{L^*}^*)^T$ of $X^*(\cdot)$.

ASYMPTOTICS FOR FUNCTIONAL LINEAR REGRESSION

Consistency for $\hat{\beta}(s,t)$ and $\widehat{Y}_{KM}(t)$ under regularity conditions
Pointwise Bands for $Y(t)$

$$\widehat{Y}^{*}_{K,M}(t) \pm \Phi(1-\alpha/2)\sqrt{\hat{\omega}_{KM}(t,t)}$$

Functional R^2:

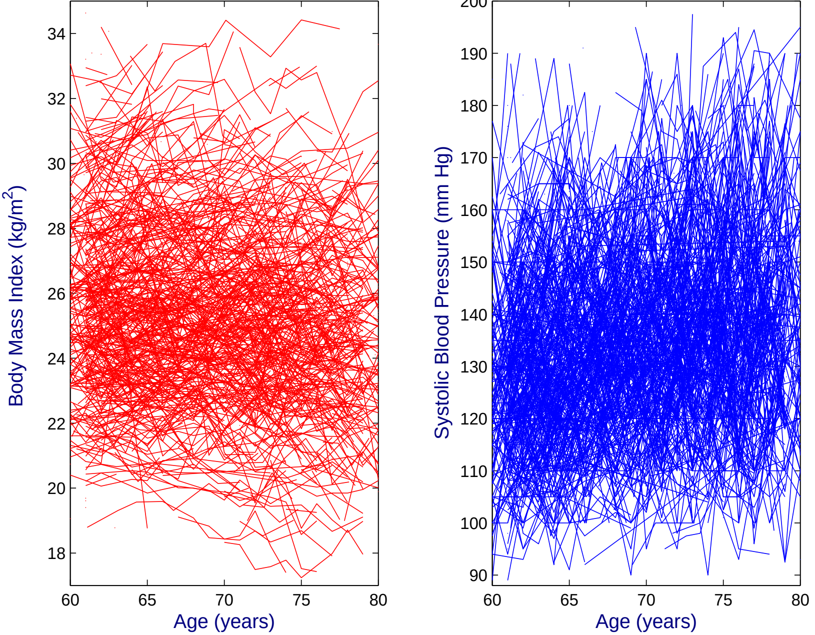
$$R^2 = \frac{\int_{\mathcal{T}} \mathrm{var}(E[Y(t)|X])dt}{\int_{\mathcal{T}} \mathrm{var}(Y(t))dt} = \frac{\sum_{k,m=1}^{\infty} \sigma_{km}^2/\rho_m}{\sum_{k=1}^{\infty} \lambda_k}$$

APPLICATION

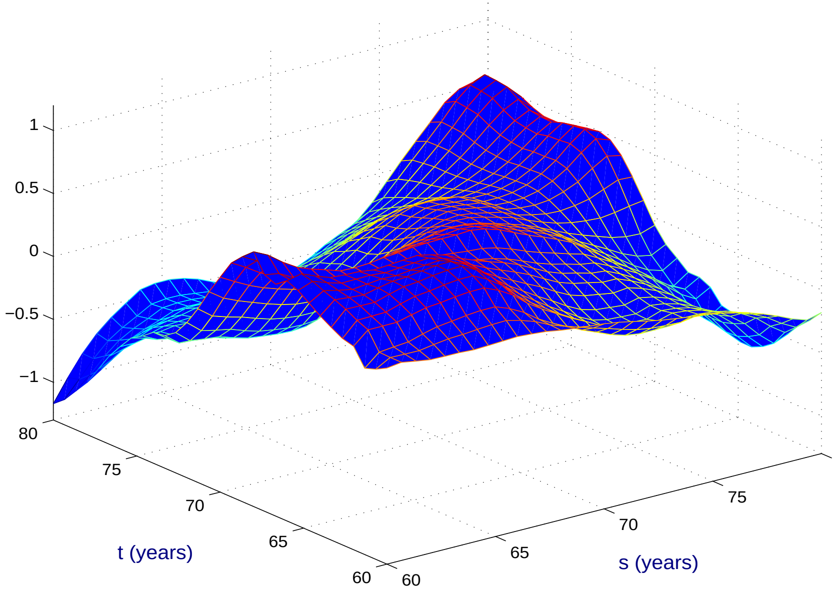
Functional Regression of Systolic Blood Pressure on Body Mass Index

Data: Body mass index (BMI) and systolic blood pressure (SBP) for 812 participants in the Baltimore Longitudinal Study on Aging

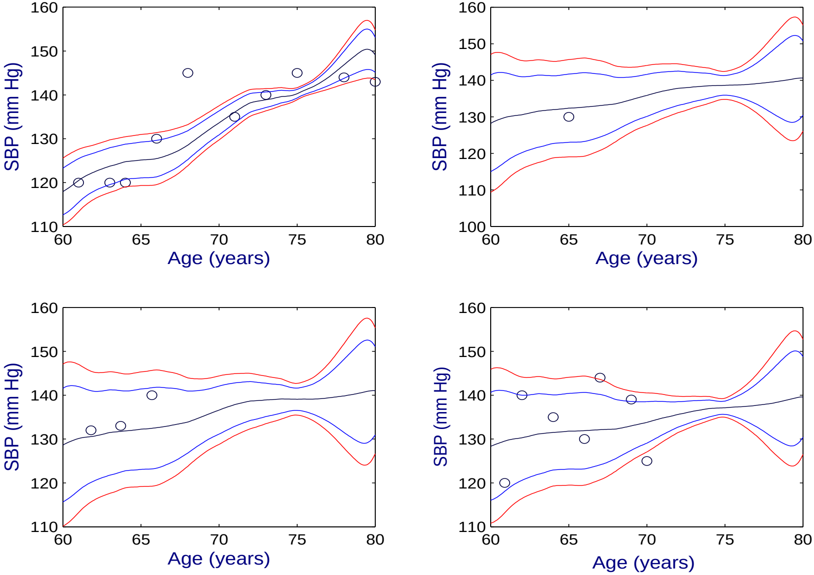
Irregular and Sparse Measurements
$R^2 = 0.13$



Observed paths of Body Mass Index (left) and Systolic Blood Pressure (right) for 812 participants.



Estimated regression function $\hat{\beta}(s, t)$, where the predictor (BMI) time is s (in years), and the response (SBP) time is t (in years).



Observed data (circles), predicted trajectories (black), 95% pointwise (blue) and simultaneous (red) bands obtained by one-leave-out analysis.

GENERALIZED FUNCTIONAL LINEAR MODEL

- Predictors $X(t) \in L^2$, Response $Y \in \mathbb{R}$
- Components: Parameter Function $\beta(\cdot)$, Link Function $g(\cdot)$, Variance Function $\sigma^2(\cdot)$

$$
\begin{aligned}
\eta_i &= \alpha + \int \beta(t) X_i(t)\, dw(t) \quad \text{linear predictors} \\
Y_i &= g(\eta_i) + e_i = \mu_i + e_i, \;\; i = 1, \ldots, n,
\end{aligned}
$$

with i.i.d. errors e_i, means $E(Y_i) = \mu_i = g(\eta_i)$ and $E(e|X(\cdot)) = 0, \mathrm{var}(e|X(\cdot)) = \sigma^2(\mu)$.

- If link function $g(\cdot)$ and variance function $\sigma^2(\cdot)$ are unknown and smooth, they can be estimated from the data.
- Applications of generalized functional linear model (GFLM): For example in classification, when Y denotes class membership and a binary link function (e.g., logistic link) is used.
- With orthonormal basis φ_j, $j \geq 1$,

$$X(t) = \sum_{j=1}^{\infty} \zeta_j \varphi_j(t), \quad \beta(t) = \sum_{j=1}^{\infty} \beta_j \varphi_j(t)$$

$$\int \beta(t) X(t)\, dw(t) = \sum_{j=1}^{\infty} \beta_j \zeta_j.$$

ANALYSIS

Writing $\zeta_j^{(i)} = \int X_i(t)\varphi_j(t)dw(t)$, consider a sequence of p_n-truncated models, $p_n \to \infty$ as $n \to \infty$,

$$
\begin{aligned}
Y_i^{(p_n)} &= g\left(\alpha + \sum_{j=1}^{p_n} \beta_j \zeta_j^{(i)}\right) \\
&+ e_i' \tilde{\sigma}\left(\alpha + \sum_{j=1}^{p_n} \beta_j \zeta_j^{(i)}\right), \quad i = 1, \ldots, n
\end{aligned}
$$

with standardized errors e_i'.

ESTIMATING EQUATION

- Given $p = p_n$, the solution of the quasi-score (estimating) equation

$$U(\beta) = \sum_{i=1}^{n} (Y_i - \mu_i) g'(\eta_i) \zeta^{(i)} / \sigma^2(\mu_i) = 0$$

  is the $(p+1)$-vector $\hat{\beta}$.
- This is the quasi-likelihood estimator for fixed p, given a single-index model with link function g, variance function $\sigma^2(\cdot)$ and predictors $\zeta_1^{(i)}, \ldots, \zeta_p^{(i)}$.
- Solution is numerically obtained by iterative weighted least squares (Newton-Raphson).

ASYMPTOTICS

Define $\Gamma = \Gamma_p = (\gamma_{kl})_{0 \le k,l \le p}, \quad \gamma_{kl} = E\left(\frac{g'^2(\eta)}{\sigma^2(\mu)}\zeta_k\zeta_l\right)$.

Under regularity conditions, as $n \to \infty$:

Theorem. For p_n-vectors $\beta = (\beta_1, \ldots, \beta_{p_n})$ and estimates $\hat{\beta}$,

$$\frac{n(\hat{\beta} - \beta)^T \, \Gamma_{p_n} (\hat{\beta} - \beta) - (p_n + 1)}{\sqrt{2\, p_n}} \xrightarrow{d} N(0, 1).$$

Theorem. For parameter functions $\beta(\cdot)$ and estimates $\hat{\beta}(\cdot)$: If

$$\sum_{j=p_n+1}^{\infty} \lambda_j^G \, [\int \beta(t)\varphi_j^G(t)\, dw(t)]^2 \, = \, o\left(\frac{\sqrt{p_n}}{n}\right),$$

then

$$\left[n d_G^2(\hat{\beta}, \beta) - (p_n + 1) \right] / \sqrt{2\, p_n} \xrightarrow{d} N(0,1).$$

Proof: Functional analysis and GLM arguments combined with CLT for Martingale Difference Schemes.

FURTHER EXTENSIONS OF THE FLM

"Classic" extensions: linear $\Rightarrow$ quadratic $\Rightarrow$ polynomial

The polynomial functional regression model (Yao & M 2010)

$$
\begin{aligned}
E(Y|X) = \alpha \quad & + \quad \int_{\mathcal{T}} \beta(t) X^c(t) dt + \int_{\mathcal{T}^2} \gamma(s,t) X^c(s) X^c(t) ds dt \\
& + \quad \int_{\mathcal{T}^3} \gamma_3(t_1,t_2,t_3) X^c(t_1) X^c(t_2) X^c(t_3) dt_1 dt_2 dt_3 + \dots \\
& + \quad \int_{\mathcal{T}^p} \gamma_p(t_1,\dots,t_p) X^c(t_1) \dots X^c(t_p) dt_1 \dots dt_p,
\end{aligned}
$$

with α as intercept and β, γ, γ_j, $3 \le j \le p$, as linear, quadratic and jth order regression parameter functions. In terms of FPCs,

$$
\begin{aligned}
E(Y|X) = \alpha \quad & + \quad \sum_{j_1 \ge 1} \beta_{j_1} A_{j_1} + \sum_{j_1 \le j_2} \gamma_{j_1 j_2} A_{j_1} A_{j_2} + \sum_{j_1 \le j_2 \le j_3} \gamma_{j_1 j_2 j_3} A_{j_1} A_{j_2} A_{j_3} \\
& + \quad \dots + \sum_{j_1 \le \dots \le j_p} \gamma_{j_1 \dots j_p} A_{j_1} \dots A_{j_p},
\end{aligned}
$$

model includes all interaction effects up to p time points.

FUNCTIONAL QUADRATIC REGRESSION

$$E(Y|X) = \alpha + \sum_{k=1}^{\infty} \beta_k A_k + \sum_{k=1}^{\infty} \sum_{\ell=1}^{k} \gamma_{k\ell} A_k A_\ell,$$

Quadratic diagonal case

$$E(Y|X) = \alpha + \sum_k \beta_k A_k + \sum_k \gamma_{kk} A_k^2.$$

With eigenvalues λ_k for X and covariance functions

$$\begin{aligned} C_1(t) &= \operatorname{cov}\{X(t), Y\} = \sum_{k=1}^{\infty} \eta_k \phi_k(t), \\ C_2(s,t) &= E\{X(s)X(t)Y\} = \sum_{k,\ell=1}^{\infty} \rho_{k\ell} \phi_k(s) \phi_k(t), \end{aligned}$$

least squares estimators are obtained via the representations
$\alpha = \mu_Y - \sum_k \gamma_{kk} \lambda_k, \quad \beta_k = \eta_k / \lambda_k, \quad \gamma_{k\ell} = \rho_{k\ell} / (\lambda_k \lambda_\ell),$
for $k < \ell, \quad \gamma_{kk} = (\rho_{kk} - \mu_Y \lambda_k)/(E(A_k^4) - \lambda_k^2).$

Can easily be implemented with PACE (`quadreg`, included in version 2.12).

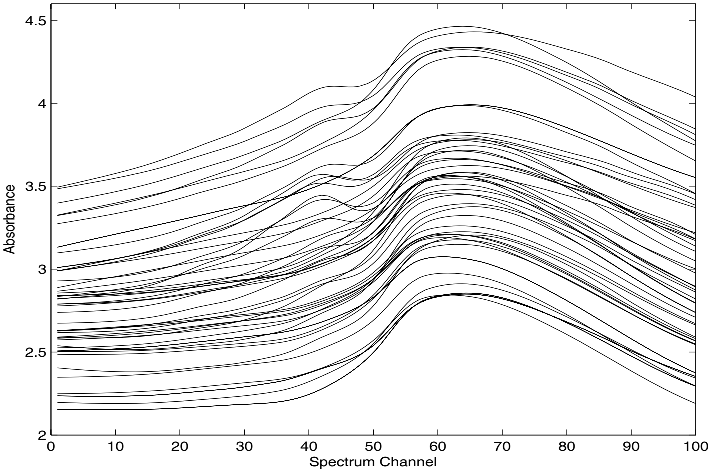
Asymptotics

Obtain consistent estimates and rates of convergence for parameter functions $\hat{\alpha} - \alpha = O_p(\alpha_n)$, $\|\hat{\beta} - \beta\| = O_p(\beta_n)$, $\|\hat{\gamma} - \gamma\| = O_p(\gamma_n)$ and for predicting new responses under either one of two assumptions:

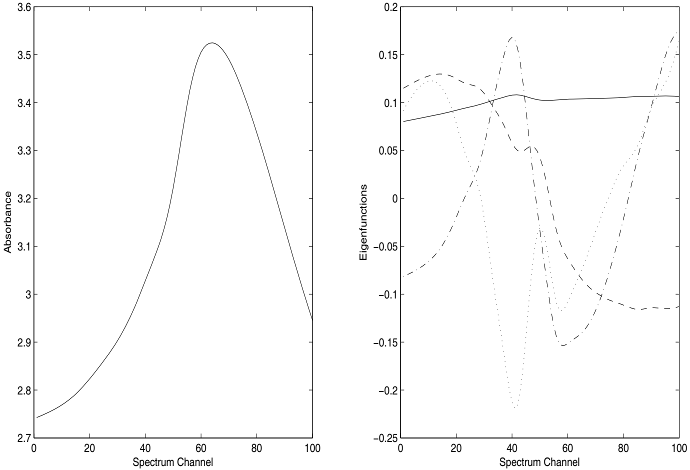
- Gaussian assumption on predictor processes X: Convergence rates for sparse irregular designs
- Densely observed functional predictors with noise; Gaussian assumption not needed for convergence rates

Note: The proofs for the two designs are quite different.

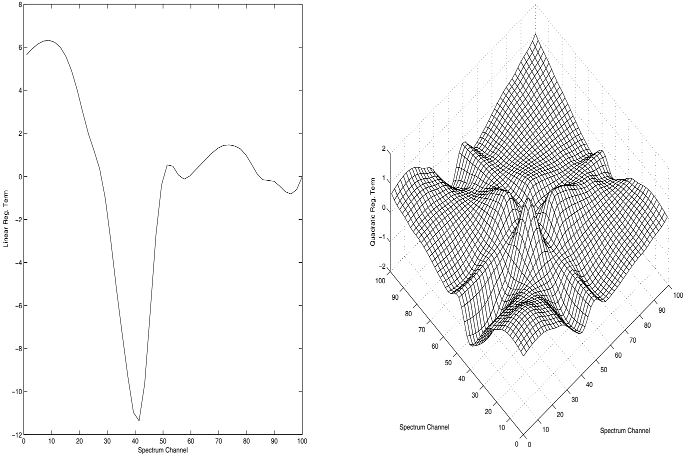
QUADRATIC FUNCTIONAL REGRESSION IN ACTION



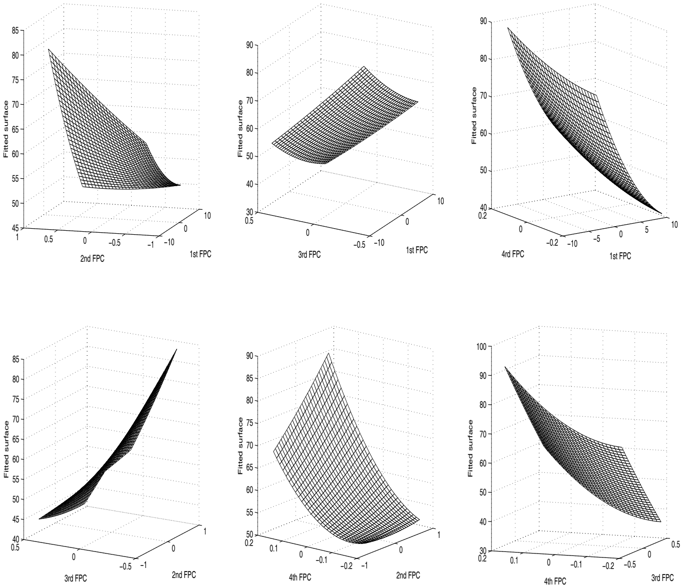
Log-transformed absorbance spectra for Tecator fat contents data, for subset of 50 meat specimen (a Chemometrics test set)



Mean function and four eigenfunctions for predictor processes



Estimates of linear parameter function β (left) and quadratic regression parameter surface γ (right). Leave-out prediction errors ranking: QFM $<$ Chemometrics-PLS $<$ FLM.



Sections through the fitted model $E(Y|A_1, A_2, A_3, A_4)$.

AN ADDITIVE EXTENSION OF THE FUNCTIONAL LINEAR MODEL (FLM)

The least squares parameter function in the FLM

$$E(Y|X) = \mu_Y + \int \beta(s)(X(s) - \mu_X(s))ds$$

has the representation

$$\beta(s) = \sum_m \sum_k \beta_k \phi_k(s) \text{ with } \beta_k = E(A_k Y)/E(A_k^2),$$

yielding

$$E(Y|X) = \sum_k \beta_k A_k.$$

This motivates the following extension:
Functional Additive Model

$$E(Y|X) = \sum_k f_k(A_k),$$

where f_k are smooth nonparametric functions; analogously for functional responses.

FUNCTIONAL ADDITIVE MODEL (FAM)

Assuming independent predictor scores A_j (automatically implied in the Gaussian case) we find

$$E(Y|A_k) = E\{E(Y|X)|A_k\} = E\{\sum_{j=1}^{\infty} f_j(A_j)|A_k\} = f_k(A_k).$$

Consequence: Functional Additive Model can be implemented simply by 1-d scatterplot smoothing of Y vs $\hat{A}_{ik}$ to obtain the defining functions f_k.

No backfitting iteration is needed: Fast and straightforward implementation with PACE. Analogously for functional regression model with scalar responses. For situations with several predictor functions within subjects: Can apply common additive model to ensemble of selected FPCs for all predictor functions.

ASYMPTOTICS FOR FAM

Employing PACE, one may show under regularity conditions that $\hat{f}_k$ is consistent for f_k and the prediction $\hat{E}(Y|X^*)$ is consistent for $E(Y|X^*)$ (M & Yao 2008)

Key steps for proof:

- Differences between A_{ik} and $\hat{A}_{ik}$ are asymptotically small enough to be negligible for the FAM smoothing steps.
- Perturbation analysis for linear operators, bounding the difference between operators A_G and $A_{\hat{G}}$.
- In the dense design case, obtain essentially 1-d rates of convergence for the component functions $\hat{f}_k$.

ADDITIVE EXTENSION OF THE FUNCTIONAL RESPONSE MODEL

Consider FLM with functional responses, with FPC representation

$$Y(t) = \mu_Y(t) + \textstyle\sum_m B_m \psi_m(t).$$

Then the least squares parameter function in the FLM

$$E(Y(t)|X) = \mu_Y(t) + \textstyle\int \beta(s,t)(X(s) - \mu_X(s))ds$$

has the representation

$\beta(s,t) = \sum_m \sum_k \beta_{km} \phi_k(s) \psi_m(t)$ with $\beta_{km} = E(A_k B_m)/E(A_k^2)$

yielding

$$E(Y(t)|X) = \sum_m \sum_k \beta_{mk} A_k \psi_m(t).$$

This motivates the following extension:

Functional Additive Model

$$E(Y(t)|X) = \sum_m \sum_k f_{km}(A_k) \psi_m(t),$$

where f_{km} are smooth nonparametric functions.

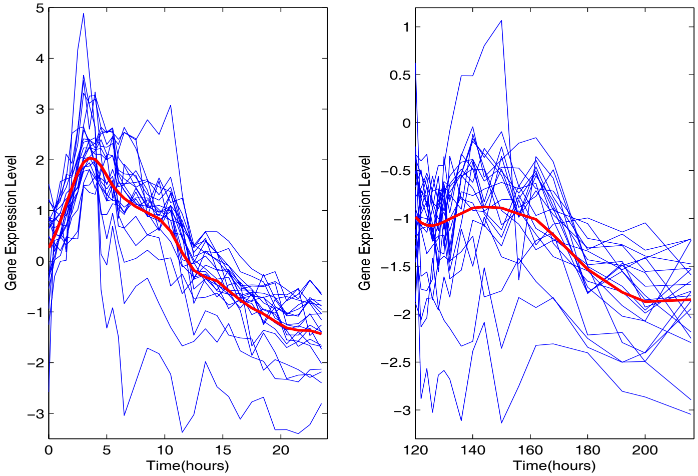
FAM FOR FUNCTIONAL RESPONSES

Assuming independent predictor scores A_j (automatically implied in the Gaussian case) we find

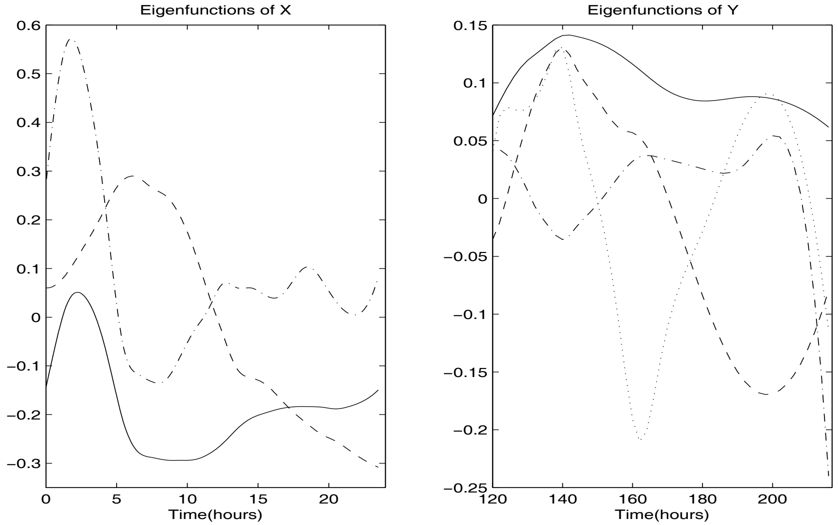
$$E(B_m|A_k) = E\{E(B_m|X)|A_k\} = E\{\sum_{j=1}^{\infty} f_{jm}(A_j)|A_k\} = f_{km}(A_k).$$

Consequence: Functional Additive Model can be implemented simply by 1-d scatterplot smoothing of $\hat{B}_{im}$ vs $\hat{A}_{ik}$ to obtain the defining functions f_{km}.

No backfitting iteration is needed: Fast and straightforward implementation with PACE. Analogously for functional regression model with scalar responses. For situations with several predictor functions within subjects: Can apply common additive model to ensemble of selected FPCs for all predictor functions.



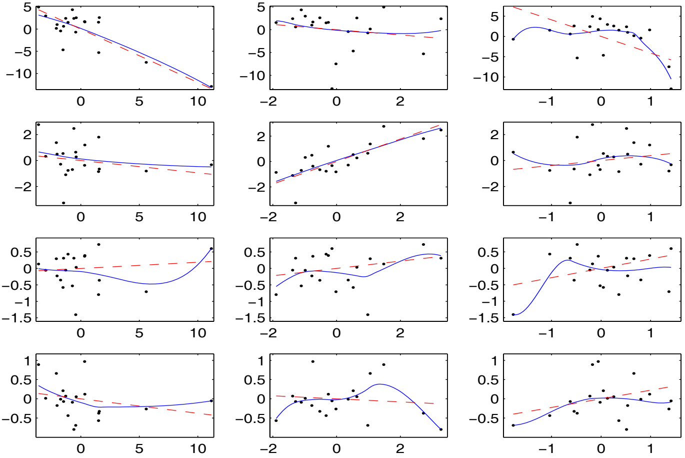
Gene time course data, zygotic genes for Drosophila for embryo phase (left) and pupa phase (right).



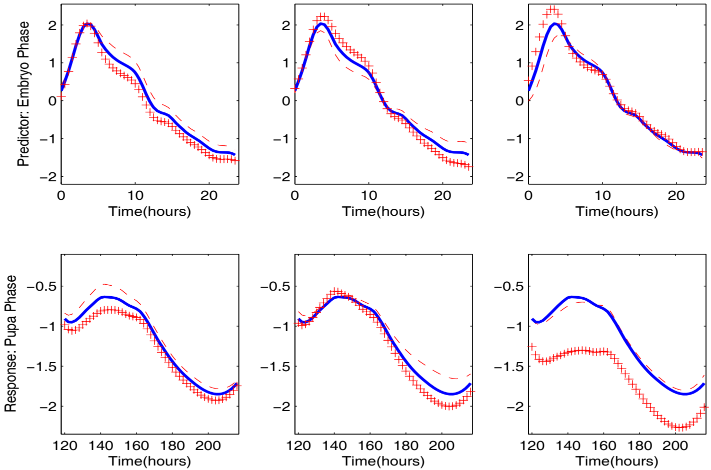
First three eigenfunctions for embryo phase (predictor) and first four eigenfunctions for pupa phase (response).

Table: Functional R^2, 25th, 50th and 75th percentiles and mean of the cross-validated observed relative prediction errors, $\text{RPE}_{(-i),f}$, comparing FAM and functional linear regression models for zygotic data.

	25th	50th	75th	Mean	R^2
FAM	.0506	.0776	.1662	.1301	0.19
LIN	.0479	.0891	.1727	.1374	0.16



Scatterplots (dots), local polynomial (solid) and linear (dashed) estimates for the regressions of estimated FPC scores of the pupa phase (y-axis) versus those for the embryo phase (x-axis).



Changes of response functions as predictor functions change in the directions of the first three eigenfunctions when fitting the Functional Additive Model.

Further Examples of Functional Regression with PAC

PACE Version 2.16, descriptions and references available at
http://anson.ucdavis.edu/∼mueller/data/pace.html

- *FPCreg*, *FPCdiag*: Let $X^c(t) = X^c(t) - \mu(t)$

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt$$

Further Examples of Functional Regression with PAC

PACE Version 2.16, descriptions and references available at
http://anson.ucdavis.edu/∼mueller/data/pace.html

- *FPCreg*, *FPCdiag*: Let $X^c(t) = X^c(t) - \mu(t)$

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt$$

- *FPCQuadReg*: (Yao and Müller 2010, Horvath and Reeder, 2012)

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt + \iint \gamma(s,t)X^c(s)X^c(t)dsdt$$

Further Examples of Functional Regression with PAC

PACE Version 2.16, descriptions and references available at
http://anson.ucdavis.edu/∼mueller/data/pace.html

- *FPCreg*, *FPCdiag*: Let $X^c(t) = X^c(t) - \mu(t)$

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt$$

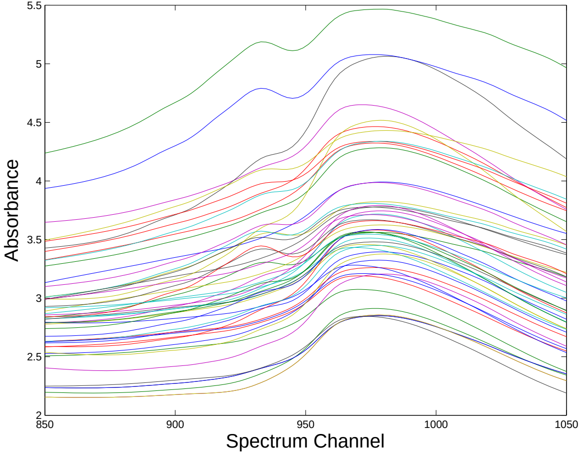
- *FPCQuadReg*: (Yao and Müller 2010, Horvath and Reeder, 2012)

$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt + \iint \gamma(s,t)X^c(s)X^c(t)dsdt$$

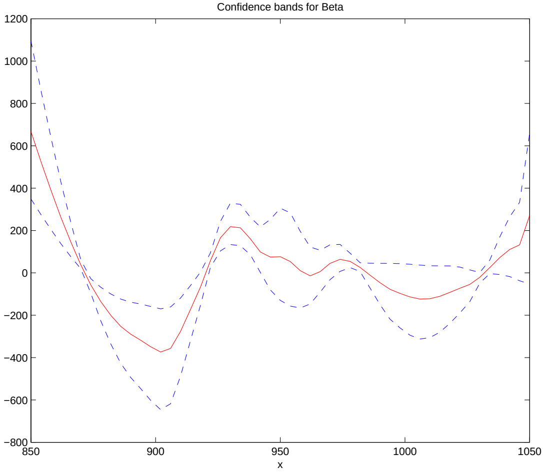
- *FPCquantile* (Chen and Müller 2012. *JRSSB.*)

$$P(Y \leq y|X) = E(I(Y \leq y)|X) = g^{-1}(\alpha(t) + \int X^c(t)\beta(y,t)dt)$$

Predictor Functions: Tecator Spectral Data

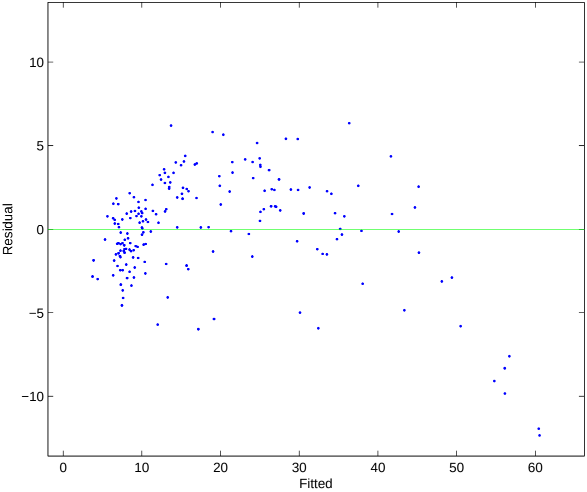


Coefficient of Linear Regression

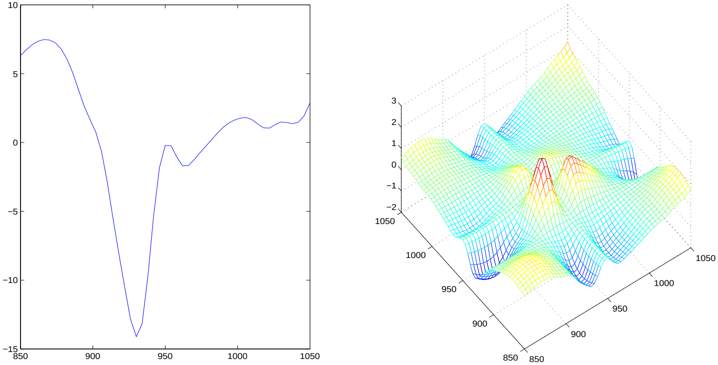


$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt$$

Residual Plot for Linear Regression

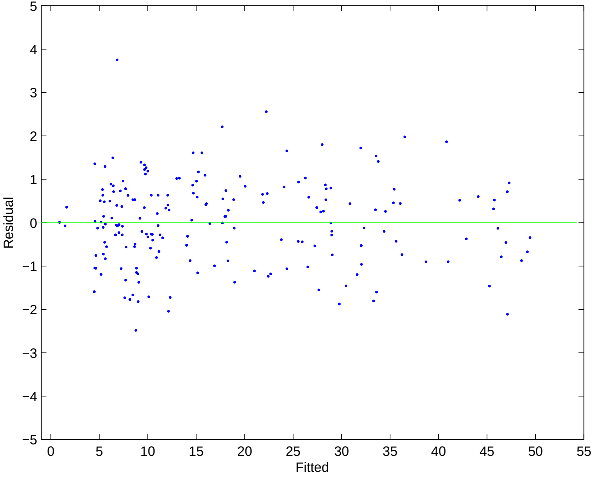


Coefficients of Quadratic Regression

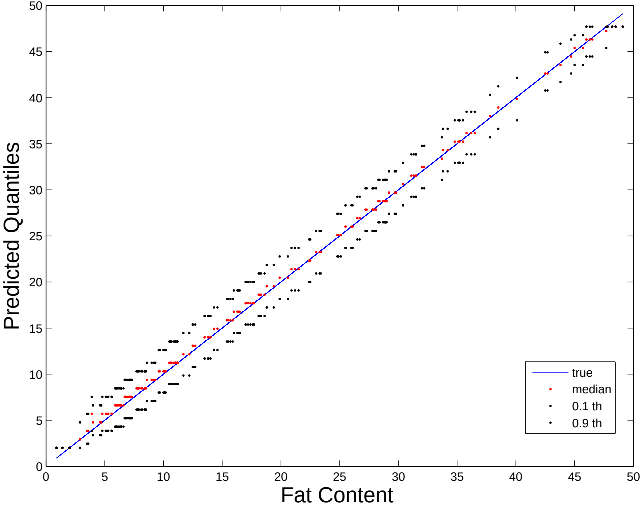


$$E(Y|X) = \alpha + \int X^c(t)\beta(t)dt + \iint \gamma(s,t)X^c(s)X^c(t)dsdt$$

Residual Plot for Quadratic Regression

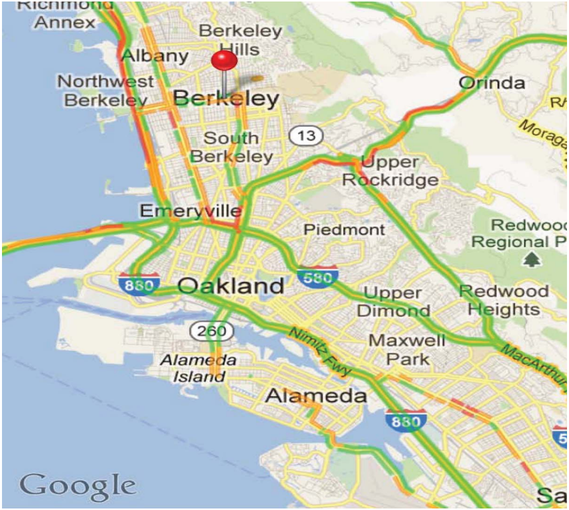


Quantiles

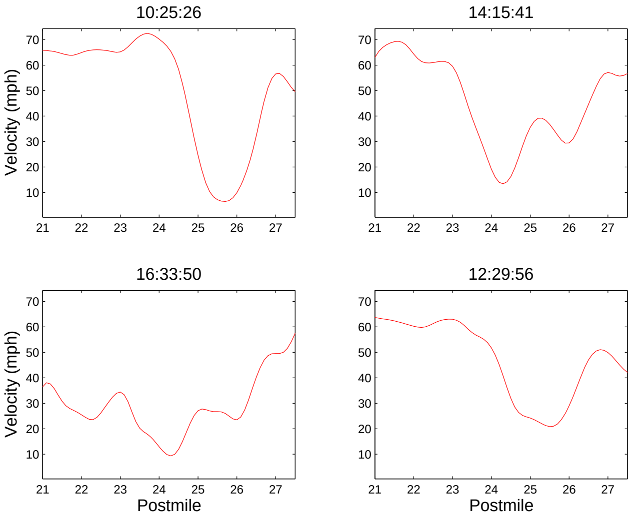


Predicted Quantiles
Fat Content
0
5
10
15
20
25
30
35
40
45
50
true
median
0.1 th
0.9 th

Illustrative Example: Traffic Data



Velocity on I-880



Prediction for Response Functions

- Y and X are both functions

Prediction for Response Functions

- Y and X are both functions
- *FPCfam*: $E(Y(t)|X) = \mu_Y(t) + \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} f_{jk}(\xi_k)\psi_j(t)$

Prediction for Response Functions

- Y and X are both functions
- *FPCfam*: $E(Y(t)|X) = \mu_Y(t) + \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} f_{jk}(\xi_k)\psi_j(t)$
- *FPCpredBands* (Chen and Müller 2012): Global prediction bands for Y conditional on X

Prediction for Response Functions

- Y and X are both functions
- *FPCfam*: $E(Y(t)|X) = \mu_Y(t) + \sum_{k=1}^{\infty} \sum_{j=1}^{\infty} f_{jk}(\xi_k)\psi_j(t)$
- *FPCpredBands* (Chen and Müller 2012): Global prediction bands for Y conditional on X
- For Gaussian process: $E(Y|X)$ and $\mathrm{cov}(Y|X)$

Prediction for Response Functions

- Y and X are both functions
- *FPCfam*: $E(Y(t)|X) = \mu_Y(t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} f_{jk}(\xi_k)\psi_j(t)$
- *FPCpredBands* (Chen and Müller 2012): Global prediction bands for Y conditional on X
- For Gaussian process: $E(Y|X)$ and $\mathrm{cov}(Y|X)$
- Common principal component assumption
  Additive assumption

  $$\begin{aligned} &\mathrm{cov}(Y(t_1), Y(t_2) \mid X) \\ &= G_{YY}(t_1, t_2) + \sum_{j=1}^{\infty}\{\sum_{k=1}^{\infty} g_{jk}(\xi_k) - \\ &\big(\sum_{k=1}^{\infty} f_{jk}(\xi_k)\big)^2\}\psi_j(t_1)\psi_j(t_2) \end{aligned}$$

Modeling the Prediction Bands

- Global prediction bands for Gaussian case:

$$P(\mu(t) - D_X(t) \leq Y_X(t) \leq \mu(t) + D_X(t) \mid X) \geq 1 - \alpha$$

  where $D_X(t) = \mathcal{C}_\alpha \left\{\mathrm{var}(Y(t)|X)\right\}^{1/2}$

- For more general random processes:

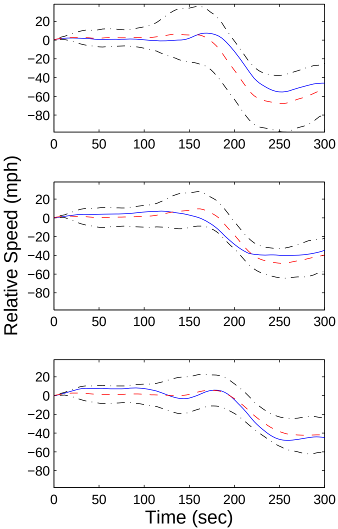
$$E\left\{P(L_X(t) \leq Y_X(t) \leq U_X(t) \mid X)\right\} \geq 1 - \alpha$$

- Find $\mathcal{C}_\alpha$ by the empirical coverage

'Mobile Century' Data

- Joint UC Berkeley - Nokia project (Herrera et al., 2010)
- Students were hired to drive on a segment of highway I-880 and send data (time, location, and speed) back through GPS enabled mobile phones.
- The follow-up project 'Mobile Millennium' is generating more data.

Estimated 90% Prediction Regions



FUNCTIONAL GRADIENTS

For functional linear regression with scalar responses:
Derivative of an operator $\Gamma : L^2 \to \mathcal{R}$ at $x = \sum_k A_{xk}\phi_k$ is a linear operator $\Gamma_x^{(1)}$: For functions u and scalars δ,

$$\Gamma(x + \delta u) = \Gamma(x) + \delta\, \Gamma_x^{(1)}(u) + o(\delta) \quad \text{as } \delta \to 0.$$

The functional derivative operator at x is characterized by the functional directional derivatives

$$\Gamma_x^{(1)}(\phi_k) = \gamma_{xk} \in \mathcal{R},\ k = 1, 2, \ldots$$

in the directions of the basis functions ϕ_k.

Representation

$$\Gamma_x^{(1)} = \sum_{k=1}^{\infty} \gamma_{xk}\, \Phi_k,$$

where $\gamma_{xk} = \Gamma_x^{(1)}(\phi_k)$ is a scalar, and Φ_k denotes the linear projection operator with

$$\Phi_k(u) = A_{uk} = \int u(t)\phi_k(t)dt, \quad \text{for all} \quad u \in L^2(\mathcal{T}).$$

Example: Functional linear model. Representing the regression parameter function β in the eigenbasis ϕ_k, $\beta(t) = \sum_k \beta_k \phi_k(t)$, $t \in \mathcal{T}$, leads to

$$\Gamma_L(X) = \mu_Y + \sum_{k=1}^{\infty} \beta_k A_{Xk} = \mu_Y + \sum_{k=1}^{\infty} \beta_k \Phi_k(X).$$

For any δ and arbitrary square integrable functions with representations $u = \sum_k A_{uk}\phi_k$ and $x = \sum_k A_{xk}\phi_k$,

$$\Gamma_L(x + \delta u) = \mu_Y + \sum_k \beta_k (A_{xk} + \delta A_{uk}) = \Gamma_L(x) + \delta \sum_k \beta_k A_{uk}.$$

Then $\Gamma_x^{(1)} = \sum_{k=1}^{\infty} \beta_k \Phi_k \quad \Rightarrow \quad \gamma_{xk} = \beta_k$. The functional derivative does not depend on x, as $\Gamma_x^{(1)}(\phi_k) = \beta_k$.

ADDITIVE MODELING OF FUNCTIONAL GRADIENTS

Consider additive functional operator

$$\Gamma_A(X) = E(Y^c|X) = \sum_{k=1}^{\infty} f_k(A_{Xk}),$$

subject to $Ef_k(A_{Xk}) = 0$, $k = 1, \ldots,$ for FPC scores A_{Xk}.

For functions $x = \sum_k A_{xk}\phi_k$ and $u = \sum_k A_{uk}\phi_k$,

$$\Gamma_A(x + \delta u) = \sum_k f_k(A_{xk} + \delta A_{uk}) = \Gamma_A(x) + \delta \sum_k f_k^{(1)}(A_{xk})A_{uk} + o(\delta),$$

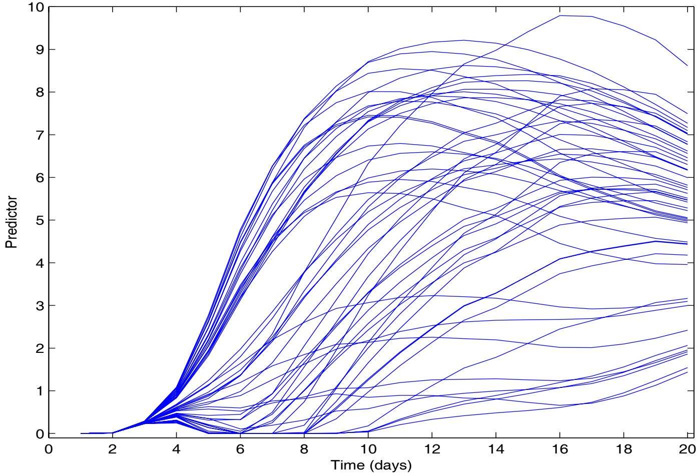
so that for the functional additive model,

$$\Gamma_{A,x}^{(1)}(u) = \sum_{k=1}^{\infty} f_k^{(1)}(A_{xk})A_{uk} = \sum_{k=1}^{\infty} \gamma_{A,xk}\Phi_k(u), \quad \gamma_{A,xk} = f_k^{(1)}(A_{xk}).$$

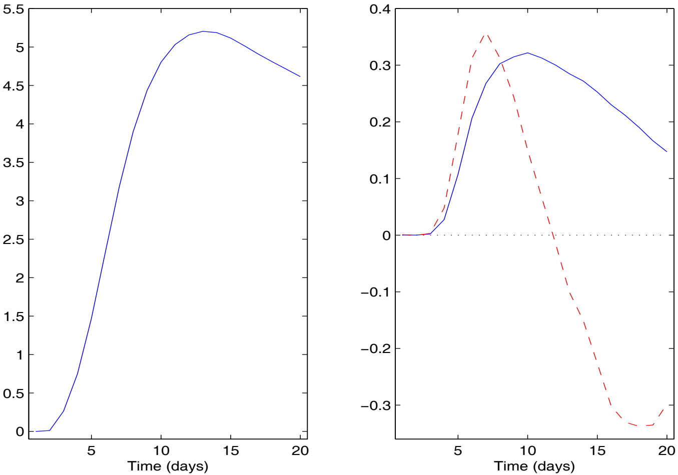
- Can easily extend to higher order derivatives due to additive structure
- Asymptotics: For densely sampled functions, may obtain derivatives through derivative estimates of the additive functions, with the 1-d rates of convergence for derivative estimation.

GRADIENTS FOR EGG-LAYING

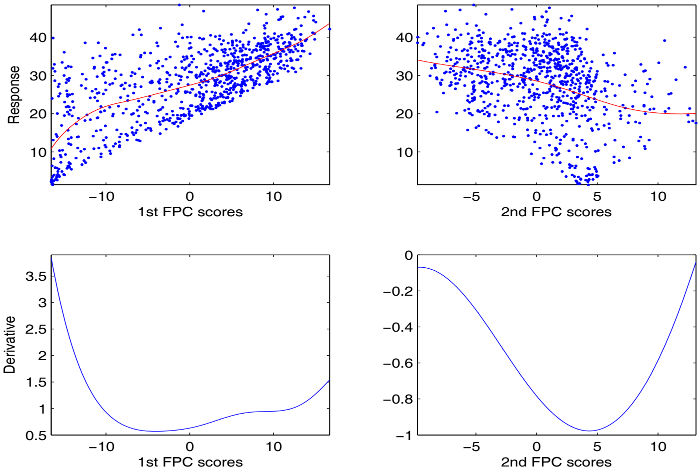
- Predictor functions: Egg-laying trajectories (daily egg counts) for cohort of 818 female medflies (Carey et al. 98) that live ≤ 20 days.
- Response: Lifetime fertility = total number of eggs laid over lifetime
- Preprocessing: Square root transformation of egg counts
- Question: How do early reproductive trajectories influence overall reproductive success.
- Tools: Gradient field and its visualization



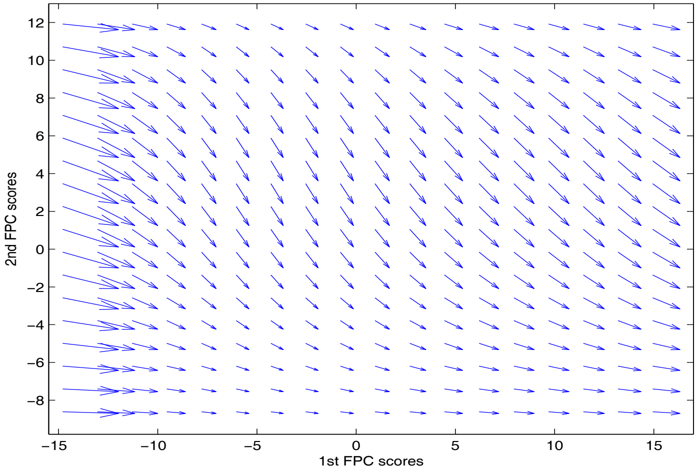
Egg-laying trajectories (predictors, smoothed) for 50 randomly selected flies, from a total of 818 medflies, for the first 20 days of their lifespan. Response=total eggs (reproductive success)



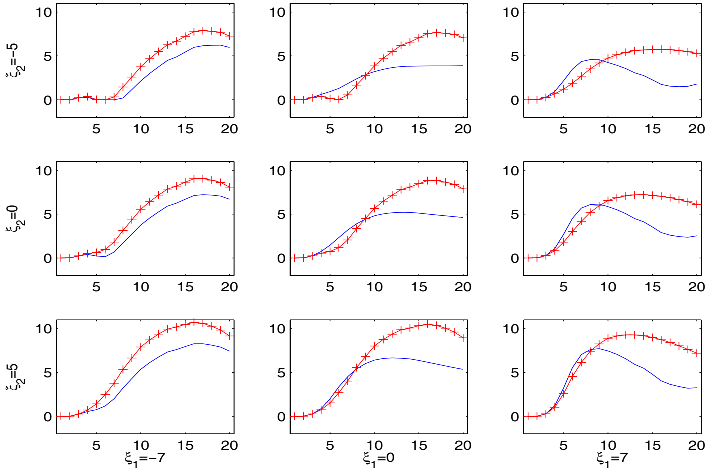
Smooth estimates of mean function (left panel) and first two eigenfunctions (right panel) of the predictor trajectories, explaining 72.1% (solid) and 18.6% (dashed) of the total variation, respectively.



Top panels: Nonparametric regression of the response (total fertility) on the first (left) and second (right) FPC scores of predictor processes. Bottom panels: Estimated derivatives of the smooth regression functions.



Estimated functional gradient field for total fertility, differentiated against the predictor process, expressed in terms of gradients of the response with respect to first (x-axis) and second (y-axis) FPC scores.



Visualization of the shape changes in fertility trajectories along the gradients: Bases are blue trajectories (9 combinations of the FPC scores $A_1 = \{-7, 0, 7\}$ and $A_2 = \{-5, 0, -5\}$, tip of the arrows red trajectories.

MODELING TIME-SPECIFIC RELATIONS

Consider $X, Y \in L^2$. Instead of modeling relation of entire functions X, Y:

Time-specific relationship of response process level $Y(t)$ on level $X(s)$ of predictor process usually of interest for $s \leq t$ (also for several predictor processes).

Nonparametric approach for fully observed functions: Obtain $E(Y(s)|X(t))$ using a smoothing approach for the observed data for all pairs (s, t) (Yao & M 2006), e.g., $E(Y(s)|X(s)) = g(X(s))$

Linear approach for sparse and irregularly observed functions under joint Gaussianity for (X, Y) (Yang & M 2010):

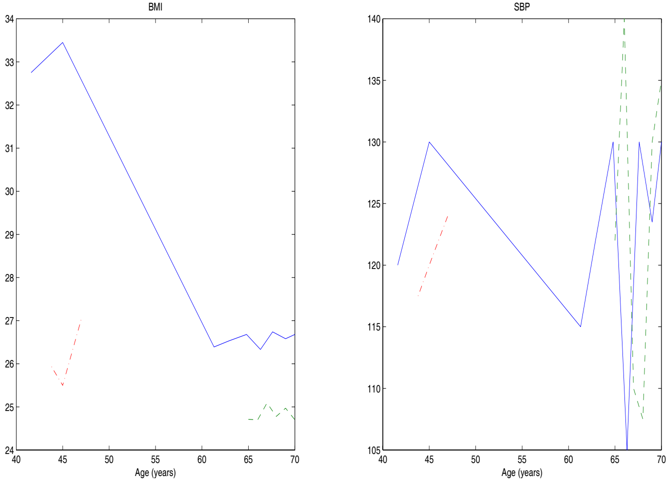
- For $X^{(c)} = \sum_k A_k \phi_k, \quad Y^{(c)} = \sum_m B_m \psi_m$, obtain

  $E(Y^{(c)}(t)|X(s)) = \gamma(s,t) X^{(c)}(s)$

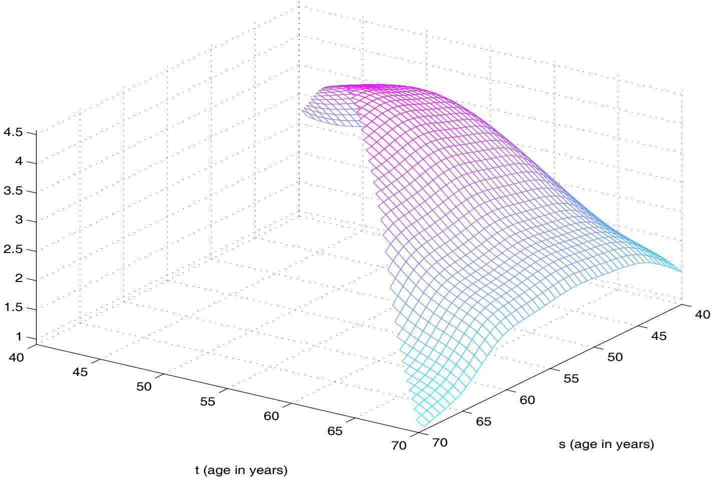
  $\gamma(s,t) = [\sum_k \sum_m E(A_k B_m) \phi_k(s) \psi_m(t)] / [\sum_k \lambda_k (\phi_k(s))^2]$.

- Derive consistent estimators for the function γ.

- Study dynamic longitudinal relationships by extending these relations to derivatives of Y and X: "Empirical Dynamics"



The observed values of body mass index (BMI) (left panel) and systolic blod pressure (SBP) (right panel) for three randomly selected subjects from Baltimore longitudinal study on aging



Estimated function $\gamma(s,t)$, $s \leq t$ linking SBP levels (response process $Y(t)$) with BMI levels (predictor process $X(s)$) through the relation $E(Y(t)|X(s)) = \mu_Y(t) + \gamma(s,t)(X(s) - \mu_X(s))$.

CONCLUSIONS

1. There is a need for functional regression models that:

   - are more flexible than the functional linear model;
   - avoid the curse of dimensionality that negatively affects structurally unconstrained models such as fully nonparametric approaches
   - accommodate more complex situations such as several predictor functions.
   - accommodate the modeling of functional gradients
   - accommodate time-specific relationships.

2. Functional principal component representations prove useful to define and analyze such models and to represent random functions as building blocks in more complex models.