Functional Data & Time Series
— A Brief Introduction —

Alexander Aue
aaue@ucdavis.edu

Department of Statistics & Graduate Group of Applied Mathematics, UC Davis

Outline

A. Functional Data

- What they are and where they show up
- How they are observed
- Adding time series context

Outline

A. Functional Data

- What they are and where they show up
- How they are observed
- Adding time series context

B. Analzying Functional Time Series

- Mean function and (auto)covariance operator
- Functional principal components
- Projections of functional autoregressive and moving average processes

Outline

A. Functional Data

- What they are and where they show up
- How they are observed
- Adding time series context

B. Analzying Functional Time Series

- Mean function and (auto)covariance operator
- Functional principal components
- Projections of functional autoregressive and moving average processes

C. Prediction and Estimation Methodology

- Predictions with functional autoregressive processes
- Estimation with functional moving average processes
- Illustrations with empirical results

Outline

A. Functional Data

- What they are and where they show up
- How they are observed
- Adding time series context

B. Analzying Functional Time Series

- Mean function and (auto)covariance operator
- Functional principal components
- Projections of functional autoregressive and moving average processes

C. Prediction and Estimation Methodology

- Predictions with functional autoregressive processes
- Estimation with functional moving average processes
- Illustrations with empirical results

D. Future Directions

A. Functional Data

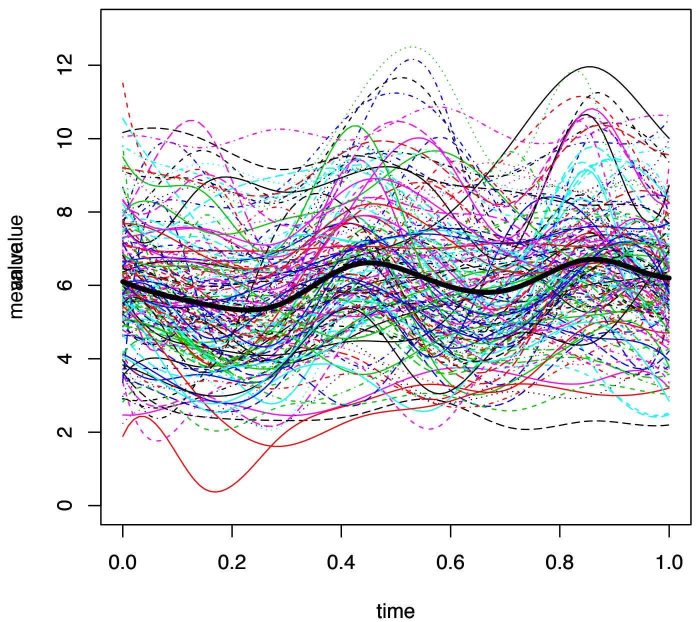
What is a Functional Observation?

A realization of a (typically smooth) random object
that takes values in an abstract function space

They often naturally arise in a times series context

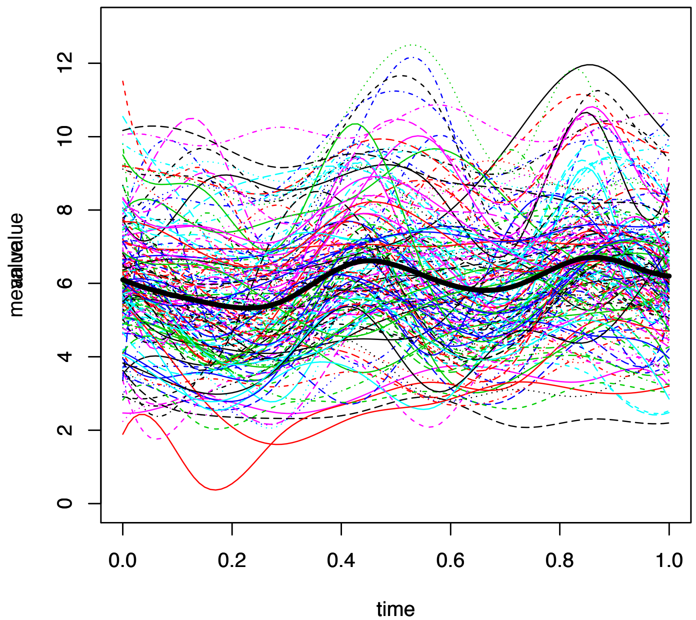
Where They Show Up: Environmental Science

- *Particulate matter:*
  - Daily PM10 curves recorded in Graz, Austria, during a winter season
  - Curves are volatile but display on average a diurnal pattern



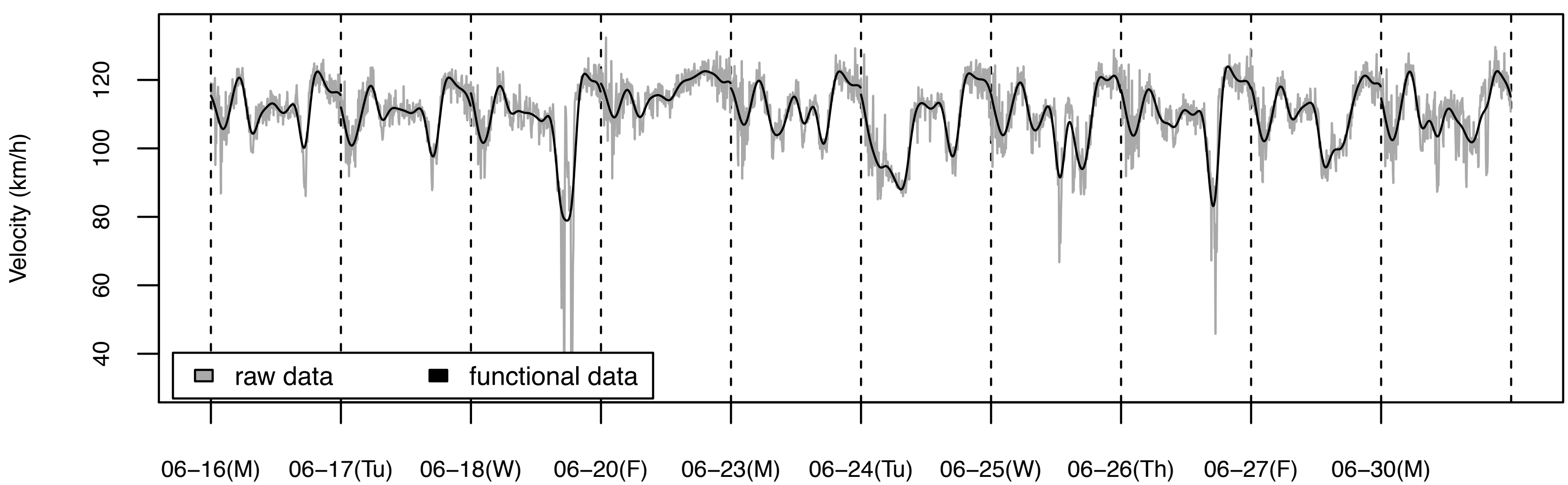
Where They Show Up: Environmental Science

- *Particulate matter:*
  - Daily PM10 curves recorded in Graz, Austria, during a winter season
  - Curves are volatile but display on average a diurnal pattern
- *Statistical Importance: Prediction problem*
  - High PM10 concentrations cause adverse health effects (cardiovascular diseases)
  - Local and EU regulation sets pollution limits, requires (local) policies to be implemented



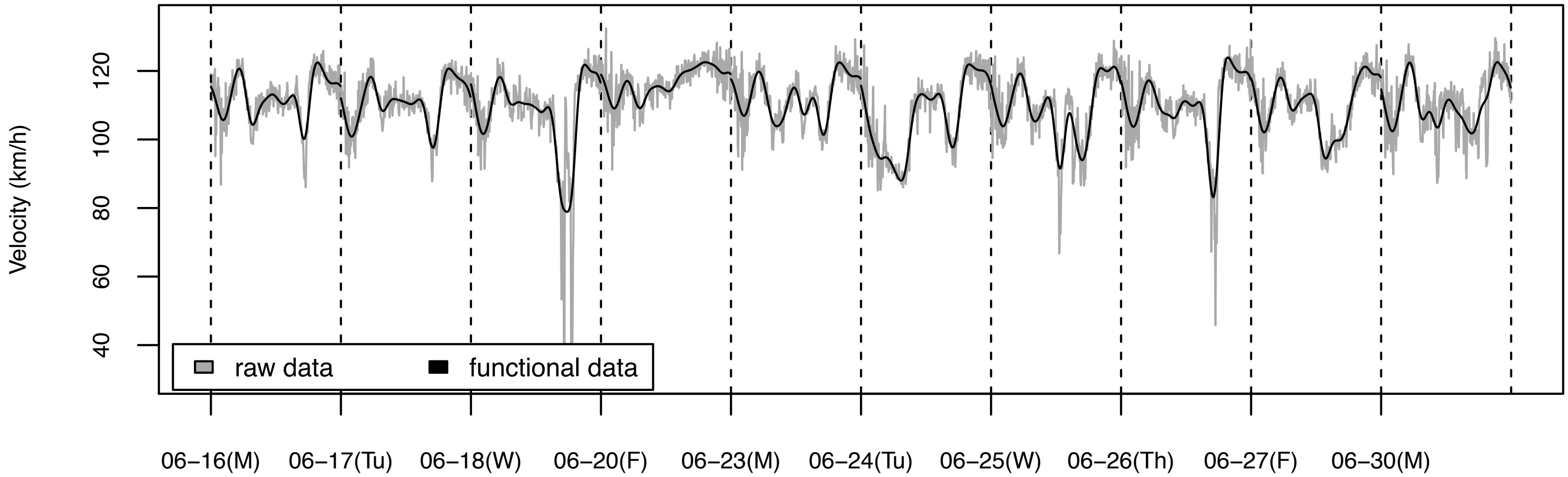
Where They Show Up: Civil Engineering

- *Traffic volume:*
    - Recorded is average velocity per minute on each of three lanes
    - Average velocities are averaged over the lanes, weighted by number of vehicles per lane



Where They Show Up: Civil Engineering

- *Traffic volume:*
  - Recorded is average velocity per minute on each of three lanes
  - Average velocities are averaged over the lanes, weighted by number of vehicles per lane
- *Importance: Estimation problem*
  - Input for macroscopic highway traffic flow model
  - Used to determine necessity of speed limits and specifics of their implementation



What They Are

- *Stylized facts:*
  - Data are typically sampled from some continuous “time” process
  - The sampled curves are envisioned as smooth [underlying low-dimensional structure?]
  - Denote a functional observation by $(x(t)\colon t \in T)$
  - Set $T = [0, 1]$
  - Important: T may not be time or univariate:
    - $x(t)$ could be the concentration of a pollutant at altitude t
    - $x(t)$ could be gray level of an image at spatial location $t \in T \subset \mathbb{R}^2$

What They Are

- *Stylized facts:*
  - Data are typically sampled from some continuous "time" process
  - The sampled curves are envisioned as smooth [underlying low-dimensional structure?]
  - Denote a functional observation by $(x(t)\colon t \in T)$
  - Set $T = [0, 1]$
  - Important: T may not be time or univariate:
    * $x(t)$ could be the concentration of a pollutant at altitude t
    * $x(t)$ could be gray level of an image at spatial location $t \in T \subset \mathbb{R}^2$
- *Definition:*
  - A random element X is a functional variable if it takes values in a function space F
  - Therefore $X = (X(t)\colon t \in T)$
  - A realization of X is denoted by $x = (x(t)\colon t \in T)$

What They Are

- *Examples of (normed) function spaces:*
  - $F = C[0,1]$, the continuous functions on the unit interval
  - $F = L^2[0,1]$, the square-integrable functions on the unit interval
  - F could be a reproducing kernel Hilbert space, RKHS
  - F could be a Sobolov space

What They Are

- *Examples of (normed) function spaces:*
  - $F = C[0,1]$, the continuous functions on the unit interval
  - $F = L^2[0,1]$, the square-integrable functions on the unit interval
  - F could be a reproducing kernel Hilbert space, RKHS
  - F could be a Sobolev space
- *Convention for this talk:*
  - Focus on $F = L^2[0,1] = L^2$
  - Under this convention, X has values in L^2
  - Formally, there is a probability space $(\Omega, \mathcal{A}, P)$ such that

    $$X : \Omega \to L^2$$

    is $\mathcal{A}$-$\mathcal{B}$-measurable, where $\mathcal{B}$ is the Borel σ-algebra generated by the open sets in L^2
  - Note: Pointwise interpretation of functions is lost

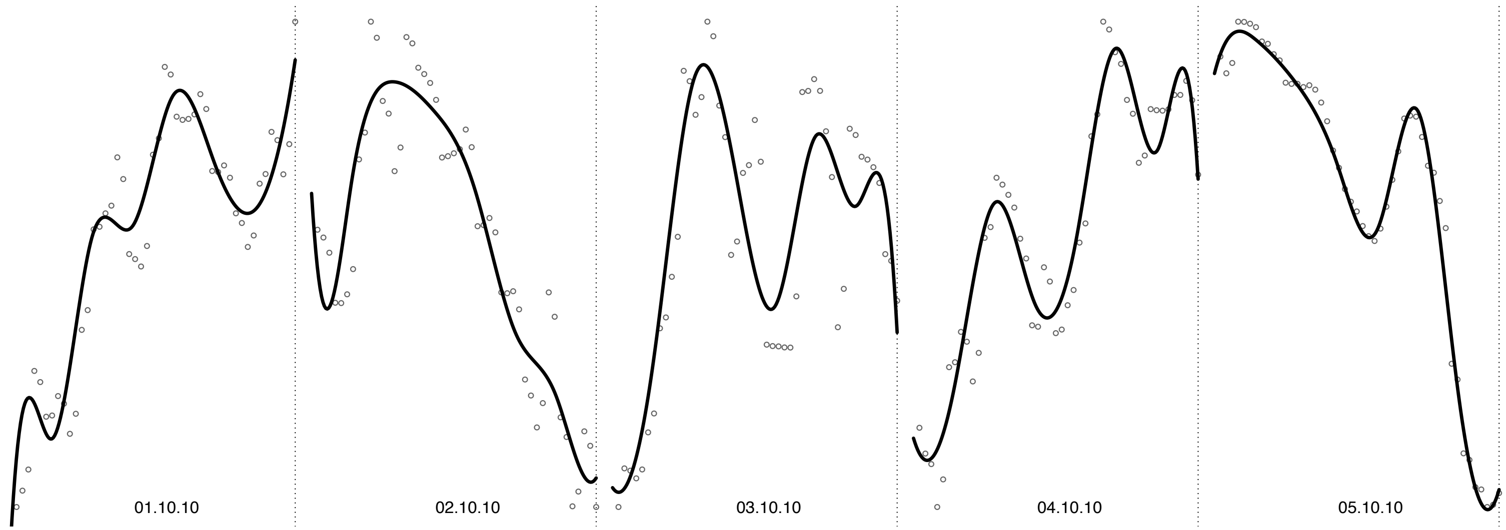
What They Are

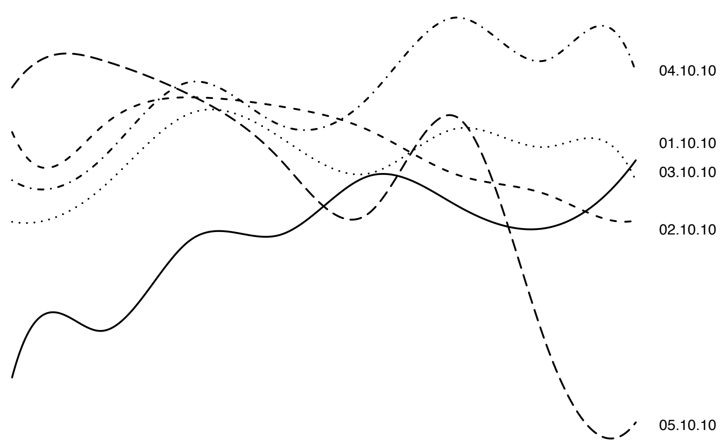
- *More stylized facts:*
  - Typically one has more than one observation
  - In many applications, functional observations are not independent
  - Often they are sampled in time
  - Leads to functional data x_j as realization of functional variable X_j, $j = 1, \dots, n$
  - There are two clocks: $X_j(t)$ has *calendar time* j and *intra-day time* t

What They Are

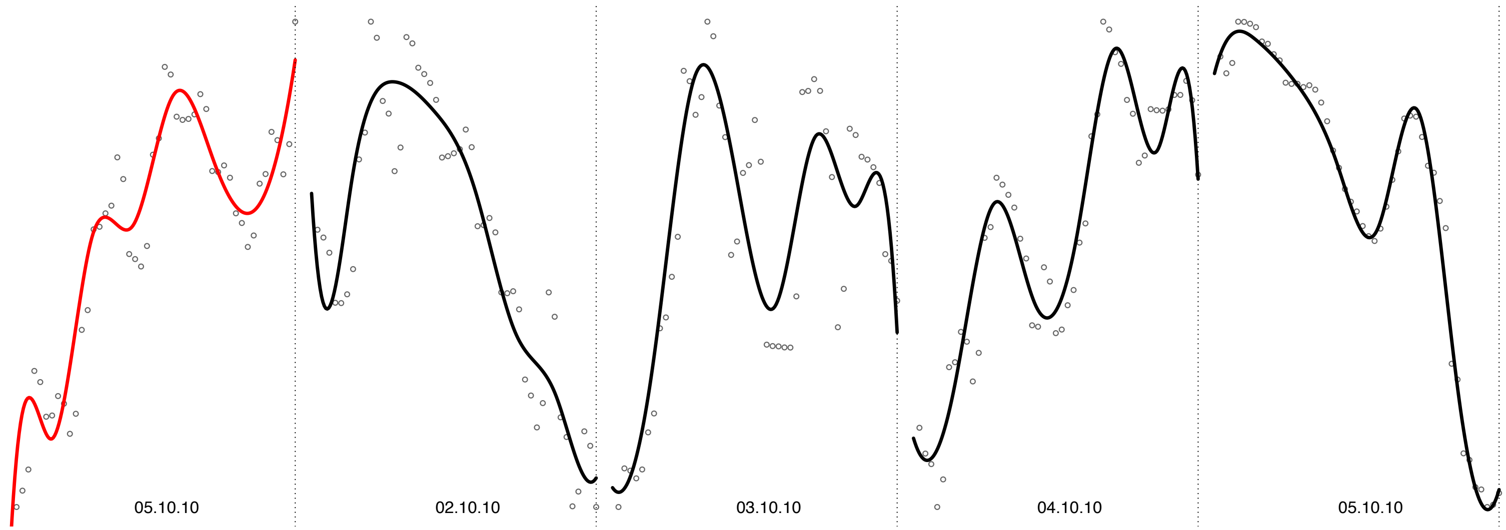
- *More stylized facts:*
  - Typically one has more than one observation
  - In many applications, functional observations are not independent
  - Often they are sampled in time
  - Leads to functional data x_j as realization of functional variable X_j, $j = 1, \ldots, n$
  - There are two clocks: $X_j(t)$ has *calendar time* j and *intra-day time* t
- *How they are observed*
  - There are no continuous measurements
  - Any realization x is observed at discrete points only: $x(t_1), \ldots, x(t_K)$ for some K
  - Measurements can be exact or contaminated with measurement error
  - High sampling frequency scheme leads to *dense functional data*
  - Low sampling frequency scheme leads to *sparse functional data*

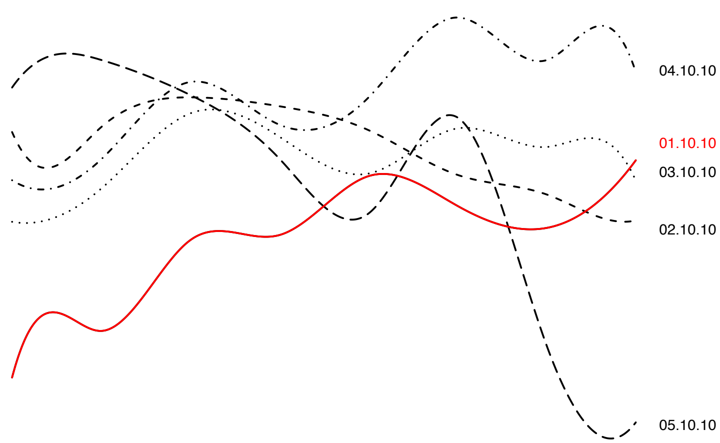
Example: Particulate Matter Data



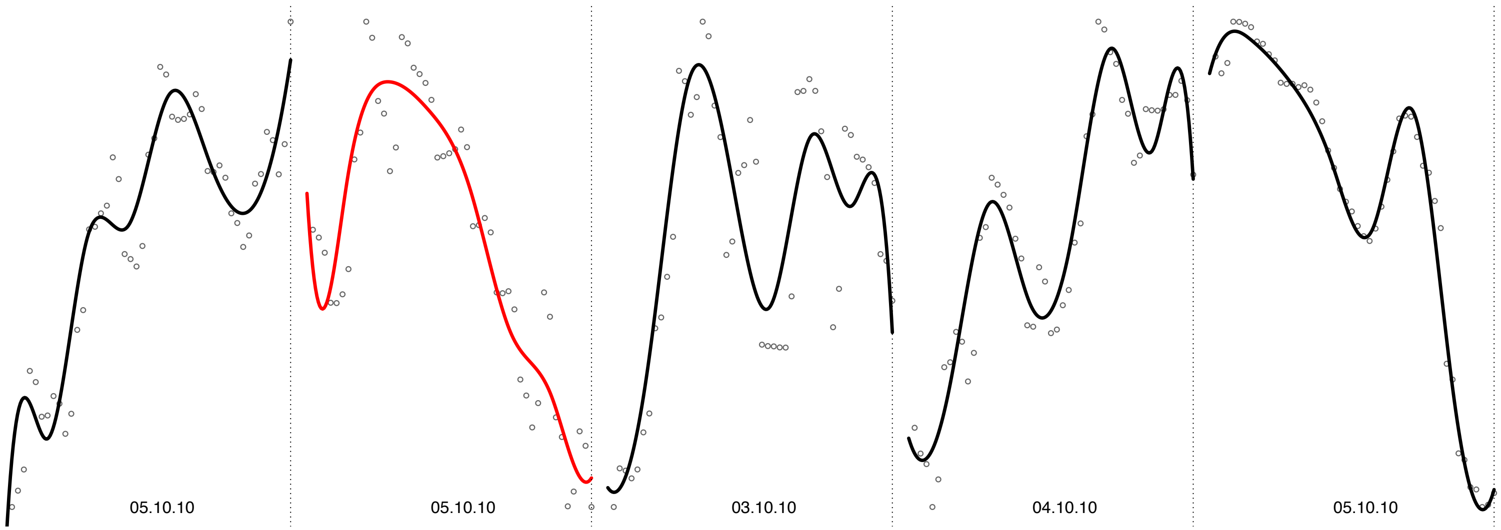


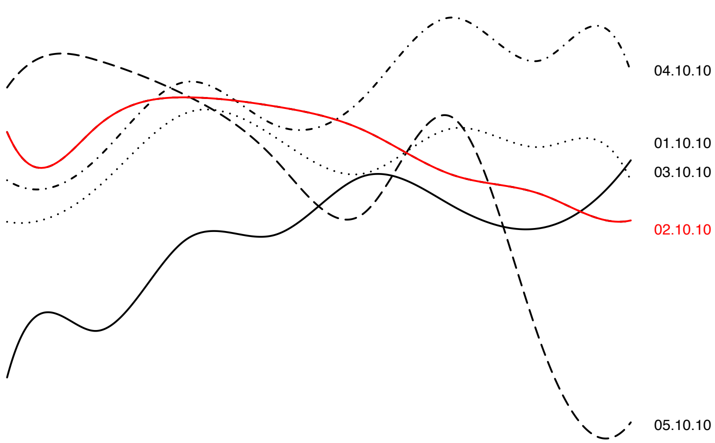
Example: Particulate Matter Data





Example: Particulate Matter Data





The Functional Time Series Context

- *Univariate and multivariate linear time series have been studied extensively*
  - Rather complete picture of strength and weaknesses of ARMA models
  - Many extensions available
  - Ready-to-use computer packages

The Functional Time Series Context

- *Univariate and multivariate linear time series have been studied extensively*
  - Rather complete picture of strength and weaknesses of ARMA models
  - Many extensions available
  - Ready-to-use computer packages
- *If observations are functions*
  - Increased complexity as infinite-dimensional objects enter
  - Some theory available
  - Much more limited time series tool box

The Functional Time Series Context

- *Univariate and multivariate linear time series have been studied extensively*
  - Rather complete picture of strength and weaknesses of ARMA models
  - Many extensions available
  - Ready-to-use computer packages
- *If observations are functions*
  - Increased complexity as infinite-dimensional objects enter
  - Some theory available
  - Much more limited time series tool box
- *Literature*
  - Focus has often been on special cases
  - First-oder functional autoregression dominates
  - Many more results are becoming available

B. Analyzing Functional Time Series

Mean Function and Covariance Operator

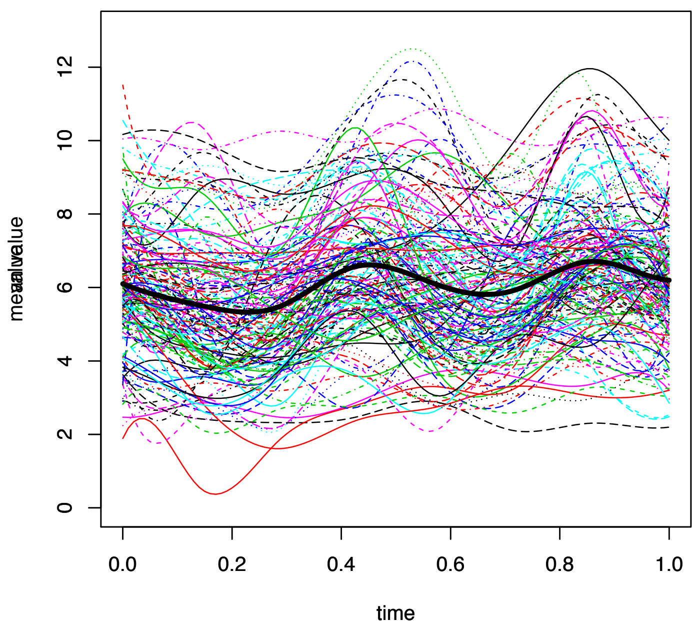
Two of the most important objects/summary statistics in multivariate statistics are the sample mean and sample covariance matrix

How can these objects be defined and analyzed in the functional context?

Mean Function

- *How to define sample and population mean functions?*
  - Forego technical definitions and background
  - Natural definition of *sample mean function* is $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$
  - Definition of *population mean function* is

$$\mu = E[X] = ((E[X])(t) \colon t \in [0,1]) = (E[X(t)] \colon t \in [0,1])$$



Covariance Operator and Spectral Decomposition

- *Definition*
  - The *covariance operator* $C\colon L^2 \to L^2$ is defined by

$$C(y) = E\big[\langle X - \mu, y\rangle (X - \mu)\big] = \int_0^1 c(s, \cdot) y(s) ds, \qquad y \in H$$

    with *covariance kernel* $c(s,t) = E[\{X(s) - \mu\}\{X(t) - \mu\}]$

  - $c(s,t)$ is symmetric and non-negative definite and describes all cross-covariances of X

Covariance Operator and Spectral Decomposition

- *Definition*
  - The *covariance operator* $C\colon L^2 \to L^2$ is defined by

$$C(y) = E\big[\langle X-\mu, y\rangle (X-\mu)\big] = \int_0^1 c(s,\cdot)y(s)ds, \qquad y \in H$$

    with *covariance kernel* $c(s,t) = E[\{X(s)-\mu\}\{X(t)-\mu\}]$

  - $c(s,t)$ is symmetric and non-negative definite and describes all cross-covariances of X

- *Spectral decomposition*
  - The kernel $c(s,t)$ allows for the spectral decomposition

$$c(s,t) = \sum_{\ell=1}^{\infty} \lambda_\ell \, e_\ell(s) e_\ell(t),$$

    where $(\lambda_\ell\colon \ell \in \mathbb{N})$ are the increasing eigenvalues with associated eigenfunctions $(e_\ell\colon \ell \in \mathbb{N})$

  - *Karhunen–Loève representation:*

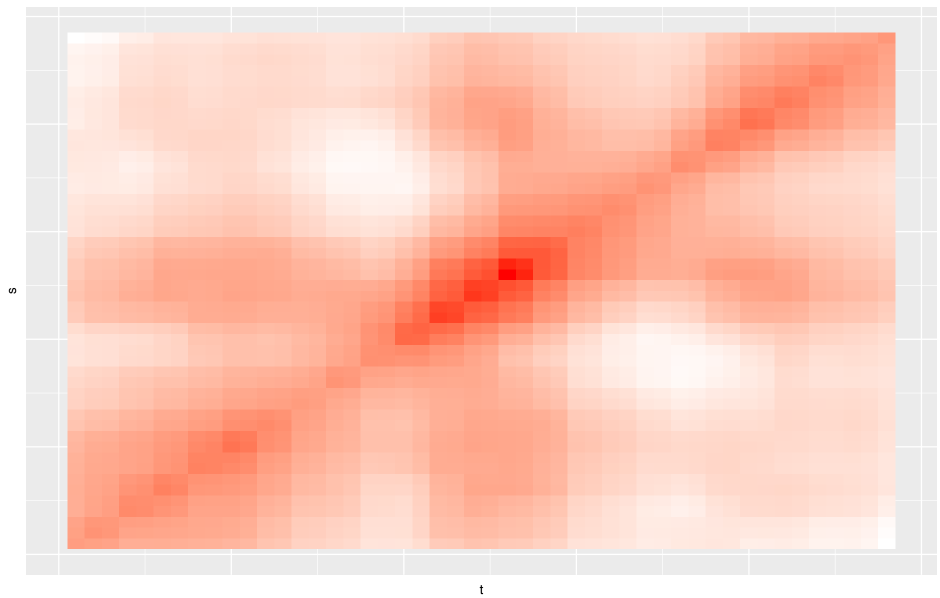
$$X_j = \sum_{\ell=1}^{\infty} \langle X_j, e_\ell \rangle e_\ell$$

Covariance Operator and Spectral Decomposition

- *Definition*
  - The *sample covariance operator* $\hat{C}_n \colon L^2 \to L^2$ is defined by

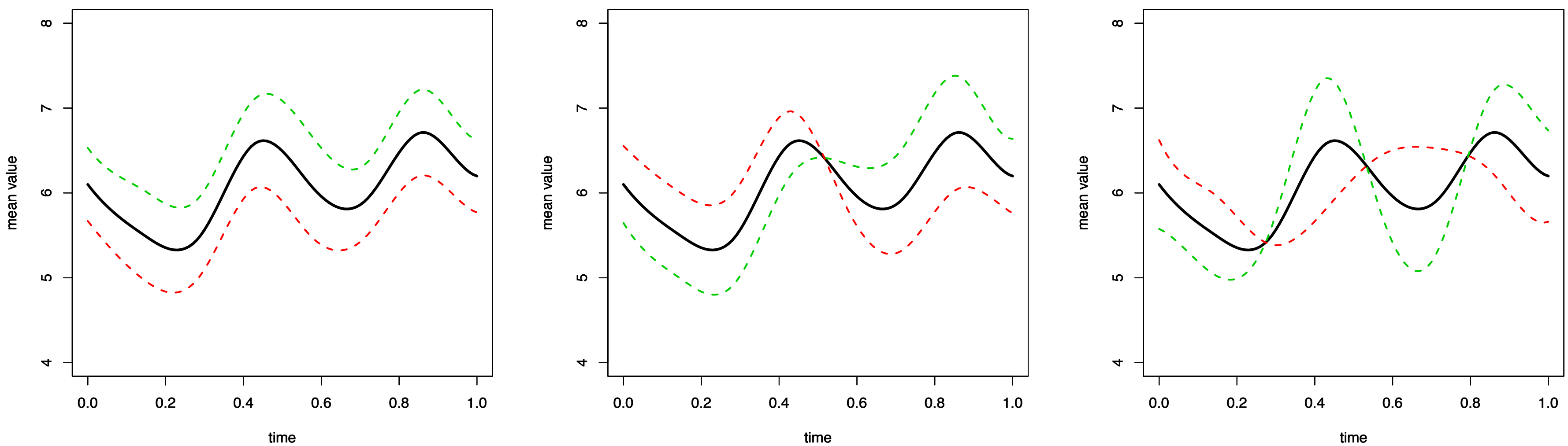
$$\hat{C}_n(y) = \frac{1}{n}\sum_{j=1}^{n}\langle X_j - \hat{X}_n, y\rangle(X_j - \hat{X}_n) = \int_0^1 \hat{c}_n(s,\cdot)y(s)ds, \qquad y \in H,$$

with *sample covariance kernel* $\hat{c}_n(s,t) = \frac{1}{n}\sum_{j=1}^{n}\{X_j(s) - \bar{X}_n\}\{X_j(t) - \bar{X}_n\}]$



Covariance Operator and Spectral Decomposition

- *Spectral decomposition*
  - $\hat{C}_n$ has at most n non-zero eigenvalues $\hat{\lambda}_\ell$ with associated sample eigenfunctions $\hat{e}_\ell$
  - Therefore only a limited number of eigenvalues and eigenfunctions can be estimated
  - Plots show effect of first three eigenfunctions for particulate matter data on mean function



Consistency Results

- *Theory in Hörmann & Kokoszka (2010)*
  - Results for wide range of stationary functional time series
  - Consistency of the mean function:

$$\sqrt{n}\|\hat{X}_n - \mu\| = \mathcal{O}_P(1)$$

  - Consistency of the covariance operator:

$$\sqrt{n}\|\hat{C}_n - C\| = \mathcal{O}_P(1)$$

  - Consistency of eigenvalues and eigenfunctions:

$$\sqrt{n} \max_{1 \le \ell \le d} \left\{ \|\hat{c}_\ell \hat{e}_\ell - e_\ell\| + |\hat{\lambda}_\ell - \lambda_\ell| \right\} = \mathcal{O}_P(1)$$

  - Random signs $\hat{c}_\ell = \mathrm{sign}(\langle e_\ell, \hat{e}_\ell \rangle)$ needed as e_ℓ is unique only up to the sign
  - But $\hat{c}_\ell$ cannot be determined from the sample
  - Any estimator or test based on eigenfunctions must not depend on signs

Autocovariance Operators

- *Linear dependence*
  - Important concept in univariate and multivariate time series analysis
  - In functional context captured by *autocovariance operators*

$$C_h(y) = E[\langle X_0 - \mu, y \rangle (X_h - \mu)], \qquad h \in \mathbb{Z},\ y \in H$$

  - Note: $C = C_0$

Autocovariance Operators

- *Linear dependence*
  - Important concept in univariate and multivariate time series analysis
  - In functional context captured by *autocovariance operators*

$$C_h(y) = E[\langle X_0 - \mu, y \rangle (X_h - \mu)], \qquad h \in \mathbb{Z},\ y \in H$$

  - Note: $C = C_0$

- *Sample autocovariance estimators*
  - C_h can be estimated by

$$\hat{C}_{h,n}(y) = \frac{1}{n} \sum_{j=1}^{n-h} \langle X_j - \hat{X}_n, y \rangle (X_{j+h} - \hat{X}_n), \qquad h \in \mathbb{Z},\ y \in H$$

  - Here only $h = 1$ will be used

Projections onto Principal Components

- *Functional PCA*
  - Idea: If complete function is too complicated work with fPC scores
  - *What happens to linear dependence after projection?*

Projections onto Principal Components

- *Functional PCA*
  - Idea: If complete function is too complicated work with fPC scores
  - *What happens to linear dependence after projection?*

- *First-order functional autoregression*
  - $X_j = \Phi X_{j-1} + \varepsilon_j$ with

$$\Phi(x) = a\big(\langle x, e_1\rangle + \langle x, e_2\rangle\big)e_1 + a\langle x, e_1\rangle e_2, \qquad x \in H,$$

    where $a \in (0,1)$ and $e_1, e_2 \in H$ orthonormal
  - Assume that $E[\langle \varepsilon_j, e_1\rangle^2] > 0$ but $E[\langle \varepsilon_j, e_2\rangle^2] = 0$
  - Then, the first fPC score series satisfies

$$\langle X_j, e_1\rangle = a\langle X_{j-1}, e_1\rangle + a^2\langle X_{j-2}, e_1\rangle + \langle \varepsilon_j, e_1\rangle$$

Projections onto Principal Components

- *Functional PCA*
  - Idea: If complete function is too complicated work with fPC scores
  - *What happens to linear dependence after projection?*

- *First-order functional autoregression*
  - $X_j = \Phi X_{j-1} + \varepsilon_j$ with

$$\Phi(x) = a\big(\langle x, e_1\rangle + \langle x, e_2\rangle\big)e_1 + a\langle x, e_1\rangle e_2, \qquad x \in H,$$

    where $a \in (0,1)$ and $e_1, e_2 \in H$ orthonormal
  - Assume that $E[\langle\varepsilon_j, e_1\rangle^2] > 0$ but $E[\langle\varepsilon_j, e_2\rangle^2] = 0$
  - Then, the first fPC score series satisfies

$$\langle X_j, e_1\rangle = a\langle X_{j-1}, e_1\rangle + a^2\langle X_{j-2}, e_1\rangle + \langle\varepsilon_j, e_1\rangle$$

- *Projection of this FAR(1) process is VAR(2) process*

C. Prediction and Estimation Methodology

C. Prediction and Estimation Methodology

A First Example

- *First-order functional autoregression*
  - The most often applied zero-mean functional time series model is

$$X_j = \Phi X_{j-1} + \varepsilon_j, \qquad j \in \mathbb{Z}$$

  - $(\varepsilon_j \colon j \in \mathbb{Z})$ are centered iid innovations and Φ a bounded linear operator satisfying $\|\Phi\|_{\mathcal{L}} < 1$

A First Example

- *First-order functional autoregression*
  - The most often applied zero-mean functional time series model is

$$X_j = \Phi X_{j-1} + \varepsilon_j, \qquad j \in \mathbb{Z}$$

  - $(\varepsilon_j \colon j \in \mathbb{Z})$ are centered iid innovations and Φ a bounded linear operator satisfying $\|\Phi\|_{\mathcal{L}} < 1$

- *Functional Yule–Walker equations; Bosq (2000)*
  - Apply $E[\langle \cdot, x \rangle X_{j-1}]$ to the model equations to obtain the functional Yule–Walker equations

$$E[\langle X_j, x \rangle X_{j-1}] = E[\langle \Phi(X_{j-1}), x \rangle X_{j-1}] + E[\langle \varepsilon_j, x \rangle X_{j-1}] = E[\langle \Phi(X_{j-1}), x \rangle X_{j-1}]$$

  - Let Φ' be the adjoint operator of Φ, given by $\langle \Phi(x), y \rangle = \langle x, \Phi'(y) \rangle$
  - This gives the operator equation $C_1(x) = C(\Phi'(x))$ and therefore

$$\Phi(x) = C_1' C^{-1}(x)$$

  - Can be estimated by smoothing techniques, gives predictor function $\tilde{X}_{n+1} = \hat{\Phi}_n X_n$

Methods Based on FPC Scores

- *Univariate and multivariate prediction methods; Hyndman & Shang (2009)*
  - This prediction technique avoids estimating operators directly

Methods Based on FPC Scores

- *Univariate and multivariate prediction methods; Hyndman & Shang (2009)*
  - This prediction technique avoids estimating operators directly
  - Step 1: Fix d. Use the data $X_1, \ldots, X_n$ to compute the vectors

$$\boldsymbol{X}_j^e = (x_{j,1}^e, \ldots, x_{j,d}^e)',$$

containing the first d empirical FPC scores $x_{j,\ell}^e = \langle X_j, \hat{e}_\ell \rangle$

Methods Based on FPC Scores

- *Univariate and multivariate prediction methods; Hyndman & Shang (2009)*
  - This prediction technique avoids estimating operators directly
  - Step 1: Fix d. Use the data $X_1, \ldots, X_n$ to compute the vectors

$$\boldsymbol{X}_j^e = (x_{j,1}^e, \ldots, x_{j,d}^e)',$$

    containing the first d empirical FPC scores $x_{j,\ell}^e = \langle X_j, \hat{v}_\ell \rangle$

  - Step 2: Fix h. Use $\boldsymbol{X}_1^e, \ldots, \boldsymbol{X}_n^e$ to determine the h-step ahead prediction

$$\hat{\boldsymbol{X}}_{n+h}^e = (\hat{y}_{n+h,1}^e, \ldots, \hat{y}_{n+h,d}^e)'$$

    for $\boldsymbol{X}_{n+h}^e$ with an appropriate multivariate algorithm

Methods Based on FPC Scores

- *Univariate and multivariate prediction methods; Hyndman & Shang (2009)*
  - This prediction technique avoids estimating operators directly
  - Step 1: Fix d. Use the data $X_1, \ldots, X_n$ to compute the vectors

$$\boldsymbol{X}_j^e = (x_{j,1}^e, \ldots, x_{j,d}^e)',$$

    containing the first d empirical FPC scores $x_{j,\ell}^e = \langle X_j, \hat{v}_\ell \rangle$
  - Step 2: Fix h. Use $\boldsymbol{X}_1^e, \ldots, \boldsymbol{X}_n^e$ to determine the h-step ahead prediction

$$\hat{\boldsymbol{X}}_{n+h}^e = (\hat{y}_{n+h,1}^e, \ldots, \hat{y}_{n+h,d}^e)'$$

    for $\boldsymbol{X}_{n+h}^e$ with an appropriate multivariate algorithm
  - Step 3: Use the functional object

$$\hat{X}_{n+h} = \hat{y}_{n+h,1}^e \, \hat{v}_1 + \ldots + \hat{y}_{n+h,d}^e \hat{v}_d$$

    as h-step ahead prediction for X_{n+h}

Methods Based on FPC Scores

- *Remarks on algorithm*
  - Gives best linear prediction (in mean square sense) of the *population* FPC scores
  - It does not assume an FAR(p) structure or any other functional time series specification
  - Standard methods such as the Durbin–Levinson and innovations algorithm can be applied
  - Alternatives include exponential smoothing and nonparametric prediction algorithms
  - Covariates can be incorporated in the prediction process

Methods Based on FPC Scores

- *Remarks on algorithm*
  - Gives best linear prediction (in mean square sense) of the *population* FPC scores
  - It does not assume an FAR(p) structure or any other functional time series specification
  - Standard methods such as the Durbin–Levinson and innovations algorithm can be applied
  - Alternatives include exponential smoothing and nonparametric prediction algorithms
  - Covariates can be incorporated in the prediction process

- *Remarks on numerical implementation*
  - Is convenient in `R`
  - In Step 1, FPC score matrix and sample eigenfunctions with `fda`
  - In Step 2, forecasting of the FPC scores with `vars`, in case VAR models are employed
  - In Step 3, combine `fda` and `vars` to obtain $\hat{X}_{n+h}$

Methods Based on FPC Scores

- *Model selection — 1; A, Dubart Norinho & Hörmann (2015)*
  - Assume $X_j = \Phi_1 X_{j-1} + \ldots \Phi_p X_{j-p} + \varepsilon_j$
  - (ε_j) i.i.d. and $\Phi_1, \ldots, \Phi_p$ Hilbert–Schmidt
  - Then

$$E\left[\|X_{n+1} - \hat{X}_{n+1}\|^2\right] \leq \sigma^2 + \gamma_d, \qquad (1)$$

where

$$\gamma_d = \left(1 + \left[\sum_{j=1}^{p} \phi_{j;d}\right]^2\right) \sum_{\ell=d+1}^{\infty} \lambda_\ell \qquad \text{and} \qquad \phi_{j;d} = \left(\sum_{\ell=d+1}^{\infty} \|\Phi_j(e_\ell)\|^2\right)^{1/2}$$

Methods Based on FPC Scores

- *Model selection — 1; A, Dubart Norinho & Hörmann (2015)*
  - Assume $X_j = \Phi_1 X_{j-1} + \ldots \Phi_p X_{j-p} + \varepsilon_j$
  - (ε_j) i.i.d. and $\Phi_1, \ldots, \Phi_p$ Hilbert–Schmidt
  - Then

$$E\left[\|X_{n+1} - \hat{X}_{n+1}\|^2\right] \leq \sigma^2 + \gamma_d, \tag{2}$$

where

$$\gamma_d = \left(1 + \left[\sum_{j=1}^{p} \phi_{j;d}\right]^2\right) \sum_{\ell=d+1}^{\infty} \lambda_\ell \qquad \text{and} \qquad \phi_{j;d} = \left(\sum_{\ell=d+1}^{\infty} \|\Phi_j(e_\ell)\|^2\right)^{1/2}$$

  - The constant γ_d bounds the additional prediction error due to dimension reduction
  - Note that $\phi_{j;d} \leq \|\Phi_j\|_{\mathcal{S}}$ for all $d \geq 0$ and $\sigma^2 = E[\|\varepsilon_{n+1}\|^2]$
  - As a simple consequence, the error in (2) tends to σ^2 for $d \to \infty$
  - Needed is a criterion to select order p and dimension d simultaneously

Methods Based on FPC Scores

- *Model selection — 2; A, Dubart Norinho & Hörmann (2015)*
  - Since the eigenfunctions e_ℓ are orthogonal and the FPC scores $x_{n,\ell}$ are uncorrelated, it follows

$$E\left[\|X_{n+1} - \hat{X}_{n+1}\|^2\right] = E\left[\left\| \sum_{\ell=1}^{\infty} x_{n+1,\ell} e_\ell - \sum_{\ell=1}^{d} \hat{x}_{n+1,\ell} e_\ell \right\|^2\right]$$

$$= E\left[\|\boldsymbol{Y}_{n+1} - \hat{\boldsymbol{Y}}_{n+1}\|^2\right] + \sum_{\ell=d+1}^{\infty} \lambda_\ell$$

(For vectors, $\|\cdot\|$ denotes Euclidean norm)

Methods Based on FPC Scores

- *Model selection — 2; A, Dubart Norinho & Hörmann (2015)*
  - Since the eigenfunctions e_ℓ are orthogonal and the FPC scores $x_{n,\ell}$ are uncorrelated, it follows

$$E\left[\|X_{n+1} - \hat{X}_{n+1}\|^2\right] = E\left[\left\|\sum_{\ell=1}^{\infty} x_{n+1,\ell}e_\ell - \sum_{\ell=1}^{d} \hat{x}_{n+1,\ell}e_\ell\right\|^2\right]$$

$$= E\left[\|\boldsymbol{Y}_{n+1} - \hat{\boldsymbol{Y}}_{n+1}\|^2\right] + \sum_{\ell=d+1}^{\infty} \lambda_\ell$$

  (For vectors, $\|\cdot\|$ denotes Euclidean norm)

  - To minimize the prediction error, set up the fFPE model selection criterion:

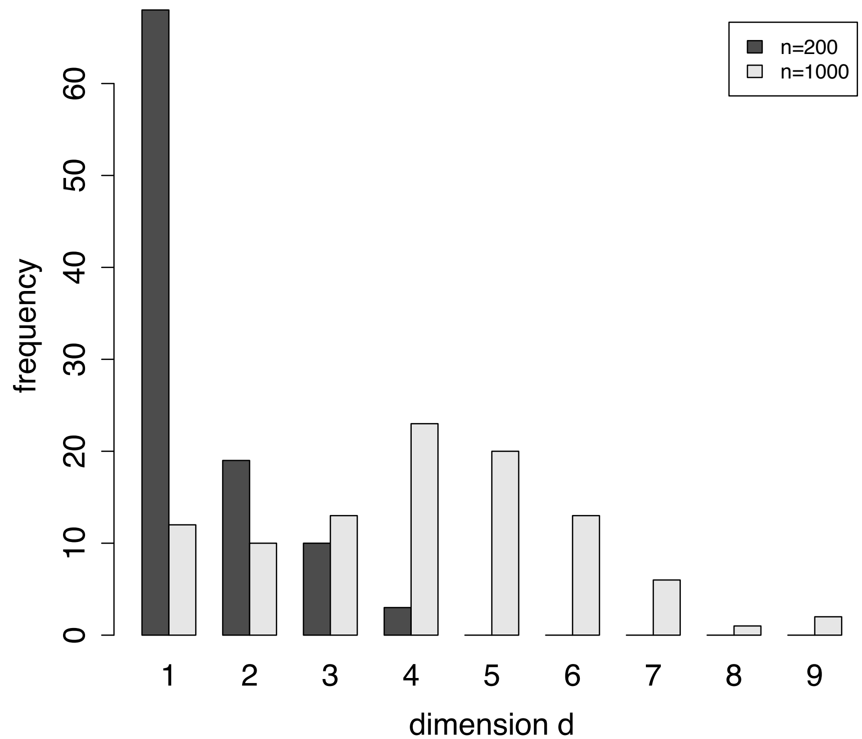
$$(\hat{p}, \hat{d}) = \arg\min_{p,d} \left\{ \frac{n + pd}{n - pd} \mathrm{tr}(\Sigma) + \sum_{\ell=d+1}^{\infty} \lambda_\ell \right\},$$

  where Σ is the covariance matrix of the residuals from a VAR(p) fit to $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$

  - Note that the multivariate FPE criterion uses the determinant instead of the trace
  - To get a fully automatic procedure, replace all population with sample quantities

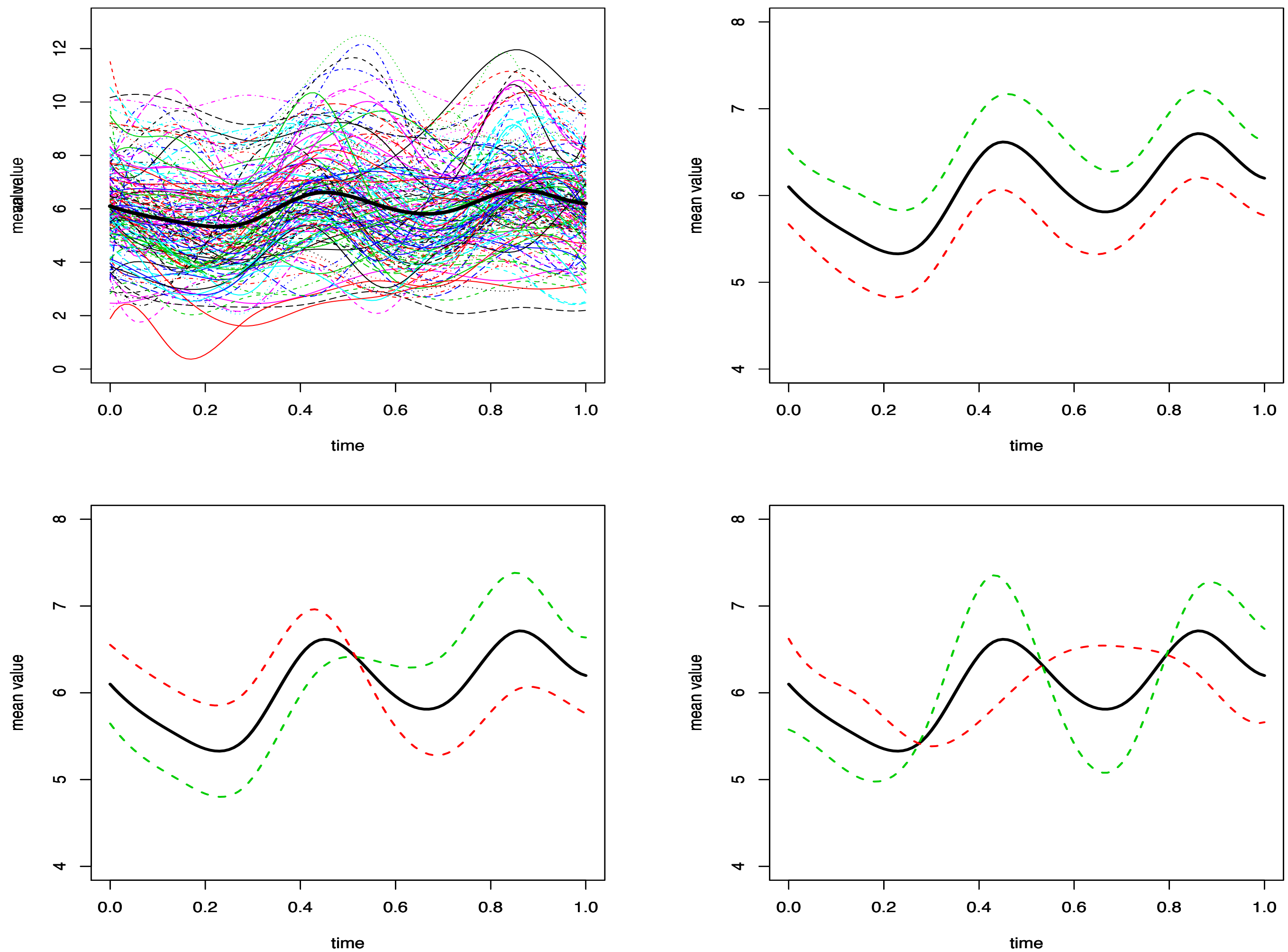
Functional FPE Criterion

- *Effect on dimension reduction*
  - Frequencies of the dimension d chosen by in 100 simulation runs for FAR(1) process
  - Plot shows that fFPE adapts to sample size



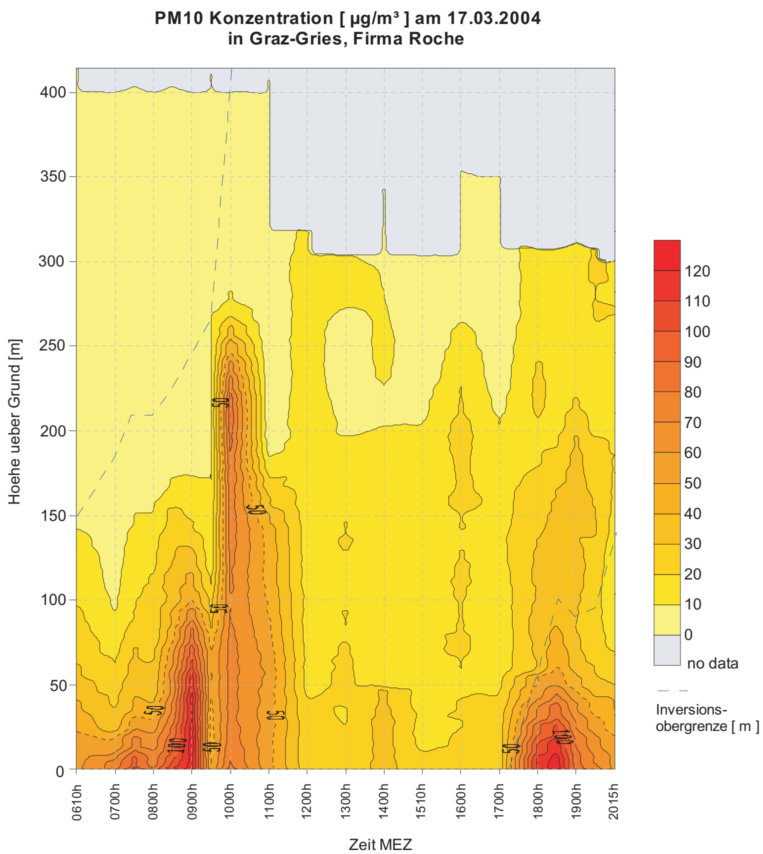
Predicting Daily Pollution Curves

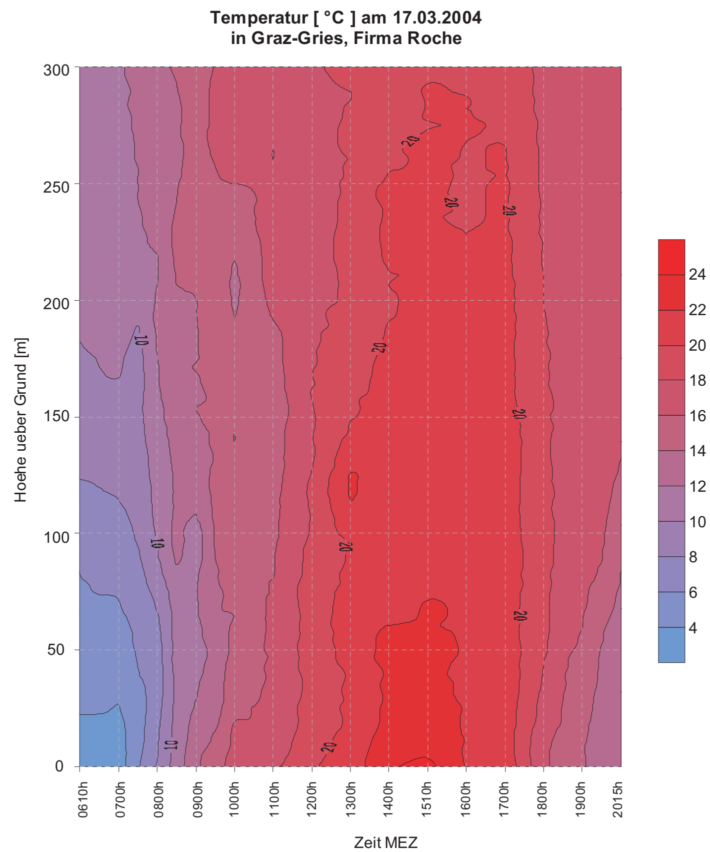
- 175 PM10 functional observations, mean function and effect of first three fPCs (90% TVE)



Predicting Daily Pollution Curves

- *Temperature difference as important covariate*
  - High PM10 concentrations are related to temperature inversions
  - Temperature difference between Graz ($350\,m$) and Kalkleiten ($710\,m$)





Predicting Daily Pollution Curves

- *Including covariates in the prediction algorithm*
  - Include temperature difference as covariate function
  - The first two FPCs describe about 92% of the variance
  - Leads to the inclusion of a two-dimensional regressor in the second step of the algorithm
  - Fit d-variate VARX(p) model to the data
  - Select d and p with covariate-adjusted fFPE criterion

$$\mathrm{fFPE}(p,d) = \frac{n+pd+r}{n-pd-r}\,\mathrm{tr}(\hat{\Sigma}_{\boldsymbol{Z}}) + \sum_{\ell>d} \hat{\lambda}_\ell \tag{3}$$

  - r is the dimension of the regressor vector (here, $r=2$)
  - $\hat{\Sigma}_{\boldsymbol{Z}}$ is the covariance matrix of the residuals when a model of order p and dimension d is fit

Predicting Daily Pollution Curves

- *Comparison of three prediction methods*
  - Subscript a (b, c) corresponds to method FPE (multiple testing, FPEX)
  - Choose five blocks of functional observations $X_{j+1}, \ldots, X_{j+100}$ for $k = 0, 15, 30, 45, 60$
  - Fit the models for the different methods
  - Make one-step ahead predictions for the functions $X_{j+100+\ell}$ and for $\ell = 1, \ldots, 15$
  - Compare through mean (MSE) and median (MED) of the 15 predictions from each block
  - Report values of p and d chosen by the respective methods

k	p_a	p_b	p_c	d_a	d_b	d_c	MSE_a	MSE_b	MSE_c	MED_a	MED_b	MED_c
0	1	1	2	3	3	3	1.33	**1.28**	1.32	1.28	1.23	**0.88**
15	3	1	3	3	3	3	2.69	5.23	**2.50**	2.38	5.34	**1.45**
30	4	1	3	3	2	3	2.05	4.05	**1.93**	1.33	2.56	**1.26**
45	3	1	3	3	2	3	2.25	2.44	**1.83**	1.34	1.67	**1.14**
60	2	1	1	3	2	5	1.22	1.82	**1.05**	1.12	1.60	**0.89**

C. Prediction and Estimation Methodology

Motivation

- *What is there*
  - Estimation can be done for several special cases
  - FAR models are covered
    - ∗ First-order case is thoroughly developed
  - Some techniques for first-order FMA models are available; Turbillon et al. (2008)
    - ∗ Procedures use restrictive assumptions

Motivation

- *What is there*
  - Estimation can be done for several special cases
  - FAR models are covered
    - ∗ First-order case is thoroughly developed
  - Some techniques for first-order FMA models are available; Turbillon et al. (2008)
    - ∗ Procedures use restrictive assumptions

- *Extension to more general setting*
  - Describe a principled way to estimate *invertible* functional time series
  - Would like to use projections but need to take into account their properties
  - Look at innovations algorithm for vector time series
  - Use concept in functional context, and for estimation

Motivation

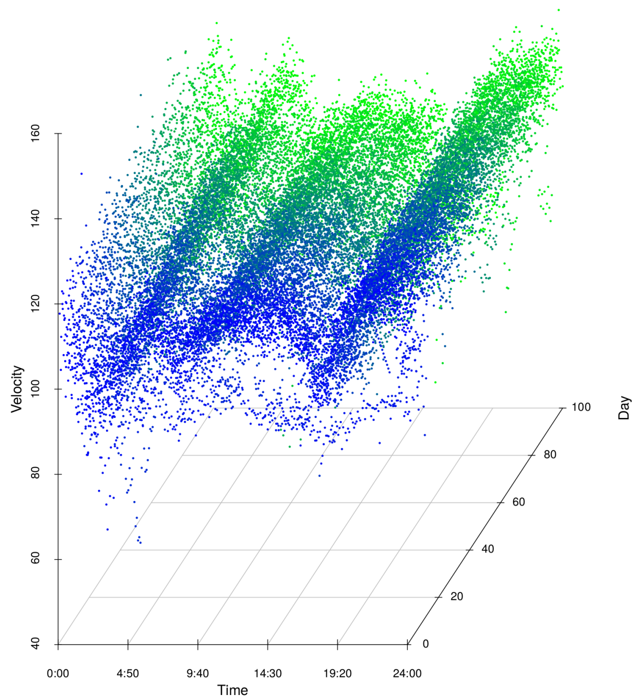
- *What is there*
  - Estimation can be done for several special cases
  - FAR models are covered
    - ∗ First-order case is thoroughly developed
  - Some techniques for first-order FMA models are available; Turbillon et al. (2008)
    - ∗ Procedures use restrictive assumptions

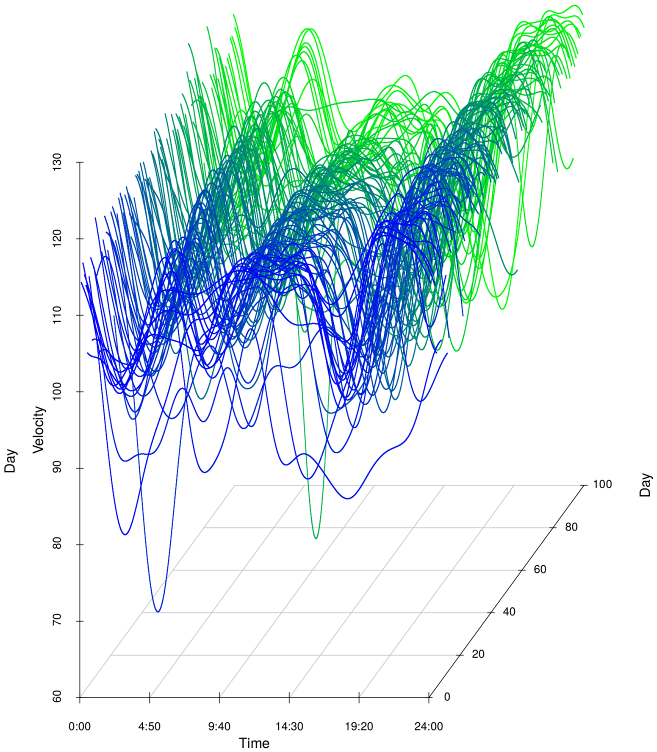
- *Extension to more general setting*
  - Describe a principled way to estimate *invertible* functional time series
  - Would like to use projections but need to take into account their properties
  - Look at innovations algorithm for vector time series
  - Use concept in functional context, and for estimation

- *For multivariate time series see Mitchell & Brockwell (1997)*

Motivation

- *Traffic volume data: Functional time series point of view*
  - Raw data organized in days (left) and corresponding functions (right)
  - Indicated periodicity in days
  - Due to double averaging process, smoothness is generated



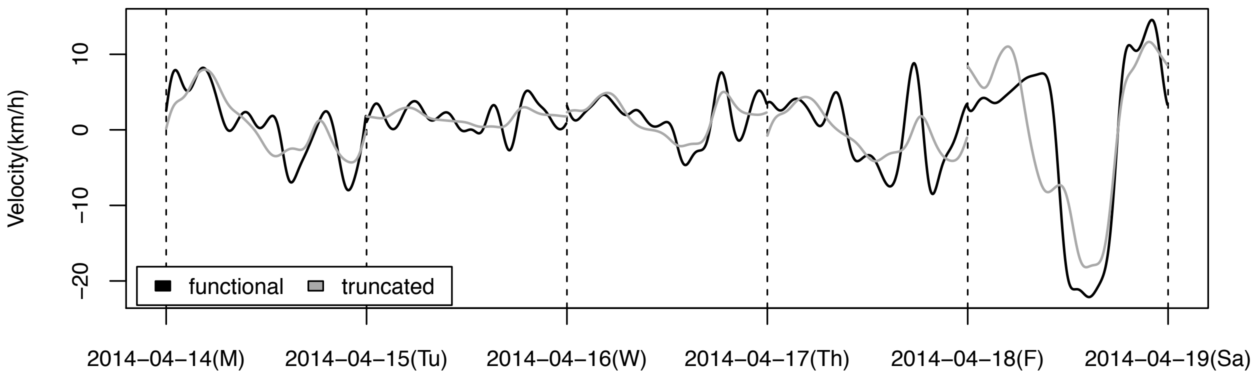


Motivation

- *Functional PCA*
  - Works for "approximable" functional time series; Hörmann & Kokoszka (2010)
  - Know: Have to be careful with description of functional and multivariate dynamics
  - Know: Invertibility is preserved under projections; Klepsch & Klüppelberg (2017)

Motivation

- *Functional PCA*
  - Works for "approximable" functional time series; Hörmann & Kokoszka (2010)
  - Know: Have to be careful with description of functional and multivariate dynamics
  - Know: Invertibility is preserved under projections; Klepsch & Klüppelberg (2017)
- *Traffic velocity data*
  - Registered centered functions (black) and four-term KL-representation (grey)
  - Use compressed functions for estimation/prediction, assess error



Main Result

- *Theorem, technical conditions suppresed; A & Klepsch (2017)*
  - $(X_j \colon j \in \mathbb{Z})$ stationary, causal and invertible functional time series
  - *Causal representation* with operators $(\Psi_\ell \colon \ell \in \mathbb{N}_0)$ given by

$$X_j = \sum_{\ell=1}^{\infty} \Psi_\ell \epsilon_{j-\ell}, \qquad j \in \mathbb{Z}$$

  - *Invertible representation* with operators $(\Pi_\ell \colon \ell \in \mathbb{N})$ given by

$$X_j = \sum_{\ell=1}^{\infty} \Pi_\ell X_{j-\ell} + \varepsilon_j, \qquad j \in \mathbb{Z}$$

Main Result

- *Theorem, technical conditions suppresed; A & Klepsch (2017)*
  - $(X_j\colon j \in \mathbb{Z})$ stationary, causal and invertible functional time series
  - *Causal representation* with operators $(\Psi_\ell\colon \ell \in \mathbb{N}_0)$ given by

$$X_j = \sum_{\ell=1}^{\infty} \Psi_\ell \epsilon_{j-\ell}, \qquad j \in \mathbb{Z}$$

  - *Invertible representation* with operators $(\Pi_\ell\colon \ell \in \mathbb{N})$ given by

$$X_j = \sum_{\ell=1}^{\infty} \Pi_\ell X_{j-\ell} + \varepsilon_j, \qquad j \in \mathbb{Z}$$

  - Recursively determine with the functional innovations algorithm the coefficients $\Theta_{k,i}$ in

$$\hat{X}_{n+1,k} = \sum_{i=1}^{k} \Theta_{k,i}(X_{d_{k+1-i},n+1-i} - \hat{X}_{n+1-i,k-i})$$

  - Then, as $k \to \infty$,

$$\|\Theta_{k,\ell} - \Psi_\ell\| \to 0$$

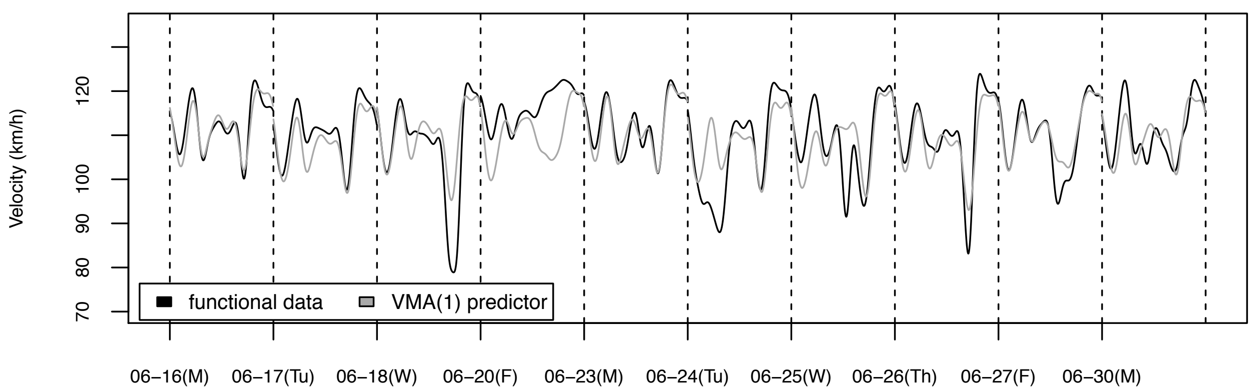
Main Result

- *Sample version*
    - There is a sample version of this result as well
    - Operators in both causal and invertible representation are consistently estimable

Main Result

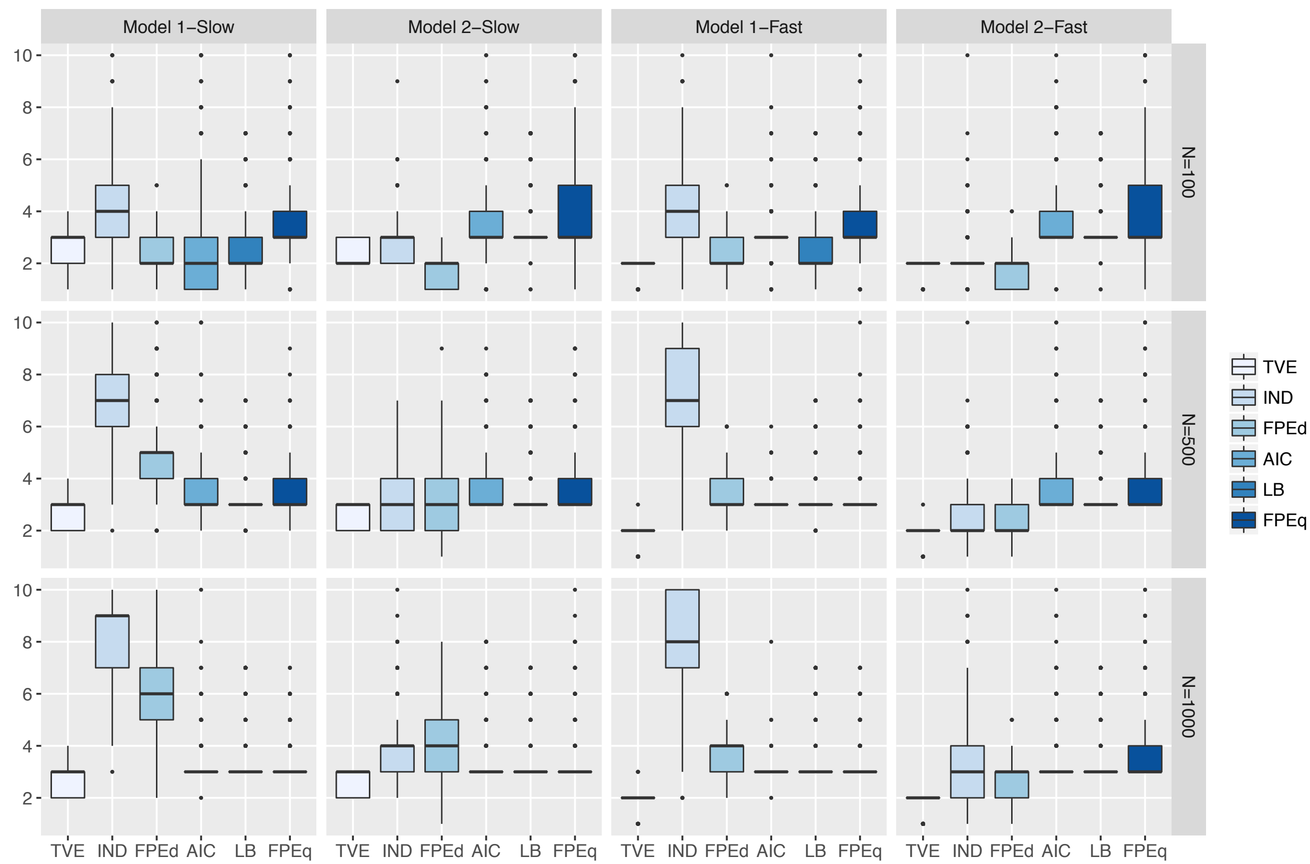
- *Sample version*
  - There is a sample version of this result as well
  - Operators in both causal and invertible representation are consistently estimable

- *Traffic velocity data*
  - One-step predictions obtained from functional innovations algorithm
  - Observed functions (black) and predictors from 10-term KL expansion



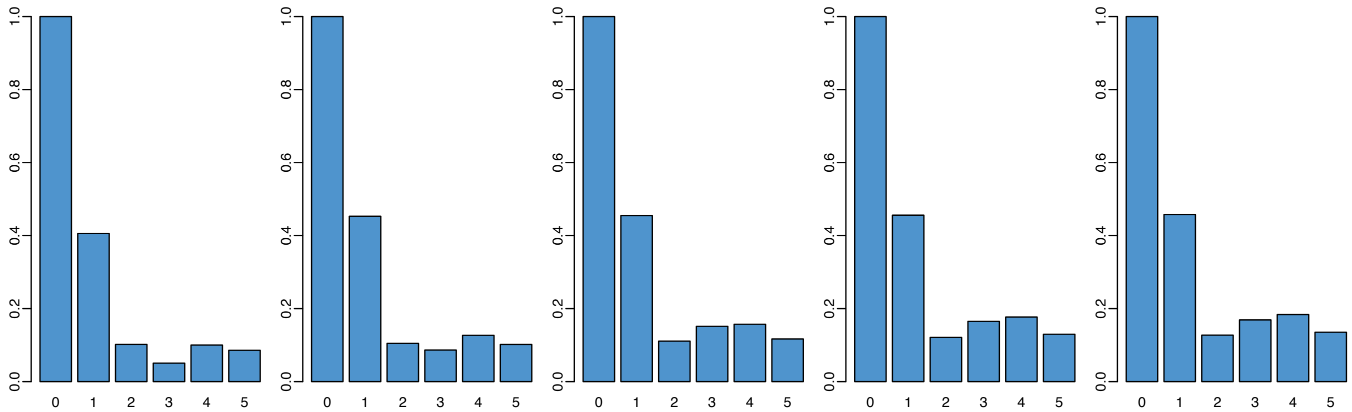
Model Selection

- *Estimating an FMA(3) process*
  - Left three boxplots are on selection of d
  - Right three boxplots on selection of q



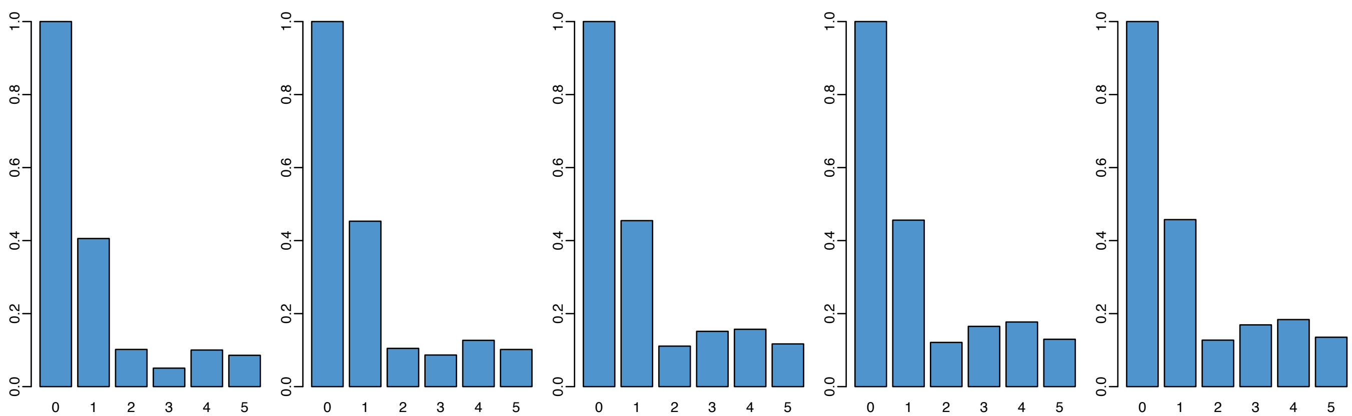
Traffic velocity data

- *Time series structure*
  - Spectral norm of estimated cross-correlation matrices for lags $h = 1, \ldots, 5$
  - Vector model based on principal subspaces of dimension $d = 1$ to $d = 5$ (left to right)



Traffic velocity data

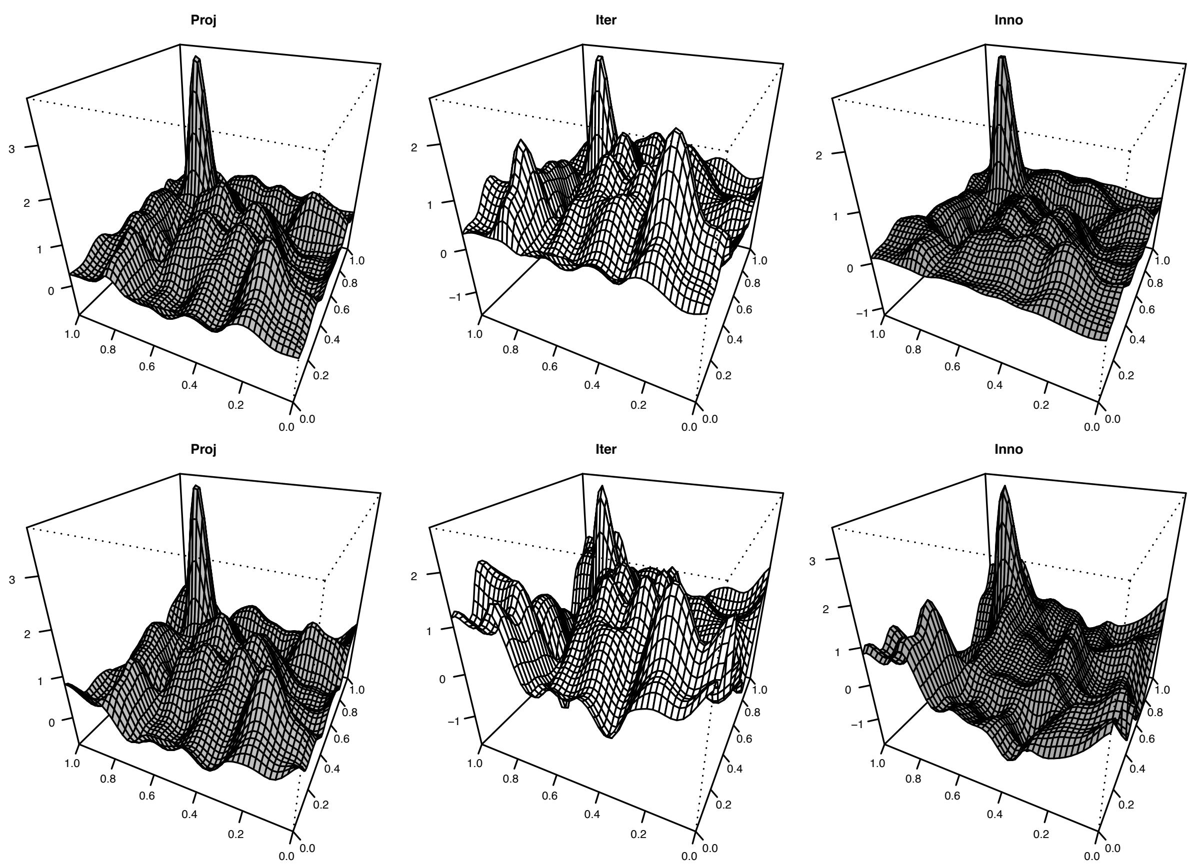
- *Time series structure*
  - Spectral norm of estimated cross-correlation matrices for lags $h = 1, \ldots, 5$
  - Vector model based on principal subspaces of dimension $d = 1$ to $d = 5$ (left to right)



- *Model selection*
  - Methods choose d between 3 and 5
  - Methods choose $q = 1$
  - This seems reasonable given the spectral norm plots

Traffic velocity data

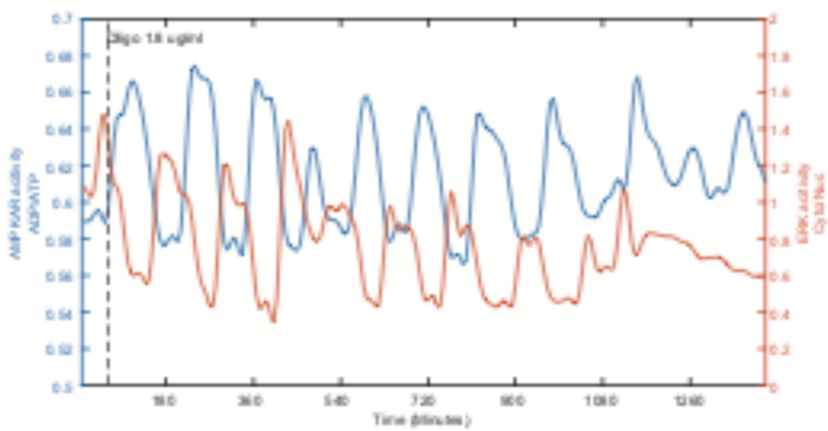
- *Estimating the moving average operator*
  - FMA(1) kernel estimated with three available methods; Turbillon et al. (2008)
  - $d = 3$ (first row) and $d = 4$ (second row)

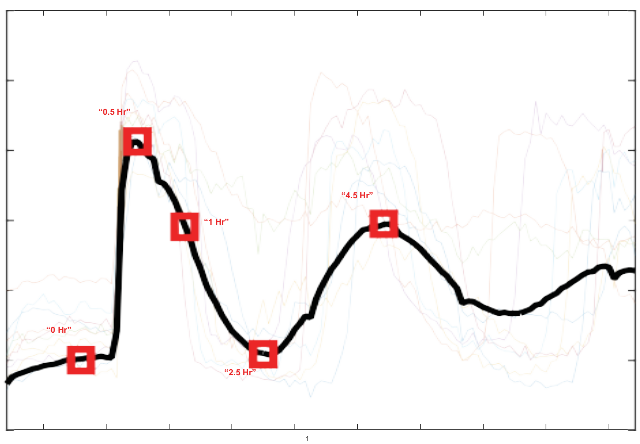


D. Future Directions

Future Directions

- *Data from single cell biology experiment*
  - Stimulating cell growth with EGF leads to “pulsing” ERK activity (red)
  - Stimulates cell metabolism measured through ATP level (blue)





Future Directions

- *Data from single cell biology experiment*
  - Stimulating cell growth with EGF leads to “pulsing” ERK activity (red)
  - Stimulates cell metabolism measured through ATP level (blue)

- *Functional time series approaches*
  - High-dimensional — graphs show one of thousands of cells (“signaling pathway”)
  - Warping — individual cells have their own clocks
  - Co-integration — groups of cells (but not all cells) seem to move together