Estimating a covariance function from fragments of functional data

Aurore Delaigle,

Joint work with Peter Hall, Wei Huang, Alois Kneip

University of Melbourne and Bonn University

Functional data: usual setting

- In statistics, we are often interested in a random variable X observed for n individuals (e.g. salary, blood pressure, age).

- Here X is a random function:

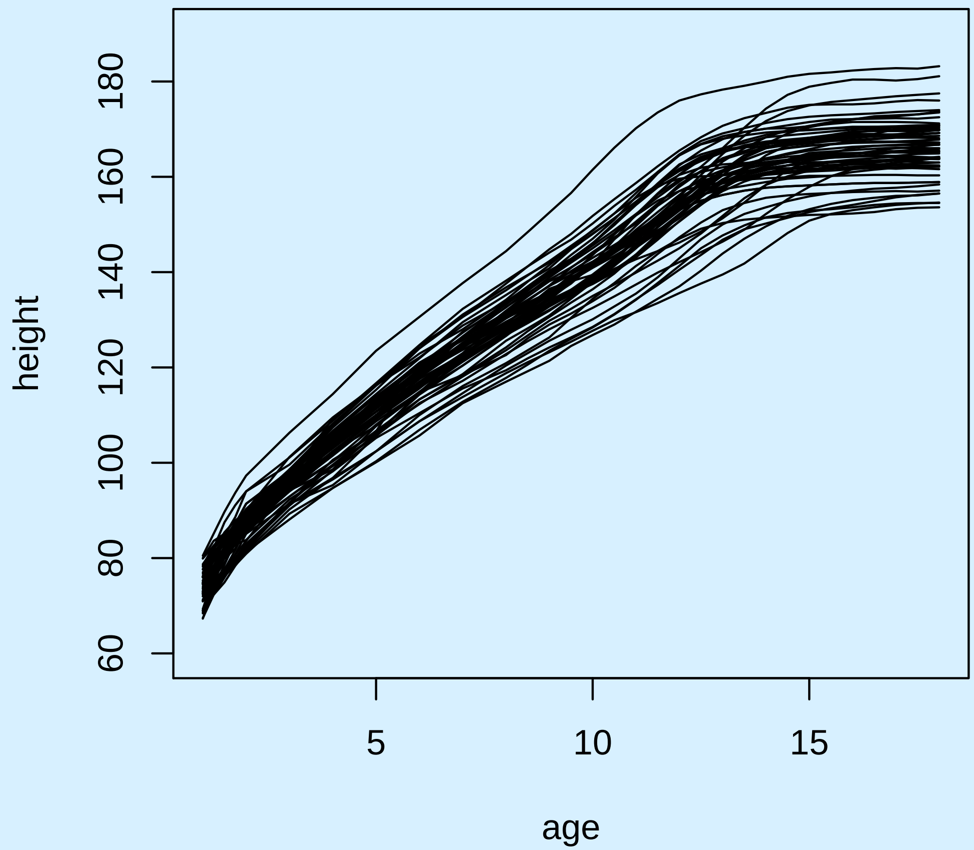
  for each of n individuals, we observe an entire curve $X(t)$, $t \in \mathcal{I}$, where $\mathcal{I}$ is a compact interval.

- We observe n i.i.d. functions $X_1, \ldots, X_n \sim X$.

Example: growth curves

Growth curves of 39 boys and 54 girls between age 1 and 18.

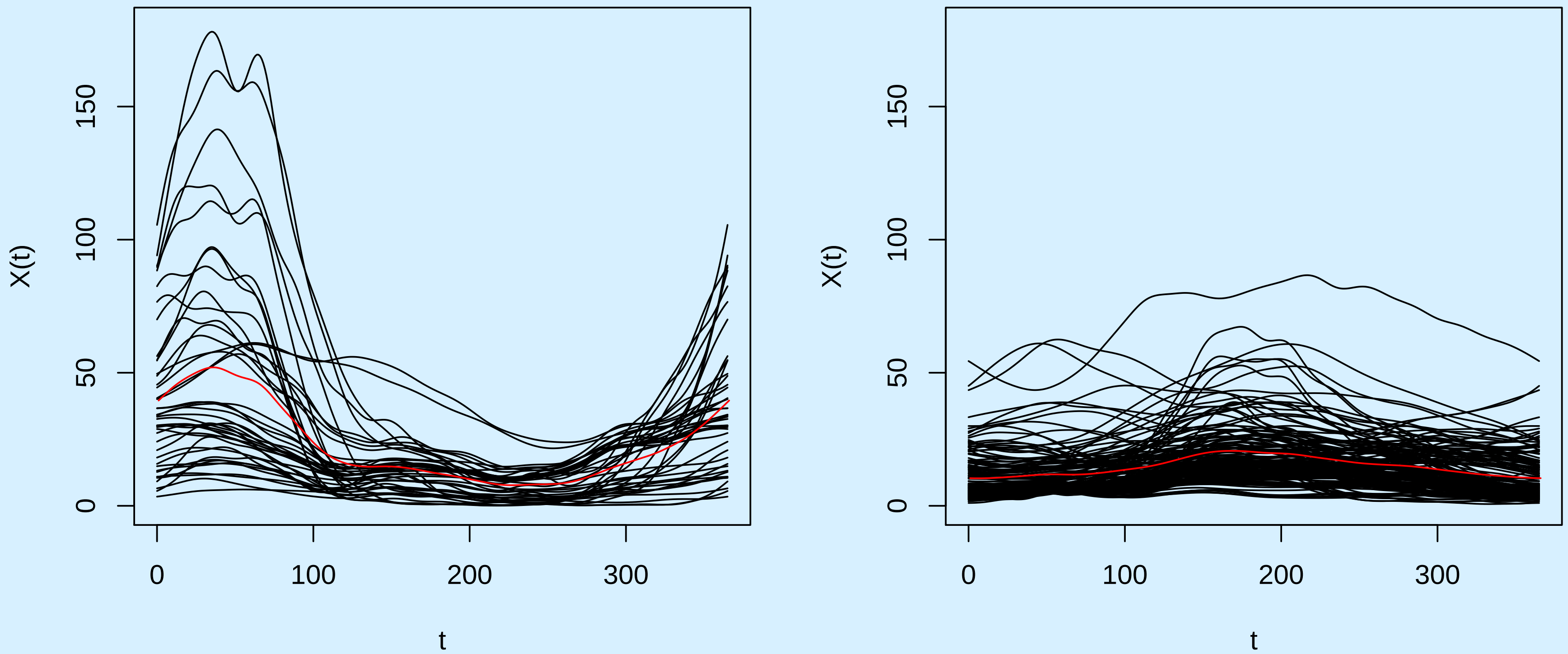
$X_i(t)$: height of ith child at age $t \in \mathcal{I} = [1, 18]$



Example: Rainfall curves

Rainfall curves from 43 northern and 147 southern Australian weather stations.

$X_i(t)$: rainfall at ith station at time $t \in \mathcal{I} = [1, 365]$



Mean and covariance

- The mean μ of X is a function defined on the interval $\mathcal{I}$ by

$$\mu(t) = E[X(t)], \quad \text{for each } t \in \mathcal{I}\,.$$

- The covariance function of X is a function $K(s,t)$ of s and t in $\mathcal{I}$, defined by

$$K(s,t) = \mathrm{cov}\{X(s), X(t)\} = E\Big(\big[X(s) - E\{X(s)\}\big]\big[X(t) - E\{X(t)\}\big]\Big).$$

It describes the variability of the population.

- Let $\phi_1, \phi_2, \ldots$ be orthogonal eigen fct. with resp eigen val. $\theta_1 \geq \theta_2 \geq \ldots \geq 0$ of the operator

$$\psi \in L^2(\mathcal{I}) \longrightarrow \int_{\mathcal{I}} K(s,t)\psi(t)\,dt\,.$$

- That is:

$$\int_{\mathcal{I}} K(s,t)\phi_j(s)\,ds = \theta_j \phi_j(t)\,.$$

- Under mild assumptions (Mercer's theorem), can decompose K into

$$K(s,t) = \sum_{j=1}^{\infty} \theta_j\, \phi_j(s)\, \phi_j(t)\,,$$

where converence is uniform over s and $t \in \mathcal{I}$.

- In practice, using the observed $X_1, \ldots, X_n$, we estimate $\mu(t)$ and $K(s,t)$ by

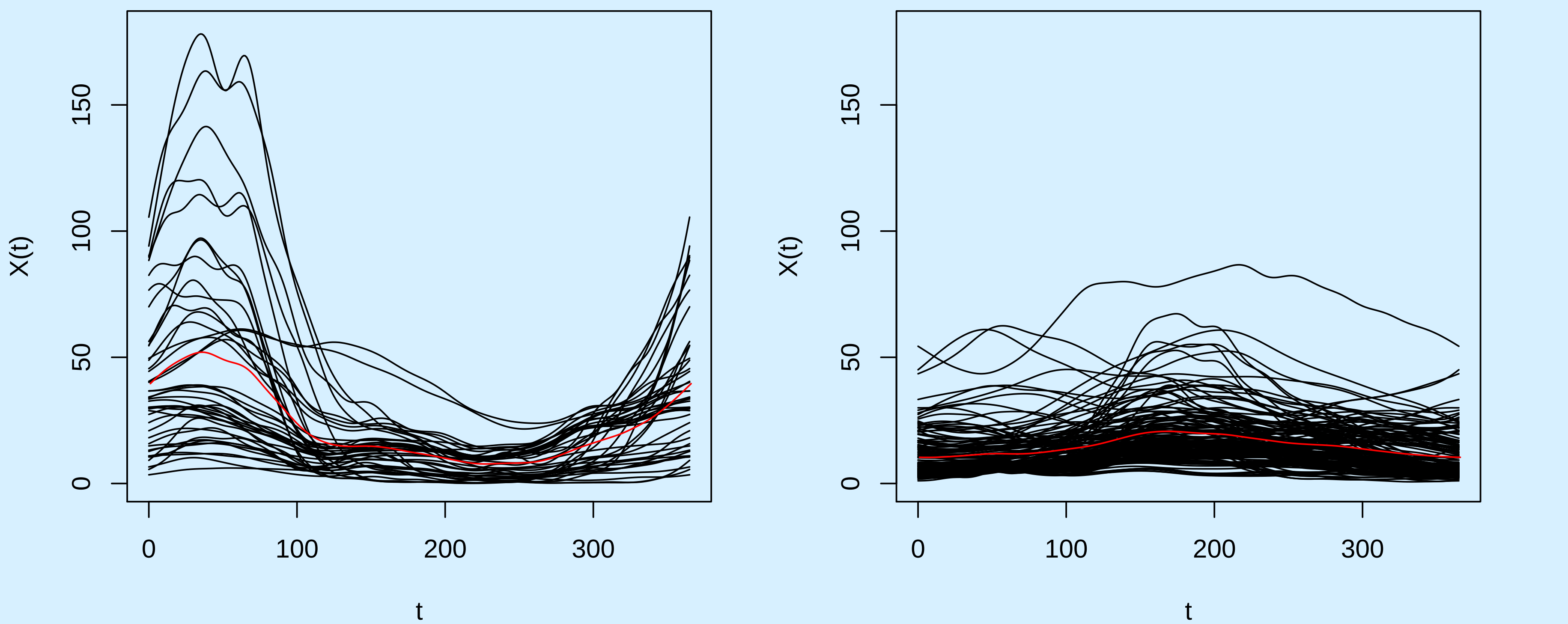
$$\hat{\mu}(t) = \bar{X}(t) = \frac{1}{n}\sum_{i=1}^{n} X_i(t)$$

$$\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(s) - \bar{X}(s)\}\{X_i(t) - \bar{X}(t)\}\,.$$

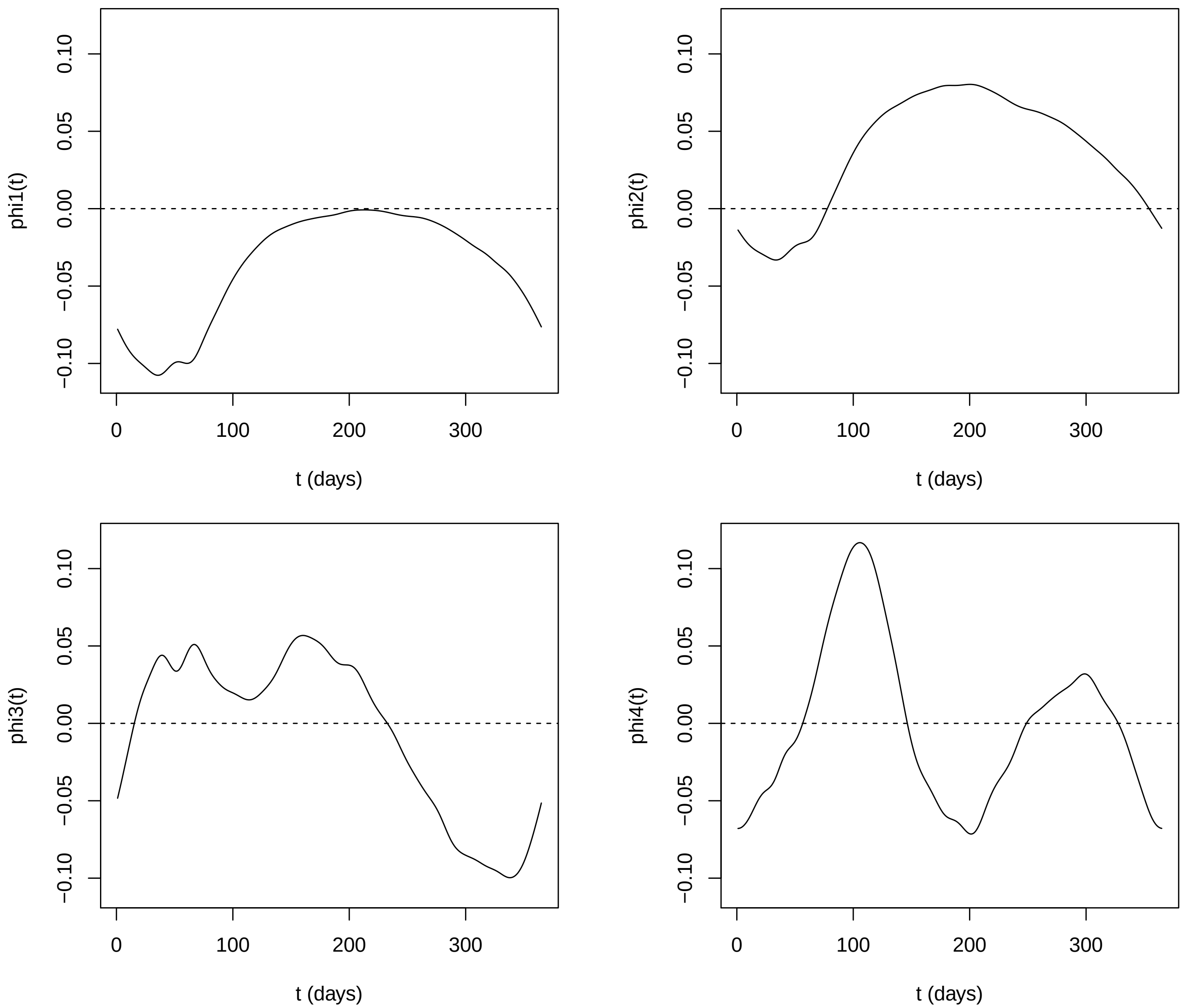
- ϕ_j's and θ_j's can be estimated empirically from the data (requires discretising integrals).

Rainfall curves from 43 northern and 147 southern Australian weather stations.

$X_i(t)$: rainfall at time $t \in \mathcal{I} = [1, 365]$



The first four eigenfunctions for the Australian rainfall data explain 70.4%, 24.8%, 3% and 1% of the variability of X:



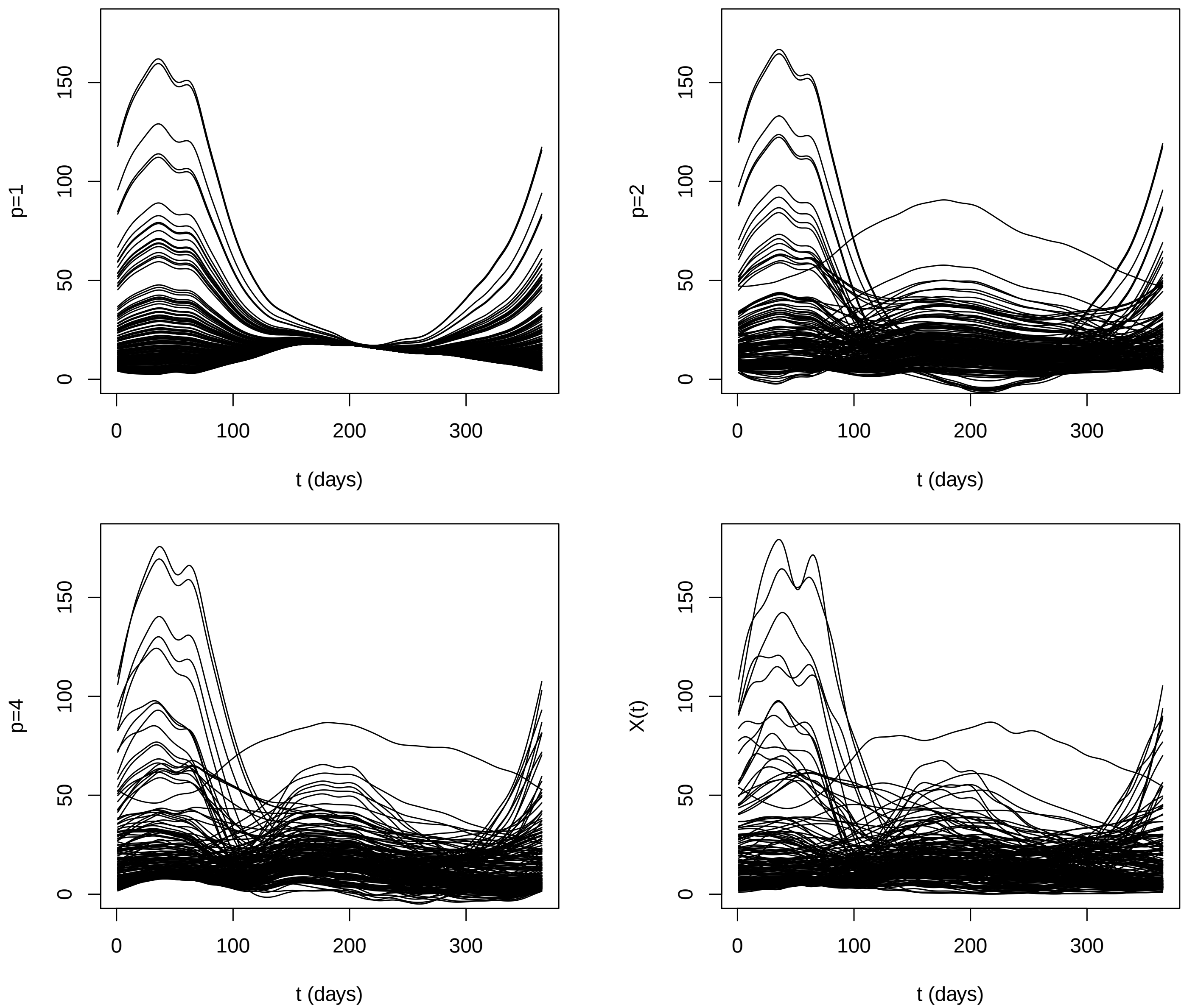
- The functions X_i can be written as

$$X_i(t) = \mu(t) + \sum_{j=1}^{\infty} \alpha_{ij}\, \phi_j(t).$$

- The ϕ_j's are such that the first few are the most important:

$$X_i(t) \approx \mu(t) + \sum_{j=1}^{p} \alpha_{ij}\, \phi_j(t).$$

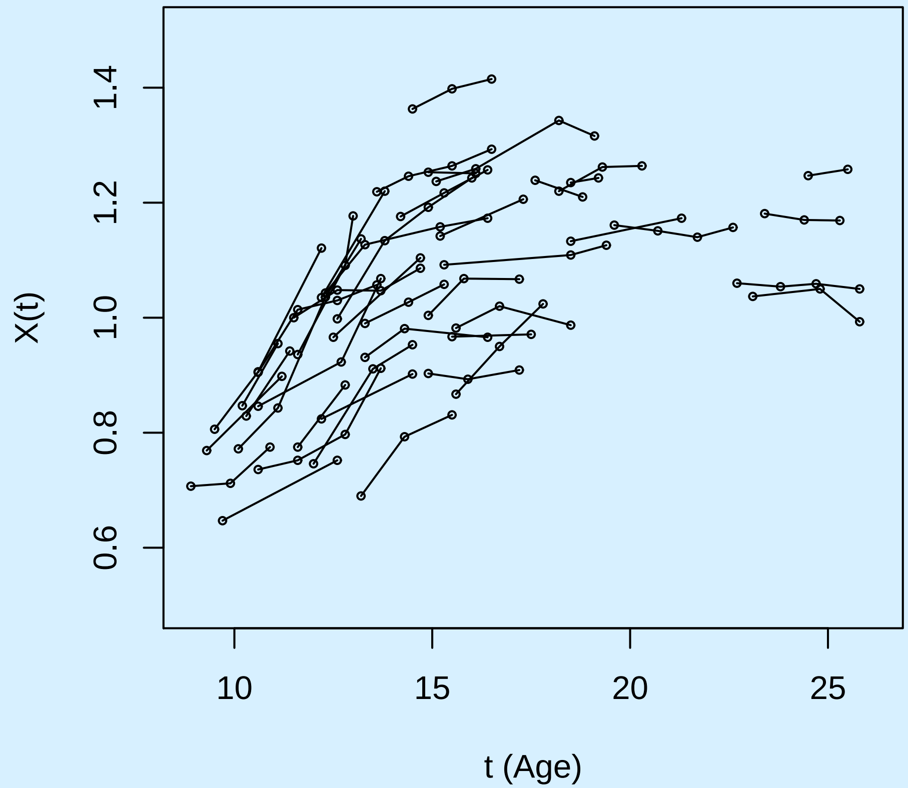
Example: rainfall data approximated by the first p terms



Functional data observed in the form of fragments

- We are interested in curves $X_1, \ldots, X_n$ defined on a compact interval $\mathcal{I} = [a, b]$.
- But: X_j observed only on a compact set $\mathcal{I}_j = [A_j, B_j] \subseteq \mathcal{I}_0$.
- Thus observe one or several fragments of curves.

Fragments – example: Growth data



partially observed growth curves of 43 females.

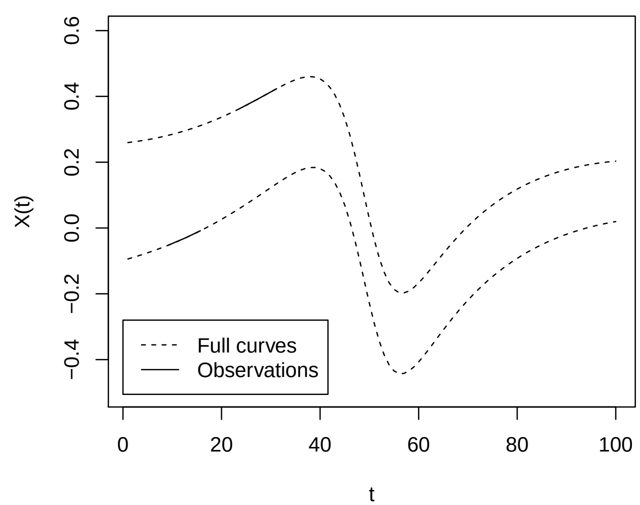
Problem

- Goal: estimate covariance function $K(s,t)$ for all $s,t \in \mathcal{I}$.

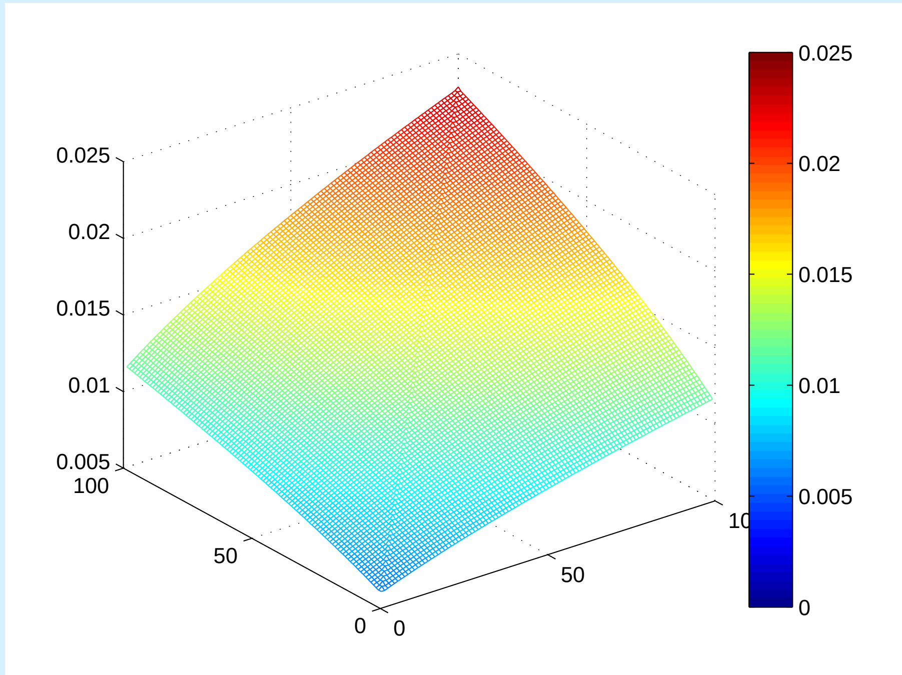
- Can compute the empirical covariance estimator only for $|s-t|$ small:

$$\hat{K}(s,t) = \frac{1}{n}\sum_{i=1}^{n}\{X_i(s) - \bar{X}(s)\}\{X_i(t) - \bar{X}(t)\}\,.$$

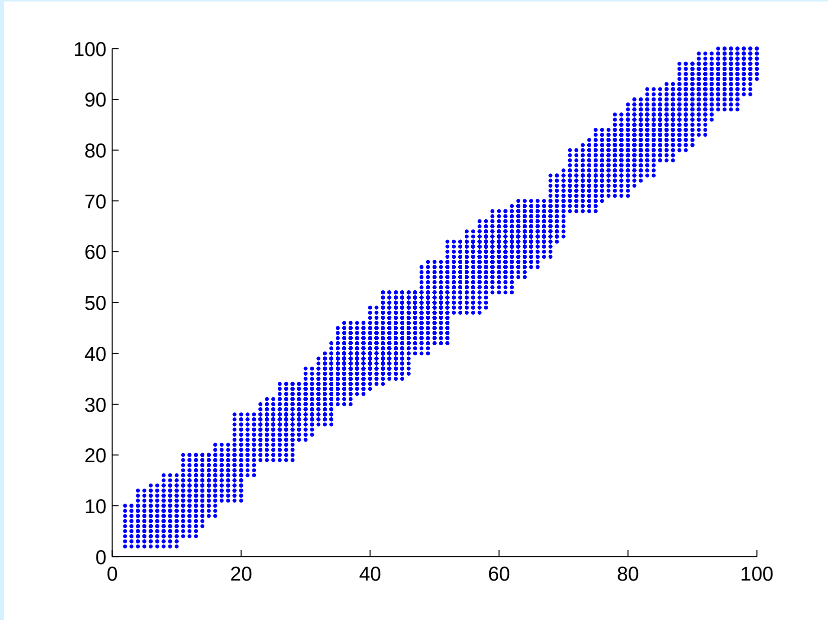
- $\hat{K}(s,t)$ requires to observe $X_i(s)$ and $X_i(t)$ and we only have that for $|s-t|$ small.



A covariance function



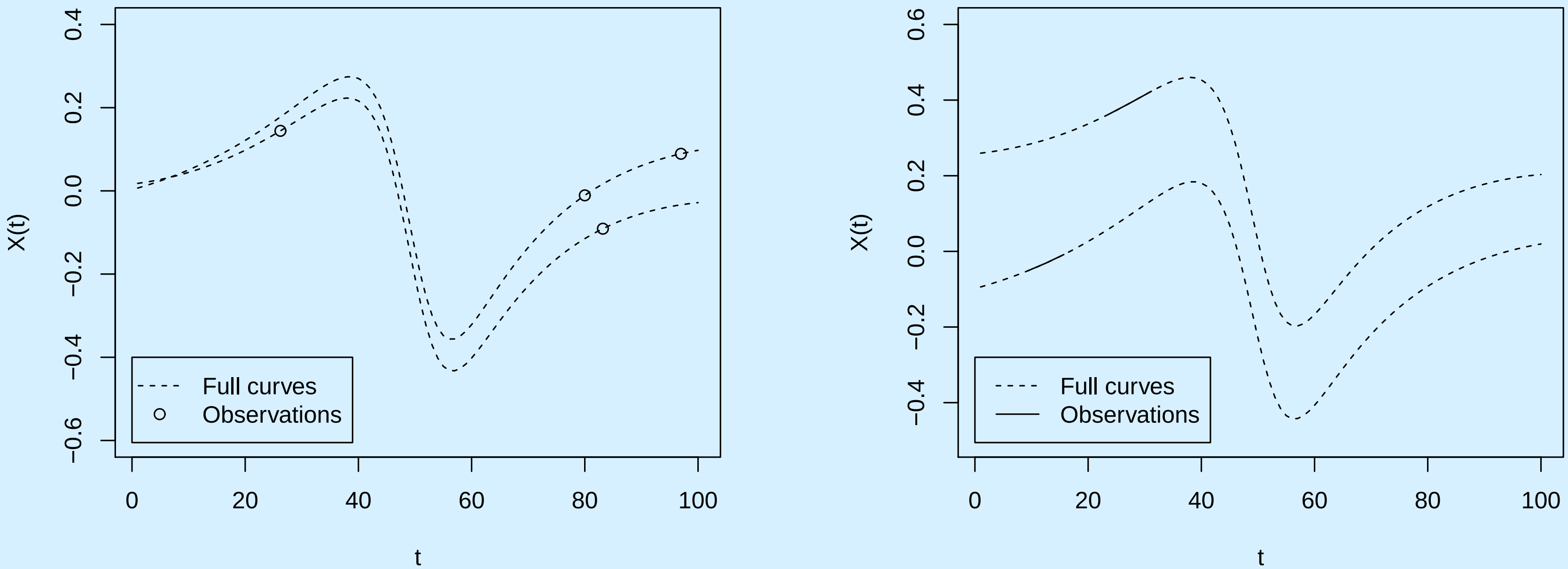
Covariance function $Cov(X(s), X(t))$ to estimate.



Scatterplot of points (s, t) at which at least one of $n = 100$ such partial curves was observed.

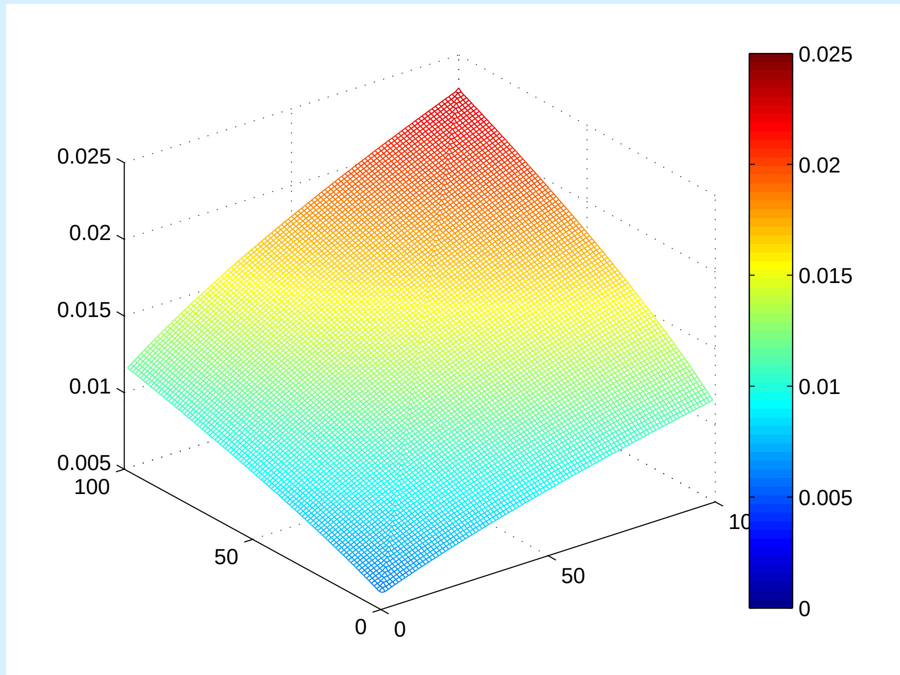
Sparse versus fragments

- Methods exist for "sparse functional data"; see Yao, Müller and Wang (2005).
- They assume ith curve is observed at a small number of points T_{ij}, where $j = 1, \ldots, N_j$.
- They assume the T_{ij}'s are i.i.d.
- Not designed for data in the form of fragments.



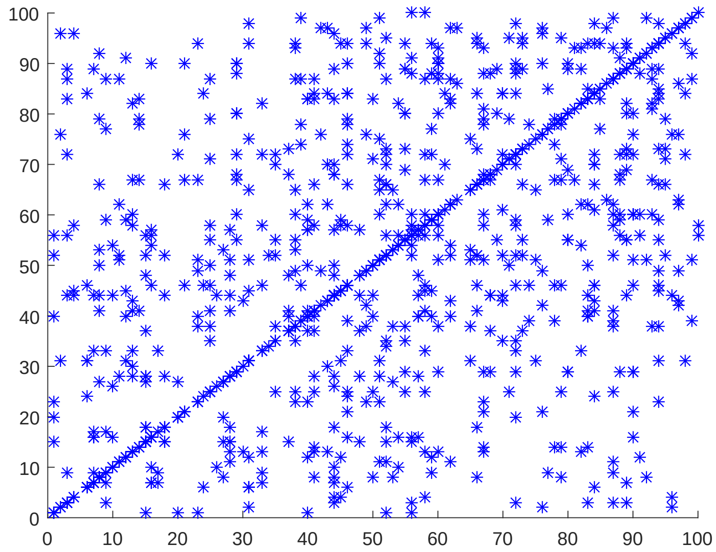
Left: sparse data. Right: fragments of curves.

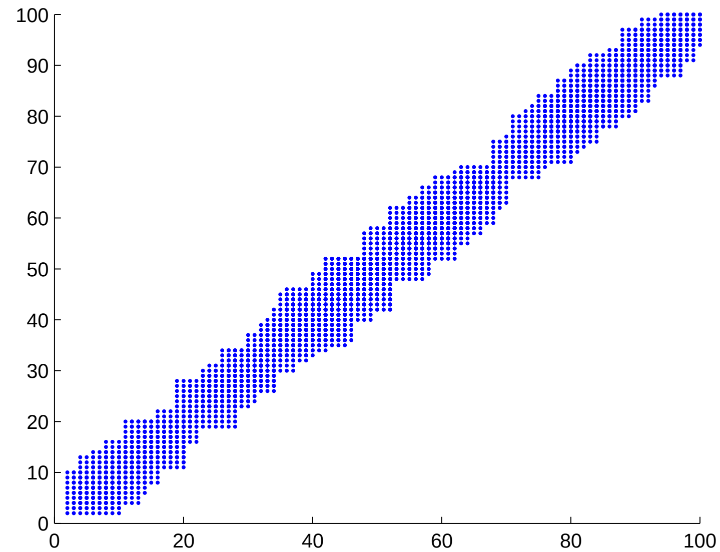
Illustration of why such methods fail for fragments



Covariance function $Cov(X(s), X(t))$ to estimate.

Illustration of why such methods fail for fragments

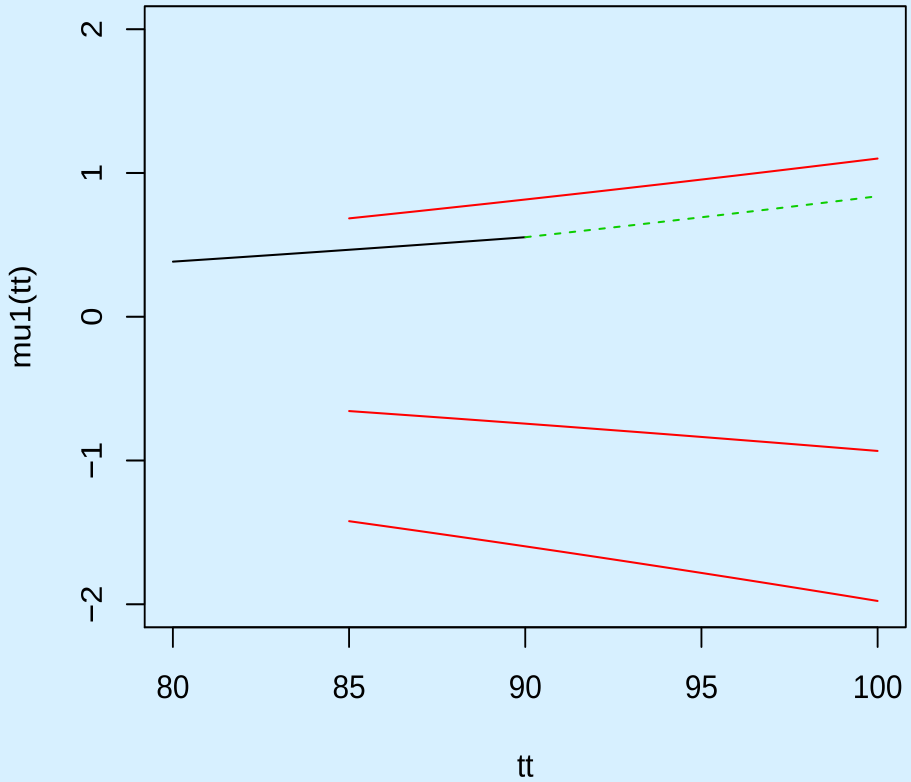




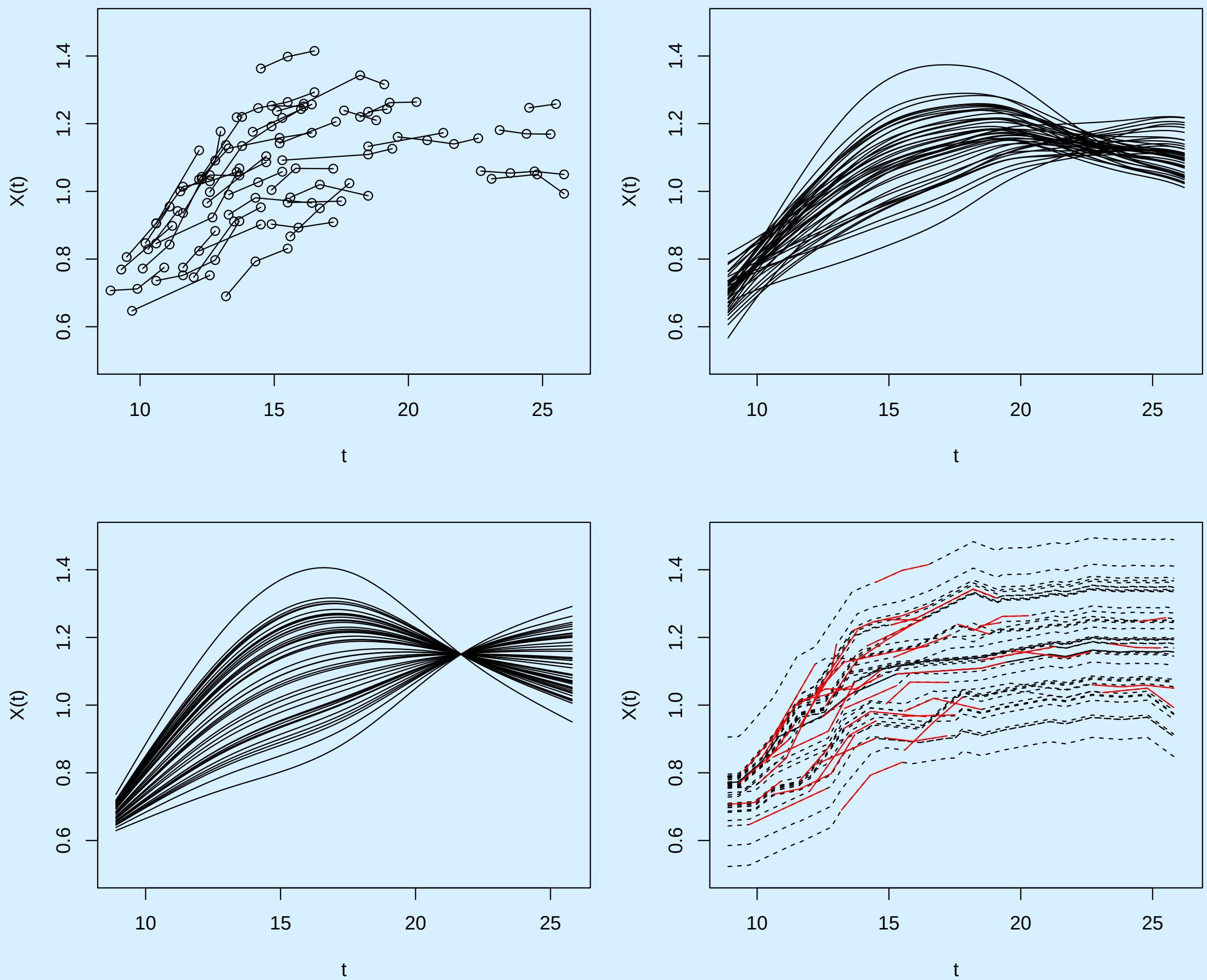
Scatterplot of points (s, t) at which at least one of $n = 100$ partial curve was observed.

Method designed for fragments

- Data are poor: cannot use sophisticated or consistent nonparam methods.
- In Delaigle and Hall (2013) we suggested a first very simple method.
- Very basic: based on copying and pasting vertically shifted fragments to each fragment.
- Then essentially use the extended fragments for inference.



Example of extension of a fragment



Yao et al.'s (2005) method with error term (top left), James and Hastie's (2001) method (bottom left), our old method (bottom right).

Method designed for fragments (part 2)

- In Delaigle and Hall (2016) model data by Markov chains.

- Discretise the process in time: $t_1, \ldots, t_{m_1}$ and space: $y_1, \ldots, y_{m_2}$.
- Reduce data $X_i(t)$, $t \in \mathcal{I}_i$, to set of point pairs $(t_j, Z_i(t_j))$.
- Assume $P\{Z(t_{k+1}) = y_\ell \mid Z(t_k), \ldots, Z(t_1)\} = P\{Z(t_{k+1}) = y_\ell \mid Z(t_k)\}$.
- For all k, j and ℓ, estimate $P\{Z(t_{k+1}) = y_\ell \mid Z(t_k) = y_j\}$.

Z observed on $[A, B] \subset [a, b]$. Markov assumption implies

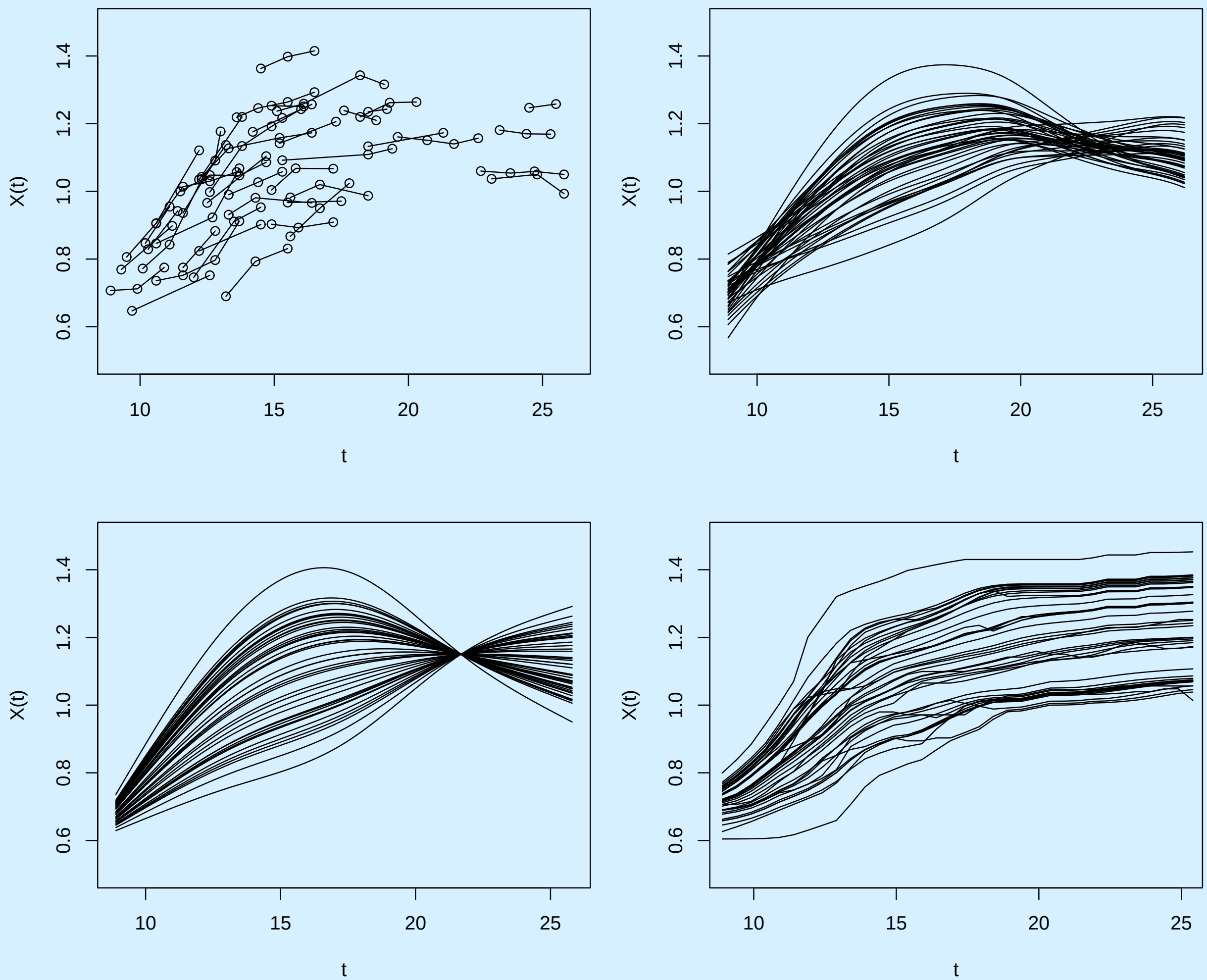
$$E\big\{Z(t)\big|Z(s)\,,\; s \in [A, B]\big\} = \begin{cases} Z(t) & \text{if } t \in [A, B] \\ E\{Z(t)\,|\,Z(A)\} & \text{if } a \le t < A \\ E\{Z(t)\,|\,Z(B)\} & \text{if } B < t \le b\,. \end{cases}$$

- We have

$$E\big\{Z(t_{j+r})\;\big|\;Z(t_j) = y_{j_1}\big\}$$
$$= \sum_{\ell=1}^{m_1} \left\{ \sum_{\text{paths from } y_{j_1} \text{ to } y_{j_r}=y_\ell} \prod_{k=1}^{r} P\big\{Z(t_{j+k}) = y_{j_{k+1}}|Z(t_{j+k-1}) = y_{j_k}\big\} \right\} y_\ell$$

- Estimate using estimated transition probabilities.
- Use matrix formulation.

Growth data example



Yao et al.'s (2005) method with error term (top left), James and Hastie's (2001) method (bottom left), our new method (bottom right).

- For covariance, use same ideas to estimate

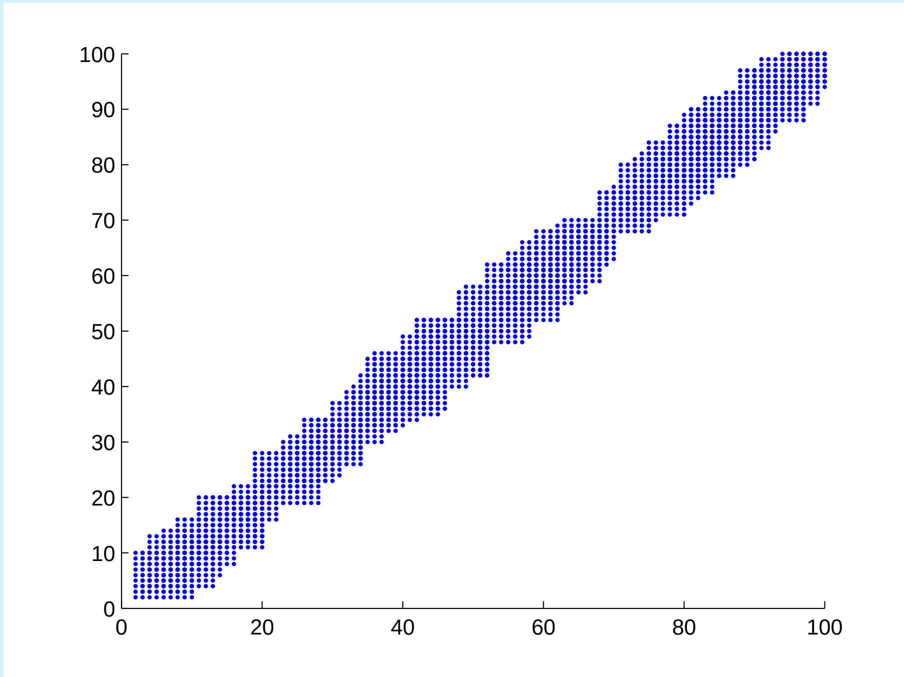
$$E\Big\{Z(t)\,Z(u)\Big|Z(s)\,,s\in[A,B]\Big\}$$
$$=\begin{cases} Z(t)\,Z(u) & \text{if } t,u\in[A,B] \\ Z(t)\,E\{Z(u)\,|\,Z(s),s\in[A,B]\} & \text{if } t\in[A,B] \text{ but } u\notin[A,B] \\ \text{etc} & \end{cases}$$

- If enough data to fit, can use higher order Markov assumption:

$$P\{Z(t_{k+r})=y_\ell\,|\,Z(t_{k+r-1}),\ldots,Z(t_1)\}=P\{Z(t_{k+r})=y_\ell\,|\,Z(t_{k+r-1}),\ldots,Z(t_k)\}.$$

Our more recent idea

- Estimating K on $\mathcal{I} \times \mathcal{I}$ from fragments is only possible if we can identify K on $\mathcal{I} \times \mathcal{I}$ by knowing K on the diagonal band where we observe data.



- In that case, we express K in an orthogonal series expansion:

$$K(s,t) = \sum_{j_1=1}^{\infty} \sum_{j_2=1}^{\infty} a_{j_1 j_2}\, \psi_{j_1}(s)\, \psi_{j_2}(t)\,. \tag{1}$$

- Then estimate the coefficients using only the diagonal band by minimising the distance between K and $\hat{K}$ computed on the diagonal band.

- Minimise under the constraint that K is symmetric and semipositive definite on $\mathcal{I} \times \mathcal{I}$.

- Then use the resulting coefficients on the whole $\mathcal{I} \times \mathcal{I}$.

Theory

- We have theoretical results giving conditions under which K is identifiable on $\mathcal{I} \times \mathcal{I}$ knowing only K on the diagonal band.
- One example is when the X_i's are real analytic, see Descary and Panaretos (2017) for similar results using another method.
- We have more general results based on fast enough decay of eigenvalues. The larger the observed band, the better.
- We can also show that even if we do not have identifiability, we can have relatively small error of approximation.

Figure 1: $n = 50$, mean fragment length: 0.5. True (left) and estimated (right).

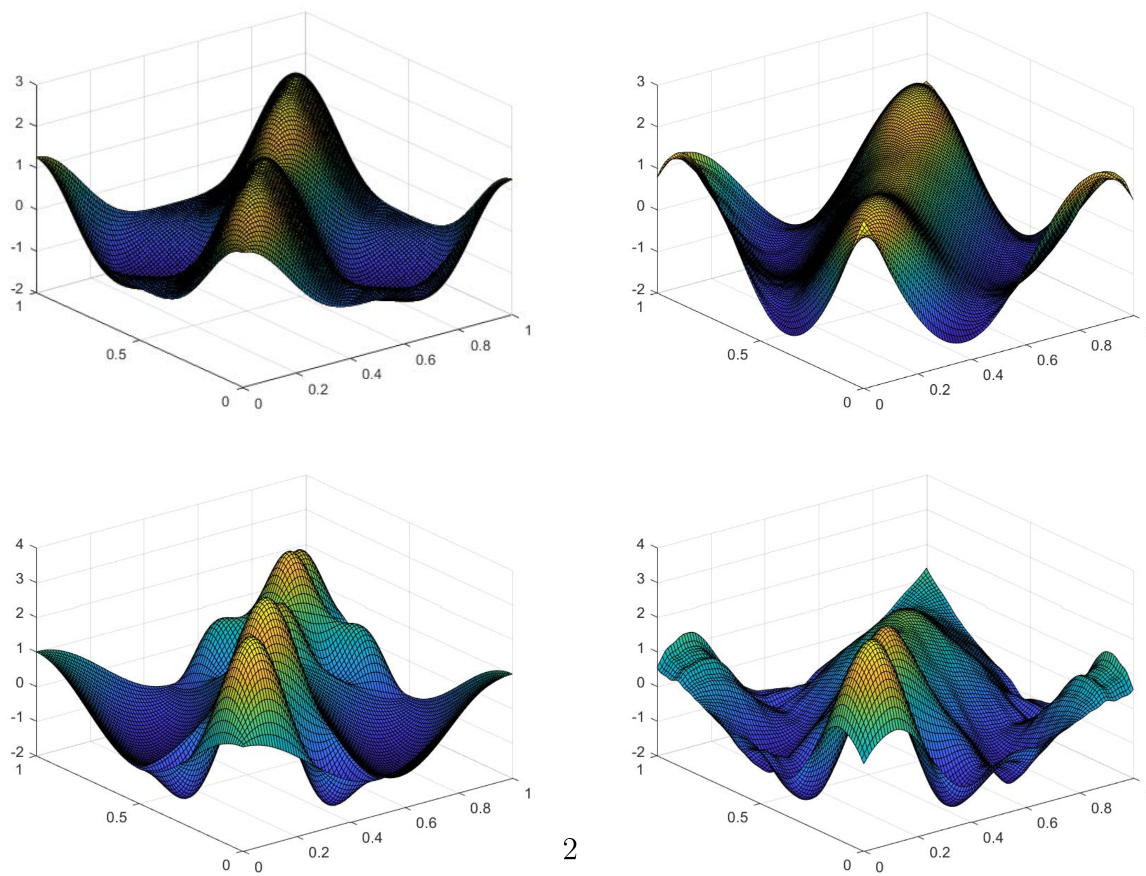
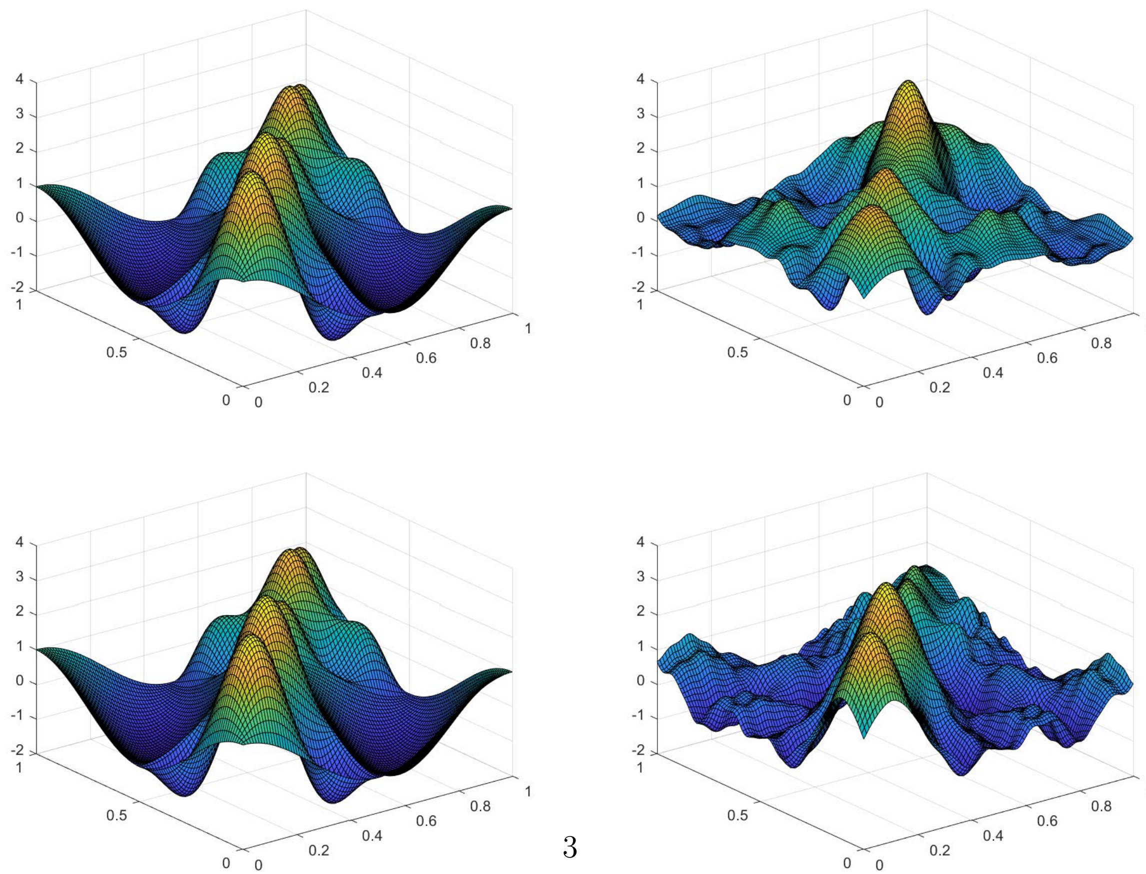


Figure 2: $n = 50$ (top) or 500 (bottom), mean fragment length: 0.2. True (left) and estimated (right).



References

- Bachrach, L. K., Hastie, T. J., Wang, M. C., Narasimhan, B. and Marcus, R. (1999). Bone mineral acquisition in healthy Asian, Hispanic, Black and Caucasian youth; a longitudinal study. *J. Clinical Endocrinology & Metabolism*, **84**, 4702–4712.
- Descary, M-H. and Panaretos, V. (2017). Recovering Covariance from Functional Fragments. *Manuscript*.
- Delaigle, A. and Hall, P. (2013). Classification using censored functional data. *J. Amer. Statist. Assoc.*, **108**, 1269–1283.
- Delaigle, A. and Hall, P. (2016). Approximating fragmented functional data by segments of Markov chains. *Biometrika*, **103**, 779–799.
- James, G. and Hastie, T. (2001). Functional linear discriminant analysis for irregularly sampled curves. *J. Roy. Statist. Soc.*, Ser. B, **63**, 533–550.
- Yao, F., Müller, H. G. and Wang, J. L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.*, **100**, 577–590