M7777 Applied Functional Data Analysis

11. Registration

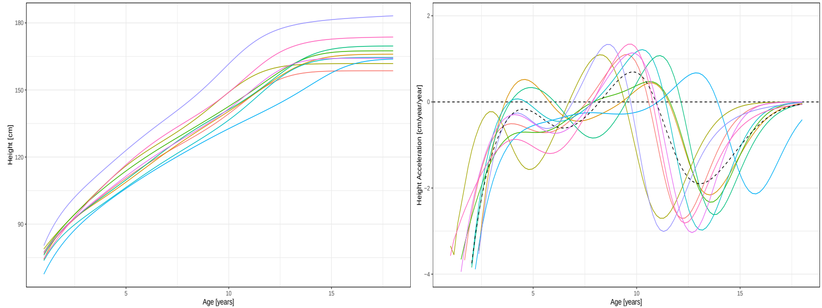
Jan Koláček (kolacek@math.muni.cz)

Dept. of Mathematics and Statistics, Faculty of Science, Masaryk University, Brno

Registration

Berkeley Growth Data

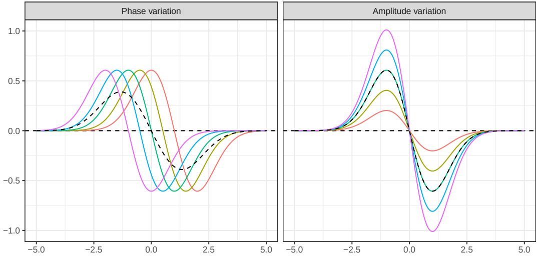
- Heights of 54 girls taken from ages 0 through 18.
- Growth process easier to visualize in terms of acceleration (2^{nd} derivative).
- Peaks in acceleration = start of growth spurts.



Sample of 10 girls.

The Registration Problem

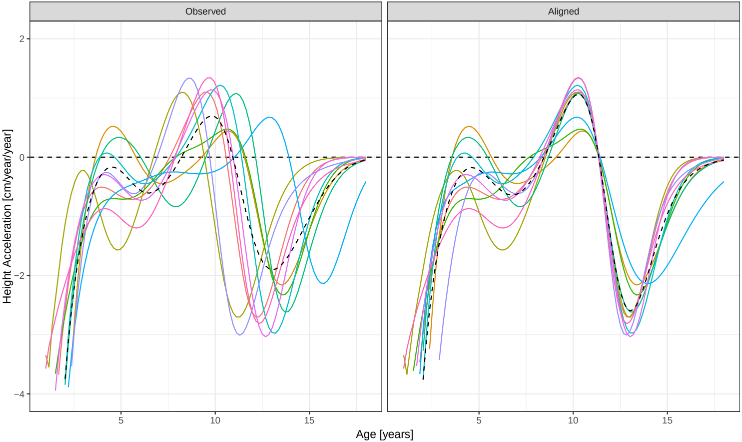
- Most analyzes only account for variation in **amplitude**.
- Frequently, observed data exhibit features that vary in time – **phase variation**.



- Mean of unregistered curves (dashed) has smaller peaks than any individual curve.
- Aligning the curves reduces variation.

Registration

Berkeley Growth Data



Sample of 10 girls.

Registration

Defining a Warping Function

Requires a transformation of **time**.

- **time-warping** function is a continuous function $h_i(t)$ defined on $[0, T]$, which is strictly increasing and $h_i(0) = 0$, $h_i(T) = T$
- **aligning** function $h_i^{-1}(t)$ is the functional inverse of $h_i(t)$, i.e.

$$h_i^{-1}(h_i(t)) = t$$

- the **registered** curves

$$x_i^*(t) = x_i(h_i^{-1}(t))$$

Registration

Landmark Registration

- For each curve $x_i(t)$ we choose points $t_{i1}, \dots, t_{iK}$
- We define a reference (usually one of the curves or mean etc.) $t_{01}, \dots, t_{0K}$
- We define constraints

$$h_i(t_{0j}) = t_{ij}, \quad i = 1, \dots, n,\ j = 1, \dots, K$$

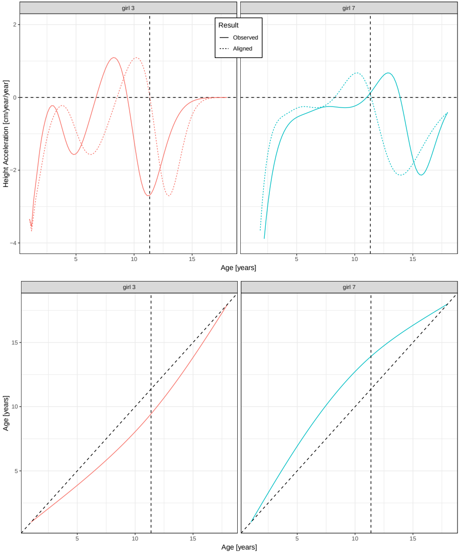
- Finally we find a constrained smooth function.

Berkeley Growth example

- Just one reference point $t_0 = 11.7$ ... average time of maximal pubertal growth spurt (acceleration crosses 0)
- t_i ... maximal pubertal growth spurt for i-th curve
- Thus the constraints

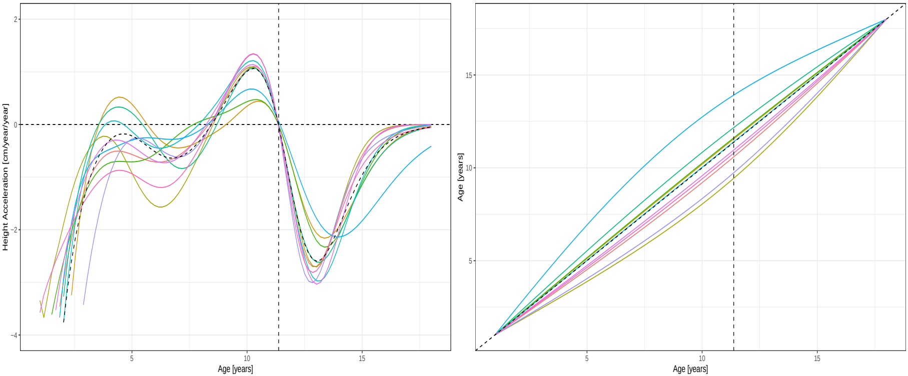
$$h_i(11.7) = t_i, \qquad i = 1, \dots, 54.$$

Registration



Aligned data with warping functions.

Registration



Aligned data for 10 girls with their warping functions.

Registration

Identifying Landmarks

Major landmarks of interest:

- where curve $x_i(t)$ crosses some value
- location of peaks or valleys
- location of inflections

Almost all are points at which some derivative of $x_i(t)$ crosses zero.

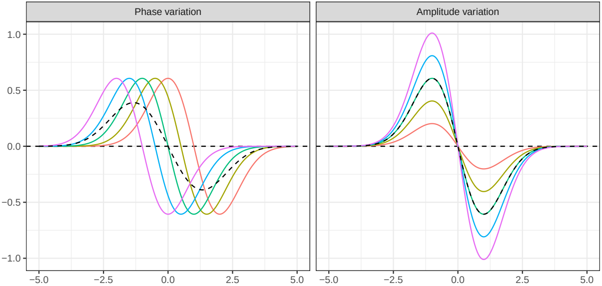
In practise, zero-crossings can be found automatically, but usually still require manual checking.

For landmark registration in R, the procedure landmarkreg is used.

Registration

Continuous Registration

- Let $x_0(t)$ be a **reference** curve (dashed)



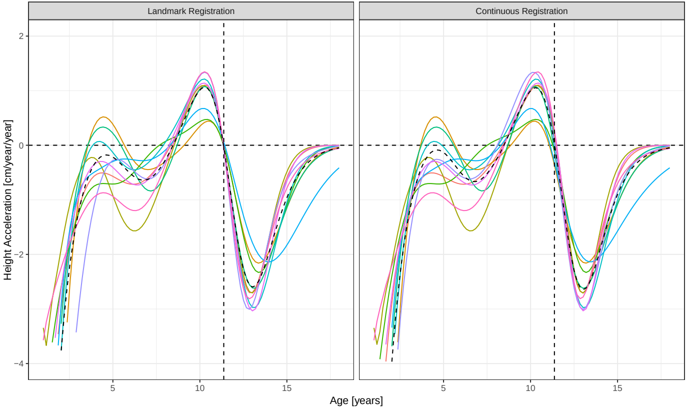
- Phase variation: curves are less correlated with $x_0(t)$,
  FPCA $\Rightarrow$ the first 3 PC's explain 55%, 39% and 5% of the variation
- Amplitude variation: curves are high correlated with $x_0(t)$,
  FPCA $\Rightarrow$ just the first PC explains 100% of the variation

Main idea: Find $h(t)$ to maximize correlation with $x_0(t)$

Registration

Practical Using

1. Do the preprocessing of data by landmark registration.
2. Do the continuous registration (we can repeat).



For continuous registration in R, the procedure register.fd is used.

Registration

Assessing the process of Registration

- The **total** MSE: $MSE_{total} = n^{-1} \sum_{i=1}^{n} \int_0^T [x_i(t) - \bar{x}(t)]^2 \, dt$
- The **amplitude** MSE: $MSE_{amp} = Cn^{-1} \sum_{i=1}^{n} \int_0^T [x_i^*(t) - \bar{x}^*(t)]^2 \, dt$
- The **phase** MSE: $MSE_{phase} = C \int_0^T \bar{x}^{*2}(t)dt + \int_0^T \bar{x}^2(t)dt,$

where $x_i(t)$ ... unregistered curves, $x_i^*(t)$ ... registered version

It can be shown

$$MSE_{total} = MSE_{amp} + MSE_{phase}$$

Thus we can define **squared multiple correlation index**

$$R^2 = \frac{MSE_{phase}}{MSE_{total}}$$

Berkeley Growth Data: unregistered × landmark: $R^2 = 0.7$
landmark × continuous: $R^2 = -0.06$

Problems to solve

1. Simulation
    - Conduct 3 steps of continuous registration on simulated data from example (use the attached script). Do it for the case of Phase variation (see Figure 1). Does it have any sense to do it for the case of Amplitude variation?
2. Pinch-Force Data
    - Load the variable `pinchraw` from the `fda` package (see Lecture 6 for data details).
    - Smooth data by B-spline basis of order 5 (with GCV-optimal λ) and plot the first derivatives of curves (see Figure 2).
    - Compute a mid point for landmark registration (see the location of the maximum of each curve) and align curves. Plot unregistered curves together with aligned curves (see Figure 3).
    - Conduct the continuous registration on aligned curves and compare results (see Figure 4).

Simulation

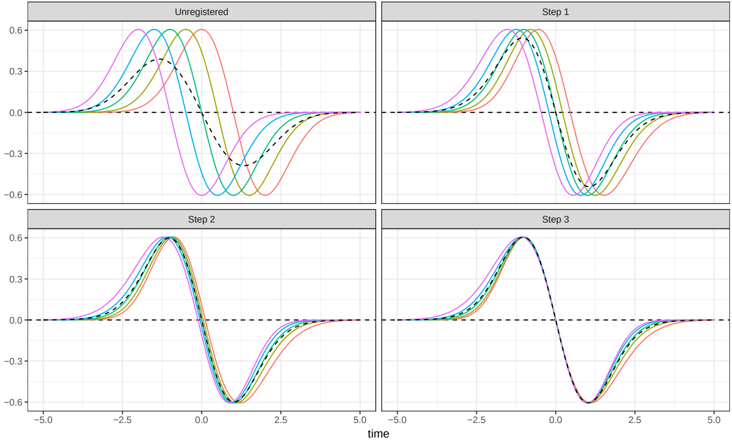


Figure 1.

Pinch-Force Data

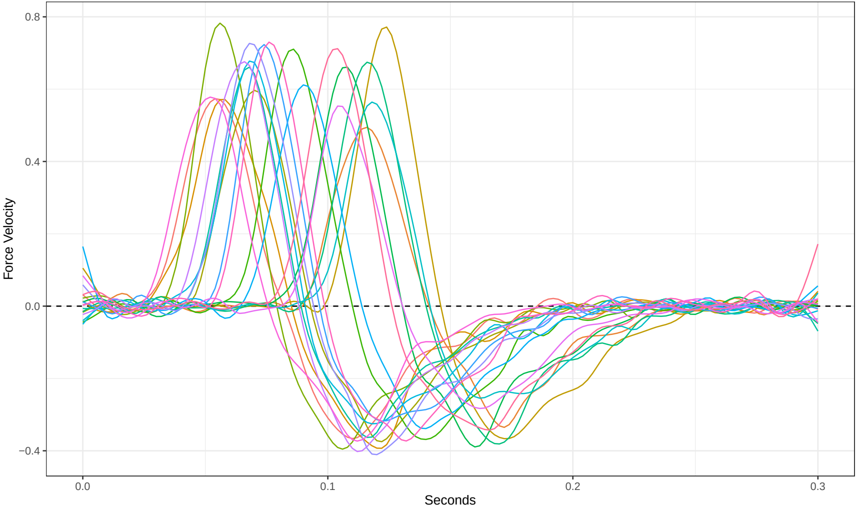


Figure 2.

Pinch-Force Data

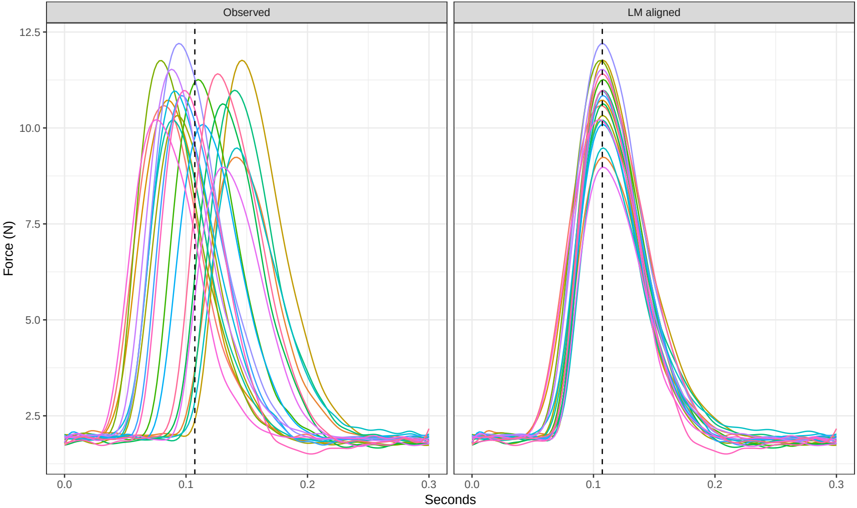


Figure 3.

Registration

Pinch-Force Data

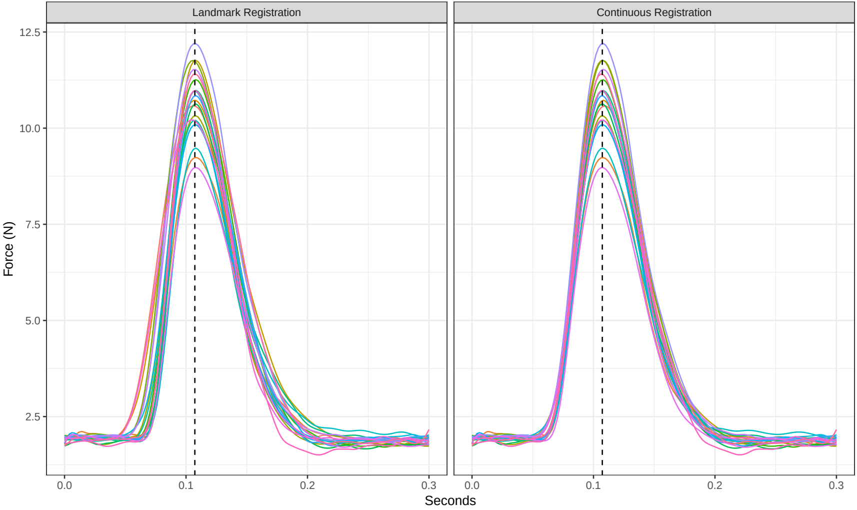


Figure 4.