An Introduction to Functional Data Analysis

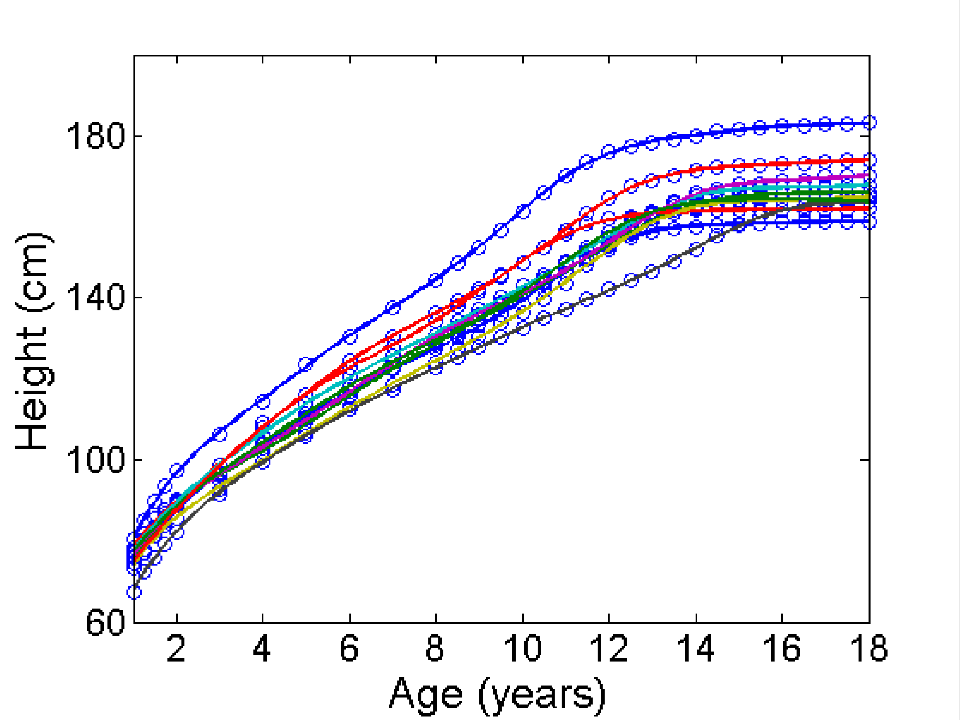
1. Overview

- What are functional data?
- Some functional data analyses
- The goals of functional data analysis
- First steps in a functional data analysis
- Using derivatives in functional data analysis

This talk follows closely the first chapter of J. O. Ramsay and B. W. Silverman, (2005) *Functional Data Analysis, Second Edition*. New York: Springer.

2. What are functional data?

Heights of ten girls



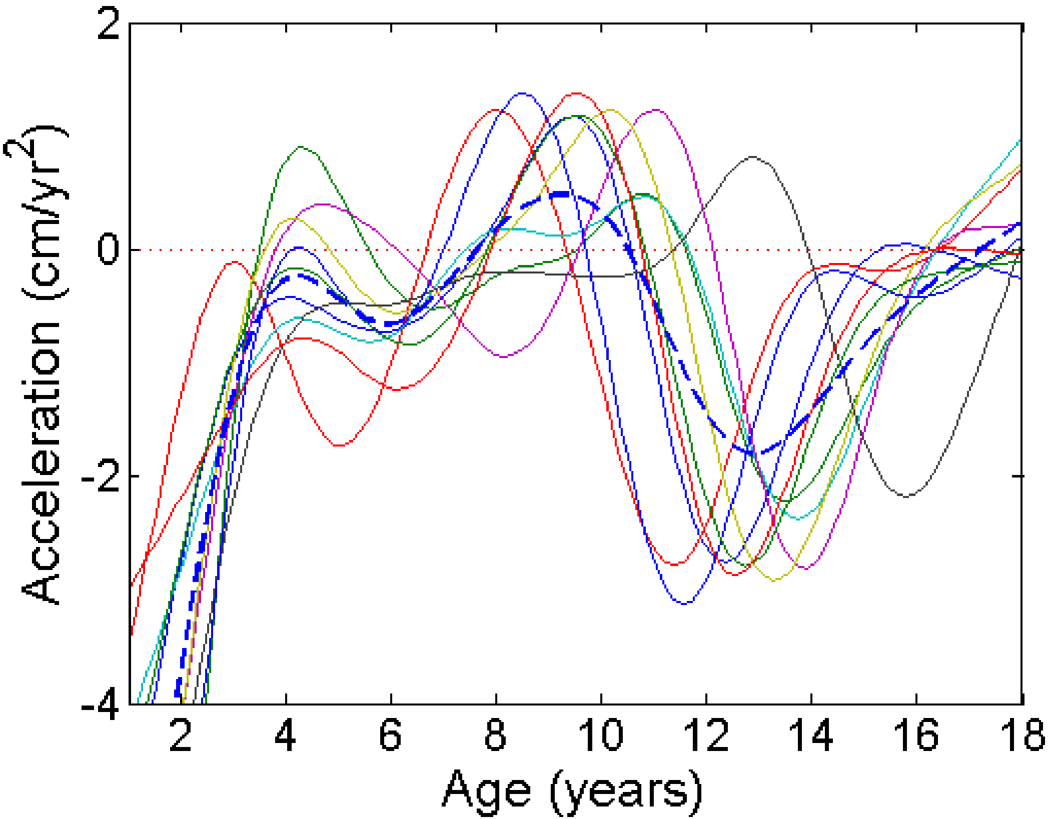
Data challenges

- We need repeated and regular access to each child for up to 20 years.
- Height changes over the day, and must be measured at a fixed time.
- Height is measured in supine position in infancy, followed by standing height. The change involves an adjustment of about 1 cm.
- Measurement error is about 0.5 cm in later years, but is rather larger in infancy. This is a signal–to–noise ratio of about 150.
- Measurements are not taken at equally spaced points in time.

Modelling challenges

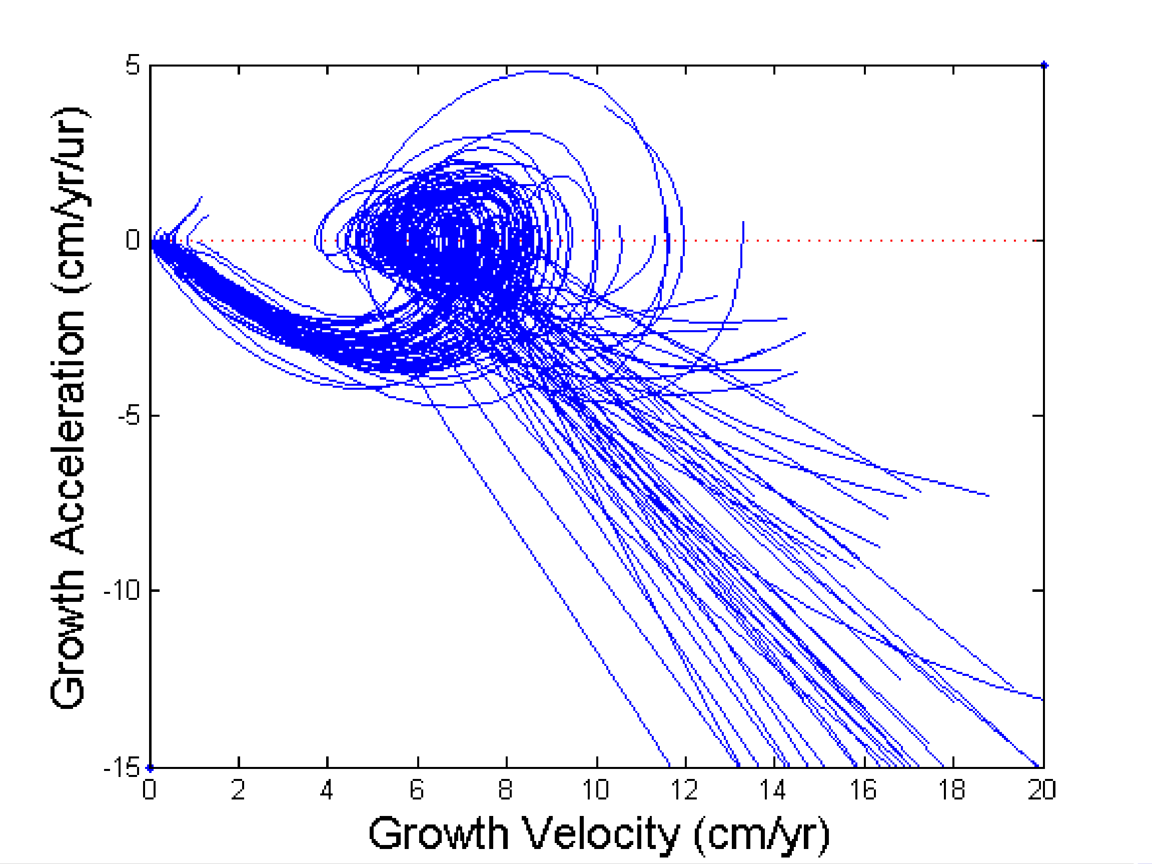
- We want smooth curves that fit the data as well as is reasonable. That is, with a typical error level that starts at about 0.7 cm but decreases to around 0.5 cm.
- In principle the curves should be monotone; i. e., have a positive derivative.
- We will want to look at velocity and acceleration, so that we want to differentiate twice and have a smooth curve.

Ten height accelerations

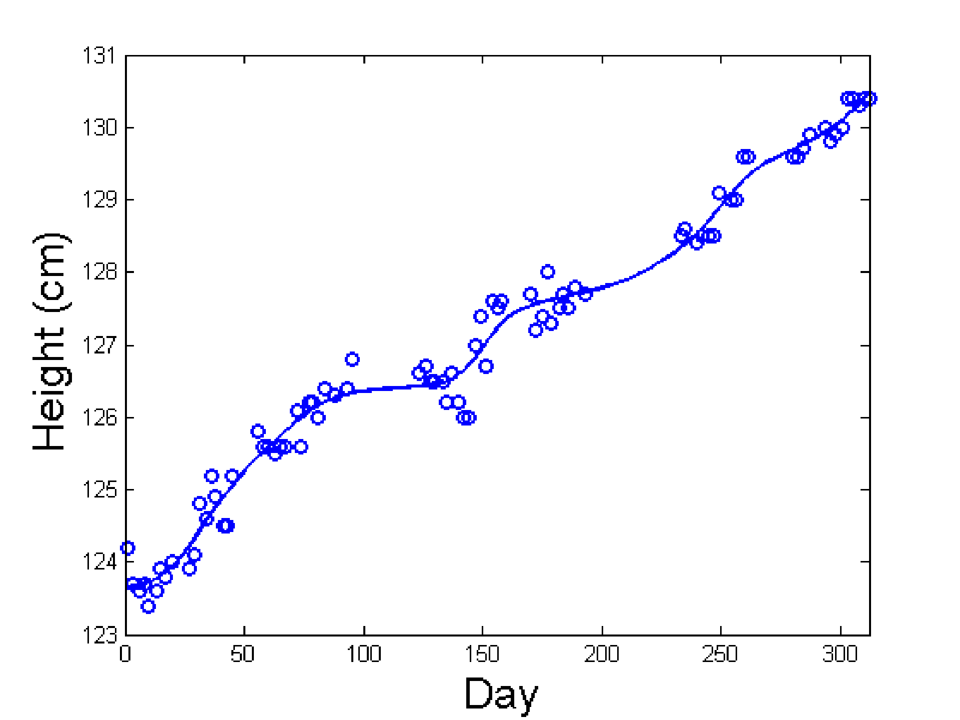


Overview
What are functional data?
Some functional data . . .
The goals of functional . . .
The first steps in a . . .
Using derivatives in . . .
Summary: What . . .
Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 7 of 33
Go Back
Full Screen
Close
Quit

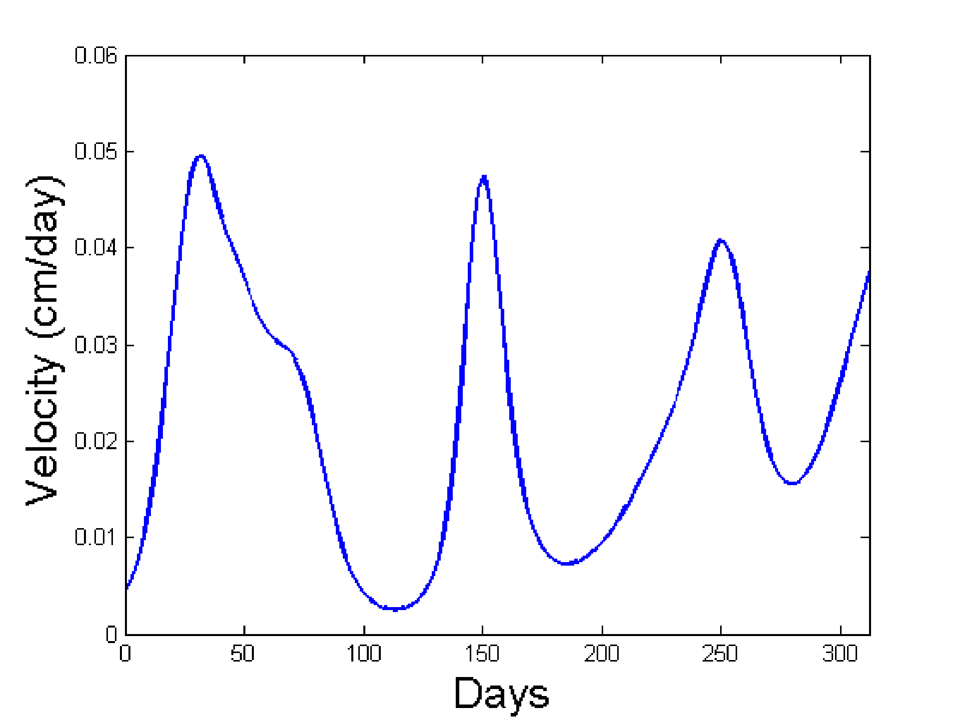
Plotting acceleration against velocity



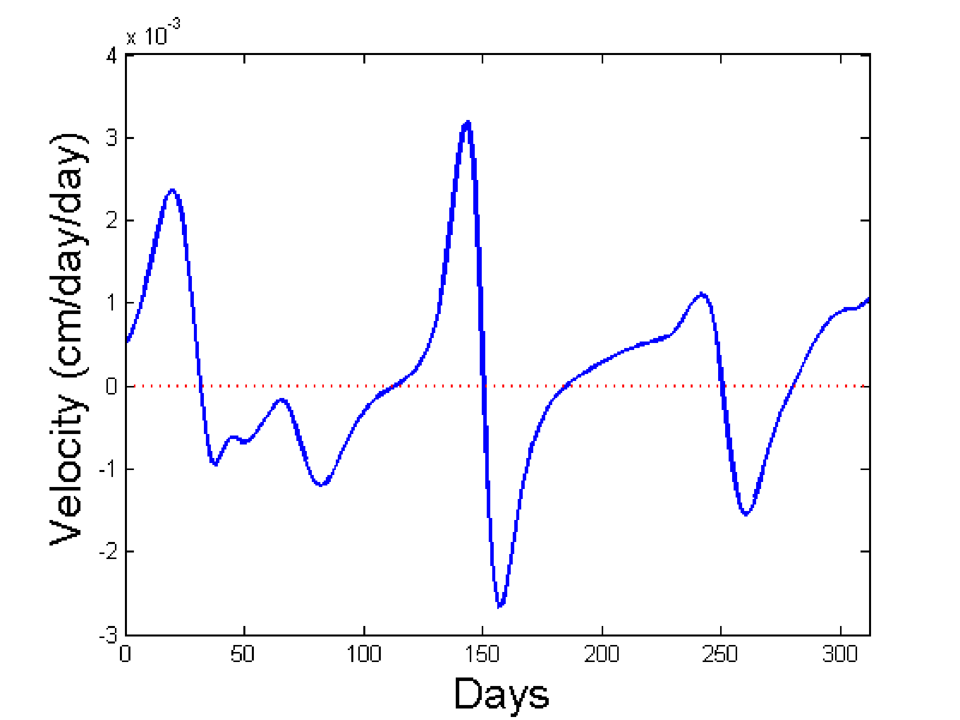
One boy’s height curve



One boy's height velocity curve



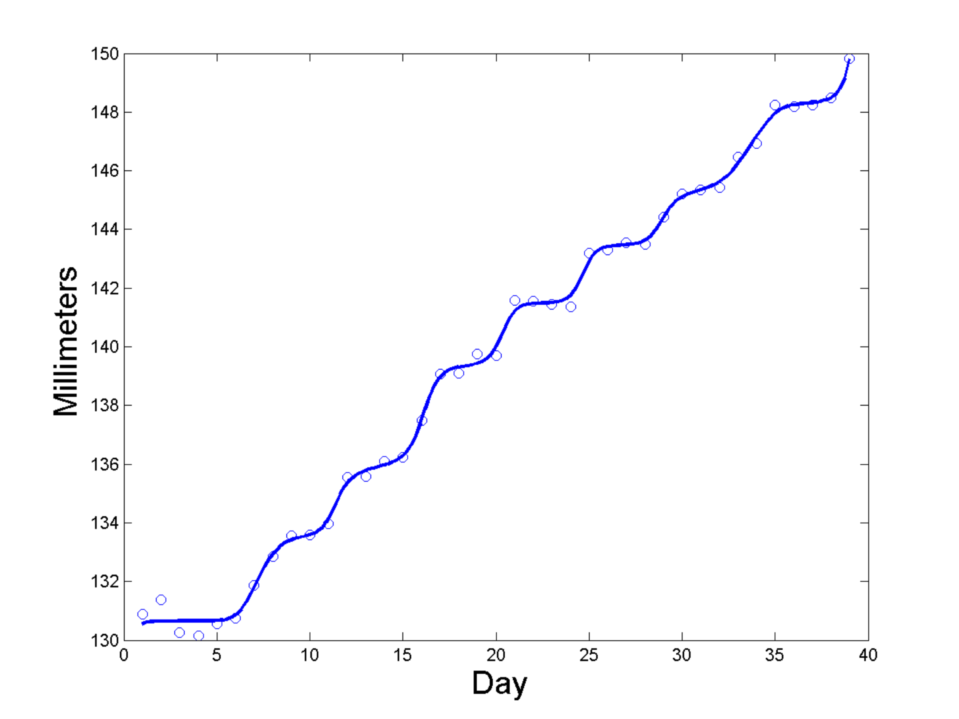
One boy's height acceleration curve



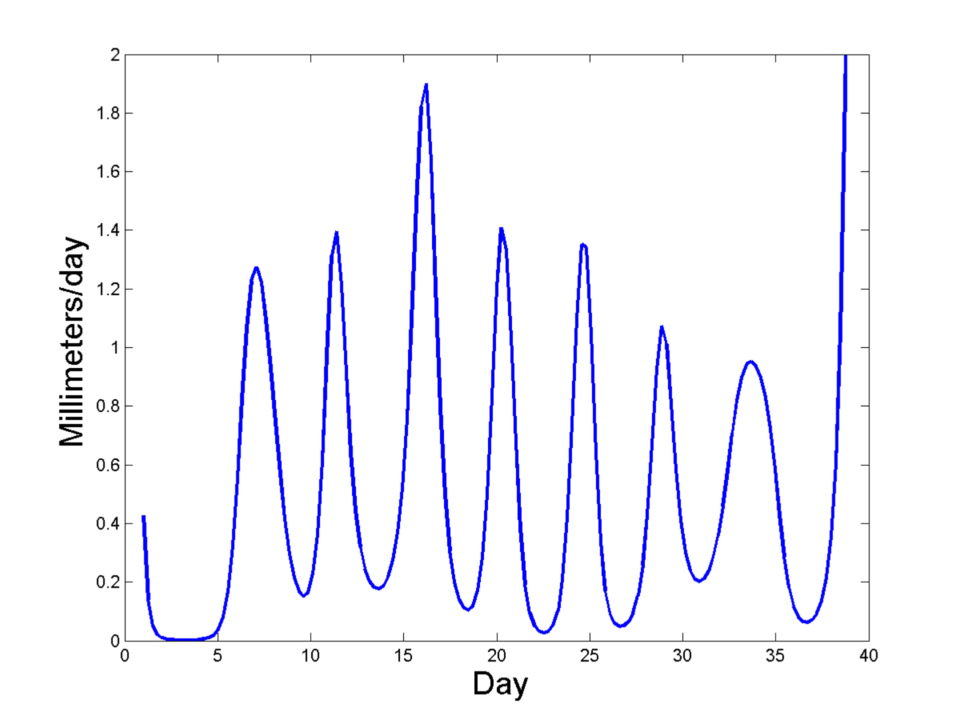
Data from one newborn baby

- Prof. Michael Hermanussen developed an instrument capable of measuring the length of the tibia of a baby (the lower leg bone) with an accuracy of about 0.1 millimeters.
- He measured newborn infant's tibias daily and hourly.

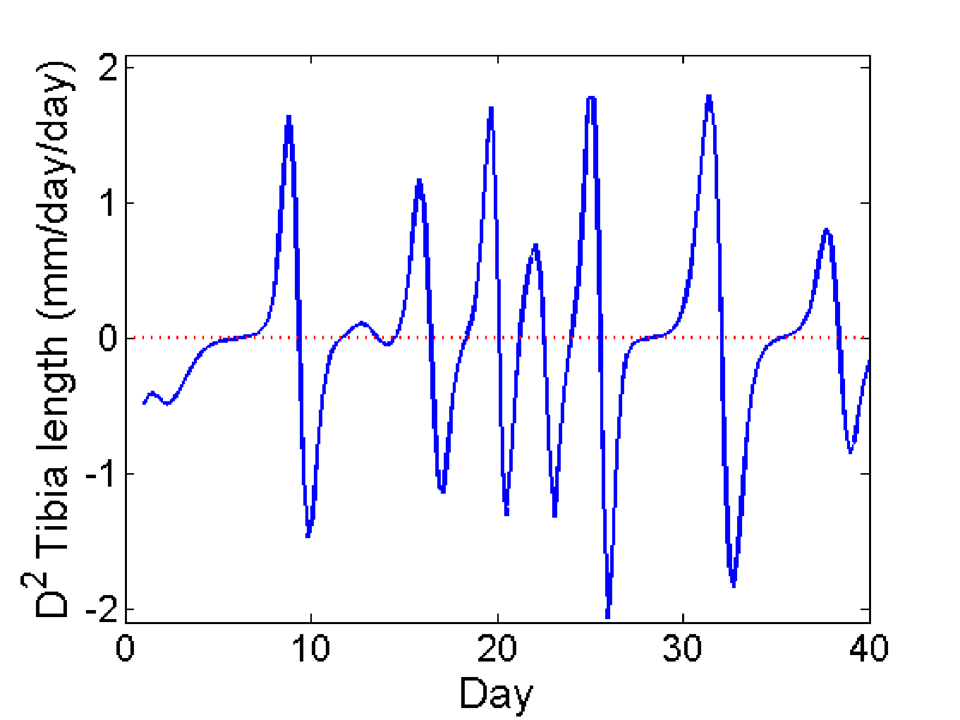
One baby's height curve



One baby's height velocity curve



One baby's height acceleration curve

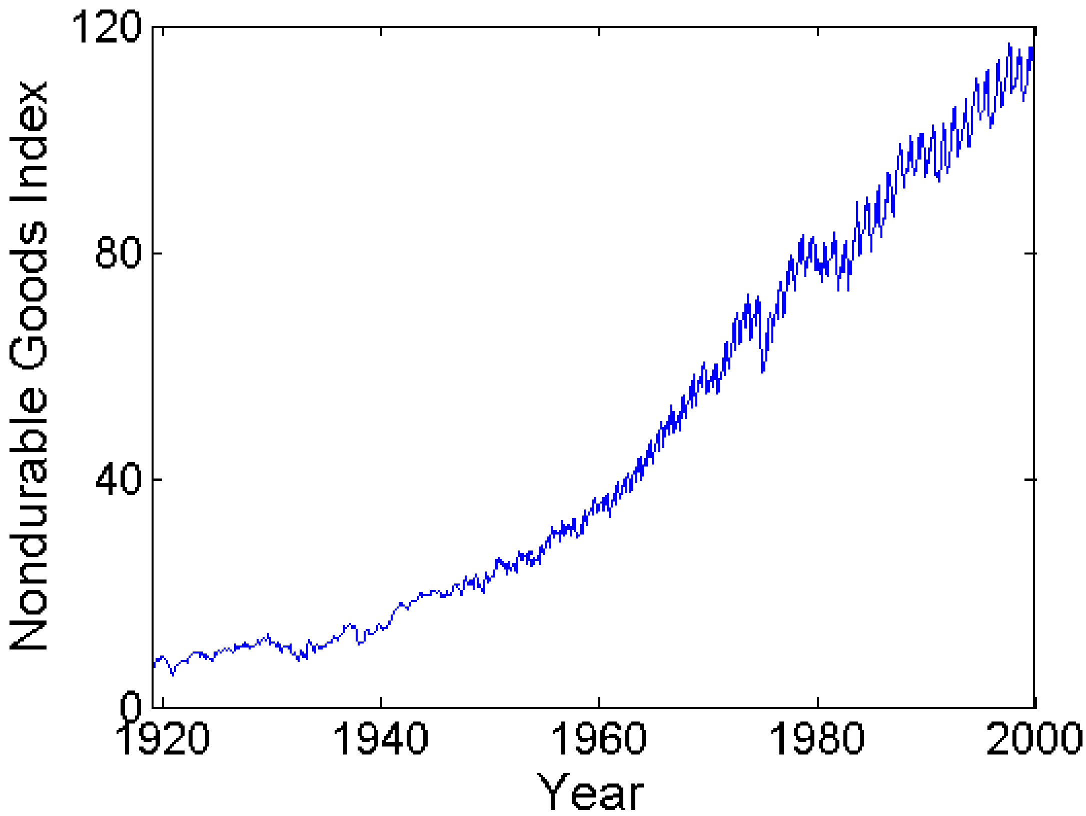


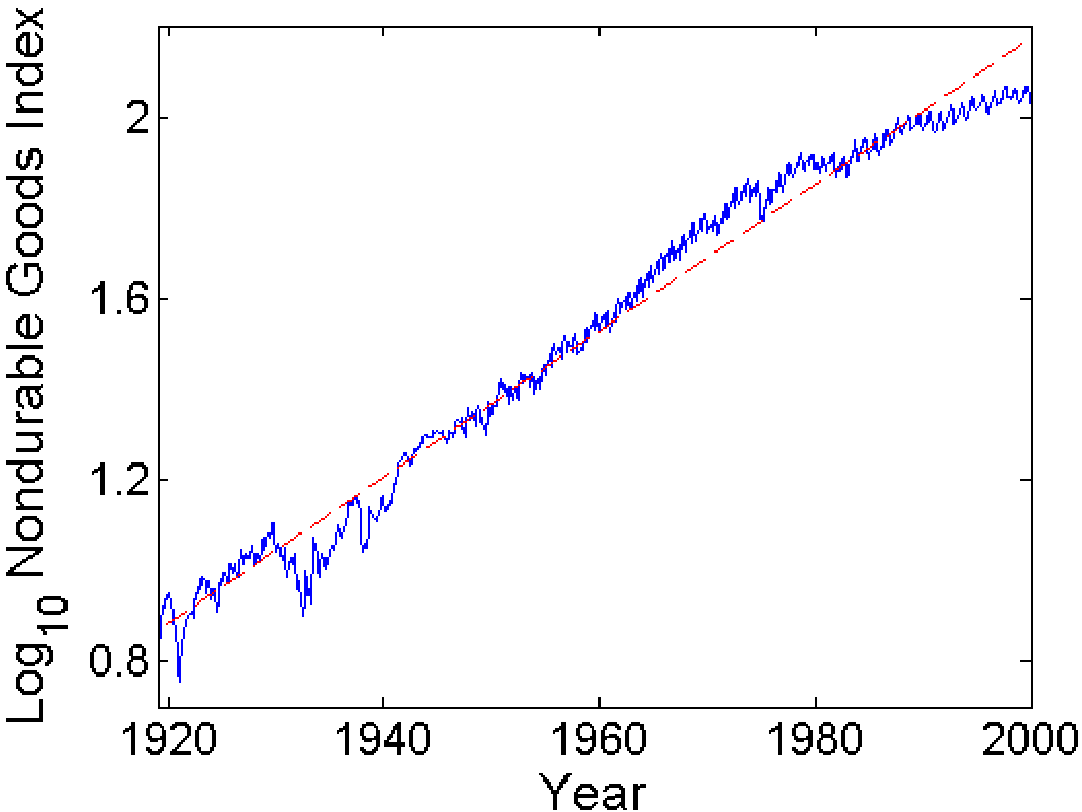
Some conclusions about growth

- Over 20 years, there is one major growth spurt, but clear evidence for at least one minor spurt.
- The timing of these spurts varies from child to child.
- Zooming in on a daily scale, at ten years of age there is a growth spurt every 100 days or so, and the amount of energy in the spurts seems to be decreasing.
- A newborn's tibia can grow at an astonishing 2 millimeters per day!
- A critical aspect of growth is what shuts it *off*.

A single long functional observation

The production of nondurable goods in the U. S.





Log_{10} Nondurable Goods Index
2
1.6
1.2
0.8
1920
1940
1960
1980
2000
Year

Multiscale variation

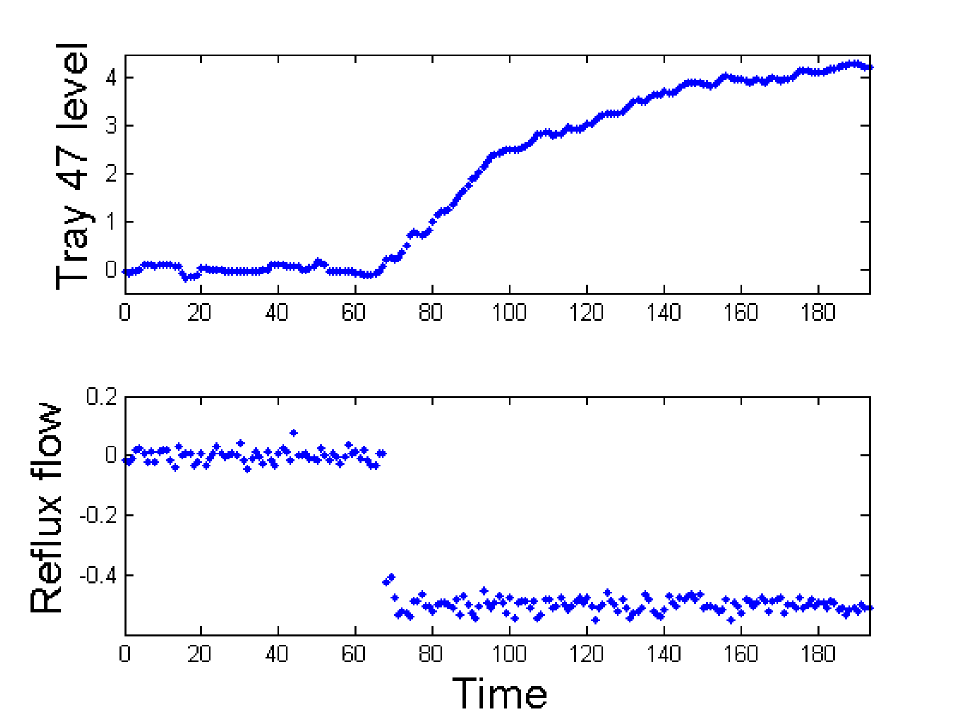
These data, after transformation, have interesting variation on four different time scales:

- **Long term**: A remarkably linear trend with a slope of 1.6.
- **Medium Term**: Multi–year changes due to the depression, World War II, the Vietnam War, and over the last decade.
- **Short Term**: Shocks like the stock market crash of 1928, the 1938 reduction of money supply and the end of the Vietnam War in 1976.
- **Seasonal Effects**: Within-year effects that we will consider later, and that evolve smoothly from one year to the next.

3. Some functional data analyses

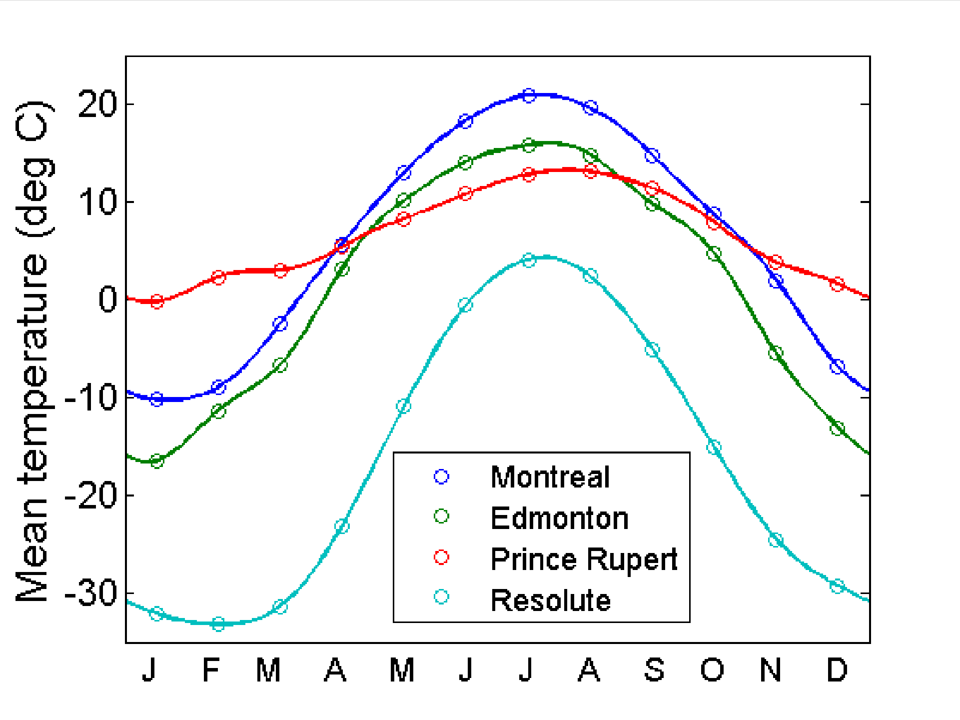
An input/output system

Tray 47 level in an oil refinery responds to a step change in input.



Can we develop a functional linear model to describe this relation?

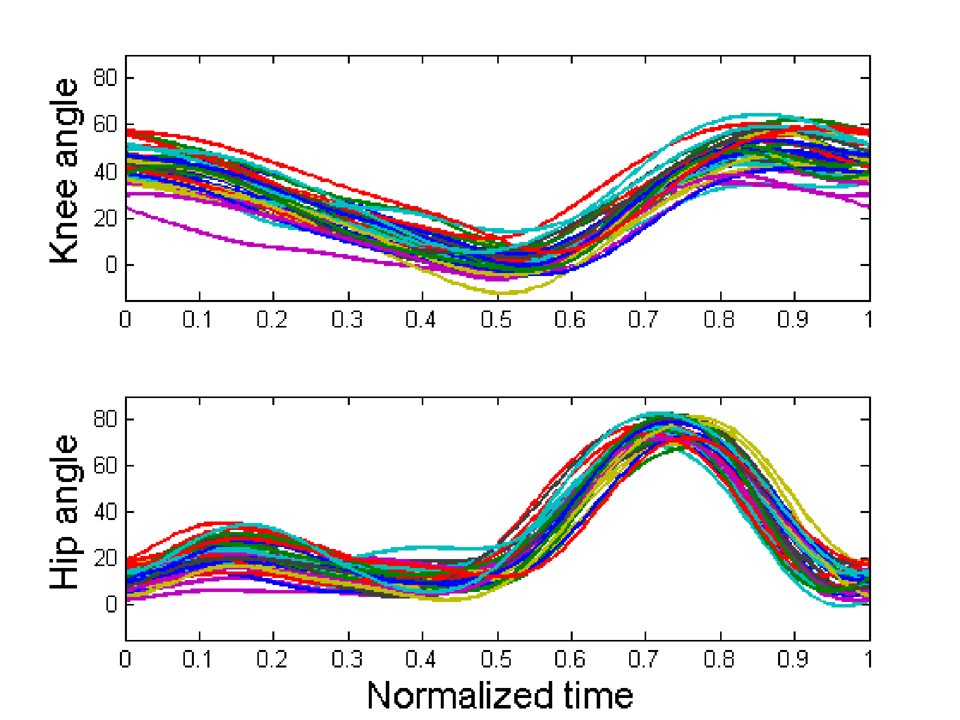
Mean annual temperatures at four weather stations



We will use principal components analysis on data from 35 weather stations.

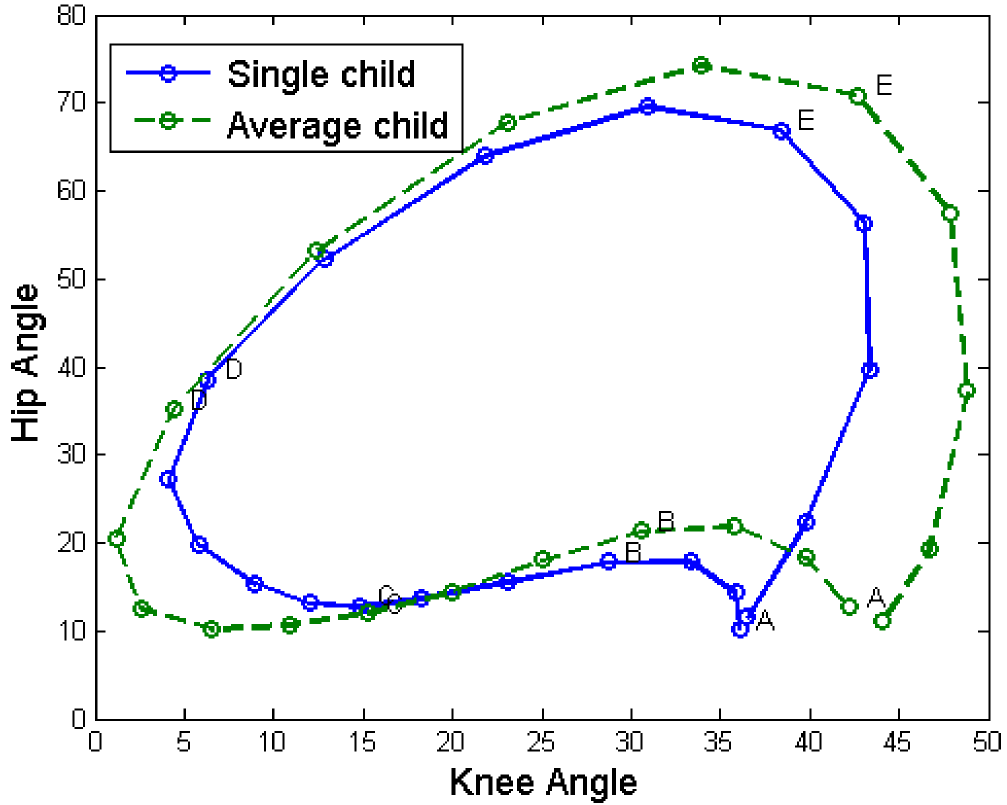
Some multivariate functional data

Angles at the knee and hip for 39 children over a single gait cycle.



Functional canonical correlation analysis will help here.

Comparing one child's cycle with the mean.



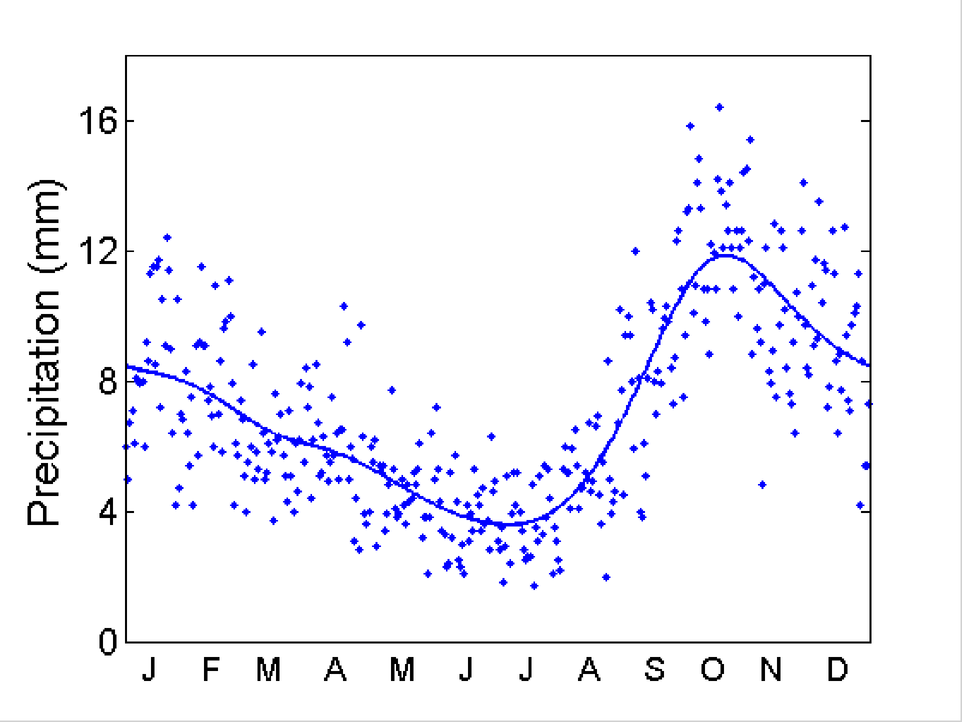
4. The goals of functional data analysis

The goals of functional data analysis are essentially the same as those of any other branch of statistics. They include:

- to represent the data in ways that aid further analysis
- to display the data so as to highlight various characteristics
- to study important sources of pattern and variation among the data
- to explain variation in an outcome or dependent variable by using input or independent variable information
- to compare two or more sets of data with respect to certain types of variation, where two sets of data can contain different sets of replicates of the same functions, or different functions for a common set of replicates.

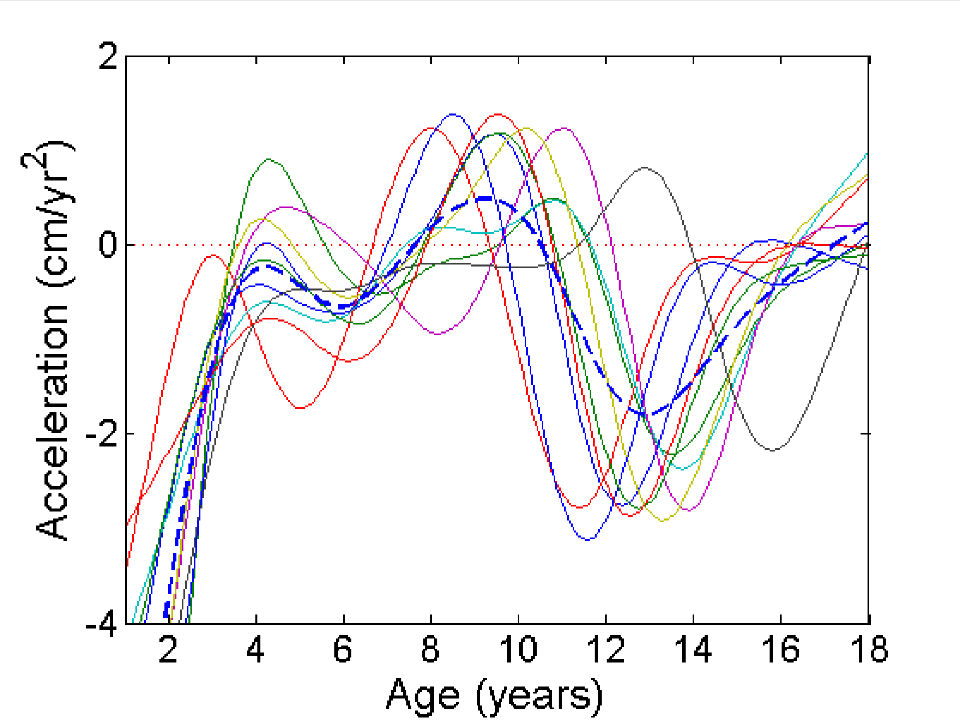
5. The first steps in a functional data analysis

Smoothing the rainfall data for Prince Rupert



The smooth line is constrained to be positive.

Data registration or feature alignment



The problem of phase variation

- Often important features in replicated curves do not occur at the same time. Like the pubertal growth spurt.
- *Phase variation* disrupts most obvious functional data analyses, which are designed for only *amplitude variation*.
- The mean curve here is a worthless summary of these growth acceleration curves.
- We must first align features, a process called *curve registration*.
- Registration separates phase and amplitude variation, which can then be studied independently, and also jointly.

6. Using derivatives in functional data analysis

The sinusoidal component of weather

- One expects temperature to be primarily sinusoidal in character, and certainly periodic over the annual cycle.
- There is much variation in level and some variation in phase.
- A model of the form

$$\mathtt{Temp}_i(t) \approx c_{i1} + c_{i2}\sin(\pi t/6) + c_{i3}\cos(\pi t/6)$$

should do rather nicely for these data.

- There are clear departures from sinusoidal or simple harmonic behavior.
- We could remove sinusoidal trend by regression, but let's use differentiation instead.
- We use $D^m x$ to refer to the mth derivative.
- We compute

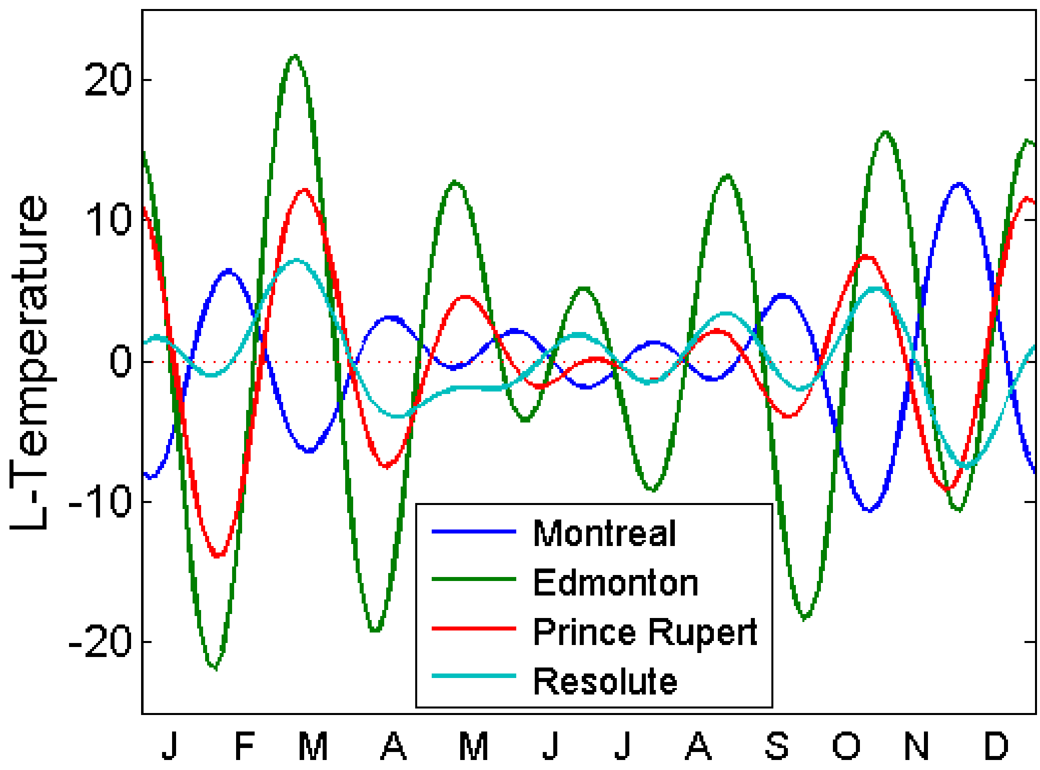
$$L\texttt{Temp} = (\pi/6)^2 D\texttt{Temp} + D^3\texttt{Temp},$$

  which will annihilate shifted sinusoids.
- L is a *linear differential operator*.
- We can define temperature as the solution to the differential equation

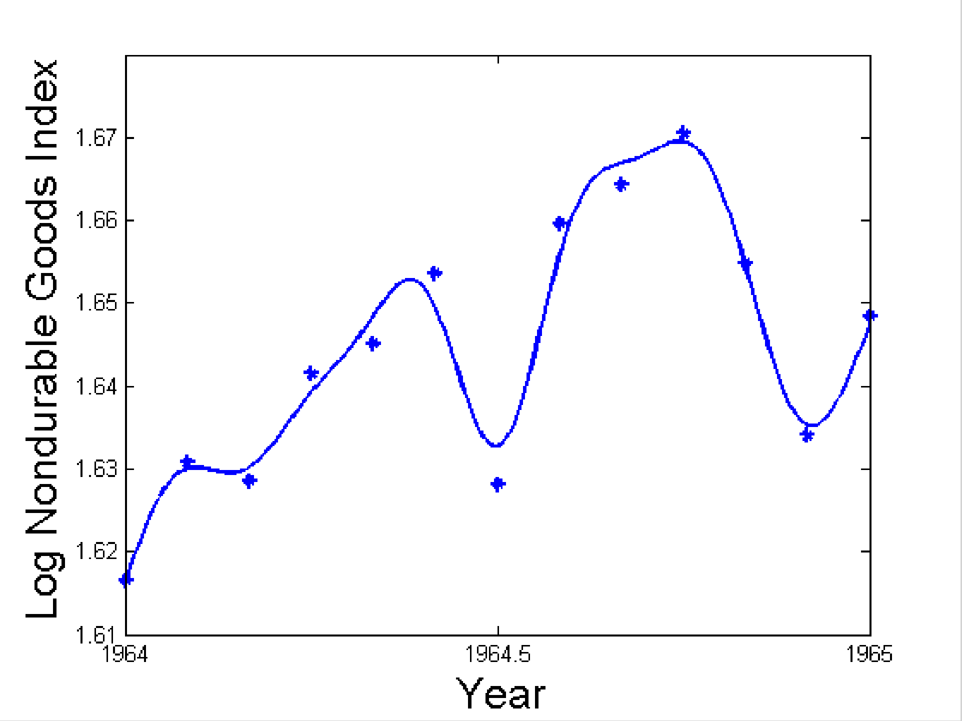
$$L\texttt{temp} = u$$

  where u is called a *forcing function*, and accounts for the non–sinusoidal effects.

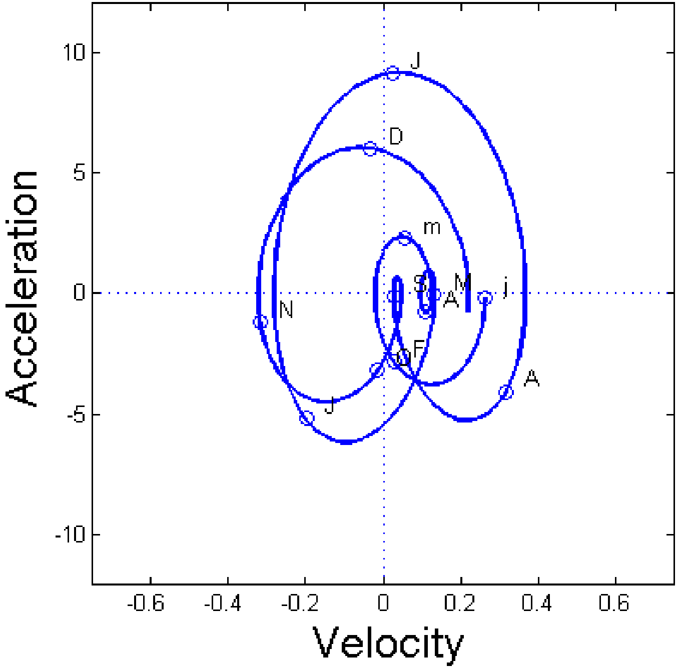
De–sined temperature



The seasonal trend for a typical year in the goods index



Displaying seasonal dynamics: the *phase-plane plot*



- Many types of functional data show strong *harmonic* variation.
- The acceleration or second derivative reflects *potential energy* in a mechanical system, like a pendulum or spring.
- The first derivative reflects its *kinetic energy*.
- A sinusoid is the prototype for such variation. Plotting its second derivative against first derivative produces a circle.
- The radius of the cycle is the total energy in the system, conserved as energy changes state.
- These ideas apply most periodic phenomena.
- The phase-plane plot is a graphic version of a *differential equation*.

7. Summary: What makes FDA different?

- Unlike time series analyses, no assumptions of stationarity are made, and data are not sampled at equally spaced time points.
- Unlike most longitudinal data, a large number of time points are available, and the signal-to-noise ratio is medium to high.
- The data can support the accurate estimate of one or more derivatives, and these play several critical roles.
- Phase variation is recognized and separated from amplitude variation.
- Familiar multivariate methods have functional counterparts, and the smoothness of functional parameter estimates is explicitly controlled.
- Differential equations are new modelling tools.

8. Where do we go for more information?

- A web site containing more information, data, sample analyses, software, news, and etc.:
- `www.functionaldata.org`
- Two books to consider:
- J. O. Ramsay and B. W. Silverman, (2005) *Functional Data Analysis, Second Edition*. New York: Springer.
- J. O. Ramsay and B. W. Silverman, (2002) *Applied Functional Data Analysis, Second Edition*. New York: Springer.