An Introduction to Functional Data Analysis

Jim Ramsay, McGill University

McGill Psychology Department
Montreal, February 23, 2009

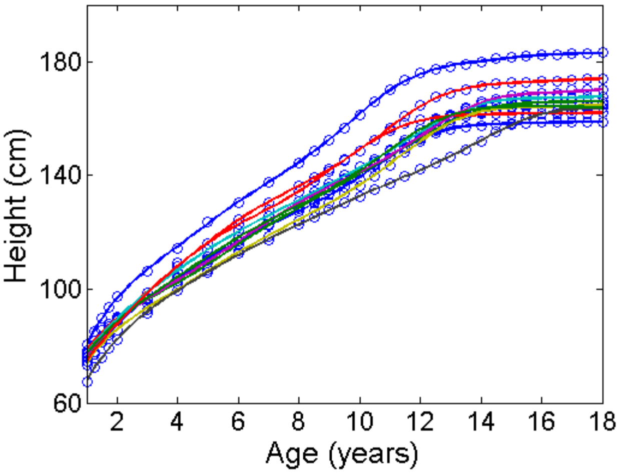
Overview

- 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.

Outline

Heights of ten girls



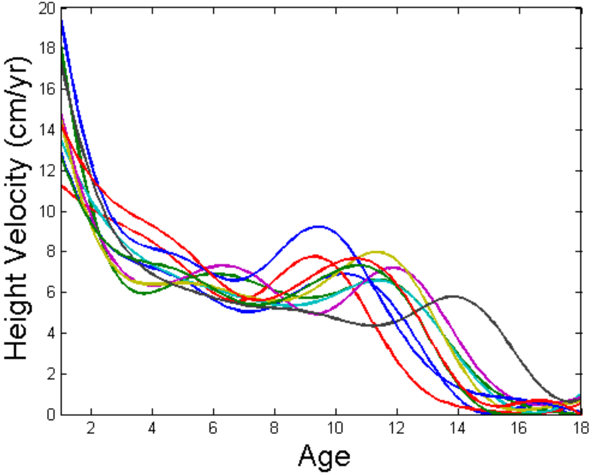
Data challenges

- 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.

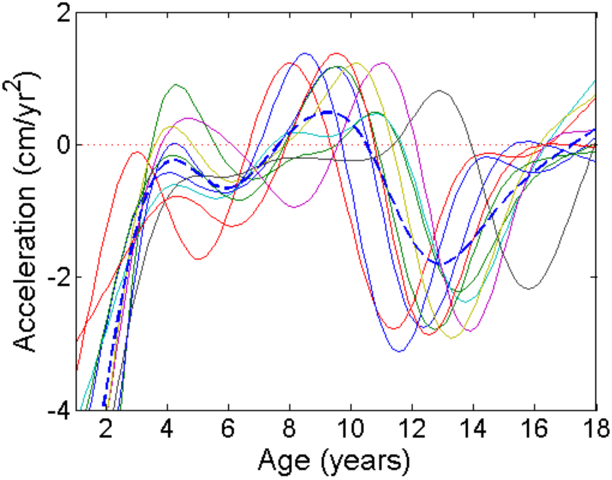
Modelling challenges

- 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.

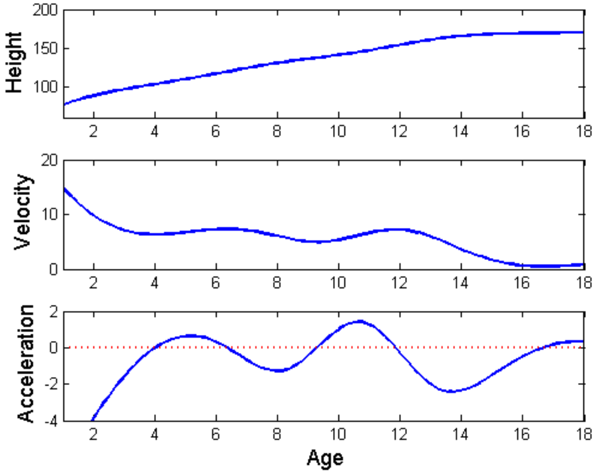
Ten height velocities



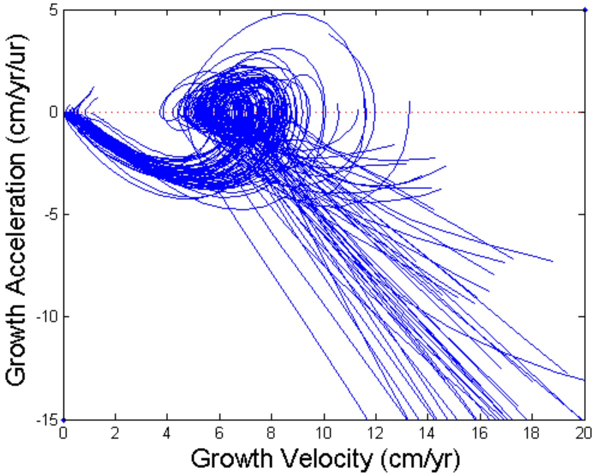
Ten height accelerations



Ten height, velocities and accelerations



Plotting acceleration against velocity



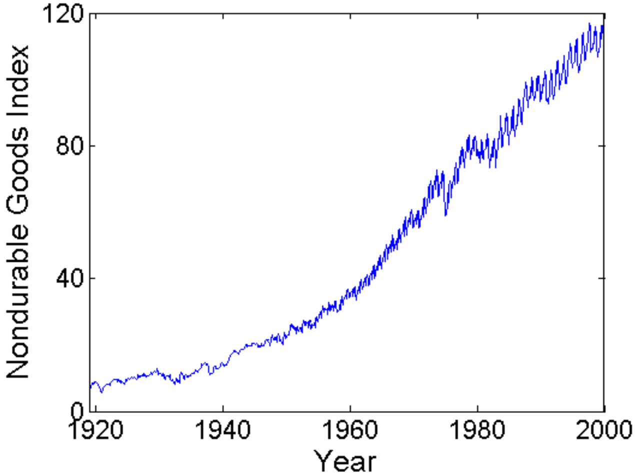
Within and between sample sizes

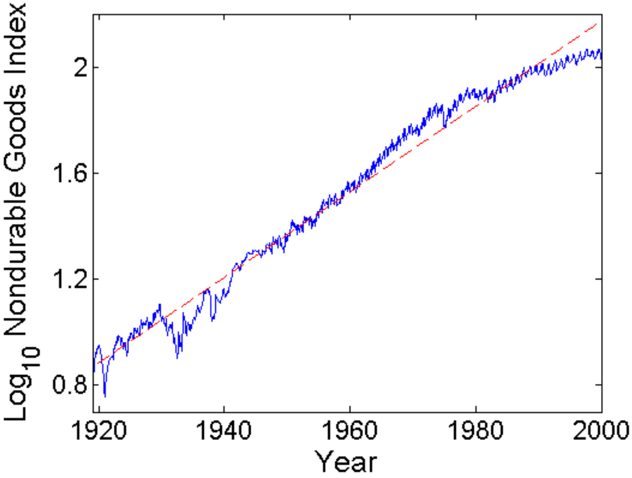
- The growth data are an example of a sample of functional observations.
- In this case, the sample size $N = 10$. They are a sub-sample of the Berkeley growth data, where the sample size for girls was 54.
- N determines how much we can learn about *between-curve variation*.
- The within record sample size is $n = 31$. This, along with
  - the accuracy of the observations,
  - and how the observations are distributed

  determines how much we can learn about *within-curve variation*.
- We now turn to a single long functional observation ($N = 1, n = 993$.).

A single long functional observation

The production of nondurable goods in the U. S.





Log10 Nondurable Goods Index
2
1.6
1.2
0.8
1920
1940
1960
1980
2000
Year

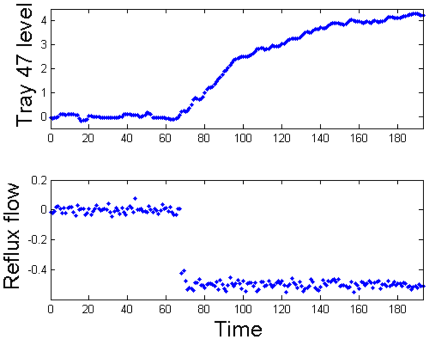
Multiscale variation

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.

An input/output system

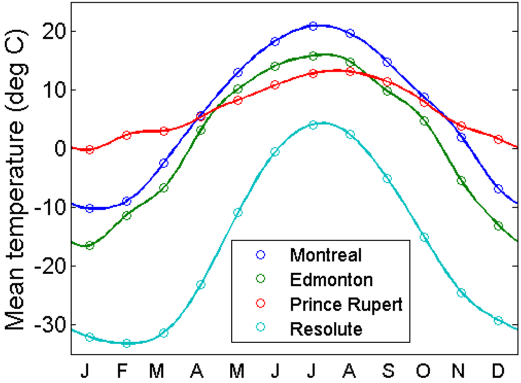
Tray 47 level in an oil refinery responds to a step change in input.



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

Outline

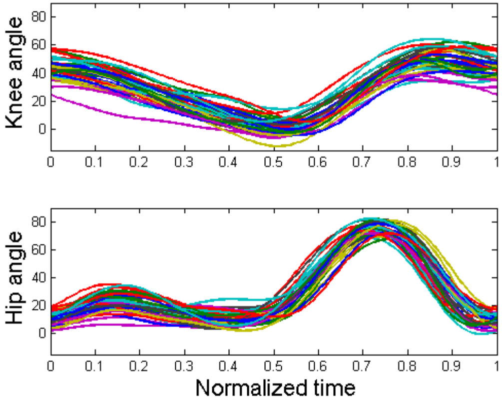
Mean annual temperatures at four weather stations



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

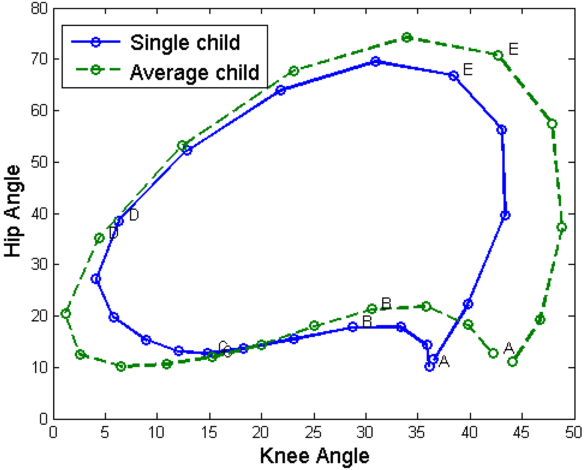
Some multivariate functional data

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

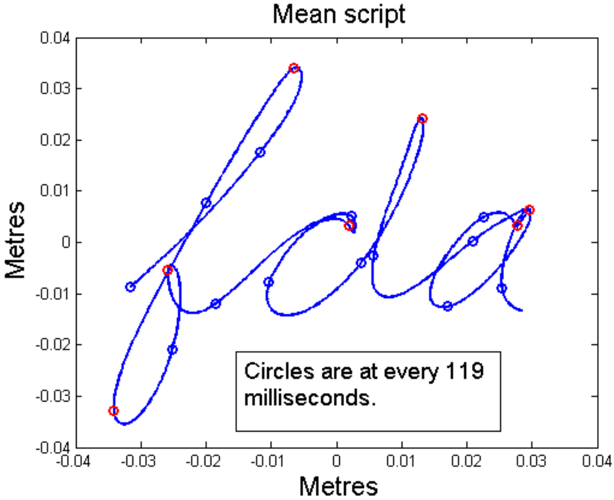


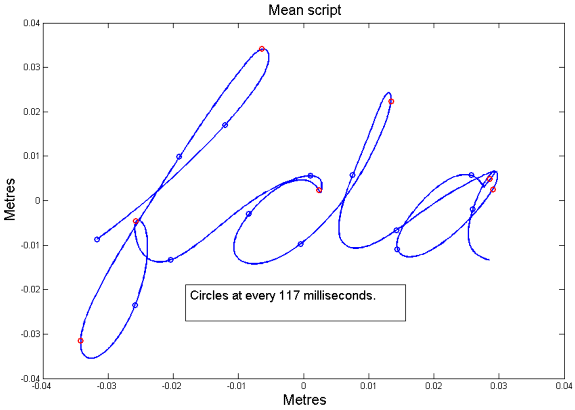
Functional canonical correlation analysis will help here.

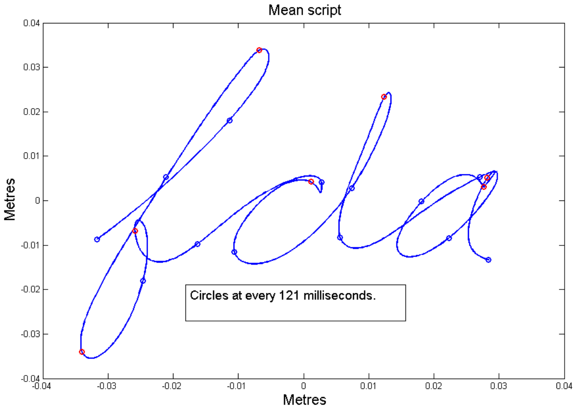
Comparing one child's cycle with the mean.



The mean over 20 replications of script “fda”

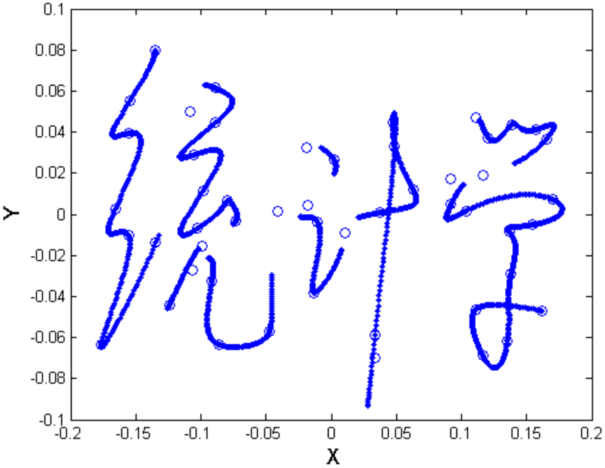



Mean script
Circles at every 117 milliseconds.
Metres
Metres
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04

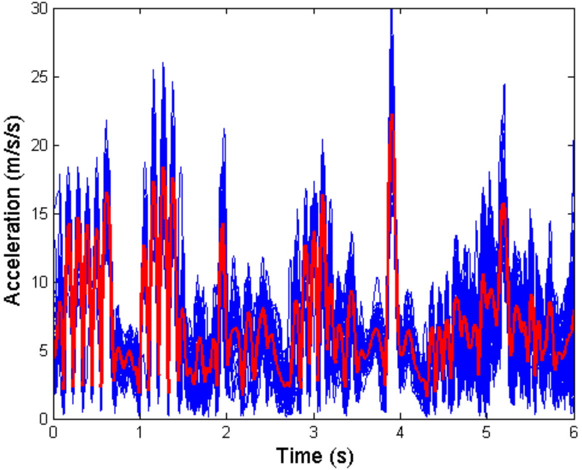

Mean script
0.04
0.03
0.02
0.01
0
-0.01
-0.02
-0.03
-0.04
-0.04
-0.03
-0.02
-0.01
0
0.01
0.02
0.03
0.04
Metres
Metres
Circles at every 121 milliseconds.

- The precision of the timing of cusps and other sharp features in handwriting seems remarkable.
- Two milliseconds either way results in misalignment.
- Fast-acting neural pathways transmit spike potentials about every 10 milliseconds.
- How do the central nervous system, the spinal cord, and the neural/muscle boundary achieve this level of precision?

Mean over 50 replications of Chinese script for “statistics”. Circles every 120 milliseconds.



Accelerations along the writing path (red = mean)



- Again we see the amazing temporal precision at the level of both the script and the acceleration.
- Accelerations along the trajectory of the handwriting reach 30 metres/second/second.
- Or about three gravitational forces.
- If sustained, this level of acceleration would launch a satellite in less than seven minutes.
- Equally impressive is that acceleration peaks are mostly separated by near zero levels of acceleration.
- The energy being managed so precisely in this behavior is incredible!

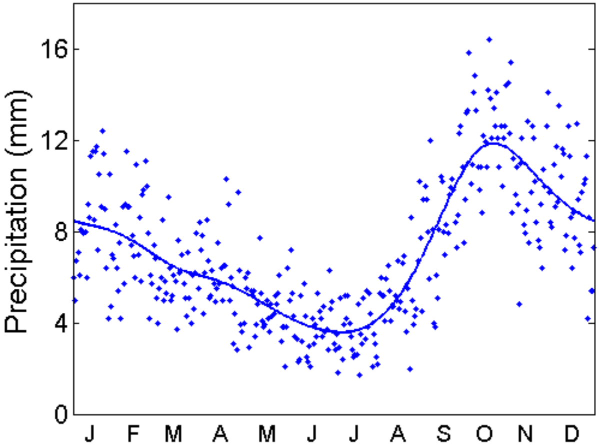
Outline

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.

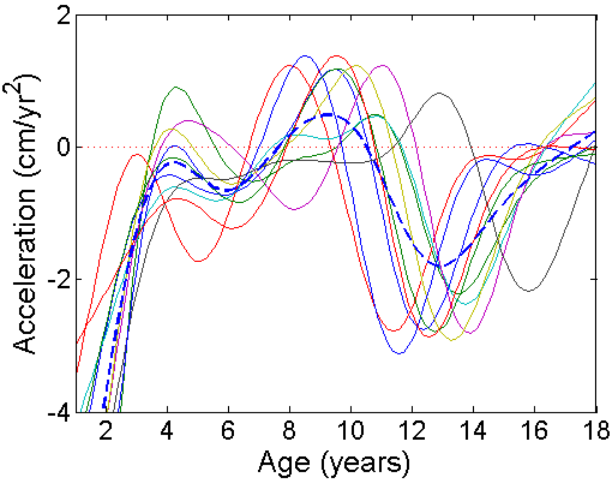
Outline

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.

Outline

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

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

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

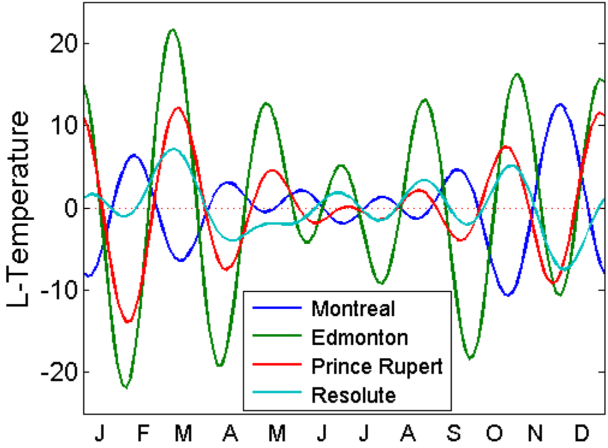
$$L\texttt{Temp} = (\pi/6)^2 D\texttt{Temp} + D^3\texttt{Temp},$$

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

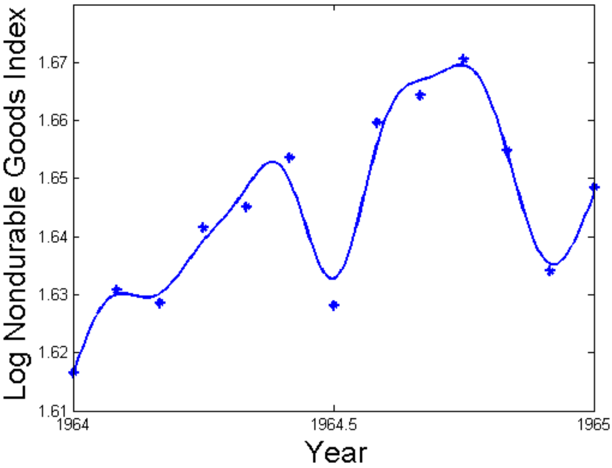
$$L\texttt{Temp} = u$$

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

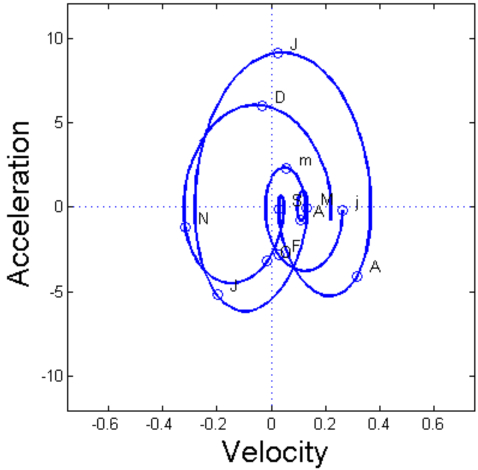
De–sined temperature



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*.

Outline

- 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.

Outline

- 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.