Functional responses, functional covariates and the concurrent model

1. Predicting precipitation profiles from temperature curves

- Precipitation is much harder to predict than temperature.
- It comes in two main forms:
  - *Drizzle*: Large low pressure zones drop moisture over many hours or days.
  - *Storms*: Convective, short violent storms with a lot of precipitation in a hurry, and spatially localized.
- Precipitation tends to be seasonal; more in the spring and fall than in the summer and winter.

A model

- We can assume that climate zone is important.
- We will predict log precipitation; logging stabilizes variance and eliminates the positivity constraint.
- We will use the difference $\texttt{TempRes}_{mg}(t)$ between a temperature profile and the mean for the climate zone as a function covariate.
- We can extend the functional ANOVA model to

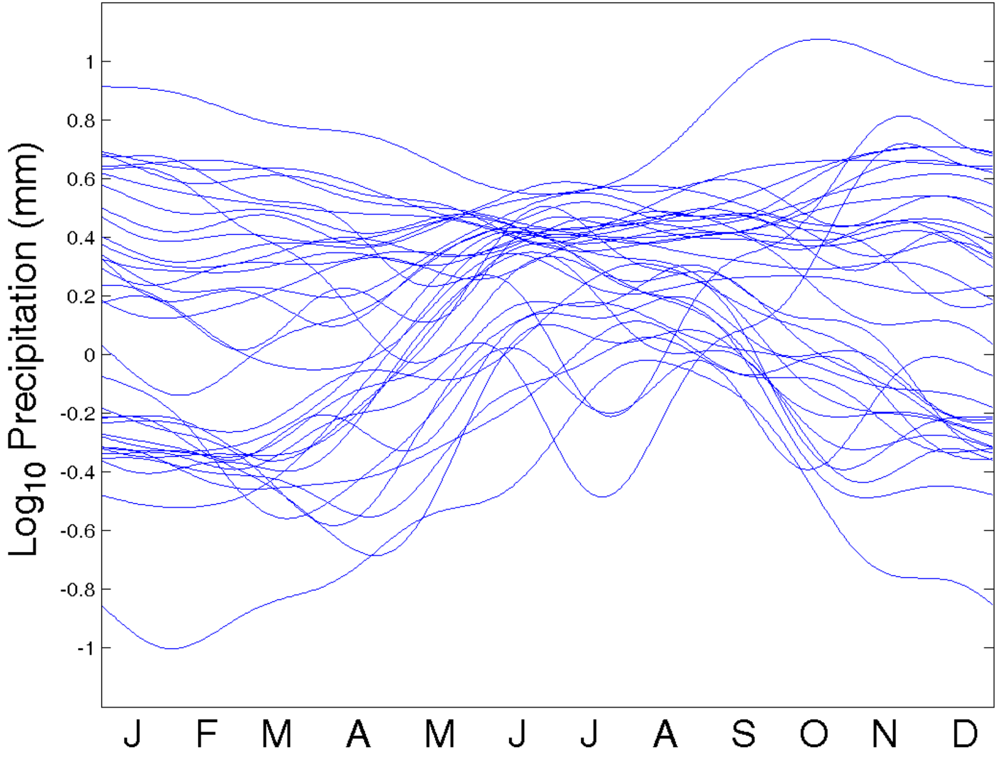
$$\log[\texttt{Prec}_{mg}(t)] = \mu(t) + \alpha_g(t) + \texttt{TempRes}_{mg}(t)\beta(t) + \epsilon_{mg}(t)$$

- We call this model *concurrent* because it assumes that the temperature today affects today's precipitation.

The functional data

- Where precipitation was recorded as 0 mm, we changed it to 0.05 mm, half the minimum positive value.
- We used 11 Fourier series basis functions for precipitation with no roughness penalty.
- We used 21 Fourier series basis functions for temperature with no roughness penalty.

Log precipitation profiles



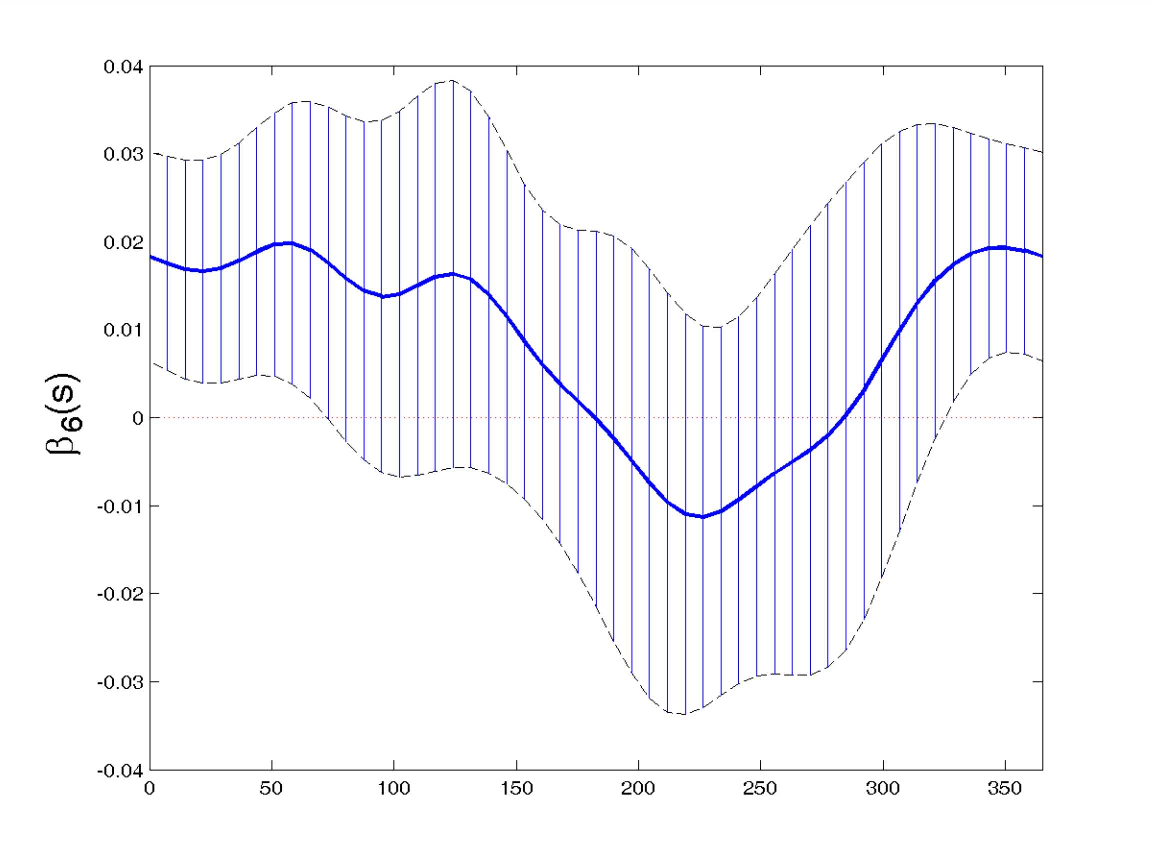
The fitting criterion and some results

- The fitting criterion was the unpenalized error sum of squares

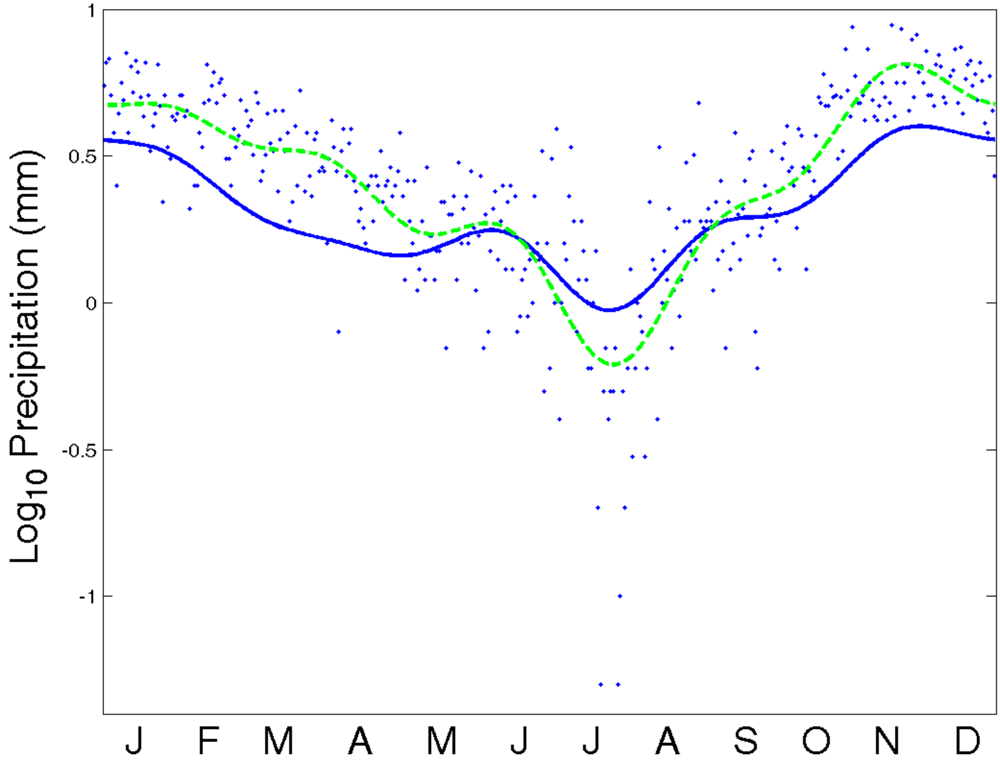
$$\texttt{LMSSE}(\mu, \alpha_g, \beta) = \int \sum_{k,g}^{N} [\texttt{LogPrec}_{kg}(t) - \mu(t) - \alpha_g(t) - \texttt{TempRes}_{kg}(t)\beta(t)]^2 \, dt$$

- The resulting root–mean–squared–residual was 0.19 mm.
- When we dropped $\texttt{TempRes}(t)$ from the model, this increased to 0.20 mm.
- As we see in the following plot, the only place where temperature appears to make a contribution is in mid–winter.

The estimated regression function $\beta(t)$



The fit to Vancouver's data



A probe for the winter effect

- The confidence limits are point–wise; we need a measure of the temperature influence accumulated over the winter months.
- Here is a probe that works:

$$\int_0^{365} \cos[2\pi(t-64.5)/365]\beta(t)\,dt = 2.32\ ,$$

- The estimated standard error of this probe is 0.77, giving a t-ratio of 3.0.
- It appears that elevated temperatures in mid-winter go along with increased precipitation.

2. Fitting the concurrent model

- Here is a general statement of the current functional/functional model:

$$y_i(t) = \sum_{j=1}^{q} z_{ij}(t)\beta_j(t) + \epsilon_i(t) \ .$$

- or in matrix notation:

$$\mathbf{y}(t) = \mathbf{Z}(t)\boldsymbol{\beta}(t) + \boldsymbol{\epsilon}(t) \ ,$$

- We will use a penalized error sum of squares criterion:

$$\begin{aligned} \mathtt{LMSSE}(\boldsymbol{\beta}) = \\ \int [\mathbf{y}(t) - \mathbf{Z}(t)\boldsymbol{\beta}(t)]'[\mathbf{y}(t) - \mathbf{Z}(t)\boldsymbol{\beta}(t)]\, dt \\ + \sum_j^p \lambda_j \int [L_j\beta_j(t)]^2\, dt \end{aligned} \quad .$$

The basis function expansions for $\beta_j(s)$

- Let regression function $\beta_j(s)$ have the expansion

$$\beta_j(s) = \mathbf{b}_j'\boldsymbol{\theta}_j(s)$$

  in terms of K_j basis functions $\theta_{jk}(s)$.

- Some of the independent variables can be scalar; in this case the basis for their $\beta_j(s)$'s is the constant basis;

$$\theta_{j1}(s) = 1$$

- Defining $K_\beta = \sum_j^q K_j$, we construct vector $\mathbf{b}$ of length K_β by stacking the coefficient vectors vertically, that is,

$$\mathbf{b} = (b_1', b_2', \ldots, b_q')' \ .$$

- Now assemble q by K_β matrix function $\boldsymbol{\Theta}$ as follows:

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\theta}_1' & 0 & \cdots & 0 \\ 0 & \boldsymbol{\theta}_2' & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \boldsymbol{\theta}_q' \end{bmatrix} \ .$$

- We can now express our model as

$$\mathbf{y}(t) = \mathbf{Z}(t)\boldsymbol{\Theta}(t)\mathbf{b} + \boldsymbol{\epsilon}(t) \ .$$

- We also need to arrange the order K_j roughness penalty matrices

$$\lambda_j \mathbf{R}_j = \lambda_j \int L\boldsymbol{\theta}_j(t) L\boldsymbol{\theta}_j'(t)\, dt$$

into the symmetric block diagonal matrix $\mathbf{R}$ of order K_β:

$$\mathbf{R} = \begin{bmatrix} \lambda_1 \mathbf{R}_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 \mathbf{R}_2 & \cdots & 0 \\ \vdots & \vdots & \cdots & \vdots \\ 0 & 0 & \cdots & \lambda_q \mathbf{R}_q \end{bmatrix} . \qquad (1)$$

Home Page
Title Page
◀◀ ▶▶
◀ ▶
Page 13 of 19
Go Back
Full Screen
Close
Quit

The normal equations

- $$[\int \boldsymbol{\Theta}'(t)\mathbf{Z}'(t)\mathbf{Z}(t)\boldsymbol{\Theta}(t)\,dt + \mathbf{R}]\mathbf{b} \;=\; [\int \boldsymbol{\Theta}'(t)\mathbf{Z}'(t)\mathbf{y}(t)\,dt]$$

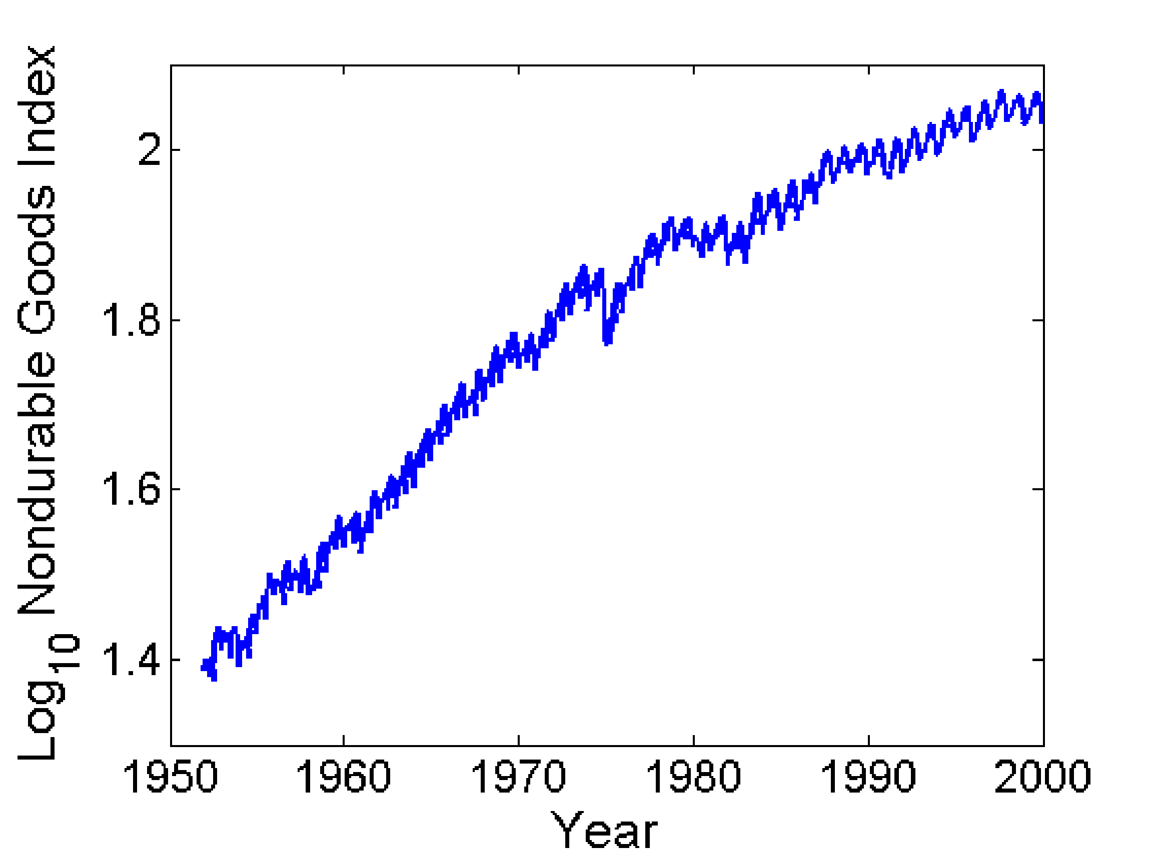
- The numerical integration in these equations is not as difficult as it seems. The scalar functions

$$\omega_{j\ell}(t) = \sum_{i}^{N} z_{ij}(t) z_{i\ell}(t)$$

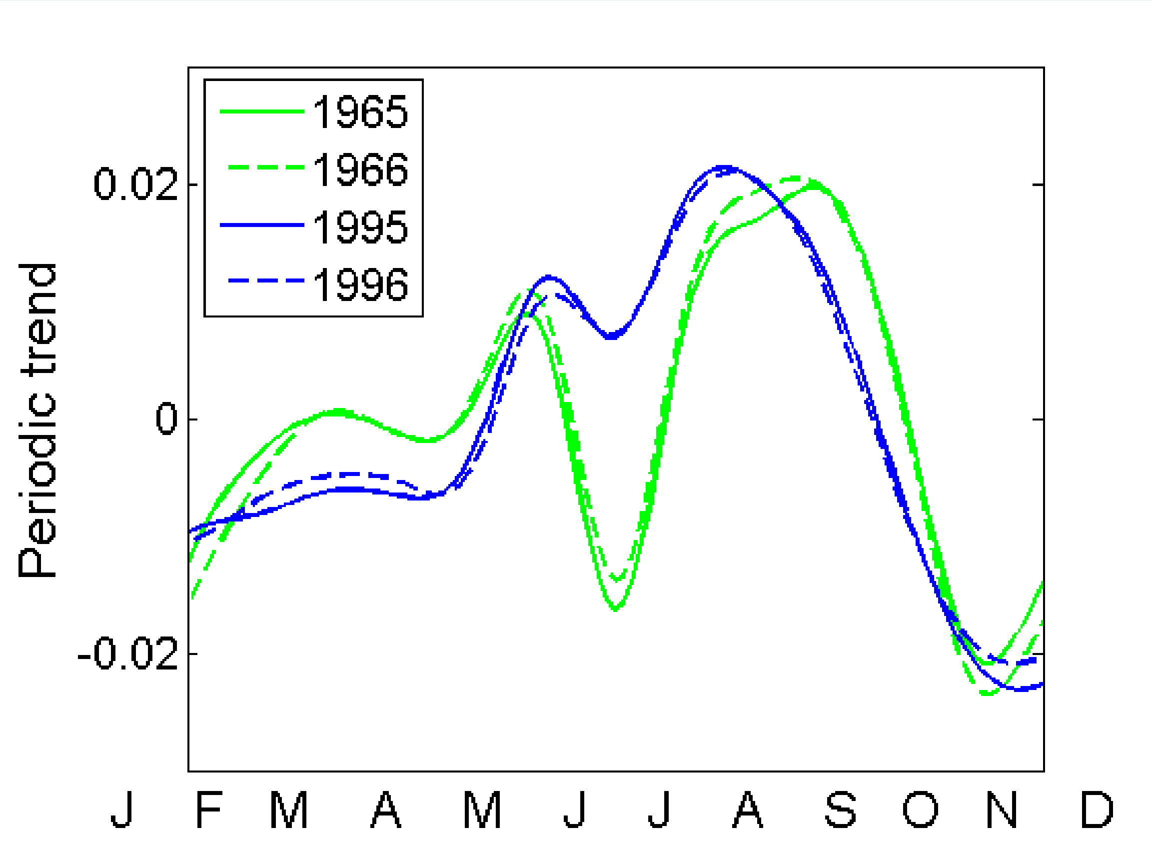
play the role of *weighting functions* for the functional inner products

$$\int \boldsymbol{\theta}_j(t)\boldsymbol{\theta}'_\ell(t)\omega_{j\ell}(t)\,dt, j, \ell = 1, \ldots, q\ .$$

3. Evolution in seasonal trend for the nondurable goods index



Four seasonal trends



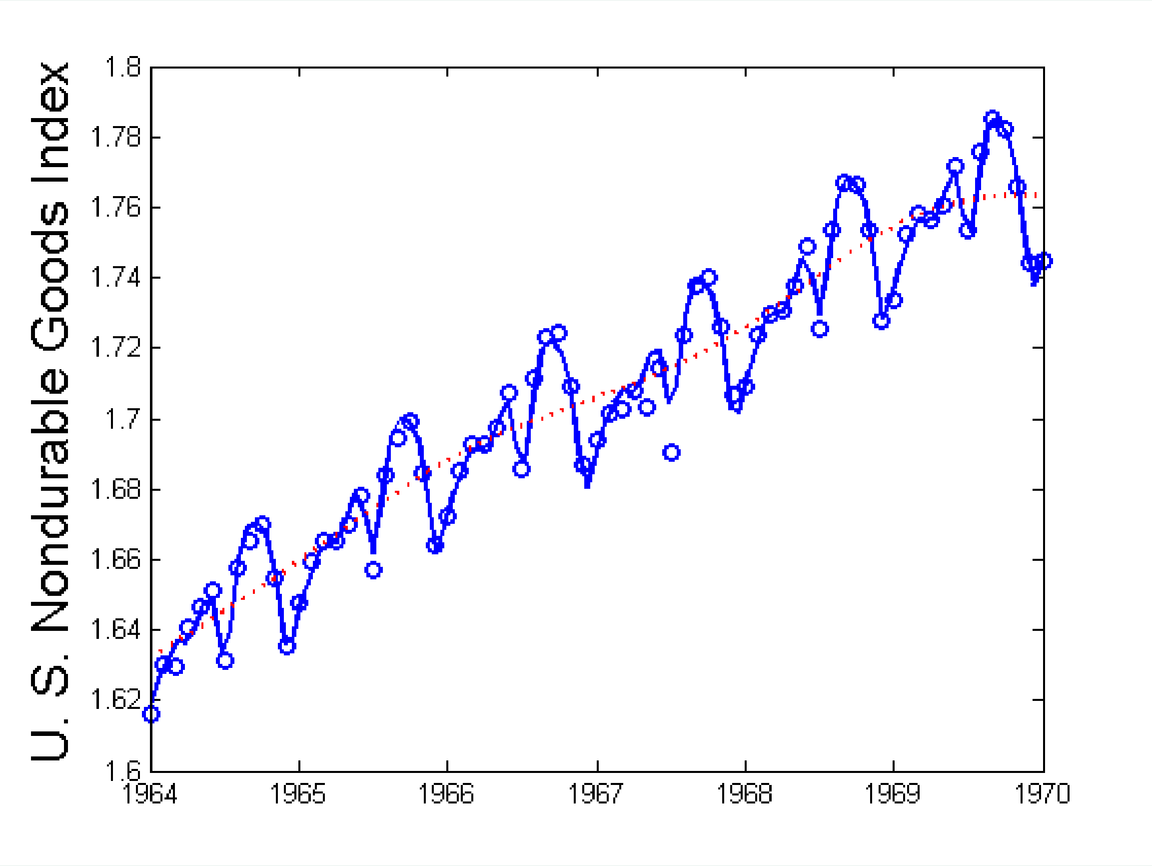
- Seasonal trends are stable over a couple of years, but evolve over a longer time span.
- We can model nonseasonal trend plus an evolving seasonal trend as follows:

$$y(t) = \alpha(t) + \beta_1(t)\sin(2\pi t/365) + \beta_2(t)\cos(2\pi t/365) + \ldots$$

$$+\beta_{p-1}(t)\sin(p\pi t/365) + \beta_p(t)\cos(p\pi t/365) + \epsilon(t)$$

- For the monthly index values from 1952 to 2000 we used $p = 10$.
- Intercept function α was modelled by B-splines with knots at each year and regularized with $\lambda = 0.01$.
- Each regression function β_j had 7 B-spline basis functions.
- A total of 121 parameters were estimated from 577 data points.

The fit to the data over seven years



4. Summary

- The concurrent functional linear model offers a simple way of relating a functional response to functional covariates.
- However, the influence is simultaneous, and does not permit a covariate to affect the outcome at any time other than the present.
- The model can also be fit to a single long time series provided that the number of parameters is kept small and/or regularization is used.

Estimating Dynamic Models

Jim Ramsay, McGill University, ramsay@psych.mcgill.ca
Giles Hooker, Cornell University, giles.hooker@cornell.edu
Jiguo Cao, Yale University , jiguo.Cao@yale.edu
David Campbell, McGill University,
campbell@math.mcgill.ca

Outline

Overview

- We want to fit data by a solution to a system of nonlinear differential equations (DIFE's).
- We ignore DIFE's so simple that they can be solved, such as linear constant coefficient systems. These are already well taken care of.
- Our approach is a generalization of smoothing methods combined with a computational approach involving a modification of profiling.
- We will show results for simulated data from two test-bed problems.
- Data from a chemical reactor producing nylon is analyzed to estimate parameters defining equations for reaction kinetics.

What differential equations do

- DIFE's model change.
- The link the behavior of one or more derivative to the behavior of the process itself and, possibly,
- to one or more exogenous inputs.
- Perhaps the grande dame of such dynamic models is $F = Ma$, connecting the rate of change of velocity a to mass M and an exogenous force F.
- Probably more people know about the closely related

$$e = mc^2$$

The notation

- Let $\mathbf{x}$ be a vector-valued function of length n varying over time t, and that has first derivative values $D\mathbf{x}(t)$.
- Let $\mathbf{u}$ be a vector containing one or more forcing functions.
- Let θ be a vector of parameters defining the DIFE.
- A general formulation is $D\mathbf{x}(t) = \mathbf{f}(\mathbf{x}, \mathbf{u}, t|\theta)$.
- Systems involving higher order derivatives $D^m\mathbf{x}$ are reducible to this form by defining new variables,

$$\mathbf{x}_1 = \mathbf{x}, \quad \mathbf{x}_2 = D\mathbf{x}_1, \quad \ldots, \quad \mathbf{x}_{m-1} = D^{m-1}\mathbf{x}.$$

Nonlinear least squares

- The usual approach is called by textbooks the *nonlinear least squares* or NLS method.
- An initial value numerical method, such as Runge-Kutta, is used to approximate the solution given
  - a trial set of parameter values
  - a trial set of initial conditions.
- The fit value, usually `SSE`, is input into an optimization algorithm to update parameter estimates and the initial conditions.

NLS problems

- NLS is computationally intensive since a numerical approximation to a possibly complex system is required for each update of parameters and initial conditions.
- The size of the parameter set is increased by the set of initial conditions needed to solve the system.
- The inaccuracy of the numerical approximation is added to noise in the data.
- NLS only produces point estimates of parameters.
- Where interval estimation is needed, a great deal more computation is required.
- The fitting criterion can have a complex surface geometry, including many local minima.

Other methods

- Simulated annealing can be used if only a few parameters are involved, but can be extremely slow.
- Local linearization combined with methods for linear systems such as the Kalman filter can be used if the nonlinearity is mild.
- Bayesian methods using MCMC are also possible, but require repeated numerical solution and also add initial values to the parameter set.

Outline

The FitzHugh-Nagumo model

- This simple two-component system is widely used to model properties of actual neural networks.
- They describe the reciprocal dependencies of the voltage V across an axon membrane and a recovery variable R reflecting outward currents, and
- The impact of a time-varying external input E.
- In the typical experiment only V will be measured, but we will consider both to be available.

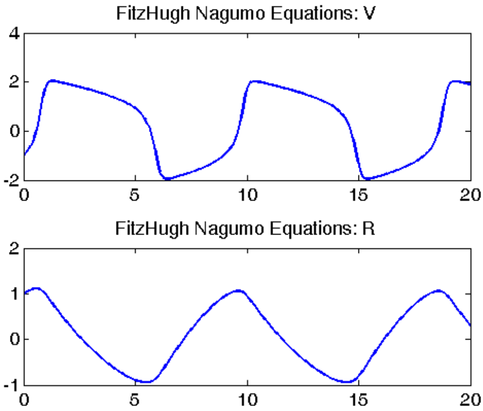
The FitzHugh-Nagumo equations

- Here is the system:

$$\dot{V} = c\left(V - \frac{V^3}{3} + R\right) + E(t)$$
$$\dot{R} = -\frac{1}{c}(V - a + bR)$$

- V is voltage across axon membrane
- R reflects outward currents
- The dynamics of the system are controlled by parameters a, b and c.
- The system would be linear except for the V^3 term.
- The van der Pol equation is a closely related system.

A FitzHugh-Nagumo solution



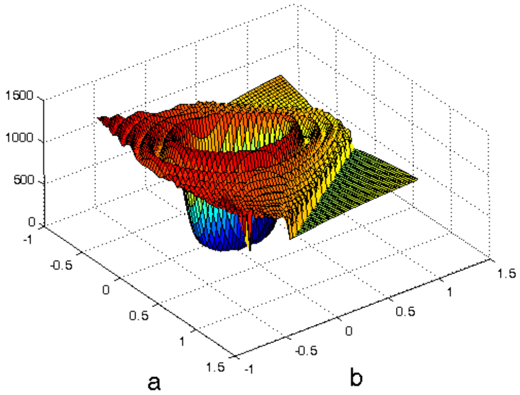
What we see

- The solution quickly reaches a steady state where it is periodic with an asymmetric pattern.
- The parameters control the amplitude and period of the response.

The response surface can be complex

- Differential equations be simple, and yet define extremely complex behavior.
- This is reflected in the response surface of these equations as a functions of parameters a and b.

A FitzHugh-Nagumo response surface



The tank reactor model

- A continuously stirred tank reactor *CSTR* consists of a tank surrounded by cooling jacket and an impeller which stirs the contents.
- It is a basic piece of equipment for a chemical engineer.

The tank reactor variables

- A fluid is pumped into the tank containing a reagent with concentration C_{in} at a flow rate F_{in} and temperature T_{in}.
- Inside the tank a reaction takes place, producing a product that leaves the tank with concentration C_{out} and temperature T_{out}.
- A coolant enters the cooling jacket with temperature T_{cool} and flow rate F_{cool}.
- Temperature T_{out} is can be cheaply measured with little delay and considerable accuracy, but concentration C_{out} requires time and money.

The tank reactor equations

$$
\begin{aligned}
DC_{out} &= -\beta_{CC}(T_{out})C_{out} + F_{in}C_{in} \\
DT_{out} &= -\beta_{TT}(F_{cool}, F_{in})T_{out} + \beta_{TC}(T_{out})C_{out} \\
&\quad + F_{in}T_{in} + \alpha(F_{cool})T_{cool}.
\end{aligned}
$$

- The concentration equation is linear and forced by C_{in}.
- The temperature equation is nonlinear because of the role of T_out in coefficient $\beta_{TC}(T_{out})$ multiplying C_{out}.

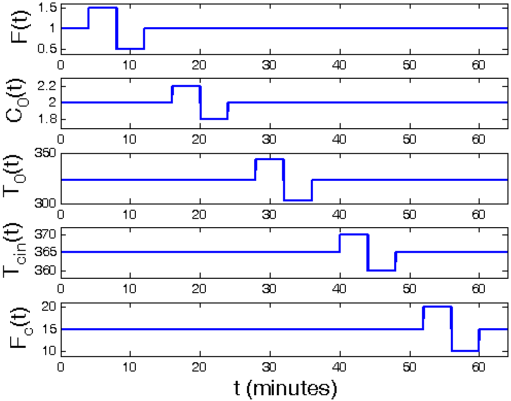
The tank reactor coefficients

- The dynamics of the system are controlled by these four coefficient functions:

$$\begin{aligned}
\beta_{CC}(T_{out}, F_{in}) &= \kappa \exp[-10^4 \tau(1/T_{out} - 1/T_{ref})] + F_{in} \\
\beta_{TT}(F_{cool}, F_{in}) &= \alpha(F_{cool}) + F_{in} \\
\beta_{TC}(T_{out}) &= 130\beta_{CC}(T_{out}, F_{in}) \\
\alpha(F_{cool}) &= aF_{cool}^{b+1}/(F_{cool} + aF_{cool}^{b}/2),
\end{aligned}$$

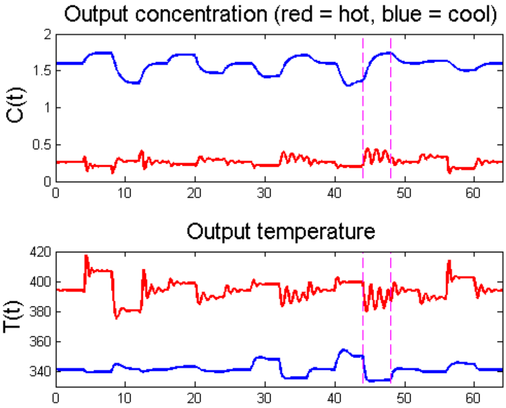
- These functions depend on two paired unknown parameters:
    - κ and τ
    - a and b

Tank reactor inputs



Each input in turn is stepped up, down and back to baseline.

Tank reactor outputs



The experiment is run at two coolant temperatures: hot and cool.

What we see

- When temperatures are moderate, the reactor responds smoothly to changes in input.
- But when temperatures are higher, sharp high frequency oscillations emerge, and are particularly troublesome for a change in coolant temperature.
- Can we predict reactor response at high temperatures from data collected and parameters estimated under the safer cool regime?
- Can we do this using only temperature measurements?

Outline

An overview

- For each variable x_i in $\mathbf{x}$, we define a basis function expansion $\mathbf{c}_i'\phi_i$, where $\mathbf{c}_i$ and ϕ_i are a coefficient vector and a vector of basis functions, respectively.
- Over 400 basis functions are used to capture the sharp variation in outputs.
- A data-fitting criterion $F(\mathbf{y}|\mathbf{x})$ is chosen that measures the fidelity of $\mathbf{x}$ to the data in vector $\mathbf{y}$, and also to the differential equations.
- The extent to which $\mathbf{x}$ is a solution of the differential equation system is assessed by the use of additional penalty terms, and
- the relative balance between these two desiderata is controlled by a set of smoothing parameters.

Structural and nuisance parameters

- There are two classes of parameters to estimate:
  - the parameters θ defining the equation, such as the four reaction kinetics parameters in the CSTR equations
  - the coefficients $\mathbf{c}_i$ defining each basis function expansion.
- The equation parameters are *structural* in the sense of being of primary interest.
- The coefficients $\mathbf{c}_i$ are *nuisance* parameters because they are not of direct interest and
- because their numbers are apt to vary with the length of the observation interval, density of observation, and other factors.
- As a rule, the number of nuisance parameters can be orders of magnitude larger than the number of equation parameters, with a ratio of about 200 applying in the CSTR problem.

Eliminating nuisance parameters

- Nuisance parameters are removed from the problem by defining them as *functions* $\mathbf{c}_i(\theta)$ of the structural parameters using a modified profiling procedure.
- The fitting criterion is then optimized with respect to the structural parameters θ alone.
- An analytic expression for the gradient is developed using the Implicit Function Theorem.
- Compared to marginaling out the nuisance parameters using MCMC, this process is
  - much faster,
  - much more stable, and
  - much easier to program.

The data fitting criterion

$$\mathtt{SSE}(\mathbf{c}|\mathbf{y}) = \sum_{i}^{n} w_i \|\mathbf{y}_i - x_i(\mathbf{t}_i)\|^2.$$

Weights w_i are defined to compensate for differences in scale in the variables.

Assessing fidelity to the equations

- x_i solves the corresponding differential equation if

$$L_i(x_i) = Dx_i - f_i(\mathbf{x}, \mathbf{u}, t|\theta) = 0.$$

- A measure of fidelity to the equation is

$$\mathrm{PEN}_i(\mathbf{x}) = w_i \int [L_i(x_i)]^2 dt.$$

- These are combined into the composite penalty term

$$\mathrm{PEN}(\mathbf{c}|\boldsymbol{\theta}, \boldsymbol{\lambda}) = \sum_i^n \lambda_i \mathrm{PEN}_i(\mathbf{x})$$

- PEN depends on θ through operator $\mathbf{L}$.

The inner optimization for estimating $\mathbf{c}(\theta)$

- Each time θ is changed, we optimize

$$G(\mathbf{c}|\boldsymbol{\theta}, \boldsymbol{\lambda}) = \mathrm{SSE}(\mathbf{c}|\mathbf{y}) + \mathrm{PEN}(\mathbf{c}|\boldsymbol{\theta}, \boldsymbol{\lambda})$$

- This profiling process *implicitly* defines the estimating function $\mathbf{c}(\theta)$.
- As $\lambda_i \to \infty$, variable x_i is forced to satisfy the differential equation more and more exactly.

The outer optimization for estimating θ

- We optimize
$$F(\mathbf{c}(\boldsymbol{\theta})|\boldsymbol{\lambda}) = \mathtt{SSE}(\mathbf{c}(\boldsymbol{\theta})|\mathbf{y})$$
- No penalty term is needed here because $\mathbf{c}(\boldsymbol{\theta})$ is already been regularized in the inner optimization.
- The iterations are greatly accelerated by computing the gradient and Hessian using the *Implicit Function Theorem*.

Outline

An overview of interval estimation

- To a first order of approximation, we can approximate $\theta(\mathbf{y}^*)$ evaluated at an alternative observation $\mathbf{y}^*$ by

$$\begin{aligned}\theta(\mathbf{y}^*) - \theta(\mathbf{y}) &\approx \frac{d\theta}{d\mathbf{y}}(\mathbf{y}^* - \mathbf{y}) \\ &= [D^2_{\boldsymbol{\theta}} F(\hat{\theta}, \hat{\mathbf{c}}|\mathbf{y})]^{-1} D^2_{\boldsymbol{\theta},\mathbf{y}} F(\hat{\theta}, \hat{\mathbf{c}}|\mathbf{y})(\mathbf{y}^* - \mathbf{y}).\end{aligned}$$

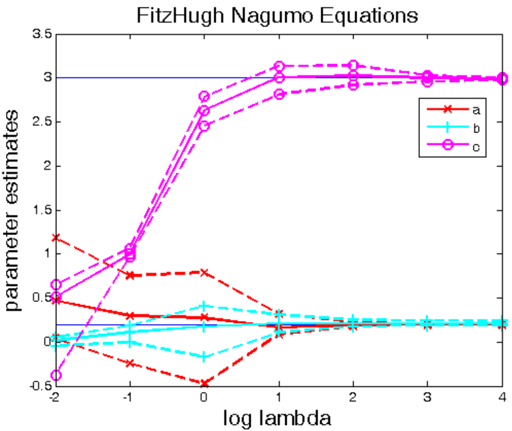
- The derivatives involved can also be computed using the Implicit Function Theorem.
- Sampling variance of θ is then obtained using the Delta method.
- An analogous procedure is used for the variance of $\mathbf{c}(\theta)$.

Outline

Results for the Fitzhugh-Nagumo equations

- The solution to be estimated was determined by $\{a, b, c\} = \{0.2, 0.2, 3\}$ and initial values $\{V(0), R(0)\} = \{-1, 1\}$.
- The paths were measured at 0.05 time units on the interval [0,20].
- Noise was then added to these values with standard deviation 0.5.
- 500 simulated samples were analyzed.

Parameter estimate variation



Both bias and sampling variance decrease as $\lambda_i \to \infty$.

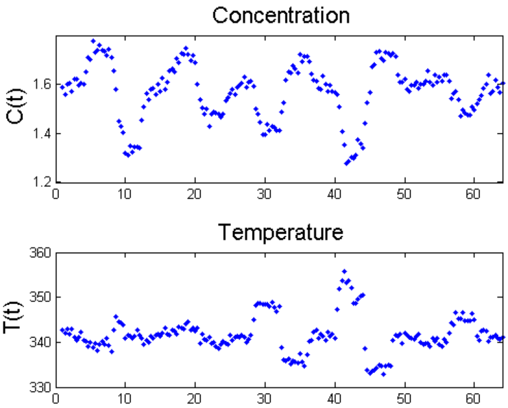
Summary statistics for parameter estimates

	a	b	c
True value	0.2000	0.2000	3.0000
Mean value	0.2005	0.1984	2.9949
Std. Dev.	0.0149	0.0643	0.0264
Est. Std. Dev.	0.0143	0.0684	0.0278
Bias	0.0005	-0.0016	-0.0051
Std. Err.	0.0007	0.0029	0.0012

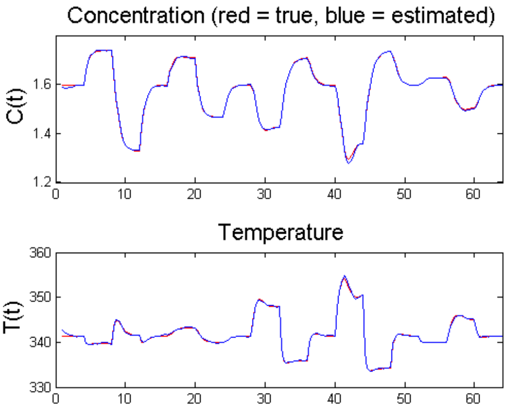
Simulations for the tank reactor equations

- Parameters and initial values for paths were set to those provided by a well known text on control engineering, Marlin (2000) *Process Control*.
- Parameter b is impossible to estimate because of its correlation with a, and therefore was fixed 0.5.
- 1000 simulated samples were analyzed.

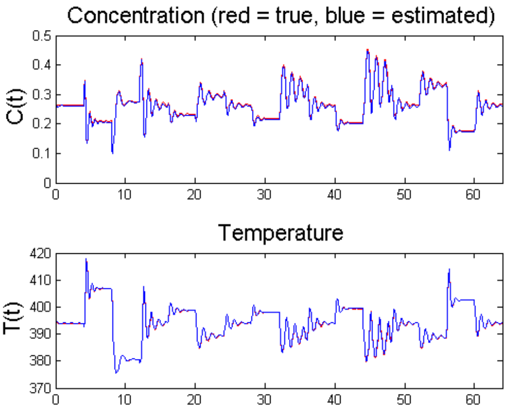
A typical set of tank reactor data



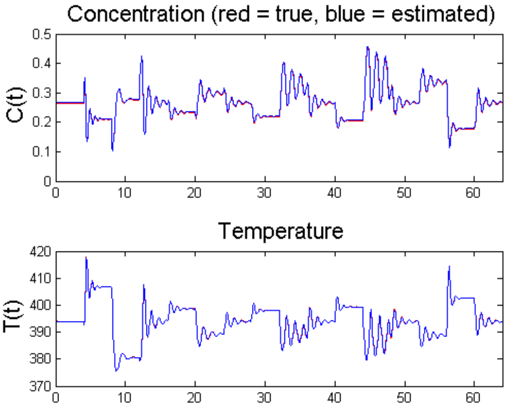
Path estimations, cool mode



Path estimations, hot mode



Path estimations, hot mode



Data for only temperature collected in the cool mode were used.

Summary statistics for parameter estimates

	κ	τ	a
True value	0.4610	0.8330	1.6780
Mean value	0.4610	0.8349	1.6745
Std. Dev.	0.0034	0.0057	0.0188
Est. Std. Dev.	0.0035	0.0056	0.0190
Bias	0.0000	0.0000	-0.0001
Std. Err.	0.0002	0.0004	0.0012

Outline

The nylon experiment

- Nylon and other polymers are created by a chemical reaction in which molecules with two special types of endings chain together to form long molecules.
- The reaction requires water to form the molecules.
- The long molecules can also be broken up, releasing water.
- Temperature and water are critical control variables.
- There were five runs of the experiment at different temperature settings.
- These data were collected in the laboratory of Prof. K. MacAuley of the Dept. of Chemical Engineering at Queen's University, Kingston, Canada.
- The concentration measurements for variables A and C cost about \$30,000 to obtain.

The variables in the nylon equations

- A: molecules with an amine group end (measured)
- C: molecules with a carboxyl group end (measured)
- L: Nylon, a long chain of molecules (a polymer) (not measured)
- W: Water, indirectly adjusted in the experiment
- The variables are related by the mass balance equation

$$A + C \rightleftharpoons L + W$$

Nylon equations

$$DA = DC = -k_p(T)\left(CA - \frac{LW}{K_a(T)}\right)$$

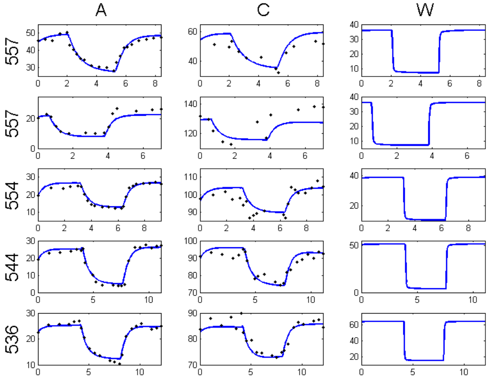
$$DW = k_p(T)\left(CA - \frac{LW}{K_a(T)}\right) - k_m\left(W - W_{eq}\right)$$

$$k_p(T) = k_{p0} \exp\left[-\frac{E}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

$$K_a(T) = \left[\frac{1 + \alpha W_{eq}}{\gamma_w / \gamma_{w0}}\right] K_{a0} \exp\left[-\frac{\triangle H}{R}\left(\frac{1}{T} - \frac{1}{T_0}\right)\right]$$

- variables and known constants are black
- parameters to be estimated are in red
- experimentally manipulated and measured constants and variables are in blue

Fits to the data



Outline

Software

- All the results were computed in Matlab.
- Matlab functional data analysis software was also used. These and a set of software routines that may be applied to any differential equation is available from the URL: http://www.functionaldata.org.

References

- A paper is available from the URL: http://www.functionaldata.org.
- J. O. Ramsay and B. W.Silverman (2005) *Functional Data Analysis*, Second Edition. New York: Springer.

Modelling Change: Incorporating Dynamic Components into Data Analysis

Jim Ramsay and Theo Koulis
Department of Psychology

McGill University

1. Introduction: Input/Output Systems

- We often collect data on units over time.
- There is an *output measure* $y_i(t)$ that reflects the status of a unit i at time t.
- There are also *input measures* $z_{ij}(t), j = 1, \ldots, p$ that indicate the status of various variables thought to affect the output measure.
- We want to study how the status of these units responds to changes in the input variables.
- We especially want to know how a *change* in an input determines the *change* in output.

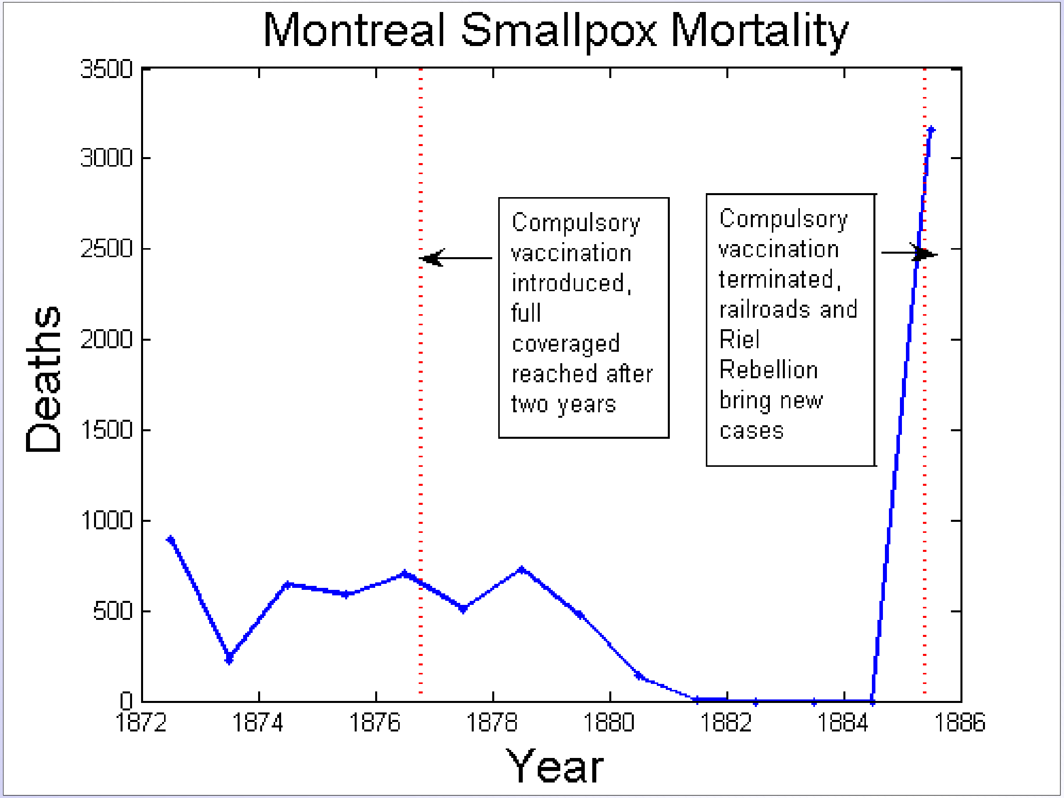
Home Page
Title Page
◀◀ ▶▶
◀ ▶

Go Back
Full Screen
Close
Quit

Examples

- How is driving performance affected by a couple drinks?
- How are golf scores affected by the purchase of a new set of clubs?
- How is pain intensity affected by a dose of morphine?
- How does tumour size respond to radiotherapy?
- How does a couple's social life respond to the birth of a child?
- How does mortality or the incidence of asthma change with an increase in ozone, particulate matter, or other airborne pollutants?

2. Montreal Smallpox Mortality



Inputs: Vaccination coverage, infection from outside the city

3. Functional Regression Analysis

- This sounds like a regression analysis problem that varies over time t.

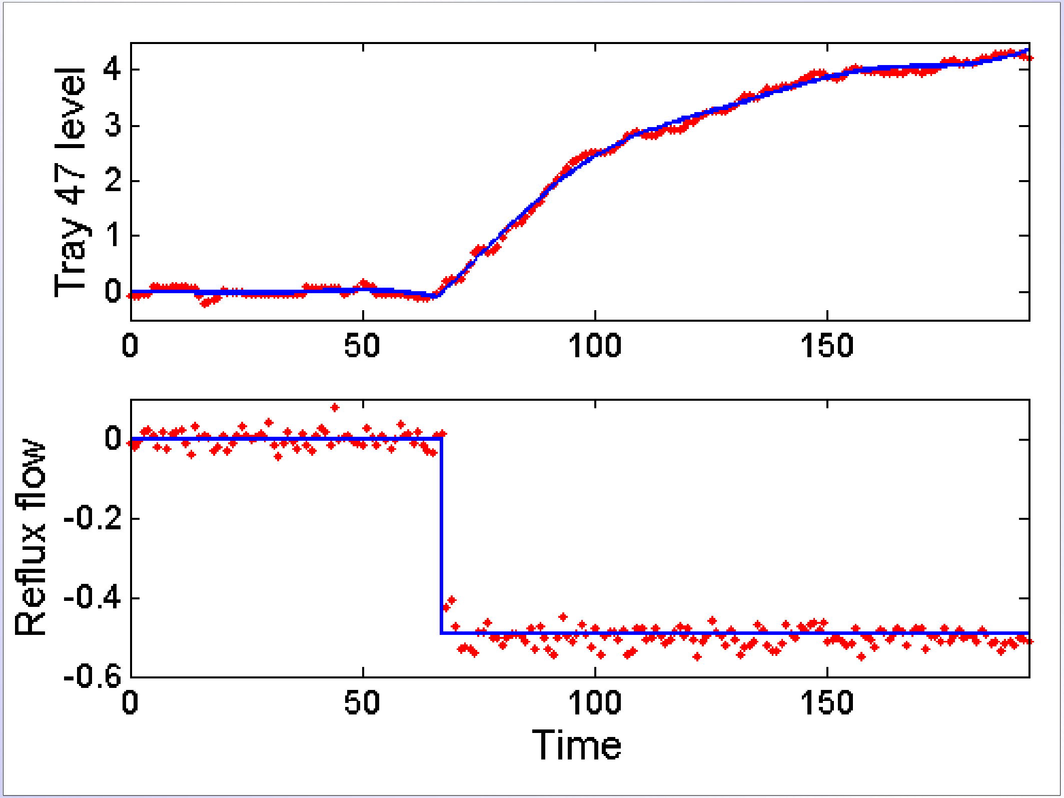
$$y_i(t) = \beta_0(t) + \beta_1(t)z_{i1}(t) + \ldots + \beta_p(t)z_{ip}(t) + \epsilon_i(t)$$

- The regression coefficients $\beta_j(t)$ are now functions of time.
- Software for estimating these regression coefficient functions is readily available. See Ramsay and Silverman (2005) *Functional Data Analysis*, Springer, and the website www.functionaldata.org.
- The model is also a variant of the generalized additive or GAM model.

4. The oil refinery data

- This is a simple input/output system in an oil refinery in Corpus Christi, Texas.
- A fluid, called reflux, flows into a tray in a distillation column in an oil refinery.
- The input variable $z(t)$ is the flow rate.
- The level of fluid in the tray is the output variable $y(t)$.

Refinery output $y(t)$ (top panel) and input $z(t)$ (bottom panel)

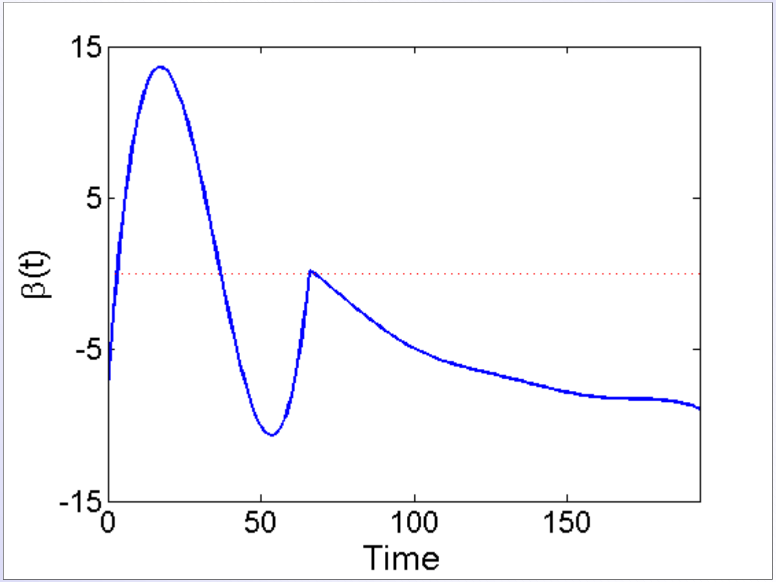


Variation on two time scales

- Over the longer scale, tray level changes from an initial level to a final level.
- But we are also interested in how rapidly the change takes place; that is, short-scale variation.

A functional regression model

$$\texttt{Tray}(t) = \beta(t)\texttt{Reflux}(t) + \epsilon(t)$$



But what does this tell us?

Adding the derivative $D\texttt{Tray}(t)$ to the output

- Can we also model the *rate of change* in the output, as reflected by the first derivative $D\texttt{Tray}(t)$?
- Suppose that we model a mixture of the *rate of change* in the output and the the output itself.
- We'll use constants for the regression functions in the hope of keeping things simple.

$$D\texttt{Tray}(t) + \gamma\texttt{Tray}(t) = \beta\texttt{Reflux}(t) + \epsilon(t)$$

- Coefficient γ controls the relative emphasis on fitting the derivative of the output versus fitting the output itself.
- We estimate $\gamma = 0.02$ and $\beta = -0.20$.

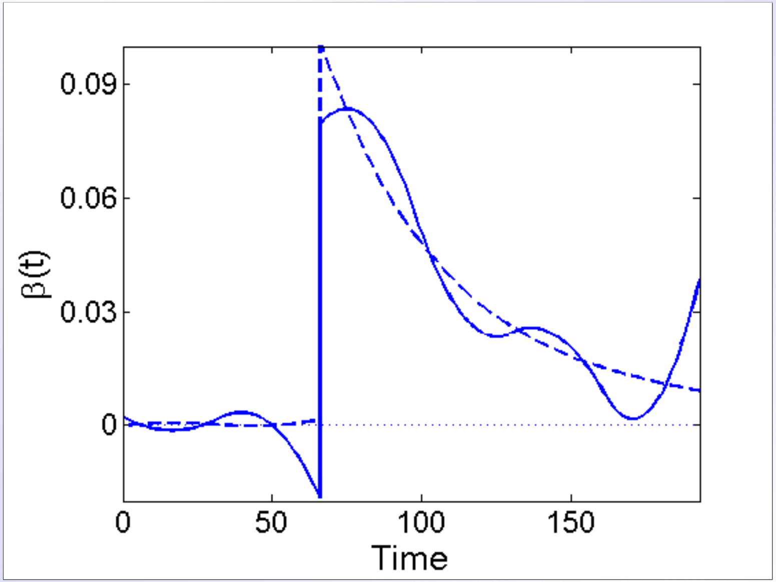
Expressing the model as a Differential Equation

- Models involving derivatives are called *differential equations*.
- They are usually expressed in this rearrangement of our model:

$$D\texttt{Tray}(t) = -\gamma\texttt{Tray}(t) + \beta\texttt{Reflux}(t) + \epsilon(t)$$

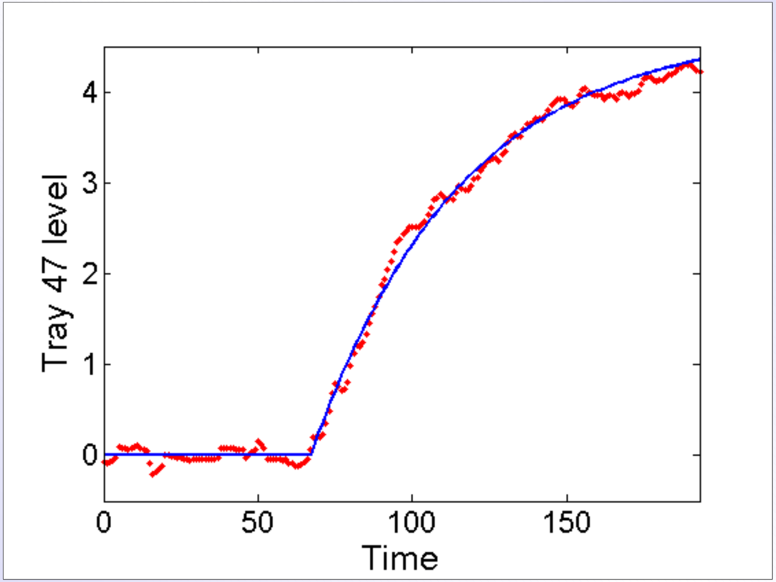
- Input $\texttt{Reflux}(t)$ is called a *forcing function*.

The derivative $D\texttt{Tray}(t)$ and its estimate



The solid line is the derivative estimated from the data, and the dashed line is the model's fit to this derivative.

The fit to the data



This seems impressive given only two parameters.

Introduction: . . .
Montreal Smallpox . . .
Functional Regression . . .
The oil refinery data
A Psychoacoustics . . .
Montreal Smallpox Again
How do I go about . . .

Home Page

Title Page

Page 13 of 100

Go Back

Full Screen

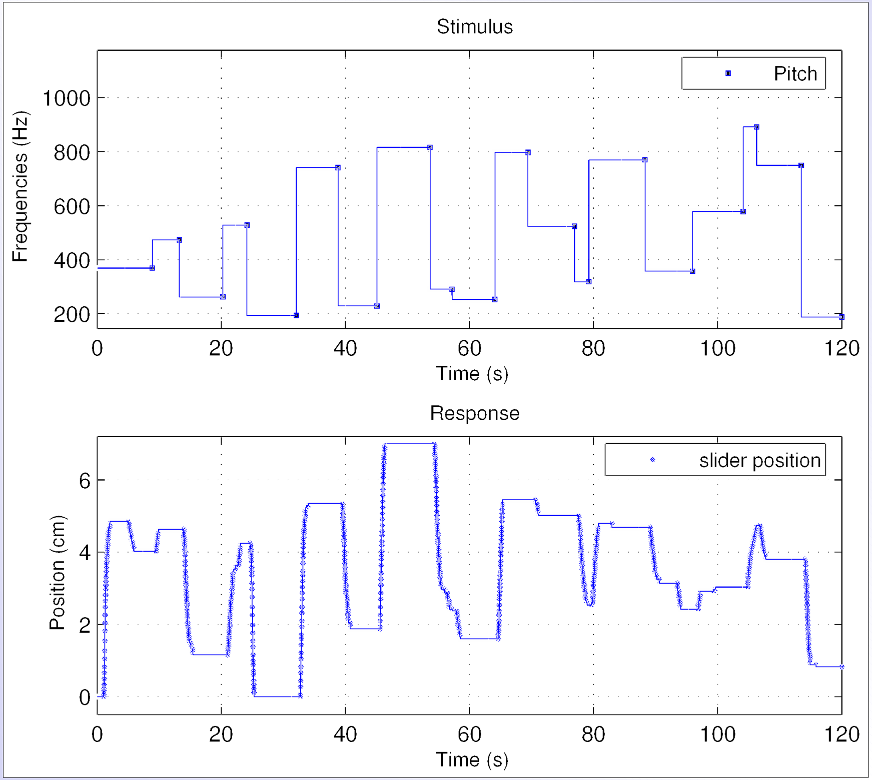
Close

Quit

5. A Psychoacoustics Experiment

- This is an input/output system in music cognition.
- Subjects are asked to follow a series of sequential pitches.
- Subjects adjust a slider on a computer input device (potentiometer).
- If the pitch increases $\longrightarrow$ slider position is increased.
- If the pitch decreases $\longrightarrow$ slider position is decreased.

Input $z(t)$ (top panel) and slider output $y(t)$ (bottom panel)

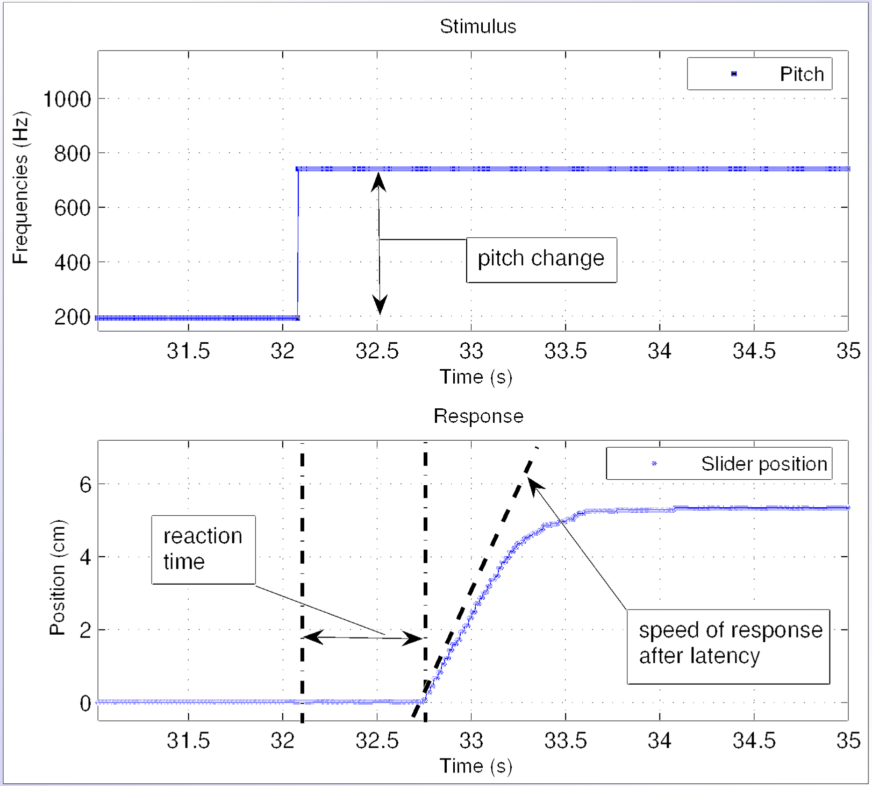


Features of the Slider Data

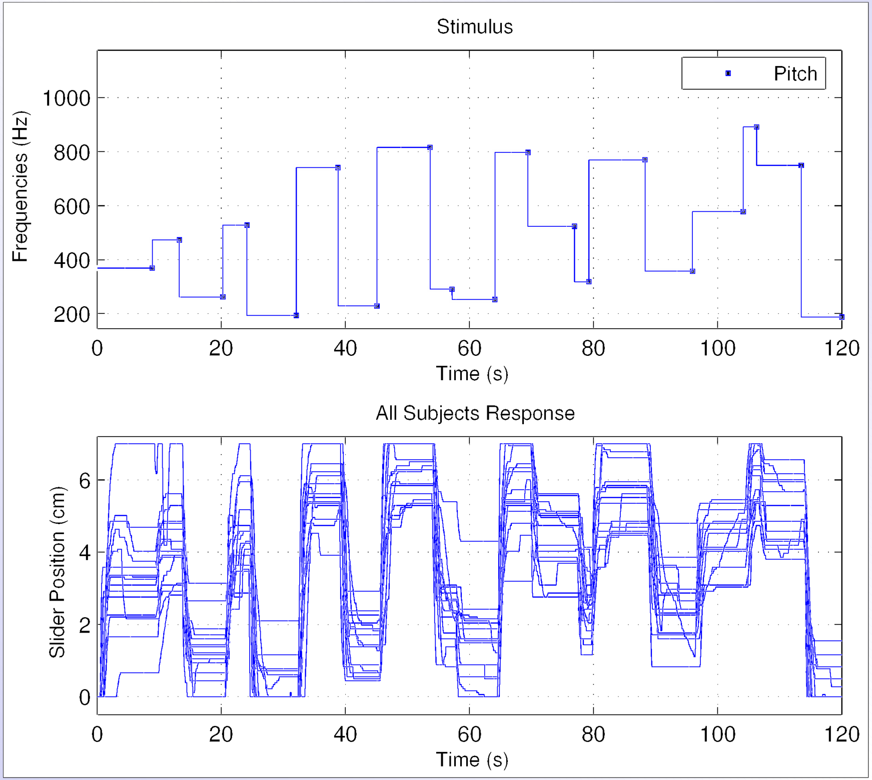
Psychologists are interested in 3 features of the data.

- Reaction Time: the latency between the onset of a fixed stimulus and the response to it.
- Response Speed: a measure of how fast a subject implements the response to the stimulus.
- Gain: the amount of “energy” required to get to a steady state. It is the ratio of “output to input”.

Features: Example from Data



Common stimulus (top panel) and output $y(t)$ (bottom panel: all subjects)



Primary Goal

- The input variable $z(t)$ is the pitch.
- The position of the slider is the output variable $y(t)$.
- There is a lot of variation across subjects.
- Both inter-subject and intra-subject variation.
- Our goal is to quantify this variation to facilitate comparisons (Calibration)

Using Derivatives: A First Approach

- A simple 3-parameter model:

$$D\texttt{Slider}(t) = -\gamma\texttt{Slider}(t) + \beta\texttt{Pitch}(t-\delta) + \epsilon(t)$$

- $\texttt{Slider}(t)$ is the output and $\texttt{Pitch}(t)$ is the forcing function
- Parameters: γ, β, δ
- How do the parameters correspond to the features of the data?
- The parameter δ is the reaction time.
- What about the other two?

Defining the Gain and Response Speed

- Consider the differential equation

$$D\texttt{Slider}(t) = -\gamma\texttt{Slider}(t) + \beta\texttt{Pitch}(t-\delta)$$

  with initial condition $Slider(0) = 0$.

- $\texttt{Pitch}(t)$ is a step function:

$$\texttt{Pitch}(t) = \begin{cases} 0 & \text{if } t < 0 \\ P & \text{if } t \geq 0 \end{cases}$$

- P is the change in pitch.

- The solution is

$$\texttt{Slider}(t) = \begin{cases} 0 & \text{if } t < \delta \\ \frac{\beta}{\gamma} P\left(1 - \exp\left\{-\gamma(t-\delta)\right\}\right) & \text{if } t \geq \delta \end{cases}$$

- The slider position starts at 0 and increases to a limiting value:

$$\texttt{Slider}^* = \frac{\beta}{\gamma} P$$

- The ratio

$$G = \frac{\beta}{\gamma} = \frac{\texttt{Slider}^*}{P}$$

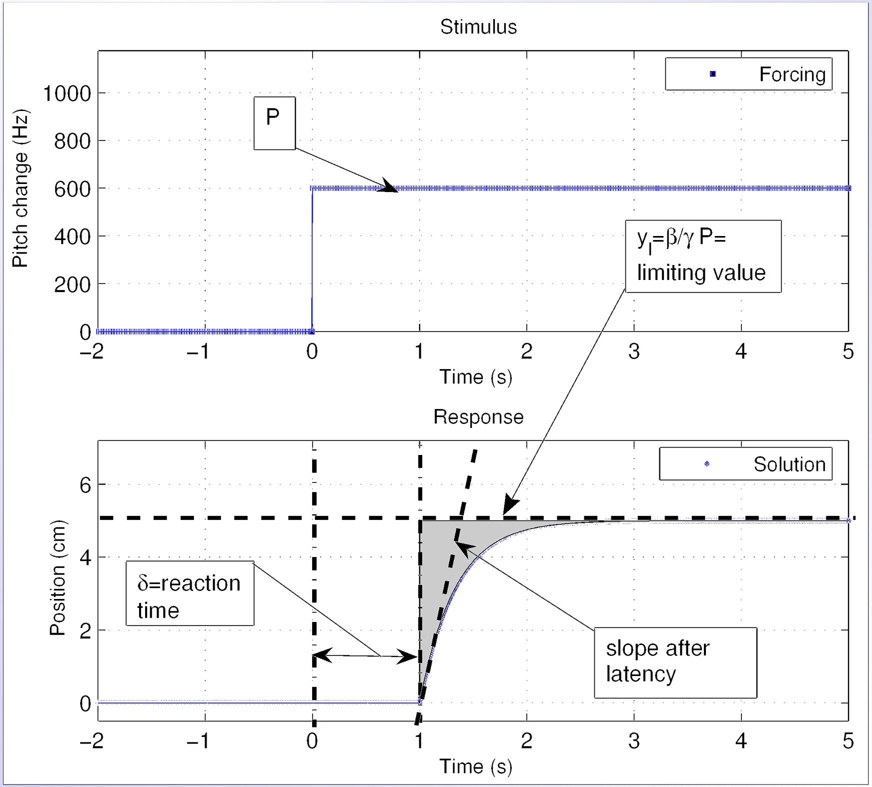
relates the input to the output. We call the ratio G the gain.

- After $\frac{1}{\gamma}$ time units, $\texttt{Slider}(t)$ has reached $2/3$ of the final value $\texttt{Slider}^*$.
- After $\frac{2}{\gamma}$ time units: $7/8$ of the final value.
- After $\frac{4}{\gamma}$ time units: 98% of the final value.
- For this reason, we call the ratio

$$\tau = \frac{1}{\gamma}$$

  the response time constant.
- The parameter γ is called the response speed.

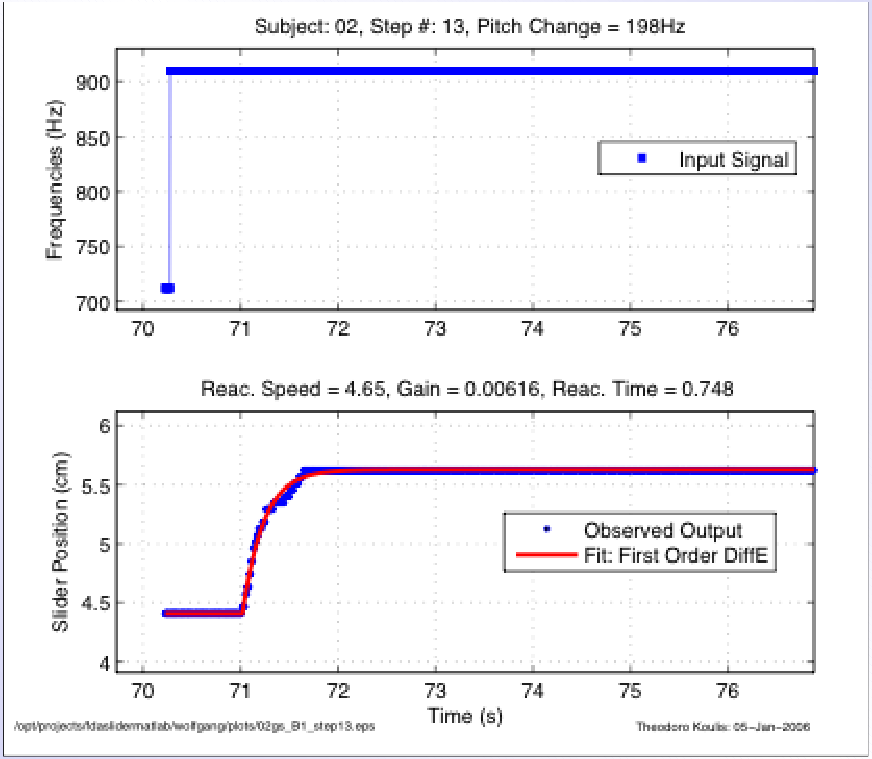
Example: $P = 600$, $\delta = 1$, $\gamma = 3$, $\beta = 0.025$



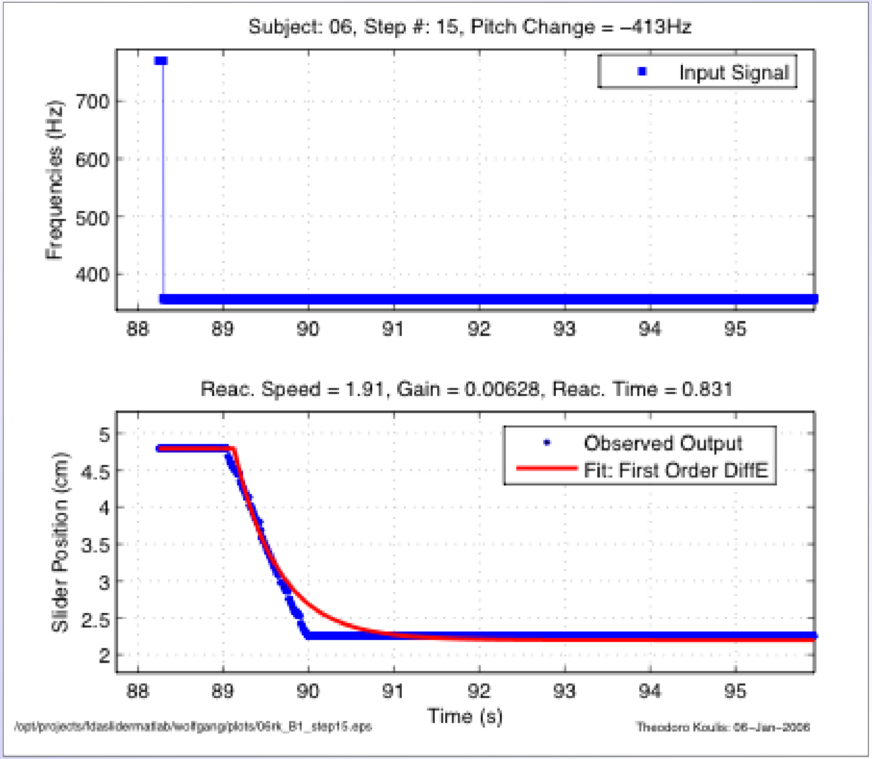
Fitting the Data

- The model does a good job at capturing the shape of the data curves.
- For most cases, the model seems adequate.
- For a few cases, the model does not fit well.
- Even so, we want to keep the simple model to make interpretation easy.

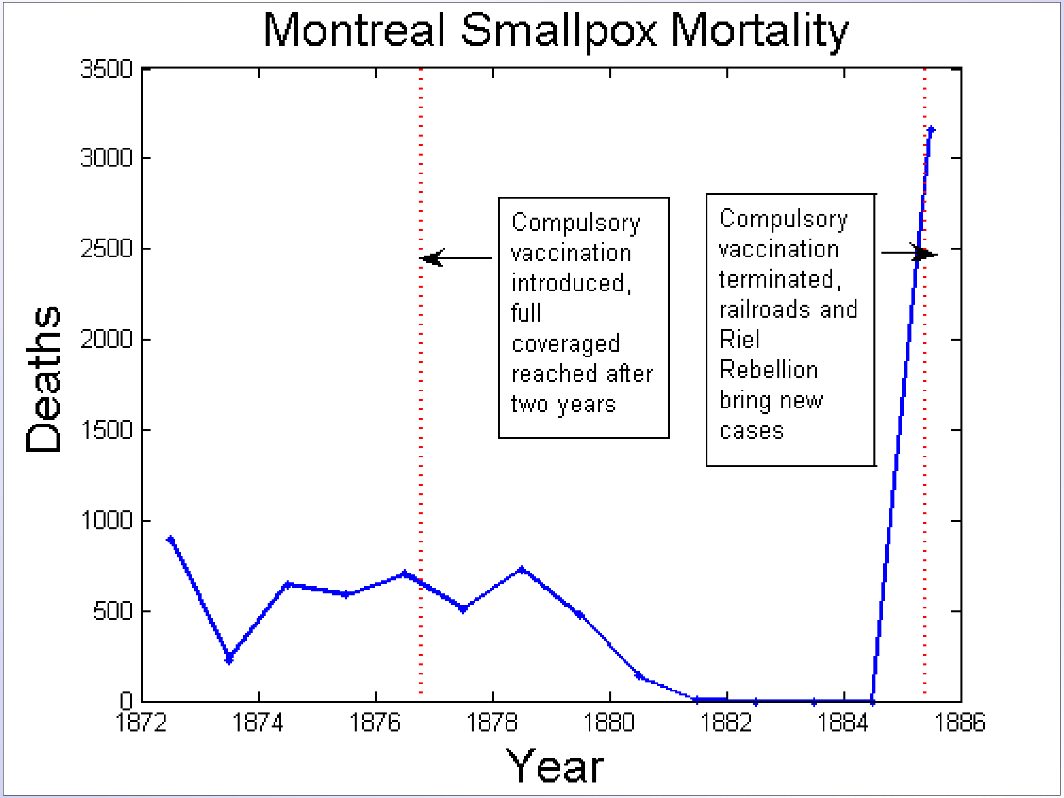
Fitting the Data: Example 1



Fitting the Data: Example 2



6. Montreal Smallpox Again



- There is a delay δ of about two years before enough vaccination coverage is reached to be effective.
- Then the disease all but disappears in two years, suggesting a time constant $\tau = 6$ months.
- The epidemic in 1885 goes from just detectable in April to full force in October, suggesting no delay and a time constant of $\tau = 1.5$ months.
- Once the epidemic was obvious to all, full vaccination coverage was almost immediate, and the disease was under control by the end of the year.
- What's most exciting about the smallpox data is the *rate of change* or *dynamics* of the system.

7. How do I go about modelling change?

Consider that there are three basic features of how a system responds to a change in input:

- How quickly does the change take place? ($4/\gamma = 4\tau$ time units)
- How much change happens? ($\beta/\gamma = \beta\tau$ output units per input unit)
- How long before the change begins? (δ time units)

There are other things to model, too, but these are the big three.
More exotic characteristics of how the output responds to a change in input might require the use of higher order derivatives, such as $D^2y(t)$ and etc.

Extending the basic regression equation

- Set down a time-varying regression equation, with the output $y(t)$ on the left side and various inputs $z_j(t)$ on the right. Some of the inputs can, of course, be constant.
- Each input is multiplied by its regression coefficient function $\beta_j(t)$, which, of course, can be constant if desired.
- Now consider replacing the output $y(t)$ by a mixture or linear combination of $y(t)$ with one or more of its derivatives, $Dy(t), D^2y(t)$ and etc. $y(t)$ and other lower-order derivatives are multiplied by weight functions $\gamma(t)$.
- Add delay parameters as required.

How do I actually fit data with a dynamic model?

- Visit the website www.functionaldata.org to find software in R, S and Matlab along with worked examples.
- Consider buying *Functional Data Analysis*; all the analyses illustrated in the book are also available on the website.

An Introduction to Functional Data Analysis

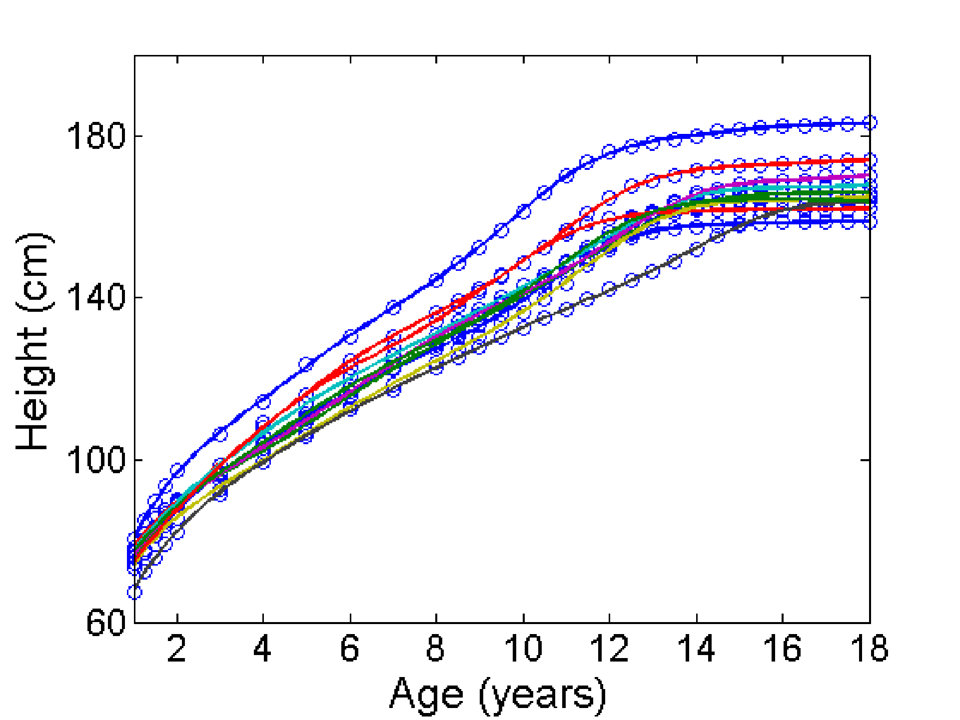
1. Overview

- 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



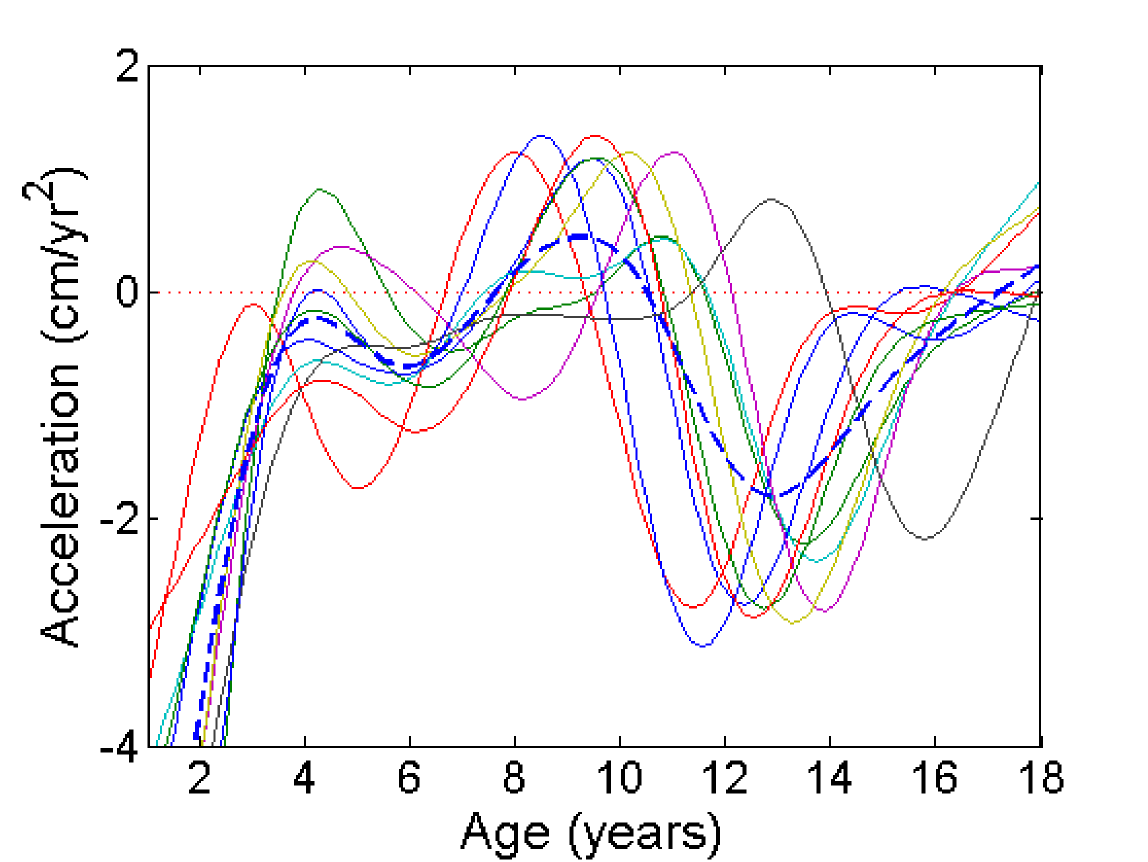
Data challenges

- 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.

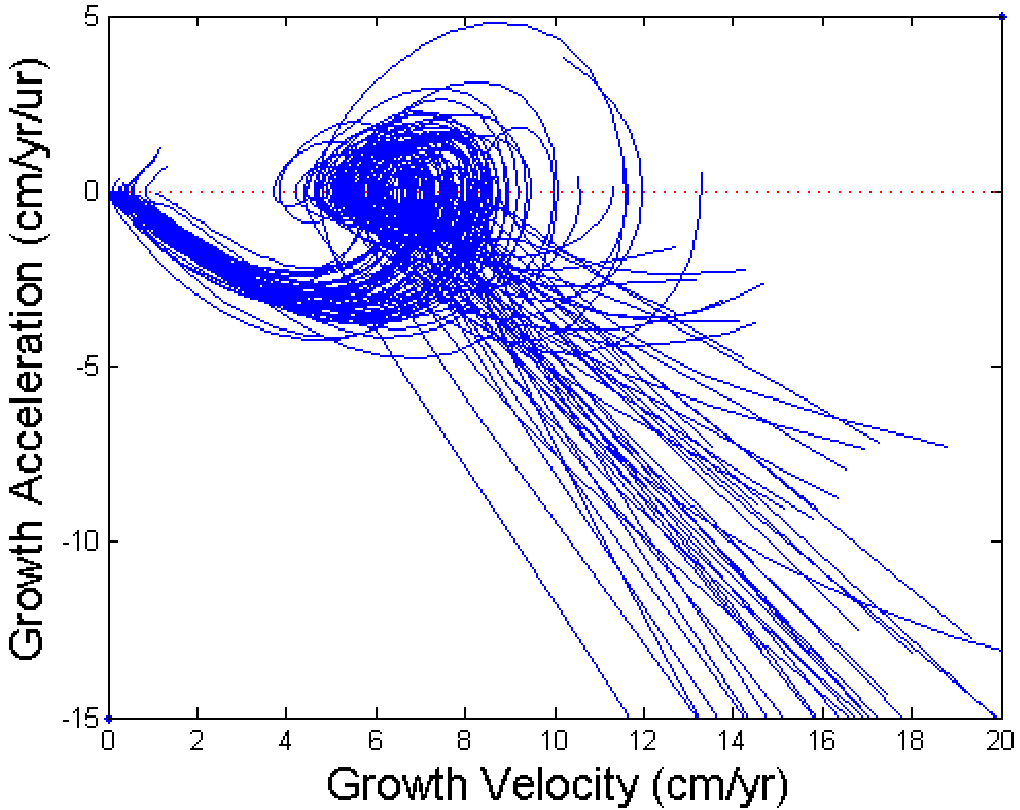
Modelling challenges

- 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

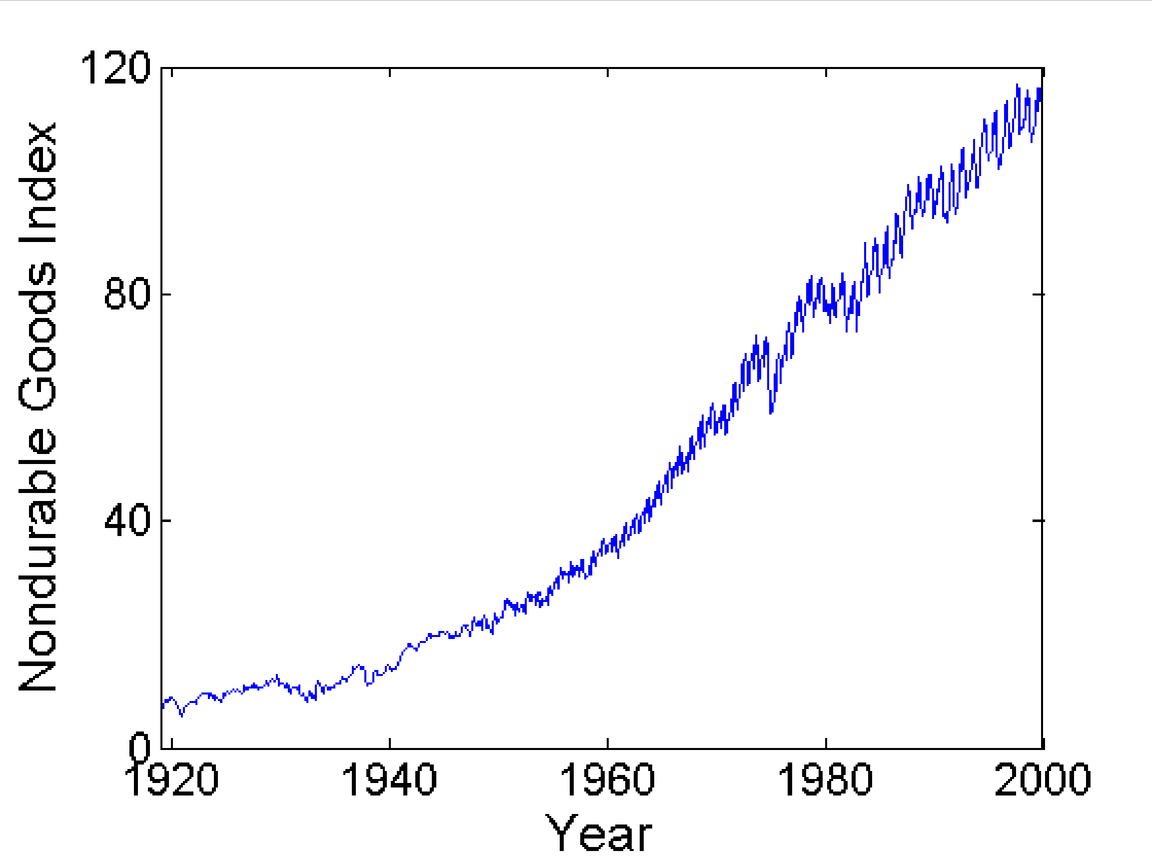


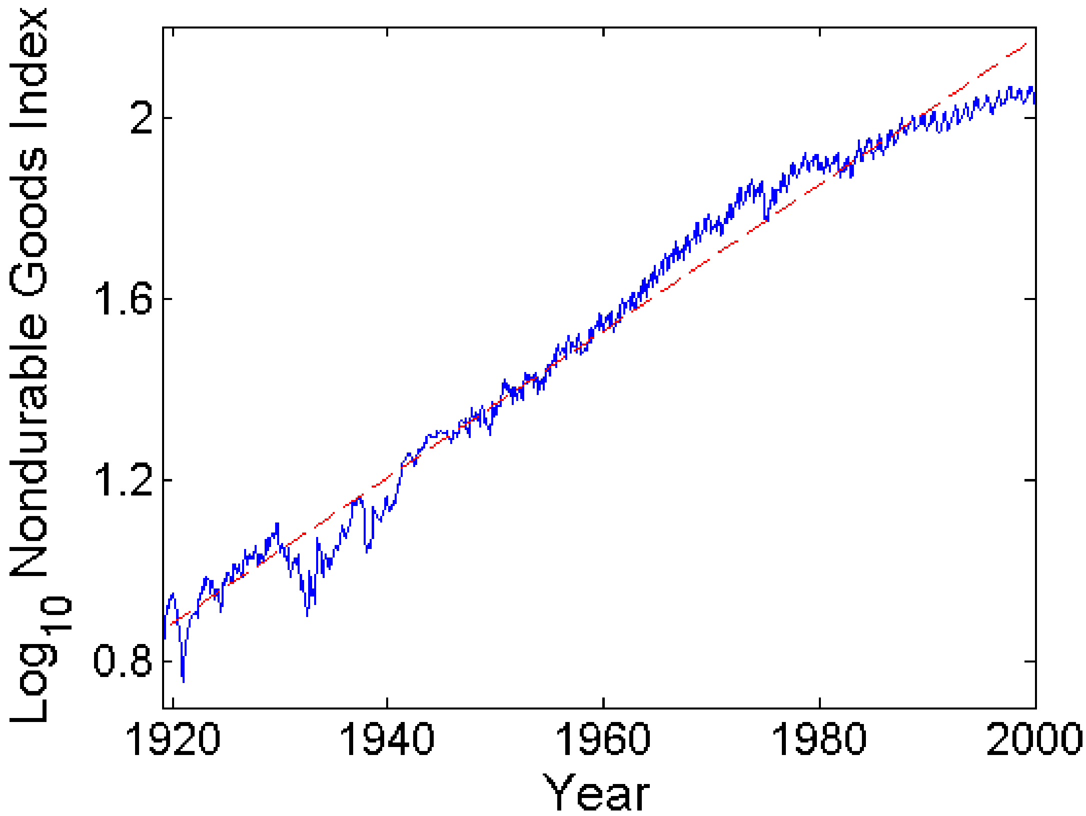
Plotting acceleration against velocity



A single long functional observation

The production of nondurable goods in the U. S.





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

Multiscale variation

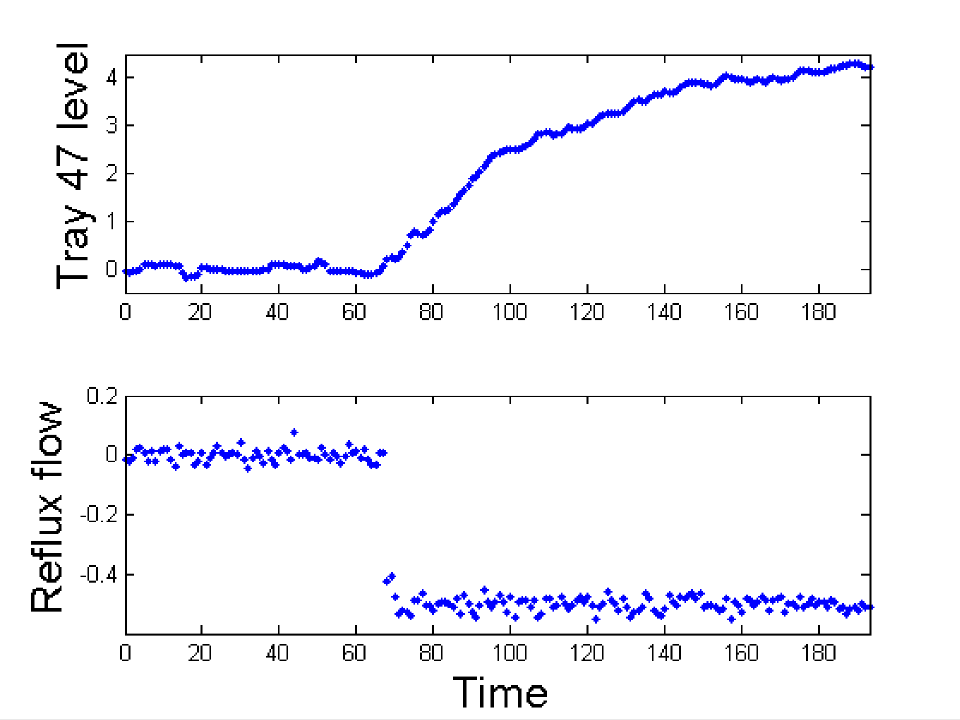
These data, after transformation, have interesting variation on four different time scales:

- **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

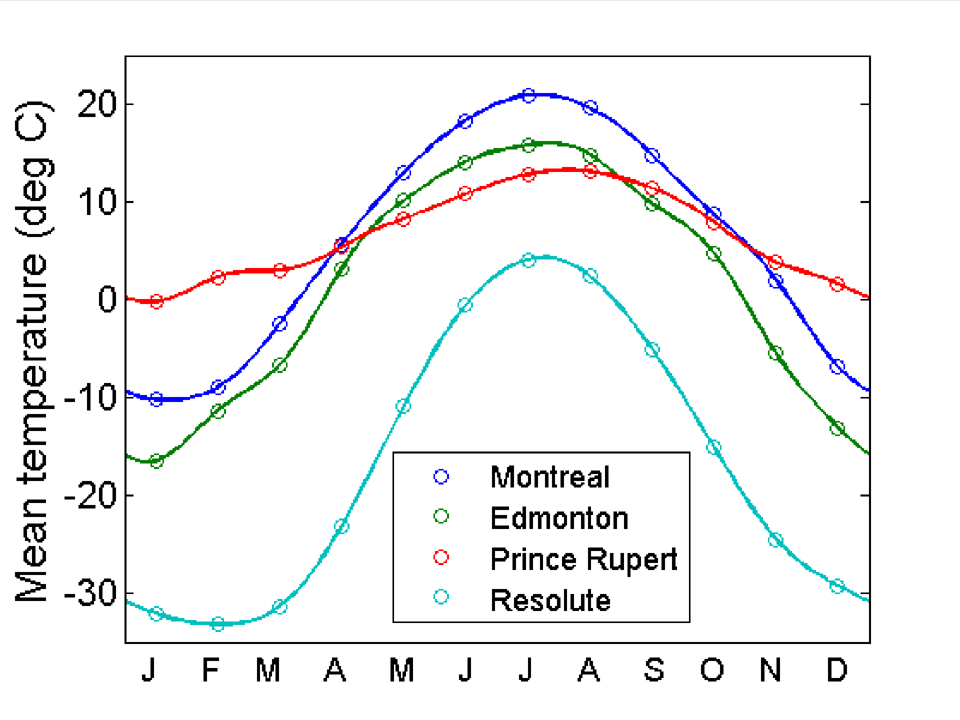
An input/output system

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?

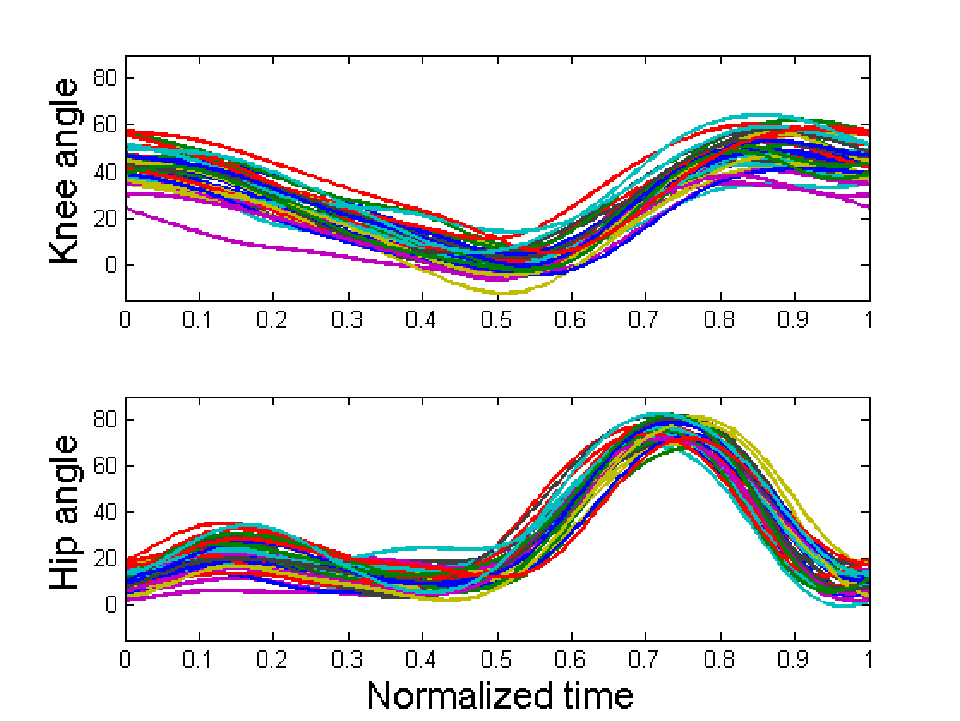
Mean annual temperatures at four weather stations



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

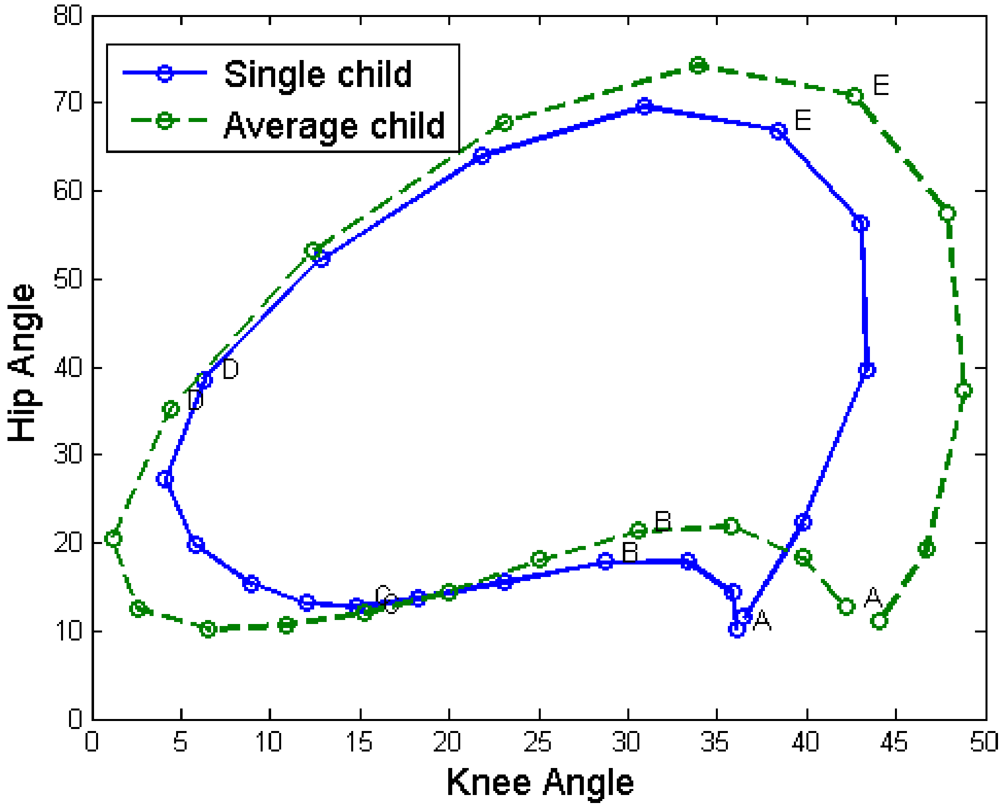
Some multivariate functional data

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.



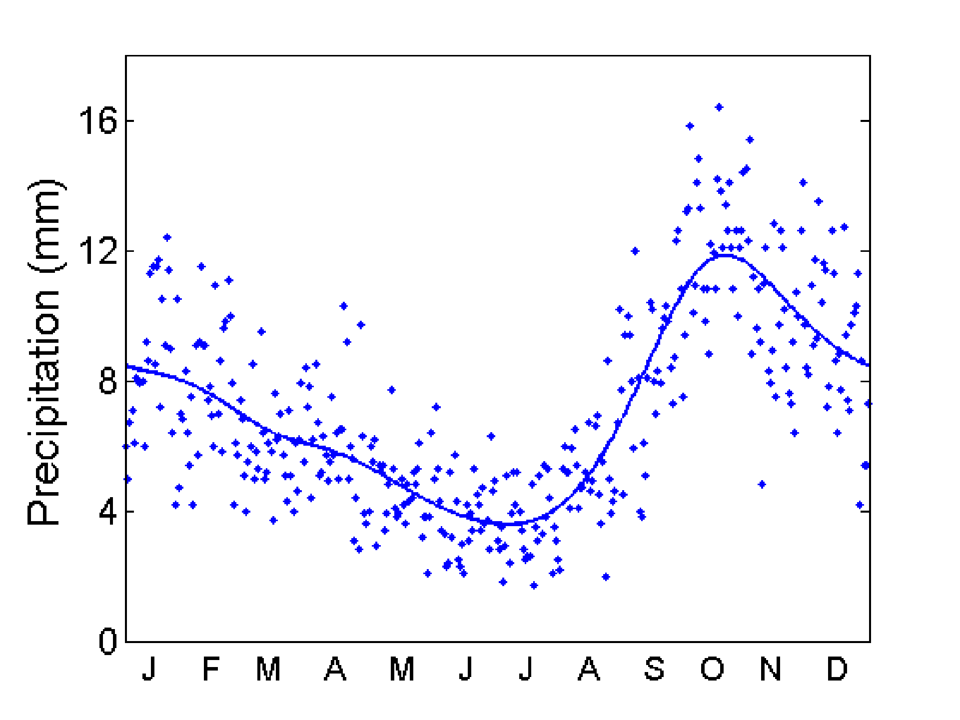
4. The goals of functional data analysis

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.

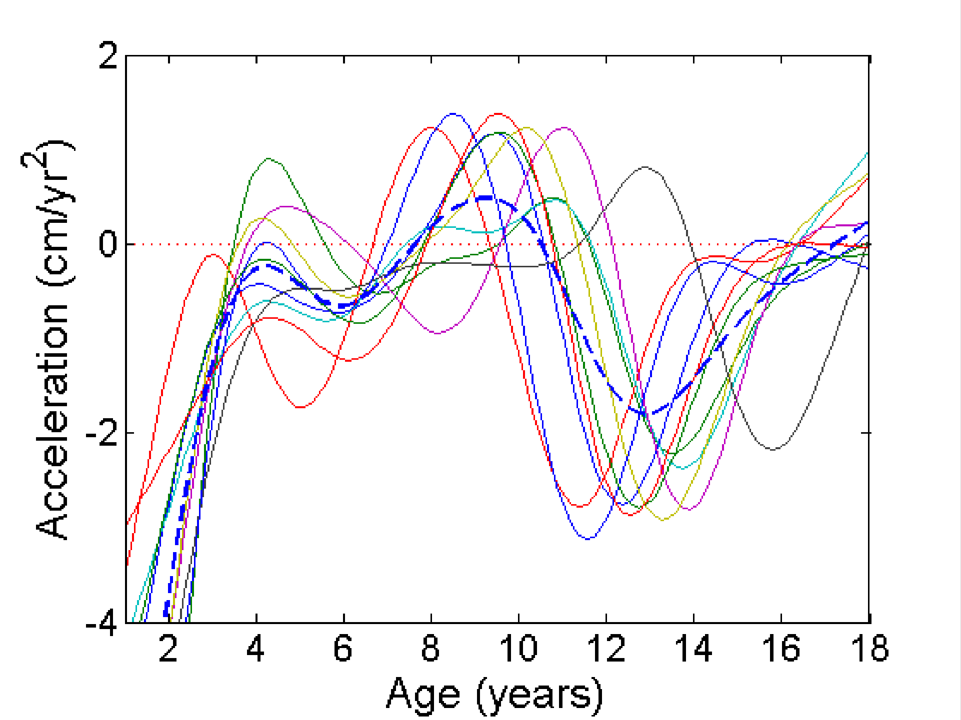
5. The first steps in a functional data analysis

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

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

  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

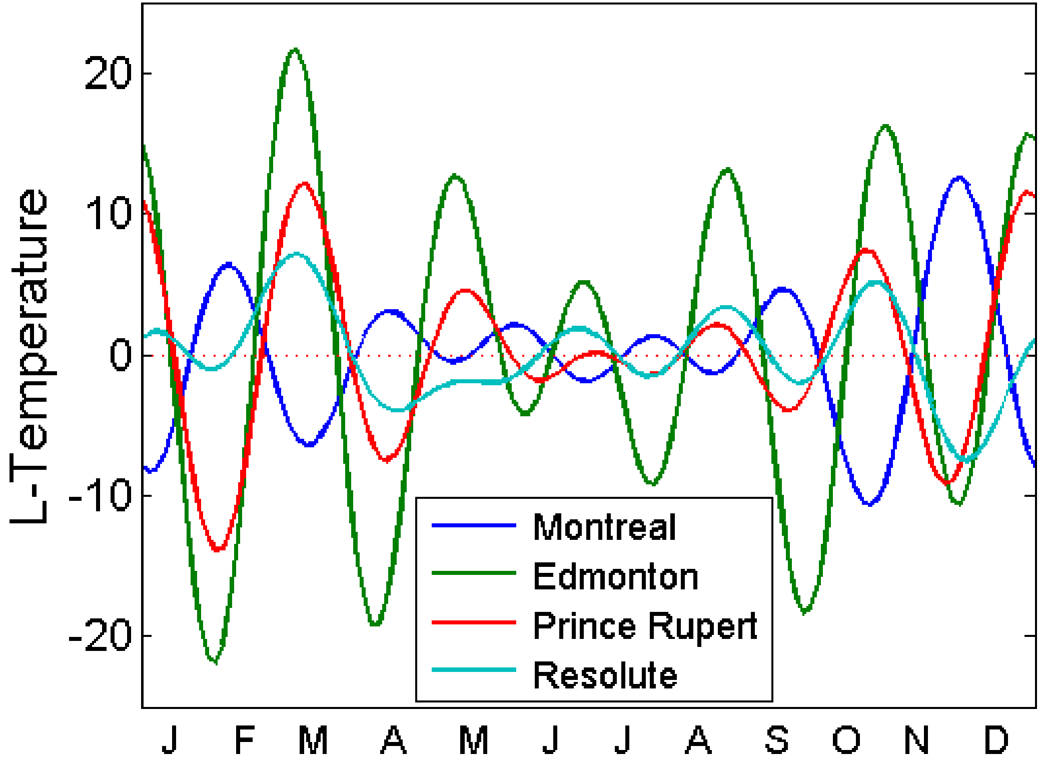
$$L\texttt{Temp} = (\pi/6)^2 D\texttt{Temp} + D^3\texttt{Temp},$$

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

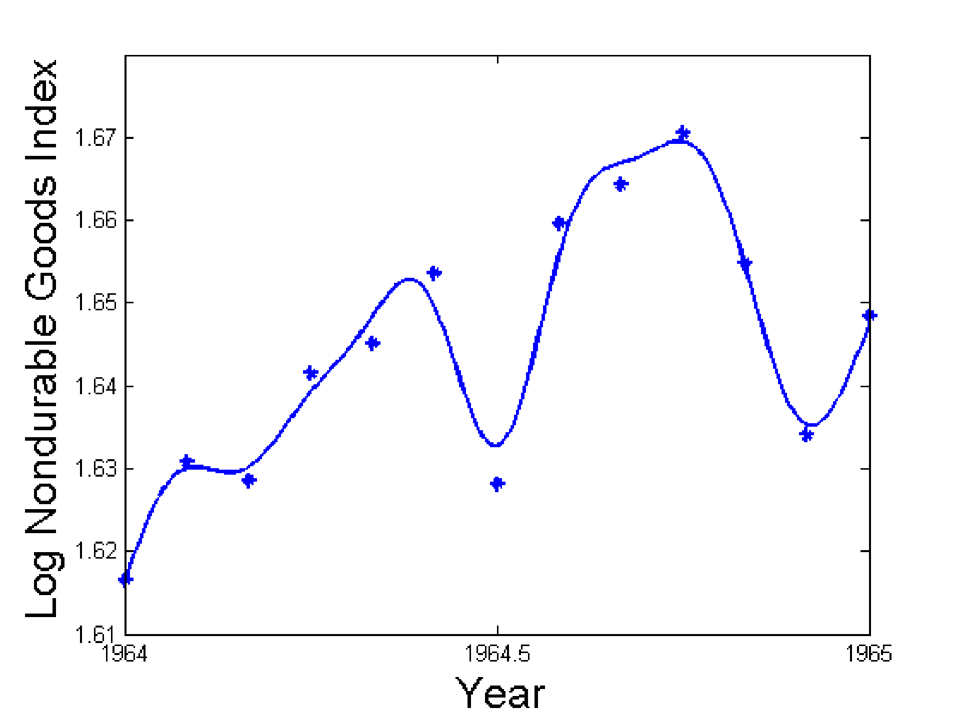
$$L\texttt{temp} = u$$

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

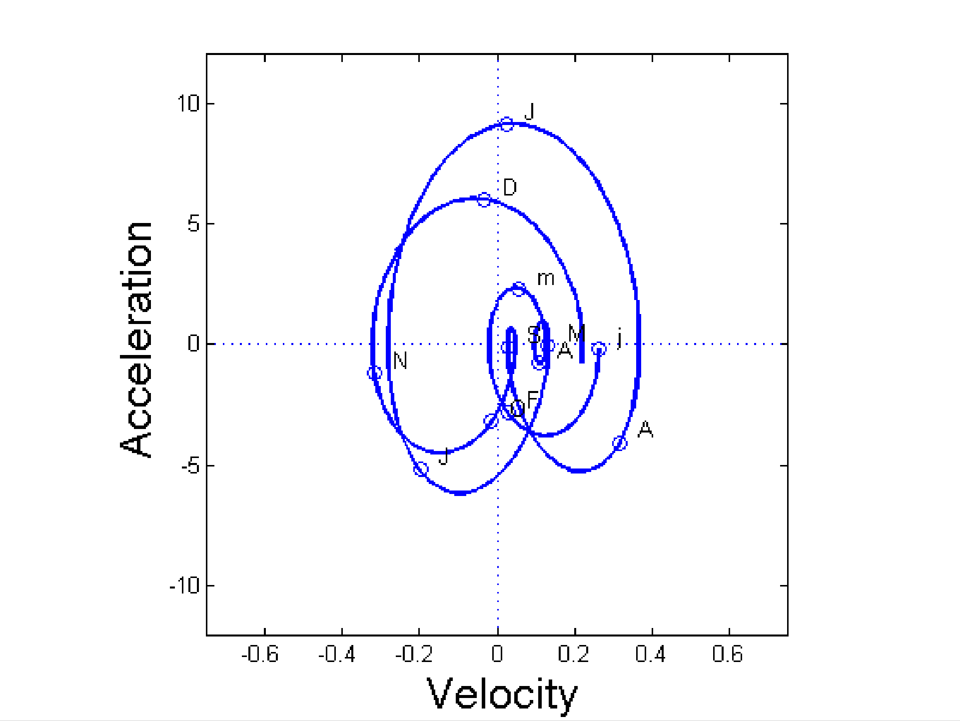
De–sined temperature



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*.

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.

An Overview of the Functional Linear Model

We want to see where these lectures on the functional linear model will go.

- Four functional linear models for the daily weather data.
- A functional ANOVA for precipitation.
- Predicting total annual precipitation from the temperature profile.
- Predicting today's precipitation from today's temperature.
- Predicting the entire year's precipitation from the year's temperature profile.
- A short–term feed–forward model for precipitation.
- A more general perspective.
- Predicting precipitation dynamics: a differential equation
- The idea of a linear model reviewed.

The average Canadian weather data

- 35 Canadian weather stations selected to cover the country.
- Daily temperatures (0.1 degrees Celsius) and precipitations (0.1 mm) averaged over the years 1960 to 1994. (Feb 29th combined with Feb. 28th).
- Canada divided into Atlantic, Continental, Pacific and Arctic weather zones.


Resolute
Inuvik
Whitehorse
Yellowknife
Iqaluit
Uranium City
Churchill
Prince Rupert
Dawson Creek
Prince George
Schefferville
Edmonton
The Pas
Prince Albert
St. John's
Kamloops
Calgary
Vancouver
Victoria
Regina
Winnipeg
Jonquière
Bagotville
Sydney
Charlottetown
Québec
Fredericton
Halifax
Thunder Bay
Ottawa
Montreal
Sherbrooke
Yarmouth
Toronto
London

1. A functional analysis of variance

- Does the precipitation profile vary from one weather zone to another?
- We have a number N_g of weather stations in each climate zone $g = 1, \ldots, 4$, and
- the model for the mth temperature function in the gth group, indicated by $\texttt{Prec}_{mg}$, is

$$\texttt{Prec}_{mg}(t) = \mu(t) + \alpha_g(t) + \epsilon_{mg}(t).$$

- $\mu(t)$ is the grand mean function, summarizing precipitation for all of Canada.
- $\alpha_g(t)$ is the functional effect of being in weather zone g
- In order to fix zone effects, we require that

$$\sum_g \alpha_g(t) = 0 \text{ for all } \; g$$

2. A scalar response and a functional independent variable

- The response is the log total annual precipitation

$$\texttt{PrecTot}_i = \int_0^{365} \texttt{Prec}_i(t)\,dt$$

- The model is

$$\log(\texttt{PrecTot}_i) = \alpha + \int_0^{365} \texttt{Temp}_i(s)\beta(s)\,ds + \epsilon_i\,.$$

- But here we have a real problem. How to avoid over-fitting the 35 scalar observations?
- We'll use regularization or roughness penalties on the estimated regression functions.

3. A functional response and a functional independent variable

This is a big topic, and breaks down into several useful special versions.

3.1. The concurrent functional model

- We might only use the temperature at the same time $s = t$ because we imagine that precipitation now depends only on the temperature now.
- Our model is

$$\texttt{Prec}_i(t) = \alpha(t) + \texttt{Temp}_i(t)\beta(t) + \epsilon_i(t)$$

- We might call this model *concurrent* or *point-wise*.
- Should we use regularization to force β to be smooth in t?

3.2. The annual or total model

- We may prefer to allow for temperature influence on $\texttt{Prec}(t)$ to extend over the whole year.
- The model expands to become

$$\texttt{Prec}_i(t) = \alpha(t) + \int_0^{365} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t)$$

- The value $\beta(s,t)$ determines the impact of temperature at time s on precipitation at time t.
- We need roughness penalties for variation in both s and t

3.3. The limited–term feed–forward model

- it may be that what counts is whether the temperature has been falling rapidly up to time t. The model expands to

$$\mathtt{Prec}_i(t) = \alpha(t) + \int_{t-\delta}^{t} \mathtt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t)$$

- Here δ is the time lag over which we use temperature information.
- Now β is only defined over the somewhat complicated trapezoidal domain: $t \in [0, 365], t - \delta \leq s \leq t$.

A functional analysis of . . .
A scalar response and . . .
A functional response . . .
Predicting derivatives
What exactly makes a . . .
Home Page
Title Page
Page 10 of 13
Go Back
Full Screen
Close
Quit

3.4. The local influence model

- Finally, we may open up the model to allow integration over s within a t-dependent set Ω_t.
- The model may therefore be

$$\texttt{Prec}_i(t) = \alpha(t) + \int_{\Omega_t} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t)$$

4. Predicting derivatives

- When the response is a derivative, then there is the potential for the function itself to be a useful covariate.
- The concurrent linear model

$$D\texttt{Prec}_i(t) = \texttt{Prec}_i(t)\beta(t) + \epsilon_i(t)$$

is a *homogeneous first order linear differential equation* in precipitation.

- If we also include an influence of temperature,

$$D\texttt{Prec}_i(t) = \texttt{Prec}_i(t)\beta_0(t) + \texttt{Temp}_i(t)\beta_1(t) + \epsilon_i(t),$$

the equation is said to be *forced* or *nonhomogeneous.*

A functional analysis of . . .
A scalar response and . . .
A functional response . . .
Predicting derivatives
What exactly makes a . . .
Home Page
Title Page
Page 12 of 13
Go Back
Full Screen
Close
Quit

5. What exactly makes a model linear?

- We see that the functional linear model has a lot more variants than it's poor multivariate cousin.
- We will need to look at a definition of a linear model that encompasses these models and many others.