From Data to Differential Equations

Jim Ramsay, McGill University

Regina Nuzzo, McGill University

Differential Equations as Models

Differential equations are often the natural way to model functional data.

- They make explicit the relationship between the derivatives of a function and the function itself. Derivatives are thus forced to behave smoothly, and are not just a byproduct of curve estimation. eg: $\omega^2 x(t) + D^2 x(t) = 0$ for harmonic motion requires that acceleration be just as smooth as the function.
- The behavior of a derivative is often of more interest than the function. Recall equations like $f = ma$ and $E = mc^2$. A derivative such as velocity or acceleration can reflect energy exchange.
- Solutions to DIFE's can exhibit behaviors nearly impossible to model directly, such as in chaotic systems. eg: The Rössler system:

$$\begin{aligned} Dx(t) &= -y(t) - z(t) \\ Dy(t) &= x(t) + ay(t) \\ Dz(t) &= b + (x(t) - c)z(t) \end{aligned}$$

- There are many more ways to introduce stochastic behavior into a DIFE model than into a nonparametric regression model.
  - stochastic forcing function
  - stochastic coefficient functions
  - stochastic initial or boundary values
  - stochastic time
  - plus the usual exogenous i.i.d. noise
- The solution to an mth order linear DIFE is an m dimensional function space, and therefore can model variation as well location for functional data.
- Natural scientists often suggest theory to engineering and biological applications in terms of DIFE systems.
- Many fields such as pharmokinetics and process control use DIFE's already, and especially when input/output systems are involved.
- Differential equation models are essential in industrial process control where feedback links are constructed so as to stabilize and optimize the system.

Representing a Input/Output System

For a single input and a single output (SISO), we can have

$$Dx(t) = \beta(t)x(t) + \alpha(t)u(t)$$

But this formulation ignores noise in the input (measurement error, small very short term fluctuations, etc). So we add a noise process $\nu(t)$.

$$Dx(t) = \beta(t)x(t) + \alpha(t)u(t) + \nu(t)$$

What about noise in the measurement of the output $x(t)$? We'd better add some noise $\epsilon(t)$, too, to the noiseless process $x(t)$ to get the observed process $y(t)$.

$$y(t) = x(t) + \epsilon(t)$$

These two equations are called the *state-space* representation of the data.

A Single Input/Single Output Example

These are real data from an oil refinery in Corpus Christi.

The single input variable is “reflux flow”, and the single output variable is “tray 47 level”.

There are 194 sampling points at which these two signals are observed.

After some experimentation, guided by some model selection tools, and with an eye to getting a good estimate of at least the second derivative of the output, we used 30 B-spline basis functions of order 6 to fit the data.

Here are the data and the smooths:

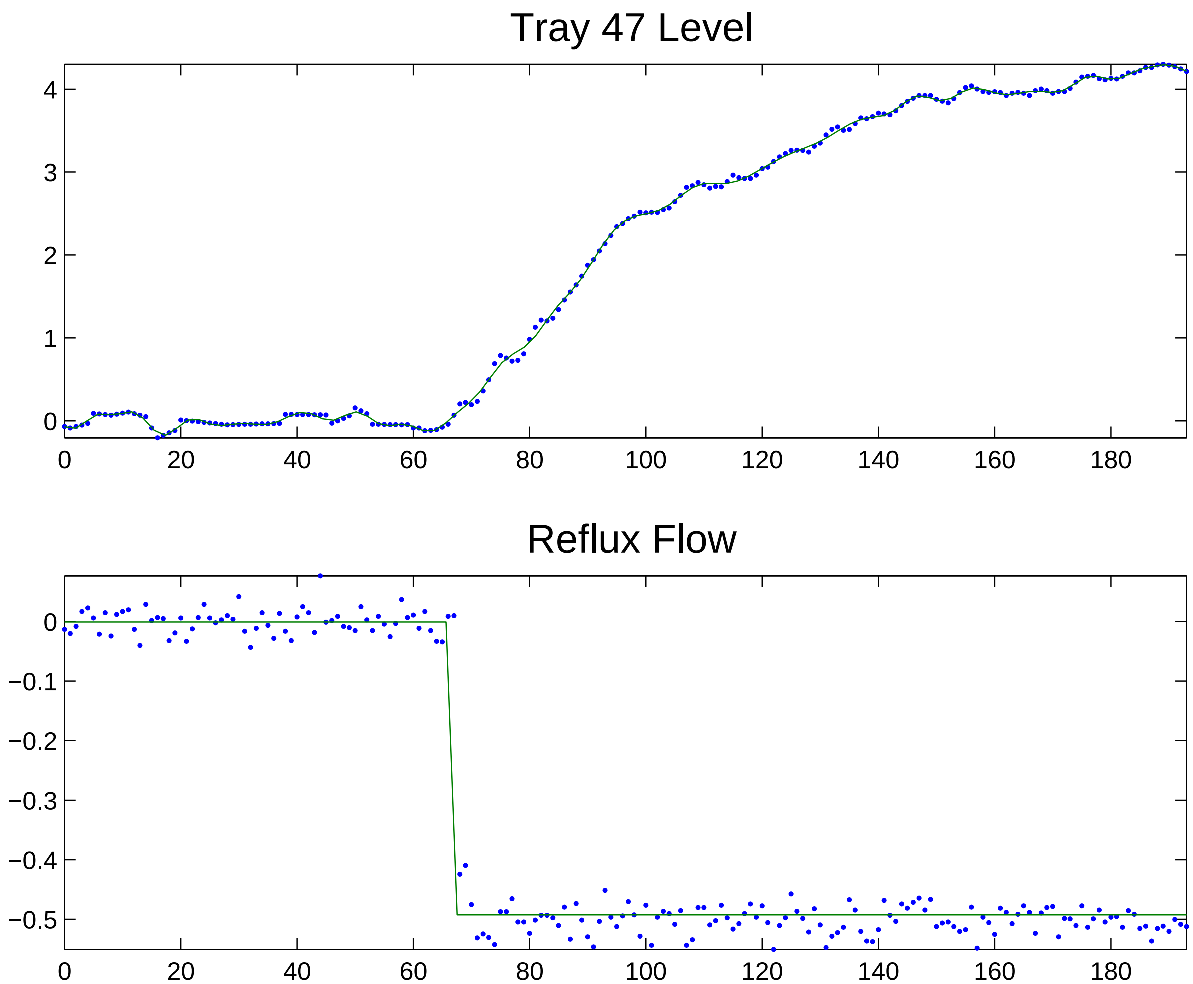
Tray 47 Level
4
3
2
1
0
0
20
40
60
80
100
120
140
160
180
Reflux Flow
0
−0.1
−0.2
−0.3
−0.4
−0.5
0
20
40
60
80
100
120
140
160
180

Figure 0.1:

DIFE For a Multiple Input/Single Output System

We have a sample of N realizations of a response function $x(t)$ and L input functions $u_\ell(t)$. N can be 1.

We want a linear differential operator of order m of the form

$$\begin{aligned} Lx(t) &= \sum_{j=0}^{m-1} \beta_j(t) D^j x(t) + D^m x(t) + \\ & \sum_{\ell=0}^{L} \alpha_\ell(t) u_\ell(t) \end{aligned}$$

where $u_0(t) = 1$.

We want the residual forcing functions $f_i(t) = Lx_i(t)$ to be of minimum norm,

$$\min_{\alpha_\ell, \beta_j} \{ \sum_{i=1}^{N} \int [Lx_i(t)]^2 \, dt \}.$$

Principal Differential Analysis (PDA)

PDA is one the more useful tools in functional data analysis. It estimates time-varying coefficients (here $\alpha(t)$ and $\beta(t)$) in systems of linear differential equations by minimizing the size of the forcing functions.

We use basis function expansions for each of the coefficient functions to be estimated. Using a common basis system, such as B-splines, with basis functions $\phi_k(t)$ that may vary in number and specification from one expansion to another, we have

$$\alpha_\ell(t) = \sum_k^{K_\ell} a_{\ell k}\phi_{\ell k}(t)$$

$$\beta_j(t) = \sum_k^{K_j} b_{jk}\phi_{jk}(t)$$

Combining Data Fitting with PDA

PDA estimates the differential operator L assuming that the functions $x_i(t)$ and their derivatives are already available.

Can we go straight from the discrete and noisy data to estimating L?

Let $y_j, j = 1 ..., n$ be noisy data and consider the least squares fit with a roughness penalty defined by L:

$$\text{PENSSE}_\lambda(x) = \sum_j^n [y_j - x(t_j)]^2 + \lambda \int [Lx]^2(t)\, dt.$$

Heckman and Ramsay (2000) considered optimizing PENSSE jointly with respect to $x(t)$ and parameters defining L.

Here we go further in alternating between

- L-spline smoothing to estimate $x(t)$ given L
- PDA to estimate the homogeneous part of L given $x(t)$

An Example

For $i = 1, \ldots, N; j = 1, \ldots, n$, let

$$y_{ij} = c_{i1} + c_{i2}t_j + c_{i3}\sin(6\pi t_j) + c_{i4}\cos(6\pi t_j) + \epsilon_{ij}$$

where the c_{ik}'s and the ϵ_{ij}'s are i.i.d. $N(0, 1)$; and $t = 0(0.01)1.$

The functional variation satisfies the DIFE

$$Lx = (6\pi)^2 D^2 x + D^4 x = 0$$

$\beta_0(t) = \beta_1(t) = \beta_3(t) = 0$ and $\beta_2(t) = 355.3.$

$$y(t) = c_1 + c_2 t + c_3 \sin(6\pi t) + c_4 \cos(6\pi t)$$

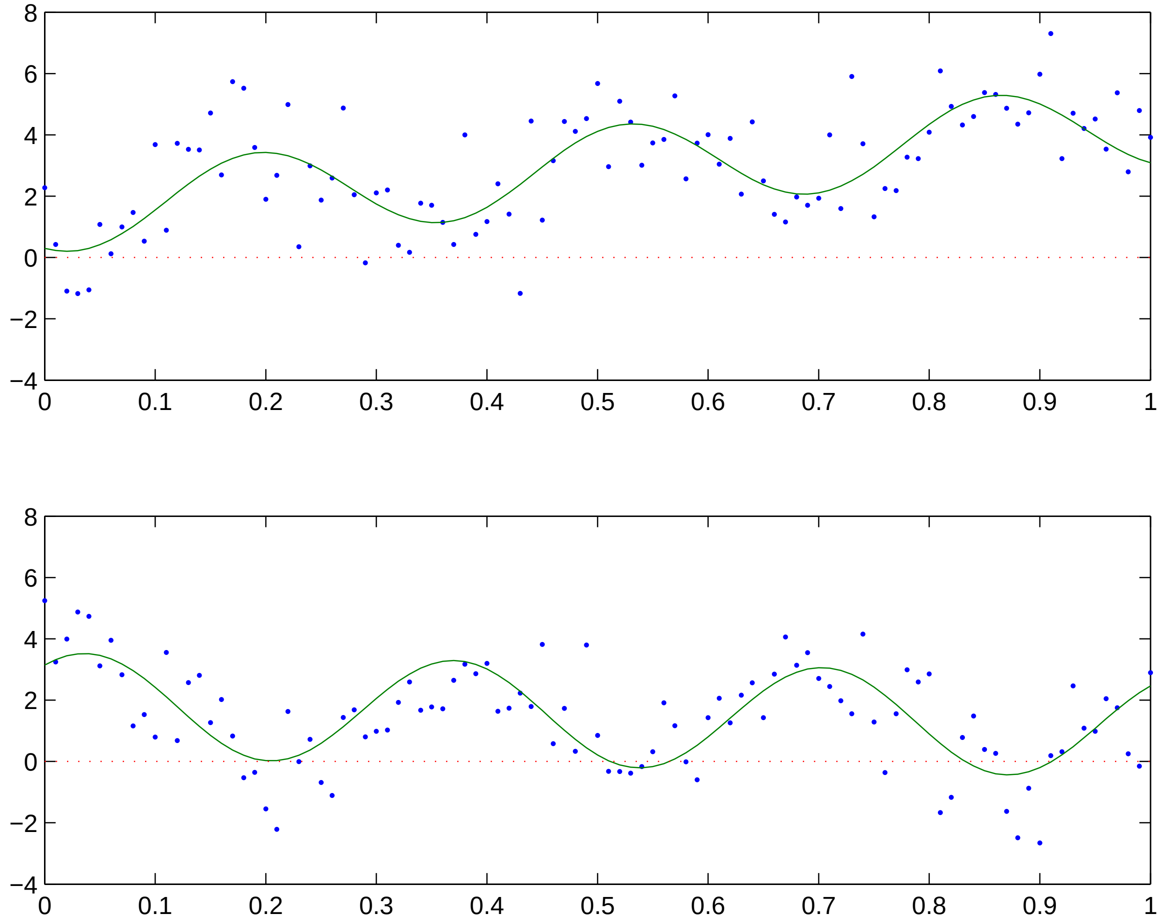

Figure 0.2:

Results

For a simulated data set with $N = 20$ curves, and using a constant basis for $\beta_0(t), \ldots, \beta_4$, we get

- for $L = D^4$, best results are for $\lambda = 10^{-10}$ and the RMISE's for derivatives 0, 1, and 2 are 0.32, 9.3, and 315.6, respectively.
- for L estimated, best results are for $\lambda = 10^{-5}$ and the RMISE's for derivatives 0, 1, and 2 are 0.18, 2.8, and 49.3, respectively,
- giving precision ratios of 1.8, 3.3, and 6.4, respectively.

β_2 is estimated to be 353.6, whereas the true value is 355.3.

β_3 is estimated to be 0.1, whereas the true value is 0.0.

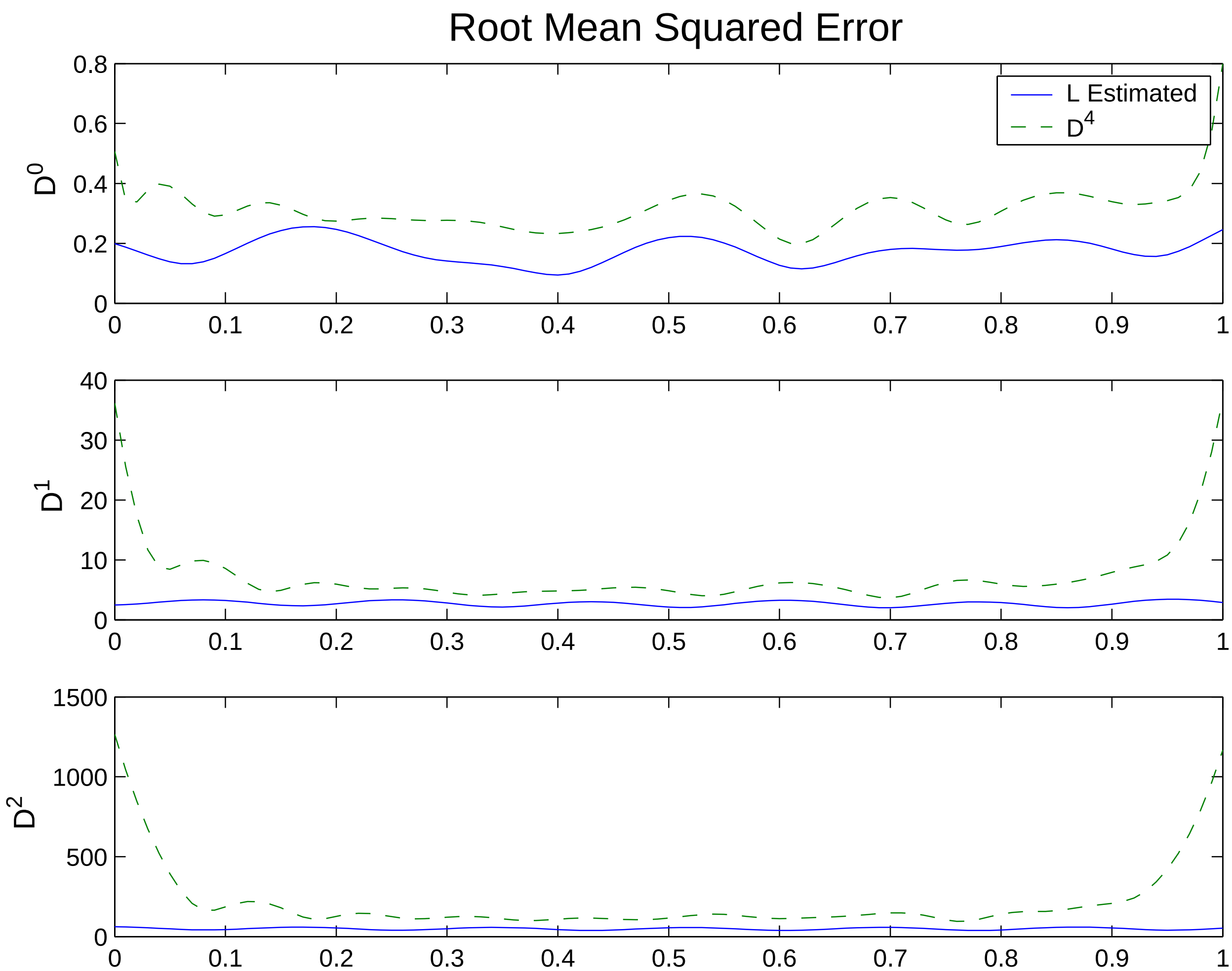
Root Mean Squared Error
L Estimated
D4
D0
D1
D2
0.8
0.6
0.4
0.2
0
40
30
20
10
1500
1000
500
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
1

Figure 0.3:

Models for the Refinery Data

We will consider the first order differential equation

$$Dx(t) = \beta(t)x(t) + \alpha(t)u(t) + \alpha_0 + \nu(t)$$

The constant α_0 compensates for changes in overall level of either variable, and only uses one degree of freedom.

We use two types of estimates for the weighting functions $\alpha(t)$ and $\beta_j(t)$:

- just a constant value, not varying with t. This is the usual practice in control engineering.
- a B-spline expansion of each coefficient using 5 B-spline basis functions of order 4. This allows the weight functions to vary, but rather slowly.

We will also include $\alpha(t) = 0$..

Goodness of Fit

We use an R^2 measure of fit

$$R^2 = \frac{\int (D^m x)^2 - \int (\nu^2)}{\int (D^m x)^2}$$

where m is the order of the equation.

How do we know when we are over-fitting the data?

The degrees of freedom in what is fit are the number of basis functions used to represent the response $x(t)$. In this case 30.

The degrees of freedom in the model are the total number of basis functions used to represent the coefficients. For a first order equation with 5 basis functions per coefficient and a constant, this is 11.

R^2 Results

$$Dx(t) = \beta(t)x(t) + \alpha(t)u(t) + \alpha_0 + \nu(t)$$

$\alpha(t)$	β constant	β varying
absent	0.34	0.66
constant	0.81	0.81
varying	0.81	0.84

It looks like

- adding the effect of the forcing function, reflux flow, improves the model substantially
- but little is gained by allowing either of the coefficients to vary.

The best model is

$$Dx(t) = -0.0204x(t) - 0.186u(t).$$

Conclusions

- We can estimate a linear DIFE that represents discrete noisy data. This involves optimizing L-spline smoothing with respect to the operator L.
- When we estimate the DIFE, the accuracy of the estimates of the functions and their derivatives improve.