An Overview of the Functional Linear Model

Jim Ramsay
McGill University

Overview

- 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.

Outline

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.

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

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

A scalar response and a functional independent variable

- The response is the log total annual precipitation

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

- The model is

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

- 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.

A functional response and a functional independent variable

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

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

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

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

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

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

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

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

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

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

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.*

What exactly makes a model linear?

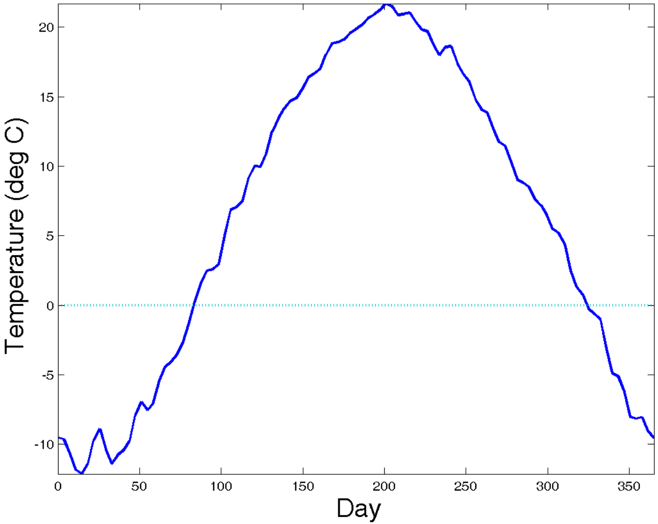
- We see that the functional linear model has a lot more variants than it's poor multivariate cousin.
- We will need to look at a definition of a linear model that encompasses these models and many others.

Outline

Predicting temperature curves from climate zones

- We have 35 weather stations distributed across four climate zones:
  - Atlantic (16)
  - Pacific (6)
  - Continental (13)
  - Arctic (4)
- The dependent variable is $\texttt{Temp}(t)$, a function representing daily temperatures averaged over 1960–1994.
- The temperature functions were obtained by expanding the original 365 discrete daily averages in terms of 65 Fourier basis functions.

Montreal's temperature profile



The functional ANOVA model

- The model is

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

- μ is the grand mean function
- α_g are the specific effects on temperature of being in climate zone g. To be able to identify them uniquely, we require that they satisfy the constraint

$$\sum_g \alpha_g(t) = 0 \text{ for all } t. \tag{1}$$

- ϵ_{mg} is the residual function showing unexplained variation specific to the kth weather station within climate group g.

Setting up the model

- Set up a 35 by 5 matrix $\mathbf{Z}$. Column 1 contains all 1's, and columns $g+1, g = 1, \ldots, 4$ contain zeros except for 1's in rows corresponding to stations in climate zone g.
- Append a final row with 0 in column 1, and 1's in the remaining columns.
- Let the functional response vector $\mathbf{Temp}(t)$ contain the 35 temperature profiles *plus* a final function that is zero for all t.
- Let functional regression coefficient vector $\beta(t)$ contain the functions $(\mu, \alpha_1, \ldots, \alpha_4)$.
- The model in matrix notation, including the zero sum constraint, is

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

Fitting the model

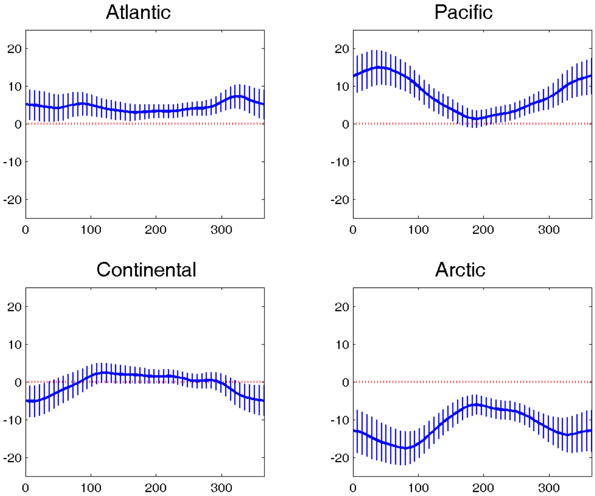
- The residual $\mathtt{Temp}_i(t) - \mathbf{Z}_i\beta(t)$ is now a function.
- The least squares fitting criterion becomes

$$\mathtt{LMSSE}(\beta) = \sum_g^4 \sum_m^{N_g} \int [\mathtt{Temp}_{mg}(t) - \sum_j^q z_{(mg),j}\beta_j(t)]^2 \, dt.$$

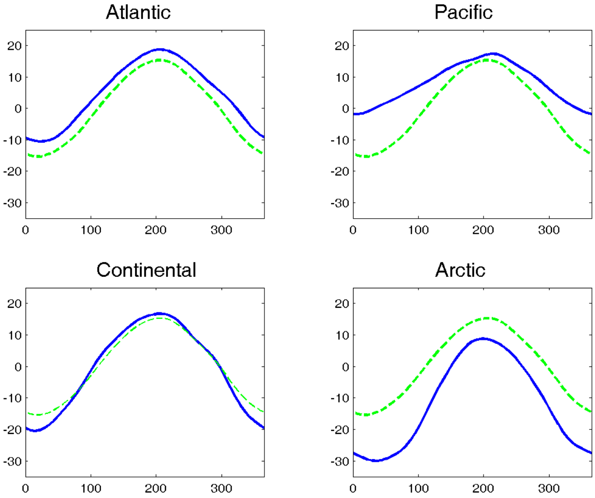
- This is minimized with respect to the regression functions by

$$\hat{\beta(t)} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Temp}(t)$$

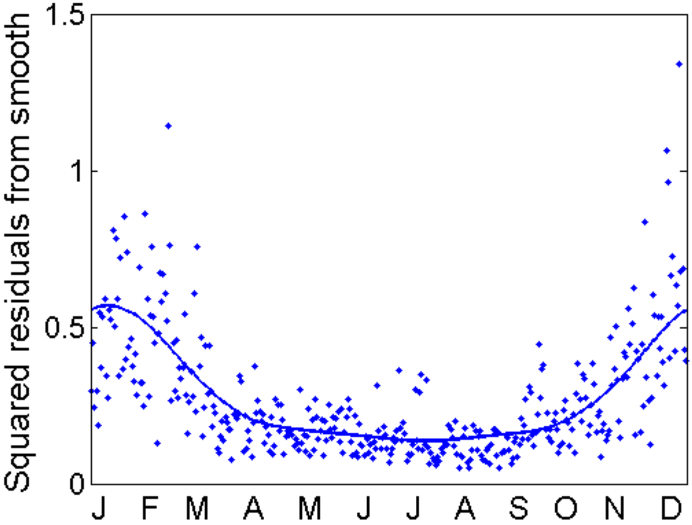
The region effects $\alpha_g(t)$



The mean plus region effects $\mu(t) + \alpha_g(t)$



The standard error of measurement function $\sigma(t)$



Assessing fit

- Is there significant variation in temperature over climate zones? Of course there is! This does not seem like an interesting question.
- On the other hand, whether the Atlantic, Pacific and Continental stations are significantly different in the summer might be.
- Interesting summaries of fit, of effects, and inferences are likely to be *local* in nature.

- It is useful to use the error sum of squares function

$$\mathtt{SSE}(t) = \sum_{mg} [\mathtt{Temp}_{mg}(t) - \mathbf{Z}_{mg}\hat{\beta}(t)]^2.$$

  to assess fit at or near time t.
- As in ordinary regression, we can compare this to the variation of the response about its mean

$$\mathtt{SSY}(t) = \sum_{mg} [\mathtt{Temp}_{mg}(t) - \hat{\mu}(t)]^2$$

- The corresponding mean squared error functions are

$$\mathtt{MSE}(t) = \mathtt{SSE}(t)/\mathtt{df}(\mathrm{error})$$

$$\mathtt{MSR}(t) = \frac{\mathtt{SSY}(t) - \mathtt{SSE}(t)}{\mathtt{df}(\mathrm{model})}$$

Multiple correlation and F-ratio functions

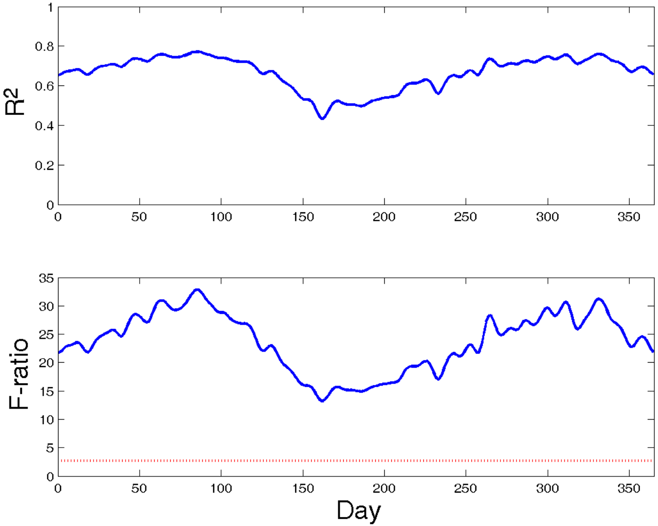
- The squared multiple correlation function is

$$\mathtt{RSQ}(t) = [\mathtt{SSY}(t) - \mathtt{SSE}(t)]/\mathtt{SSY}(t).$$

- and the F-ratio function is

$$\mathtt{FRATIO}(t) = \frac{\mathtt{MSR}(t)}{\mathtt{MSE}(t)}.$$

R^2 and F–ratio plots



Estimating the regression functions $\beta_j(t)$

- We want a general framework for estimating functional parameters in this and other linear models.
- We want to be able to penalize the roughness of any parameter β_j.
- We also want the capacity to estimate confidence intervals for a parameter,
- and for functionals $\rho(\beta_j)$ of a parameter.

Some basis function expansions for $\beta_j(t)$

- Let the regression coefficient vector $\boldsymbol{\beta}(t)$ have the expansion

$$\boldsymbol{\beta}(t) = \mathbf{B}\boldsymbol{\theta}(t)$$

  where matrix $\mathbf{B}$ is q by K_β and the K_β basis functions $\theta_\ell(t)$ are contained in vector $\boldsymbol{\theta}(t)$.
- In the temperature example, it would be natural to use a certain number K_β of Fourier basis functions.

A roughness penalty for $\beta_j(t)$

- If the response curves in $\mathbf{y}(t)$ are rough, we may want to impose some smoothness on the estimated β_j's.
- Let L be a linear differential operator, such as $L = D^2$, that defines variation $L\boldsymbol{\beta}(t)$ that we wish to penalize.
- Our roughness penalty on $\boldsymbol{\beta}(t)$ is

$$\mathtt{PEN}(\beta) = \int [L\boldsymbol{\beta}(s)]'[L\boldsymbol{\beta}(s)]\, ds \ .$$

Functional probes or contrasts

- Estimating the entire regression function $\beta_j(t)$ is fine, but we want to focus our attention on local or specific shape features of $\beta_j(t)$.
- Perhaps, for example, we want to examine the behavior of the temperature coefficient functions in mid-winter.
- A functional *probe* or *contrast* is of the form

$$\rho(\beta) = \int \xi(s)\beta_j(s)\, ds$$

- $\xi(s)$ is a weight function that we choose so as to concentrate our attention on a local region, or to look for specific patterns of variation in $\beta_j(t)$.
- There no particular need for $\xi(s)$ to integrate to 0.

Some examples

- *Point evaluation*:

$$\xi(s) = \delta(s - t)$$

  This simply produces the function value $\beta(t)$.

- *Local behavior*: Assuming that β is periodic, we can use

$$\xi(s) = \exp[(s - t)^2/(2\sigma)]$$

  to assess the behavior of β in a neighborhood of t of a size determined by constant σ.

How do I work out confidence limits for these probes?

- The random element in a linear model is the residual function value
$$\epsilon_i(t_j) = y_{ij} - x_i(t_j).$$
- Any linear function of the data inherits it's variance from the variance of the data.
- The variance of the data conditional on the model is the variance of the residuals.

- We have two tasks:
  - Estimate the variance of the residuals for a single response. (The mean can usually be taken to be 0.) Let's call this $\boldsymbol{\Sigma}_e$.
  - Assuming independence of the observations, the variance of the whole response data matrix is

$$\texttt{Var}[\text{vec}(\mathbf{Y})] = \boldsymbol{\Sigma}_e \otimes \mathbf{I}.$$

  - Work out the linear mapping from the data to the probe $\rho(\beta_j)$ that is being estimated. Let us call this $\mathbf{M}_j$.
- The rest is easy:

$$\texttt{Var}[\rho(\beta_j)] = \mathbf{M}_j'(\boldsymbol{\Sigma}_e \otimes \mathbf{I})\mathbf{M}_j$$

Some cautionary notes

- These sampling variances would only be “exact” if we knew Σ_e. The value of our confidence limit estimates depends critically on the quality of the estimate of Σ_e. There are many open questions about how to do this.
- We are assuming that the distribution of a probe is well summarized by its mean and variance.
- Our estimates are all conditioned on how many basis functions we use for both $y_i(t)$ and $\beta_j(t)$, namely K_y and K_β. Since we never know exactly how many to use, these should be regarded as random quantities, and a Bayesian treatment seems to be indicated.
- We should back up the use of these “delta method” confidence regions by bootstrapping and simulations.

Summary

- Regressing a functional response on multivariate independent variables or on a design matrix is not much different from the conventional regression analysis.
- One important difference is that we want to do *local* inference and interval estimation.
- We have, too, the capacity to smooth estimated functional parameters.
- But the number of basis functions that we use is not a fixed parameter in the traditional sense.

Outline

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

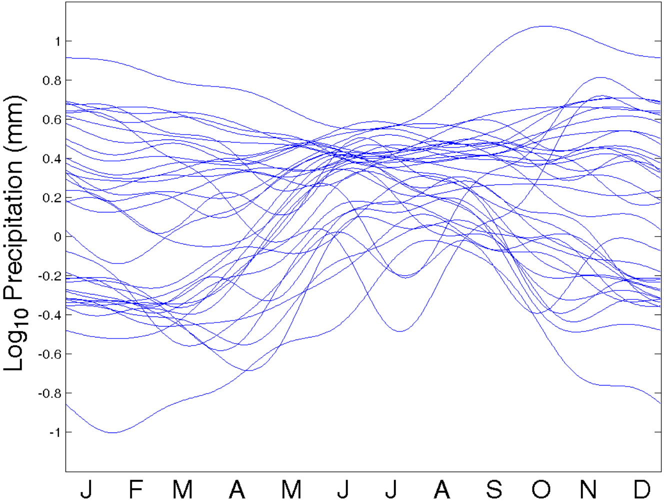
$$\log[\texttt{Prec}_{mg}(t)] = \mu(t) + \alpha_g(t) + \texttt{TempRes}_{mg}(t)\beta(t) + \epsilon_{mg}(t)$$

- 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



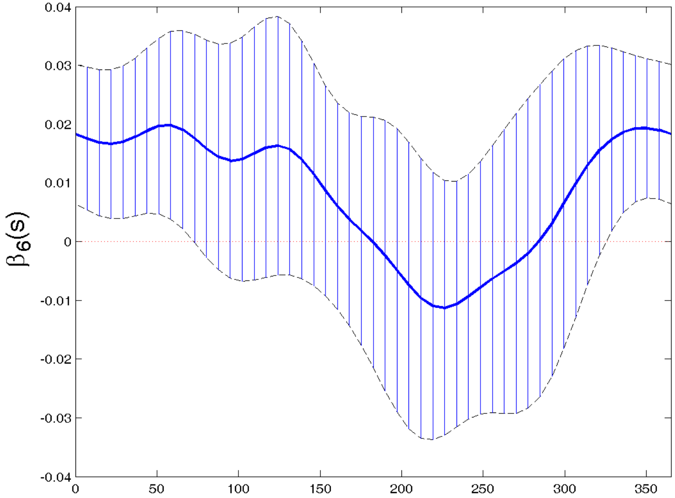
The fitting criterion and some results

- The fitting criterion was the unpenalized error sum of squares

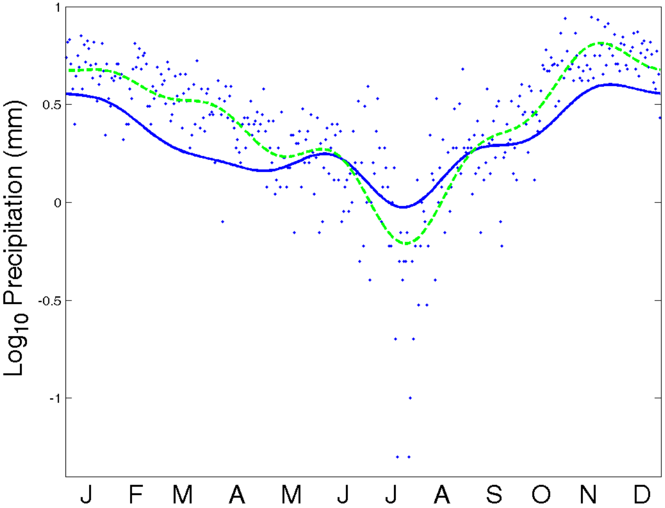
$$\begin{aligned} \texttt{LMSSE}(\mu, \alpha_g, \beta) = \\ \int \sum_{k,g}^{N} [\texttt{LogPrec}_{kg}(t) - \mu(t) - \alpha_g(t) \\ - \texttt{TempRes}_{kg}(t)\beta(t)]^2 \, dt \end{aligned}$$

- 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



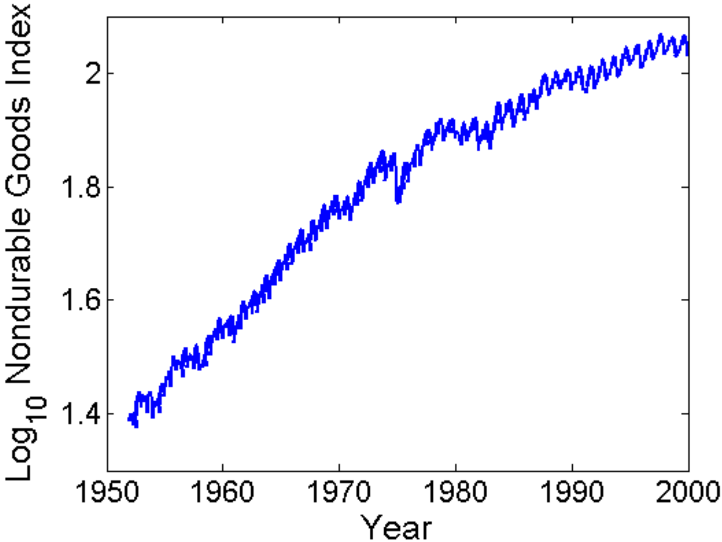
A probe for the winter effect

- 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:

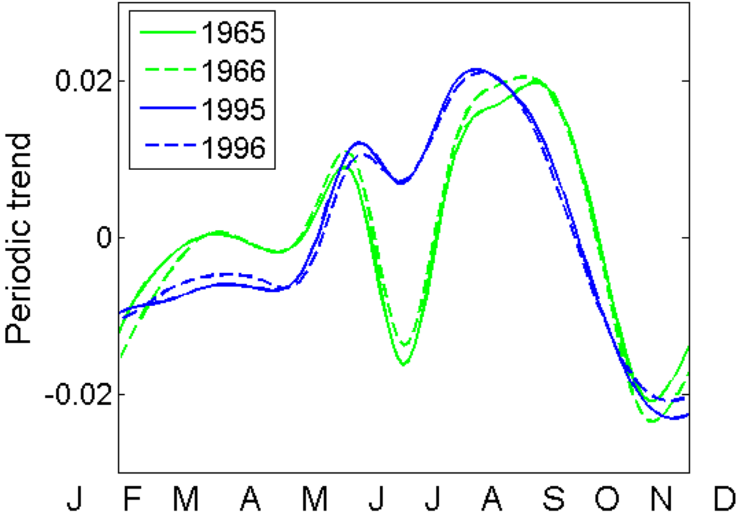
$$\rho(\beta) = \int_0^{365} \cos[2\pi(t - 64.5)/365]\beta(t)\,dt = 2.32 \ ,$$

- 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.

Evolution in seasonal trend for the nondurable goods index



Four seasonal trends



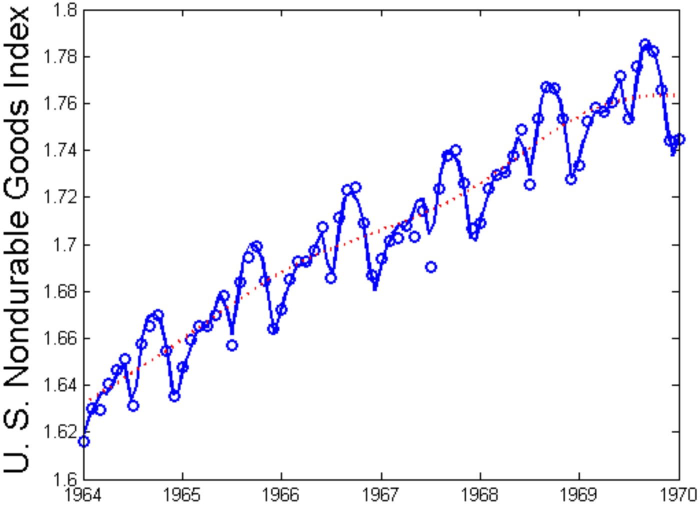
- Seasonal trends are stable over a couple of years, but evolve over a longer time span.
- We can model nonseasonal trend plus an evolving seasonal trend as follows:

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

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

- 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



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.

Outline

- With functional responses and multivariate independent variables we could estimate the regression coefficient functions without necessarily needing to use roughness penalties.
- The same with functional responses, functional independent variables and the concurrent model.
- Now we look at a scalar response predicted by a functional independent variable, and discover that a roughness penalty or *regularization* is indispensable.

A model for total annual precipitation

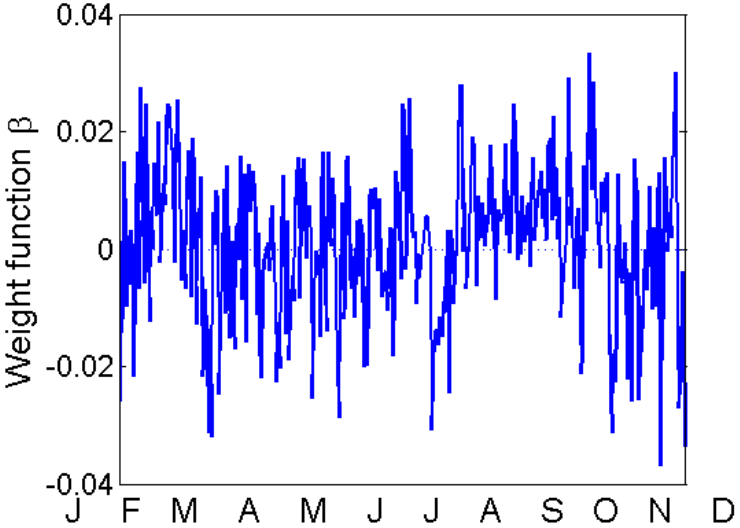
- Let $y_i = \texttt{LogPrec}_i$ be the logarithm of total annual precipitation at weather station i.
- Here is our model:

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

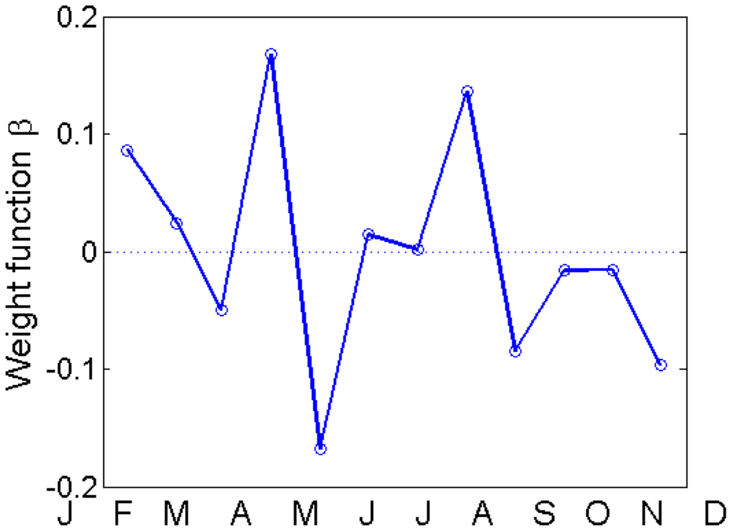
- We can think of the function values $\texttt{Temp}(s)$ associated with each fixed s as a separate scalar independent variable.
- If so, we have enough fitting power at our disposal to fit any number of responses, and certainly only 35 of them.

A bad idea

- If we use the discrete daily temperature averages, we have 365 plus 1 for constant α independent variables to fit 35 responses.
- Using the Moore-Penrose generalized inverse to keep us out of trouble, we get the following estimate of β.


0.04
0.02
0
-0.02
-0.04
Weight function β
J F M A M J J A S O N D

Using only monthly values doesn't help much.



Estimating $\beta(s)$ with a roughness penalty

- We could impose smoothness on $\beta(s)$ by expanding it in terms of a small number (< 35) of basis functions.
- Using a roughness penalty, however, gives us continuous control over smoothness and other advantages.
- Here is the penalized least squares criterion:

$$\begin{aligned} \mathtt{PENSSE}_\lambda(\alpha, \beta) &= \sum_{i=1}^{N} [y_i - \alpha - \int z_i(s)\beta(s)\, ds]^2 \\ &+ \lambda \int [L\beta(s)]^2\, ds \ , \end{aligned}$$

Choosing a roughness penalty

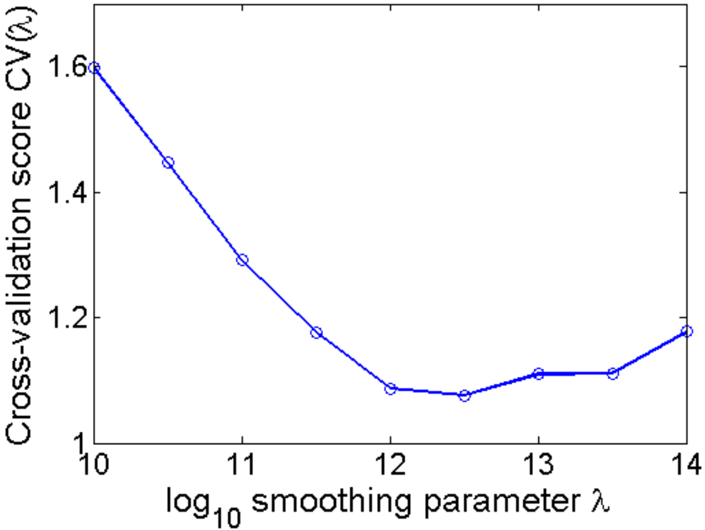
- Penalize *harmonic acceleration* because $\beta(s)$ is periodic:

$$L\beta(s) = (\frac{2\pi}{365})^2 D\beta(s) + D^3\beta(s)$$

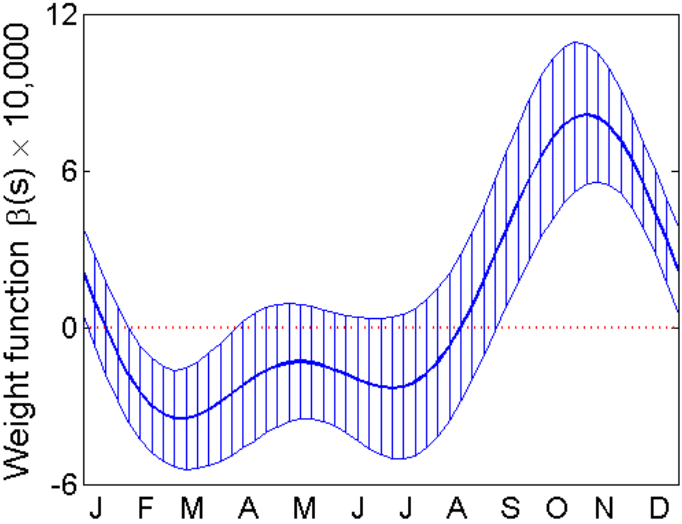
- We choose the smoothing parameter λ by minimizing the cross-validation criterion.
- Let $\alpha_\lambda^{(-i)}$ and $\beta_\lambda^{(-i)}$ be the estimates using all the responses except y_i.
- The criterion to be minimized is

$$\mathtt{CV}(\lambda) = \sum_{i=1}^{N} [y_i - \alpha_\lambda^{(-i)} - \int z_i(s)\beta_\lambda^{(-i)}(s)\, ds]^2$$

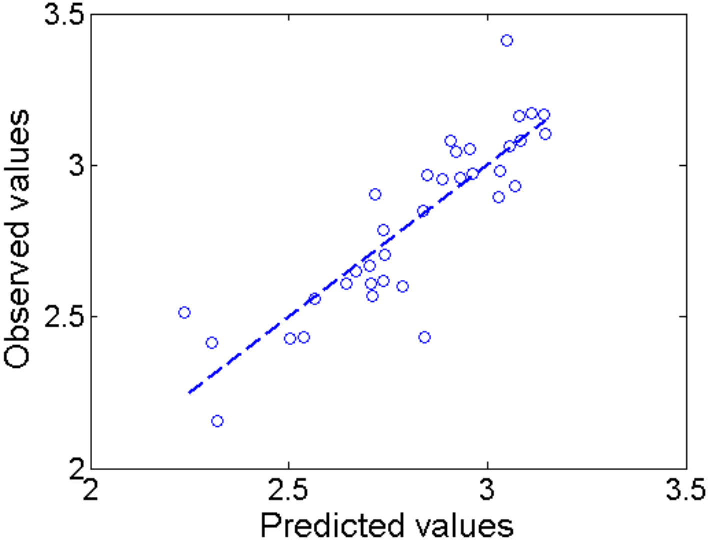
A plot of $\mathrm{CV}(\lambda)$



$\beta(s)$ for $\log_{10} \lambda = 12.5$



A plot of the fit



Summary

- Either dimension reduction or regularization is essential when the dimensionality of the covariate exceeds the dimensionality of the response.
- Functional covariates for scalar responses has been heavily researched.
    - The group STAPH at U. Toulouse uses functional PCA to reduce covariate dimensionality. Go to `www.lsp.ups-tlse.fr/FP/Ferraty/staph.html` to learn more.
    - Gareth James, the group at the U. of Granada, and Doug Clarkson at Insightful Corp. have introduced functional covariates into the generalized linear model.

Outline

Overview

- The fundamental question is this: If we have a covariate $z(s)$, how much of its variation over s should we use to fit the response $y(t)$ at fixed t?
- In the simplest case, we only use $z(t)$. We call this the *concurrent model* because it predicts y at time t by z at time t. We might call this "now-casting."
- If we want to use the behavior of z over an interval of values s, or over all values, things are more complex because this model has, effectively, an infinite amount of fitting power.

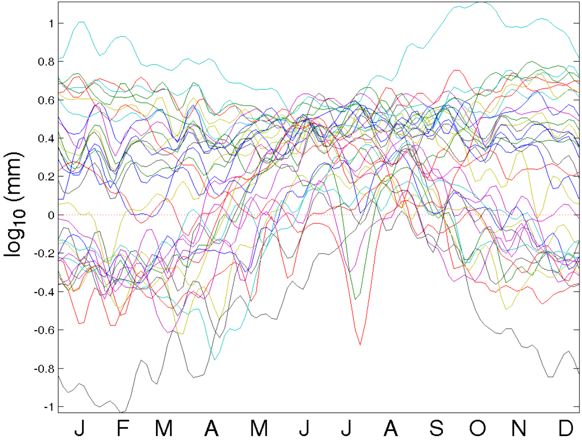
The full model for log precipitation

- We predict the log precipitation profile $\texttt{LogPrec}_i(t)$ at time t from the entire temperature profile $\texttt{Temp}_i(s)$.
- The fitting criterion is

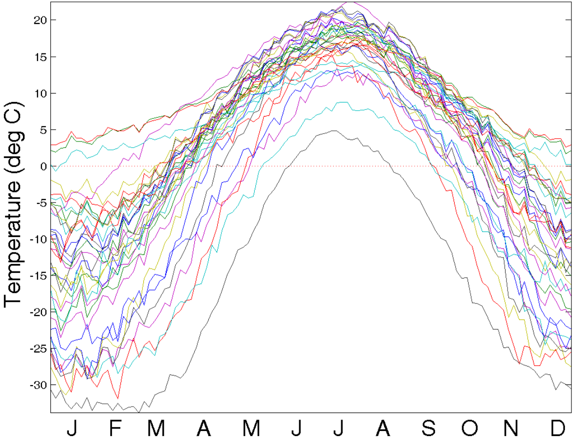
$$\texttt{LogPrec}_i(t) = \alpha(t) + \int_0^{365} \texttt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t)\ .$$

- $\beta(s,t)$ indicates the influence of temperature at time s on precipitation at time t.
- We can use the whole temperature profile because the data are periodic.
- We have already learned from predicting total log precipitation that we will have to apply a roughness penalty to $\beta(s,t)$ as a function of s.
- What about its variation as a function of t?

Log precipitation functions

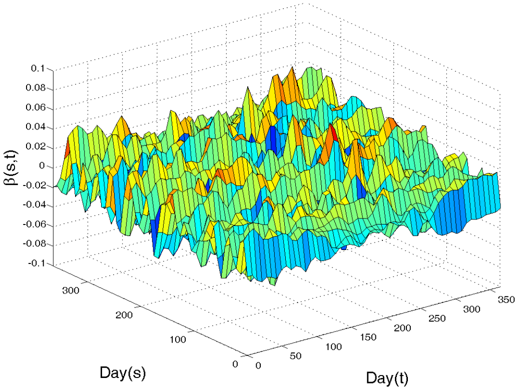


Temperature functions



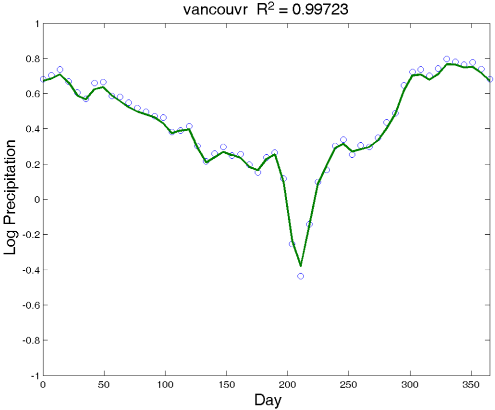
- We apply two harmonic acceleration roughness penalties to $\beta(s, t)$, one for its variation in s, and one for its variation in t.
- Let's see what happens with fairly light penalties on both types of variation.
- We'll look at $\beta(s, t)$ and at the fit to the log precipitation data for Vancouver.

$\beta(s, t)$ has light penalties on s and t



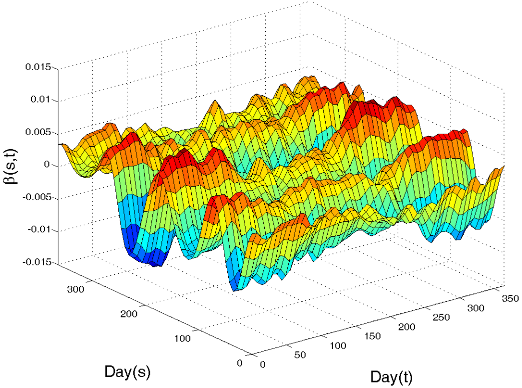
- $\beta(s, t)$ is impossible to interpret.

$\beta(s,t)$ has light penalties on s and t



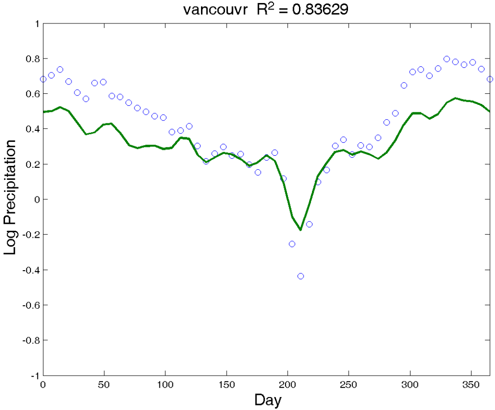
- And we seem to have over–fitted Vancouver's data.
- Let's increase the smoothing parameter for s.

$\beta(s,t)$ has heavy penalty on s and light on t



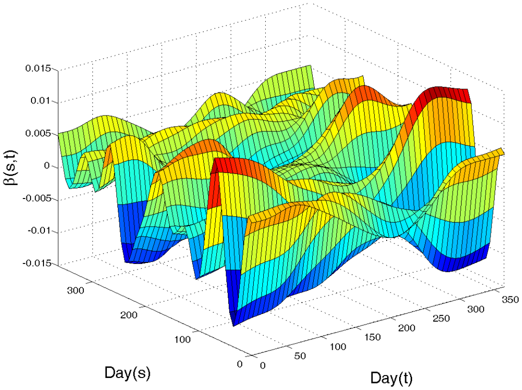
- $\beta(s,t)$ is interpretable as a function of s but impossible to understand in t.

$\beta(s,t)$ has heavy penalty on s and light on t



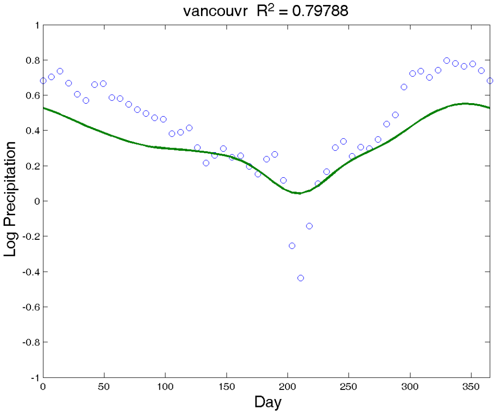
- We now have a more reasonable fit to Vancouver's data, but the fitting function is too rough.
- Let's increase smoothing parameters for both s and t.

$\beta(s,t)$ has heavy penalties on both s and t



- $\beta(s,t)$ is now smooth in both s and t.

$\beta(s, t)$ has heavy penalties on both s and t

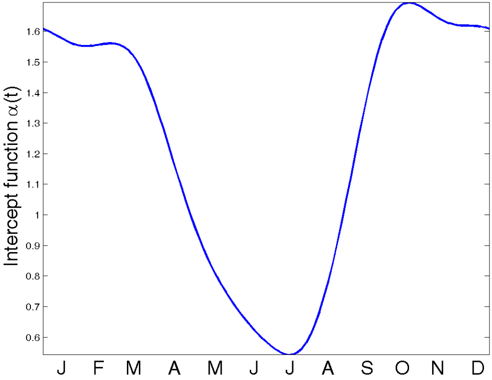


- The fit is reasonable and also smooth.

What we see

- Penalizing the roughness of $\beta(s,t)$ as a function of s prevents over-fitting.
- Penalizing the roughness of $\beta(s,t)$ as a function of t allows us to see how the influence of temperature on precipitation varies from one time to another.
- We can now see that temperature is much more influential in the winter than in the summer.
- The rapid oscillation in s suggests that it is a derivative of temperature that really influences precipitation.

The intercept function $\alpha(t)$



The historical model and other possibilities

- We were able to use all of $z(s)$ to predict $y(t)$ in the weather example because the data were periodic.
- In nonperiodic situations, it would only be meaningful to use the values of s *up to* t.

$$y(t) = \alpha(t) + \int_{t-\delta(t)}^{t} z(s)\beta(s,t)\, ds + \epsilon(s)$$

- For the lower limit of integration $t - \delta(t)$, the width $\delta(t)$ of the interval of integration can vary over t or can be constant.
- We can call this the *historical linear model*.
- See Ramsay and Silverman (2002) for an example.

More generally, we can integrate over a set Ω_t that varies with t, and also permit the covariate function z to vary over t as well,

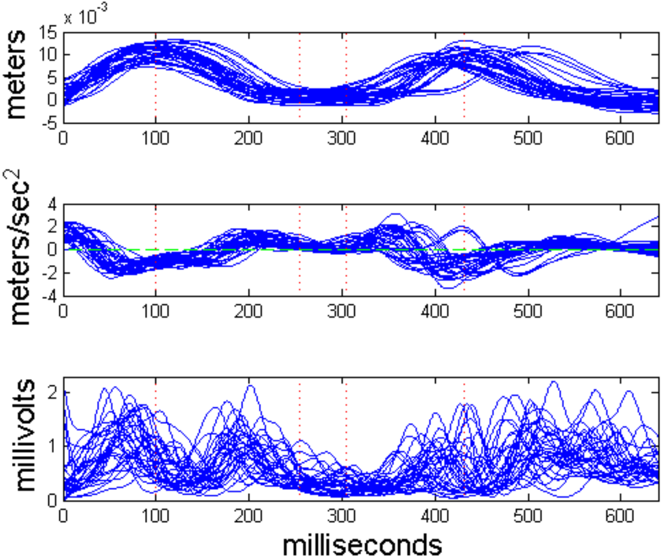
$$y(t) = \alpha(t) + \int_{\Omega_t} z(s,t)\beta(s,t)\,ds + \epsilon(s).$$

- How do we construct basis function systems for complex regions of integration?
- The triangular mesh algorithms used in finite element methods to solve partial differential equations are natural here.
- Triangular meshes adapt well to nonstandard boundaries.
- Triangular basis functions also can be regularized.
- The PDEtools toolbox in Matlab contains powerful functions for mesh generation.

Lip acceleration predicted from EMG signal

- Malfait and Ramsay (2004) studied how the acceleration of the lower lip while saying "bob" was related to the EMG signal record from the depressor lip muscle. (see also Ramsay and Silverman, 2002)
- Each of 32 replications lasted for about 0.7 seconds.
- EMG is an indirect indication of muscle activation.
- It was hoped to learn something about the brain controls speech production.

32 lip position, lip acceleration and EMG signals



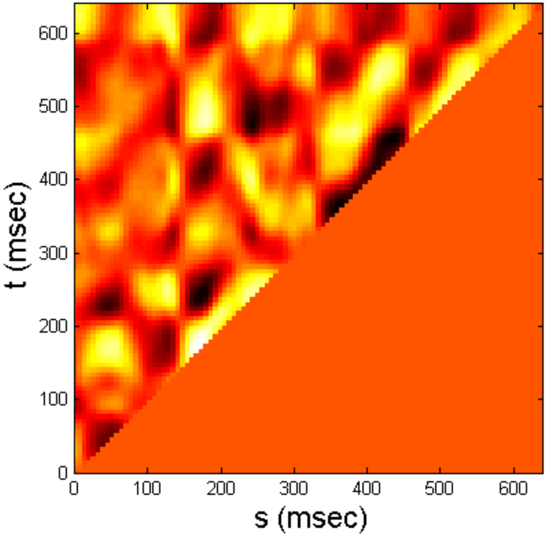
The feed–forward model

- Only feed–forward effects of EMG on acceleration are of interest.
- The model is

$$lip(t) = \alpha(t) + \int_{t-\delta}^{t} EMG(s)\beta(s,t)\,ds + \epsilon(s)$$

- Regression function $\beta(s,t)$ is defined over the triangular region $0 \leq s \leq t;\ 0 \leq t \leq 0.7$.
- How far back should we allow for an influence? How large should the lag δ be?

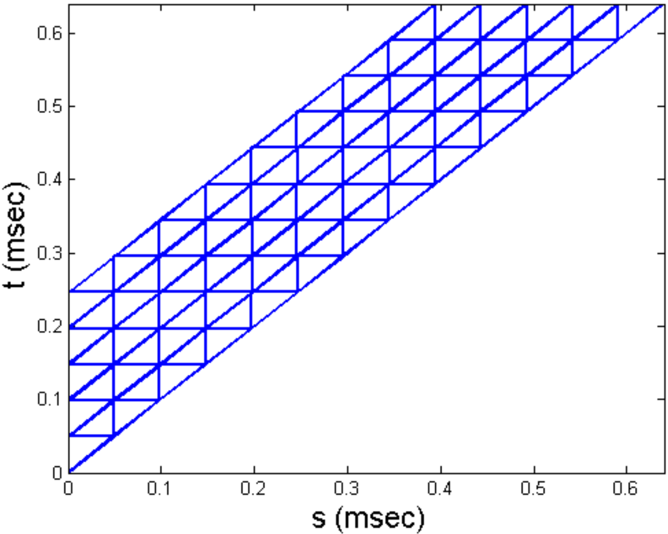
The correlation surface



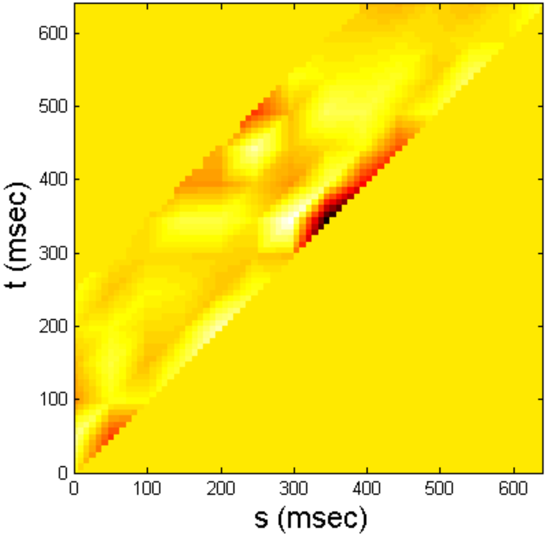
The finite element basis for $\beta(s, t)$

- The finite element basis for functions defined over two arguments uses piece–wise linear functions defined over a triangular mesh.
- Triangular meshes can easily adapt to complex boundaries. Our problem here is particularly easy.
- The coefficient matrices become more and more sparse as the number of triangles increases.
- Sparse matrix computation, available in Matlab, makes for fast solutions.

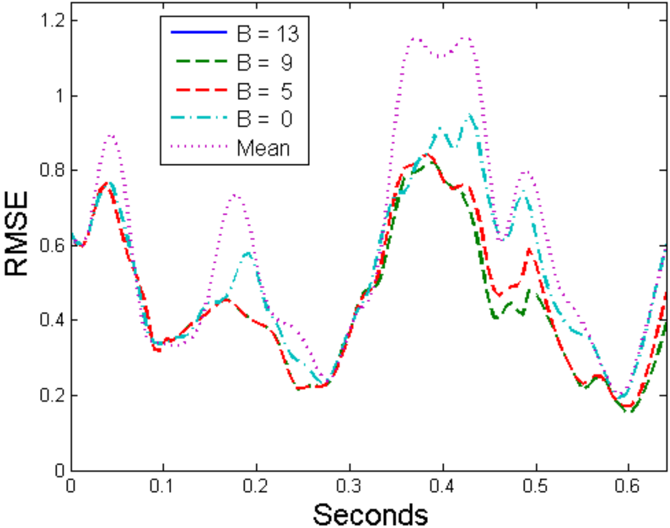
A triangular mesh for the lip/EMG problem



The regression function surface for lag $\delta = 5$ triangles



Standard error of estimate functions for various lags



Summary

- When both response and covariates are functional, there are a lot of modelling possibilities.
- Using the full variation in a functional covariate $z(s)$ is only useful if we regularize the solution, either at the level of the regression coefficient $\beta(s, t)$, or at the level of the response $y(t)$.
- Using the full variation in the covariate usually only makes sense for periodic problems.
- A feed–forward model is more likely for nonperiodic functional regressions.
- Estimating the amount of functional history to use is an important issue.
- There is a lot more work to do in this exciting area!

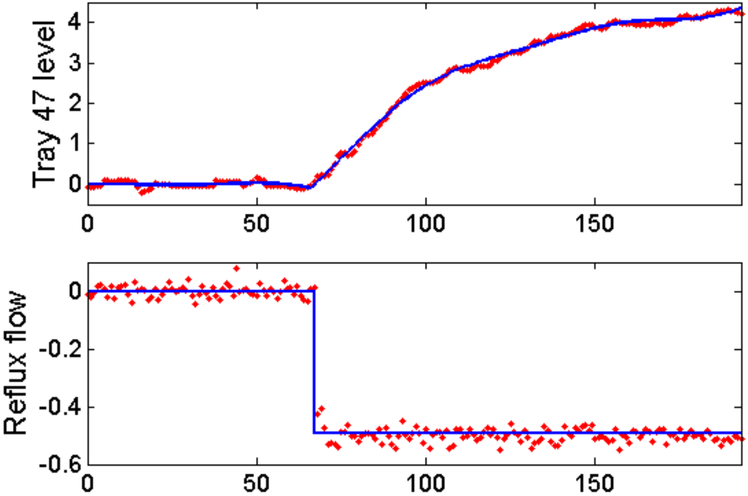
Where to we go from here?

- What if we want to use derivatives, $Dy(t)$ and $Dz(t)$, in our models?
- What about functional variables that interact with each other, such as we find in input/output systems with feedback?
- Nonlinear functional models?
- We have only tickled the ear of the elephant.

Outline

The oil refinery data

Refinery output $x(t)$ (top panel) and input $u(t)$ (bottom panel)



- This is a simple input/output system in an oil refinery in Corpus Christi, Texas.
- A fluid, called *reflux*, flows into *tray 47* in a distillation column in an oil refinery.
- The input variable $u(t)$ is the flow rate.
- The fluid level in the tray is the output variable $x(t)$.

Variation on two time scales

- Over the longer scale, tray level changes from 0 to around 4.
- But we are also interested in how rapidly the change takes place; that is, short-scale variation.
- It looks like about 2/3 of the change takes about 50 minutes, and the final value is reached in 200 minutes or so.

A concurrent functional regression

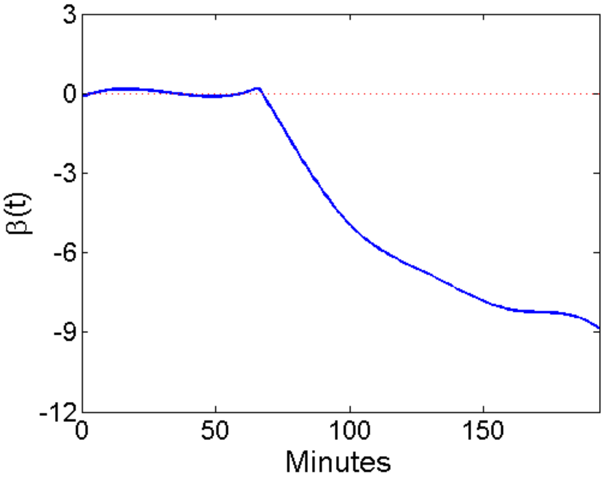
- We can model the output *state* $\texttt{Temp}(t)$ as a simple time-varying regression on input state $\texttt{Prec}(t)$:

$$\texttt{Temp}(t) = \texttt{Prec}(t)\beta(t) + \epsilon(t)$$

- In functional data analysis, we call this a *concurrent* regression because only the simultaneous influence of the input on the output is modelled.
- A least squares estimate $\hat{\beta}(t)$ of the regression coefficient function minimizes

$$\texttt{SSE} = \int [\texttt{Temp}(t) - \texttt{Prec}(t)\beta(t)]^2 \, dt$$

- $\hat{\beta}(t)$ is modelled with an expansion in terms of several B-spline basis functions.



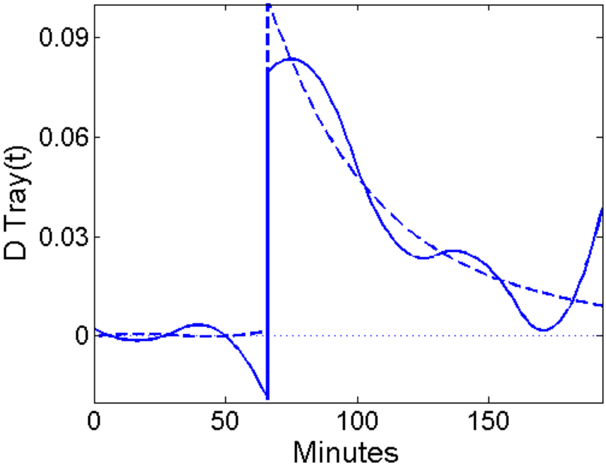
But does this tell us anything new? The regression function is just as complicated as the output.

Making the derivative $D\texttt{Temp}(t)$ the output

- We now model the *rate of change* $D\texttt{Temp}(t)$, using the output state $\texttt{Temp}(t)$ and the input $\texttt{Prec}(t)$ as covariates.
- We'll use constants for the two regression functions:

$$D\texttt{Temp}(t) = -\beta_1 \texttt{Temp}(t) + \beta_2 \texttt{Prec}(t) + \epsilon(t)$$

- This is an example of a *first order differential equation with constant coefficients*.
- We see that it is just another form of concurrent functional linear model, now with two covariates, but with regression functions β_1 and β_2 that are constant.

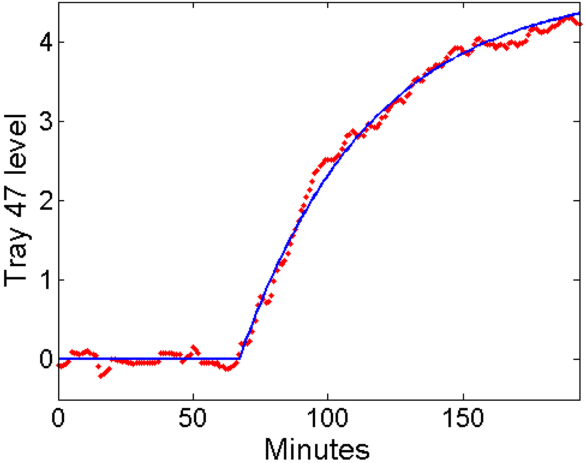


The solid line is the derivative estimated from the data, and the dashed line is the model's fit to this derivative. $\hat{\beta}_1 = 0.02$ and $\hat{\beta}_2 = 0.19$

- A differential equation that is this simple has an explicit solution.

$$x(t) = e^{-\beta_1 t}[x(0) - (\beta_2/\beta_1) \int_0^t e^{\beta_1 s} u(s)\, ds].$$

- $\beta_1 \approx 0.02$ is the rate constant, and therefore controls the rate of change of `Temp` level. About 2/3 of the change takes $1/\beta_1$ time units, and the final level is nearly reached in $4/\beta_1$ time units. β_1 models the *dynamic behavior* of `Temp`.
- $\beta_2 \approx 0.19$, along with β_1, defines the ultimate change; the long-term *gain* per unit increase in Reflux flow is $\beta_2/\beta_1 \approx 9.5$.

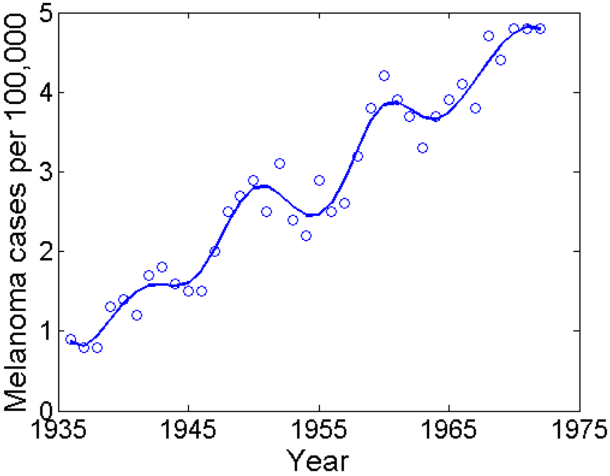


Modelling the rate of change $D\texttt{Temp}(t)$ directly produces a fine fit to the data with only two parameters.

The melanoma data

Age-adjusted melanoma incidences for Connecticut.

The solid line is a spline smooth with penalty on $D^4 x$.



Estimating a differential operator L

- A differential operator L is just a re-arranged differential equation.
- Can we smooth the data and estimate the operator at the same time?
- Will this give us a better fit to the data with fewer degrees of freedom used up?
- We will try the operator

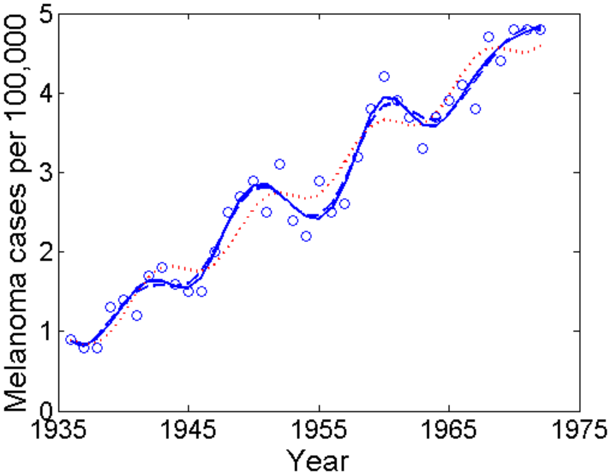
$$Lx = \beta_1 D^2 x + D^4 x$$

- This operator a tilted line plus sinusoidal trend with the period to be estimated from the data.

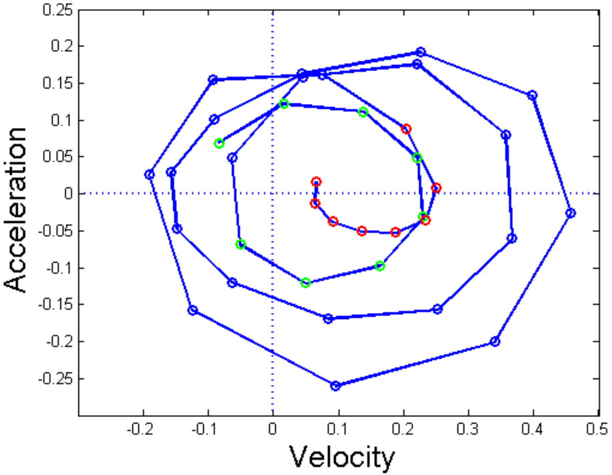
The algorithm

- Start with $\beta_1 = 0$ and $Lx = D^4 x$, and estimate derivatives up to order 4, choosing λ to minimize the GCV criterion.
- Carry out concurrent functional regression to estimate β_1.
- Re-smooth using Lx, again re-computing derivatives and minimizing GCV.
- Continue until the parameter estimates converge.

Solid blue line is smooth using converged Lx penalty. Dashed line is smoothing using D^4 penalty. Dotted line is a solution to differential equation $Lx = 0$.



Phase-plane plot



Summary

- By using a derivative as the dependent variable, we can model the rate of change of an output variable.
- The resulting differential equation can be solved to provide a model for the output variable itself.
- We effectively get two or more models for the price of one.
- We are simultaneously modelling the *level* or *state* of a system and its *dynamics*.