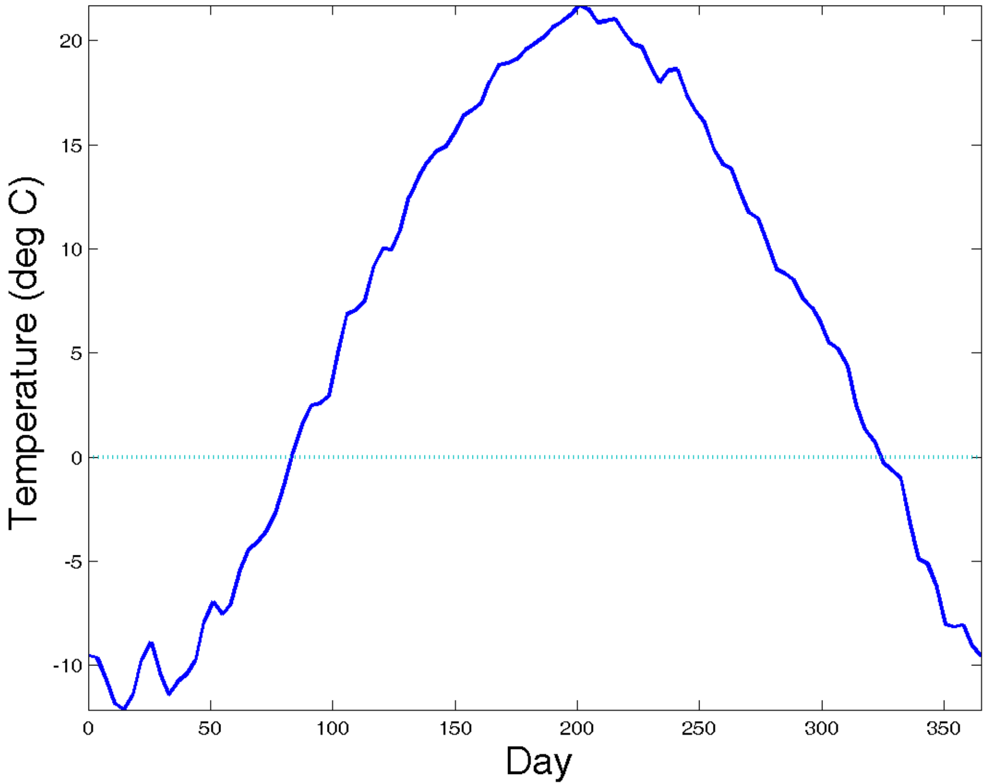
Modeling functional responses with multivariate covariates

1. 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.
- 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 $\boldsymbol{\beta}(t)$ contain the functions $(\mu, \alpha_1, \ldots, \alpha_4)$.
- The model in matrix notation,, including the zero sum constraint,is

$$\mathbf{Temp} = \mathbf{Z}\boldsymbol{\beta} + \boldsymbol{\epsilon},$$

Fitting the model

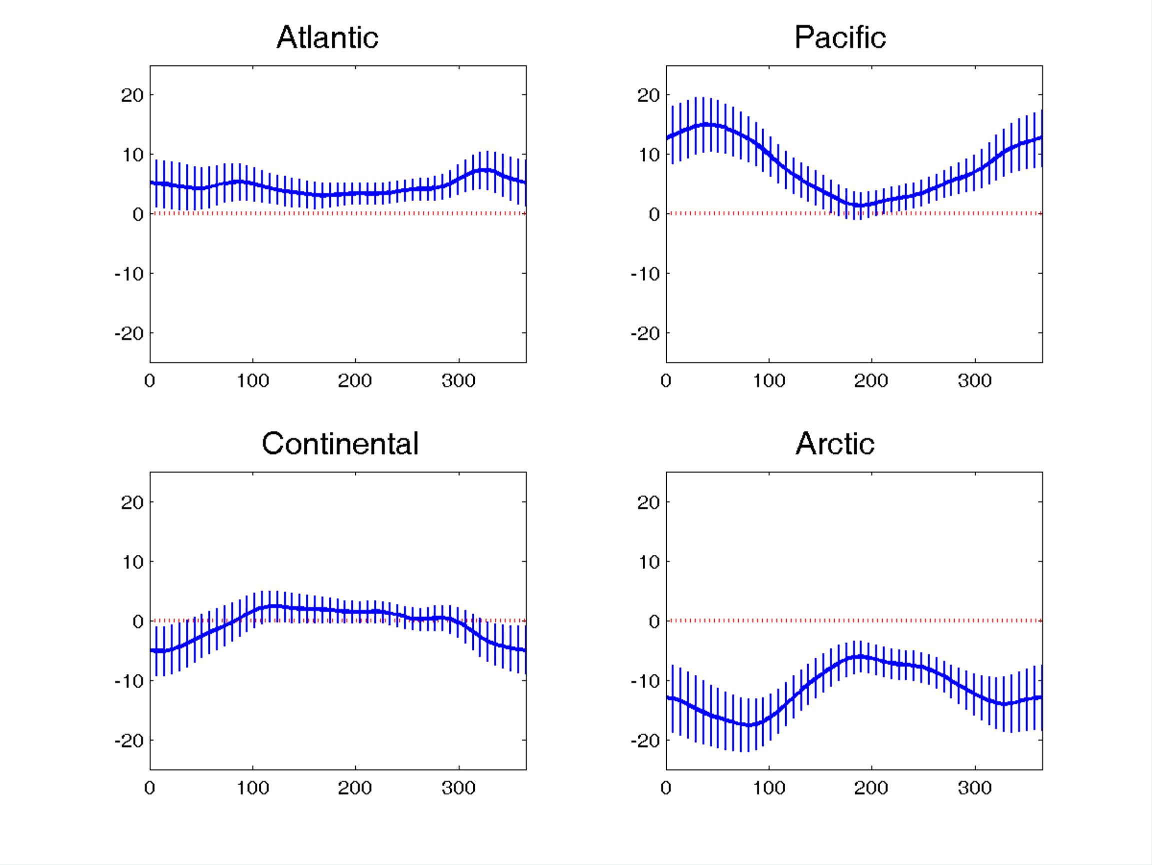
- The residual $\texttt{Temp}_i(t) - \mathbf{Z}_i\boldsymbol{\beta}(t)$ is now a function.
- The least squares fitting criterion becomes

$$\texttt{LMSSE}(\boldsymbol{\beta}) = \sum_{g}^{4}\sum_{m}^{N_g}\int [\texttt{Temp}_{mg}(t) - \sum_{j}^{q} z_{(mg),j}\beta_j(t)]^2\, dt.$$

- This is minimized with respect to the regression functions by

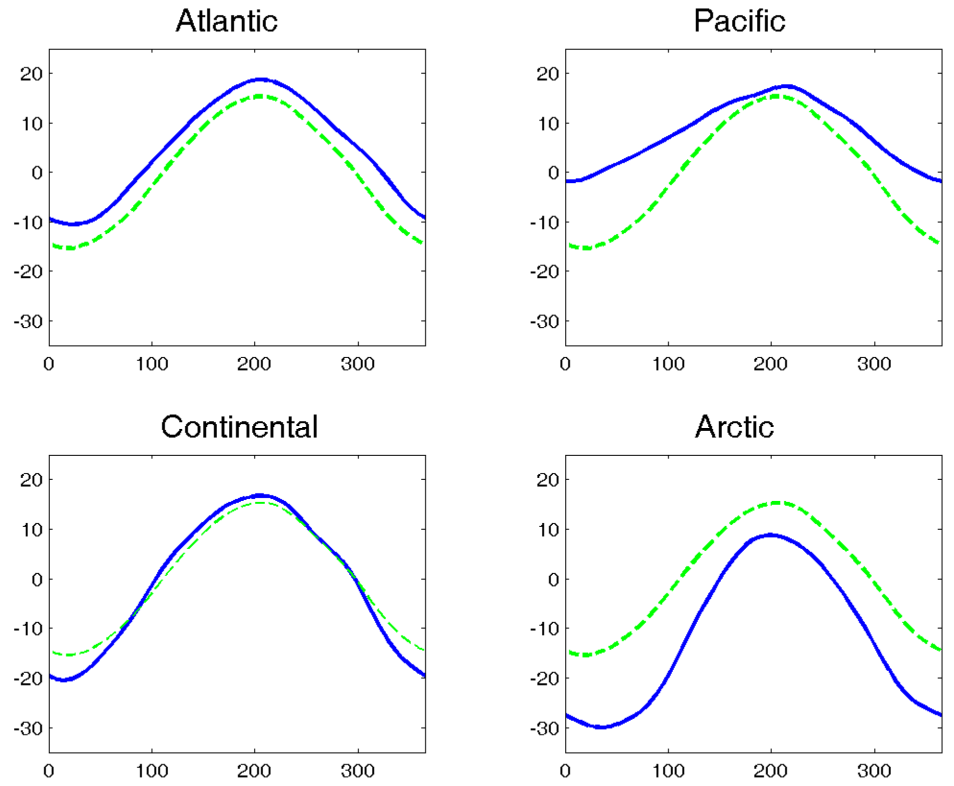
$$\hat{\boldsymbol{\beta}}(t) = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Temp}(t)$$

The region effects $\alpha_g(t)$



The mean plus region effects

$$\mu(t) + \alpha_g(t)$$



2. 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

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

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

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

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

Multiple correlation and F-ratio functions

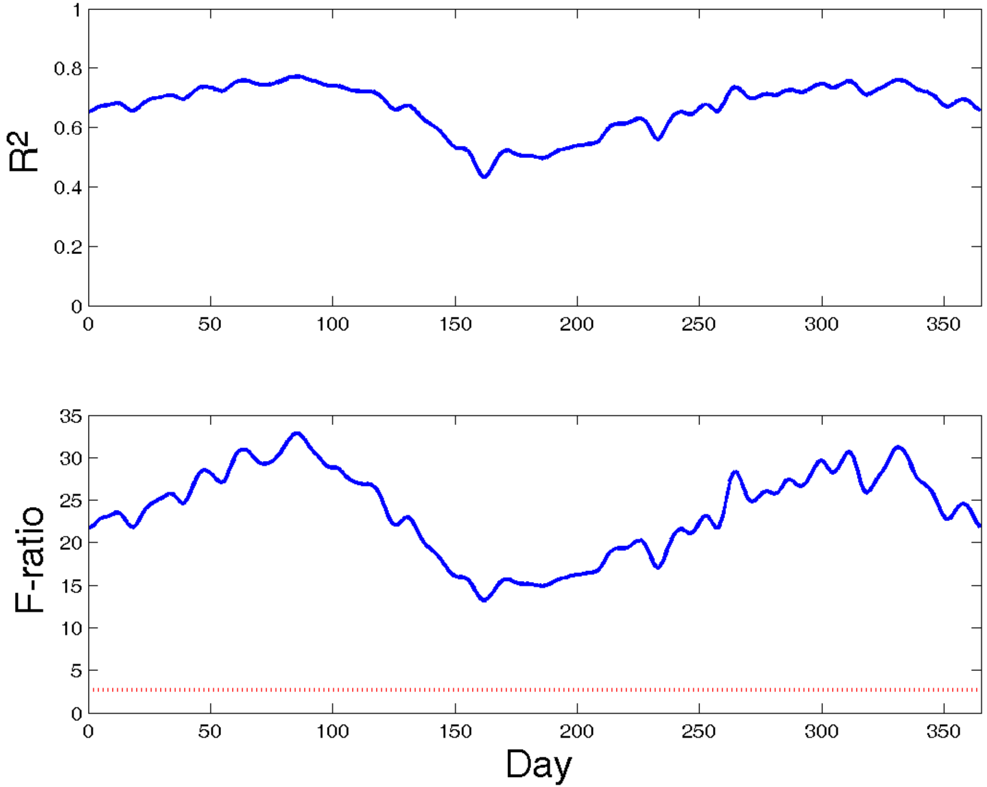
- The squared multiple correlation function is

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

- and the F-ratio function is

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

R^2 and F–ratio plots



3. Estimating the regression functions $\beta_j(t)$

- We want a general framework for estimating functional parameters in this and other linear models.
- A critical issue is being able to penalize the roughness of the parameter in a sense that we want to be able to define.
- We also want the capacity to estimate confidence intervals for the parameter or
- for functionals $\rho(\beta_j)$ of the 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

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

The penalized least squares criterion

- Let the response function vector $\mathbf{y}(t)$ have the basis function expansion in terms of K_y basis functions $\phi_k(t)$:

$$\mathbf{y}(t) = \mathbf{C}\boldsymbol{\phi}(t)$$

- Then the penalized least squares function is

$$\text{PENSSE}(y|\boldsymbol{\beta}) = \int (\mathbf{C}\boldsymbol{\phi} - \mathbf{ZB}\boldsymbol{\theta})'(\mathbf{C}\boldsymbol{\phi} - \mathbf{ZB}\boldsymbol{\theta}) + \lambda \int (L\mathbf{B}\boldsymbol{\theta})'(L\mathbf{B}\boldsymbol{\theta}) \,.$$

Penalized least squares in matrix terms

- we need to define these three matrices:

$$\mathbf{J}_{\phi\phi} = \int \boldsymbol{\phi}\boldsymbol{\phi}' \,, \quad \mathbf{J}_{\theta\theta} = \int \boldsymbol{\theta}\boldsymbol{\theta}' \,, \quad \mathbf{J}_{\phi\theta} = \int \boldsymbol{\phi}\boldsymbol{\theta}'$$

- and this roughness penalty matrix

$$\mathbf{R} = \int (L\boldsymbol{\theta})(L\boldsymbol{\theta})' \,.$$

- The fitting criterion now can be expressed as

$$\begin{aligned} \text{PENSSE}(y|\boldsymbol{\beta}) \;=\; & \text{trace}\,(\mathbf{C}'\mathbf{C}\mathbf{J}_{\phi\phi}) + \text{trace}\,(\mathbf{Z}'\mathbf{Z}\mathbf{B}\mathbf{J}_{\theta\theta}\mathbf{B}') - \\ & 2\,\text{trace}\,(\mathbf{B}\mathbf{J}_{\theta\theta}\mathbf{C}'\mathbf{Z}) + \lambda\,\text{trace}\,(\mathbf{B}\mathbf{R}\mathbf{B}') \,, \end{aligned}$$

The normal equations for B

- Taking the matrix derivative with respect to $\mathbf{B}$ and setting it to 0 gives

$$\left(\mathbf{Z}'\mathbf{Z}\mathbf{B}\mathbf{J}_{\theta\theta} + \lambda\mathbf{B}\mathbf{R}\right) = \mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta} \ .$$

- We can use the Kronecker product to convert expressions of the form $\mathbf{ABC}'$ to

$$\text{vec}\left(\mathbf{ABC}'\right) = (\mathbf{C} \otimes \mathbf{A})\text{vec}\left(\mathbf{B}\right) \ ,$$

and consequently the normal equations become

$$[\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda\mathbf{I}]\text{vec}\left(\mathbf{B}\right) = \text{vec}\left(\mathbf{Z}'\mathbf{C}\mathbf{J}_{\phi\theta}\right) \ .$$

- The estimate of $\mathbf{B}$ is therefore

$$\begin{aligned} \text{vec}\,(\hat{\mathbf{B}}) &= [\mathbf{J}_{\theta\theta} \otimes (\mathbf{Z}'\mathbf{Z}) + \mathbf{R} \otimes \lambda\mathbf{I}]^{-1}(\mathbf{J}_{\phi\theta} \otimes \mathbf{Z}')\text{vec}\,(\mathbf{C}) \\ &= \mathbf{S}_\beta \text{vec}\,(\mathbf{C}). \end{aligned}$$

4. Defining 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 $\beta_j(t)$, ignoring other aspects of the function.
- Perhaps, for example, we want to examine the behavior of the temperature coefficient functions in mid-winter.
- A functional contrast is of the form

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

  where $\xi(s)$ is a suitable weight function.
- There no particular need for $\xi(s)$ to integrate to 0.

- When $\beta_j(s)$ has the basis function expansion

$$\beta_j(s) = \mathbf{B}_j \boldsymbol{\theta}(s),$$

where $\mathbf{B}_j$ is the jth row of $\mathbf{B}$, the contrast becomes

$$\rho(\beta) = \mathbf{B}_j \int \xi(s)\boldsymbol{\theta}(s)\, ds$$

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 $\mathbf{\Sigma}_e$.
  - Assuming independence of the observations, the variance of the whole response data matrix is

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

  - 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:

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

How do I work out mapping $\mathbf{M}_j$?

- In the examples given, $\rho(\beta_j)$ is three linear mappings removed from the data:
  - The linear mapping from the raw data in matrix $\mathbf{Y}$ to the coefficient matrix $\mathbf{C}$ defining the smooth functions in $\mathbf{y}(t)$. This is

$$\text{vec}\,(\mathbf{C}) = (\mathbf{S}_y \otimes \mathbf{I})\text{vec}\,(\mathbf{Y}).$$

  - The linear mapping from $\mathbf{C}$ to the regression coefficient function coefficient vector $\mathbf{B}'_j$. We worked this out already, and called it $\mathbf{S}_\beta$.
  - The linear mapping from $\mathbf{B}'_j$ to the value of the probe. This is

$$\mathbf{U} = \int \xi(s)\boldsymbol{\theta}'(s)\,ds$$

- Now we have it, namely

$$\mathbf{M}_j = \mathbf{U}_j \mathbf{S}_\beta (\mathbf{S}_y \otimes \mathbf{I})$$

- This process is easy to extend to probes $\boldsymbol{\xi}(s)$ involving all regression coefficients.
- For example, the variance of vec $[\hat{\boldsymbol{\beta}}(\mathbf{t})]$ where $\mathbf{t}$ is a vector of values of t, is

$$(\boldsymbol{\Theta} \otimes \mathbf{I})\mathbf{S}_\beta(\mathbf{S}_y \otimes \mathbf{I})(\boldsymbol{\Sigma}_e \otimes \mathbf{I})(\mathbf{S}'_y \otimes \mathbf{I})\mathbf{S}'_\beta(\boldsymbol{\Theta} \otimes \mathbf{I})' \, .$$

where $\boldsymbol{\Theta}$ is the matrix of values of $\boldsymbol{\theta}$ at $\mathbf{t}$.

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.

5. 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.