16

Functional linear models for functional responses

16.1 Introduction: Predicting log precipitation from temperature

The aim of Chapter 15 was to predict a scalar response y from a functional covariate z. We now consider a fully functional linear model in which both the response y and the covariate z are functions. For instance, in the Canadian weather example, we might wish to investigate to what extent we can predict the complete log daily precipitation profile `LogPrec` of a weather station from information in its complete daily temperature profile `Temp`.

Because all the functions in this example are intrinsically periodic, we can expand both the log precipitations and the temperatures in Fourier series. We preprocessed the data by fitting a Fourier series with 65 terms, applying a roughness penalty smoother by tapering the series to eliminate very local variation.

We are now interested in the functional linear model

$$\mathtt{LogPrec}_i(t) = \alpha(t) + \int_0^{365} \mathtt{Temp}_i(s)\beta(s,t)\,ds + \epsilon_i(t)\ . \tag{16.1}$$

In contrast to the concurrent model discussed in Chapter 14, the regression function β is now a function of both s and t. We can interpret the regression function $\beta(s,t)$ for a fixed value of t as the relative weight placed on the temperature at day s that is required to predict log precipitation on day t.

Temperatures on day s quite far from day t may be important for predicting $\texttt{LogPrec}_i(t)$. For example, continental weather stations, where the winters are cold and the summers hot, have most of their precipitation in the summer months, in contrast to Atlantic stations, which tend to have a fairly even distribution of precipitation over the whole year. Also, the influence of temperature at a time $s > t$ is allowed here since we assume that the weather patterns are periodic, and therefore that the information "wraps around" so that we can predict rainfall in January by using temperature information in December.

The function α plays the part of the constant term in the standard regression setup, and allows for a different functional origin for the log precipitation curves than the origin for the temperature curves. In effect, the second term involving $\beta(s,t)$ indicates the advantage of using temperature information over what could be achieved by using mean log precipitation as a predictor, which is what $\alpha(t)$ would be without any covariate.

The unweighted fitting criterion is the integrated residual sum of squares that we already used in Chapter 14:

$$\texttt{LMSSE}(\alpha,\beta) = \int \sum_{i=1}^{N} [\texttt{LogPrec}_i(t) - \alpha(t) - \int \texttt{Temp}_i(s)\beta(s,t)\,ds]^2\,dt. \quad (16.2)$$

16.1.1 Fitting the model without regularization

We consider the expression of β as a double expansion in terms of K_1 basis functions η_k and K_1 basis functions θ_ℓ to give

$$\beta(s,t) = \sum_{k=1}^{K_1}\sum_{\ell=1}^{K_2} b_{k\ell}\,\eta_k(s)\theta_\ell(t) = \boldsymbol{\eta}(s)'\mathbf{B}\boldsymbol{\theta}(t), \quad (16.3)$$

where $\mathbf{B}$ is a $K_1 \times K_2$ matrix of coefficients $b_{\ell k}$, or, more compactly, as

$$\beta = \boldsymbol{\eta}'\mathbf{B}\boldsymbol{\theta}\ .$$

We will also use the basis vector $\boldsymbol{\theta}$ to expand the intercept function α as

$$\alpha(t) = \sum_{\ell=1}^{K_2} a_\ell\theta_\ell(t) = \boldsymbol{\theta}'(t)\mathbf{a}. \quad (16.4)$$

The unweighted fitting criterion (16.2) now becomes

$$\texttt{LMSSE}(\mathbf{a},\mathbf{B}) = \int \sum_{i=1}^{N} [\texttt{LogPrec}_i(t) - \boldsymbol{\theta}'(t)\mathbf{a} - \int \texttt{Temp}_i(s)\boldsymbol{\eta}(s)'\mathbf{B}\boldsymbol{\theta}(t)\,ds]^2\,dt. \quad (16.5)$$

We defer any further discussion of how to estimate the coefficient vector a for α and the coefficient matrix $\mathbf{B}$ for β to Section 16.4.

As our first attempt to fit the model, we used the same 65 basis functions used to expand the log precipitation and temperature functions for both

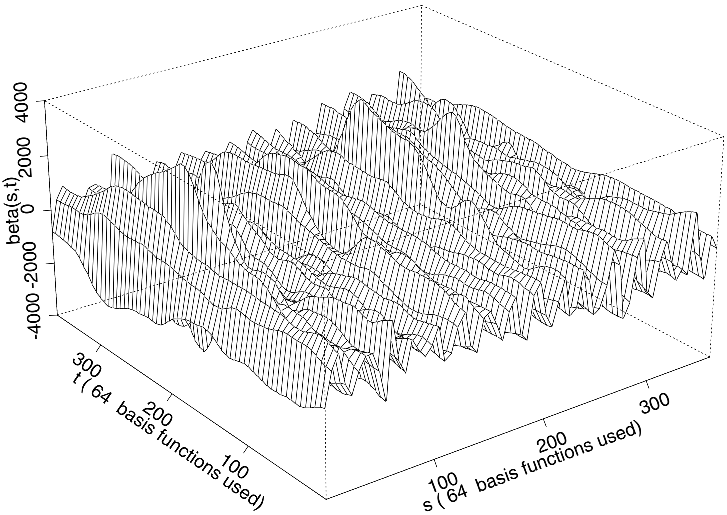

Figure 16.1. The functional parameter function β for the prediction of log precipitation from temperature, estimated direct from the data. The value $\beta(s,t)$ shows the influence of temperature at time s on log precipitation at time t.

basis systems $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. We minimized the fit criterion (16.5) to obtain the estimated function β plotted in Figure 16.1.

We see that the function β estimated by this method is extremely variable. It also turns out that this β gives perfect prediction of the given data in `LogPrec`. This does not make physical sense; whatever influence temperature patterns may have on precipitation patterns, it is naive to imagine that the precipitation pattern at a place can be entirely accounted for by its temperature pattern.

The reason for this over-fitting is an extension of the discussion in Chapter 14 on the concurrent linear model. Consider any fixed t: as in Section 15.2, we can find a number α_t and a function β_t such that, for all i,

$$\mathtt{LogPrec}_i(t) = \alpha_t + \langle \mathtt{Temp}_i, \beta_t \rangle$$

without any error. Just as in Chapter 15, we must somehow regularize the functional predictor variable. Regularization by limiting the number of basis functions is discussed in the next section, and the use of roughness penalties is taken up in Section 16.4.

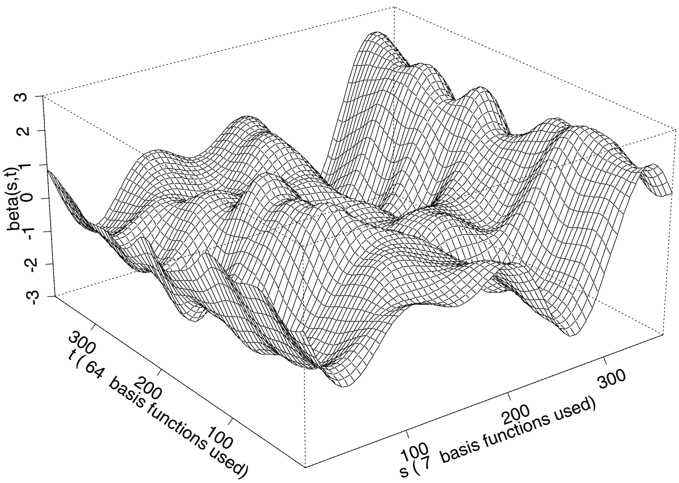

Figure 16.2. Perspective plot of estimated β function truncating the basis for the temperature covariates to 7 terms.

16.2 Regularizing the fit by restricting the bases

We now consider the effect of restricting each of the bases $\boldsymbol{\eta}$ and $\boldsymbol{\theta}$ in turn for expanding $\beta(s,t)$.

16.2.1 Restricting the basis $\boldsymbol{\eta}(s)$

We regularize β by setting the number of basis functions for its variation over argument s at $K_1 = 7$. Figure 16.2 shows the resulting estimated β function. The resulting prediction of the annual *pattern* of log precipitation at four selected stations is demonstrated in Figure 16.3. In this figure, both the original data and the predictions for log precipitation have their annual mean subtracted, to highlight the pattern of precipitation rather than its overall level. The precipitation pattern is quite well predicted except for Edmonton, which has a precipitation pattern different from other weather stations with similar temperature profiles.

Although the plot of the estimated β function demonstrates a more plausible influence of temperature pattern on precipitation pattern, it is not easy to interpret. As a function of t for any fixed s it is irregular, and this irregularity is easily explained. Because every Fourier coefficient of log precipitation is allowed to be predicted by the temperature covariate, the prediction contains frequency elements at all levels. By the arguments

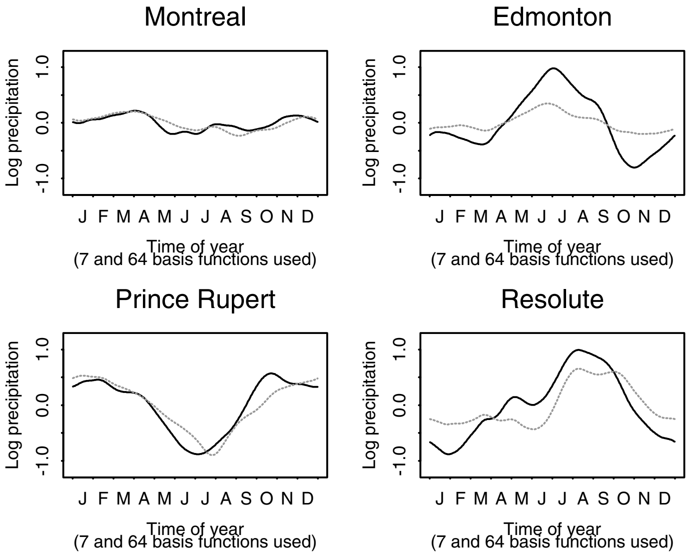

Figure 16.3. Original data (solid) and predictions (dashed) of log precipitation relative to annual mean for each of 4 weather stations. The prediction is carried out using an estimated β function with the temperature covariate truncated to 7 terms.

given in Chapter 15, we expect that each individual Fourier coefficient will be predicted sensibly as a scalar response. However, in putting these together to give a functional prediction, the high-frequency terms are given inappropriate weight. From a common-sense point of view, we cannot expect overall temperature patterns to affect a very high frequency aspect of log precipitation at all. To address this difficulty, we consider the idea of restricting or truncating the $\boldsymbol{\theta}$ basis in terms of which the functional response variable is expanded.

16.2.2 Restricting the basis $\boldsymbol{\theta}(t)$

In this section, we consider the approach of truncating the $\boldsymbol{\theta}$ basis, allowing the prediction of only low-frequency aspects of the response variable. In our example, this would correspond to the idea that the very fine detail of log precipitation could not be predicted from temperature. For the moment, suppose that we do not truncate the $\boldsymbol{\eta}$ basis, but that we allow only $K_0 = 7$ terms in the expansion of the y_i, with corresponding adjustments to the matrices $\mathbf{C}$ and $\mathbf{B}$. Figures 16.4 and 16.5 show the resulting β functions and sample predictions. The predictions are smooth, but otherwise very close to the original data. The function β is similar in overall character to

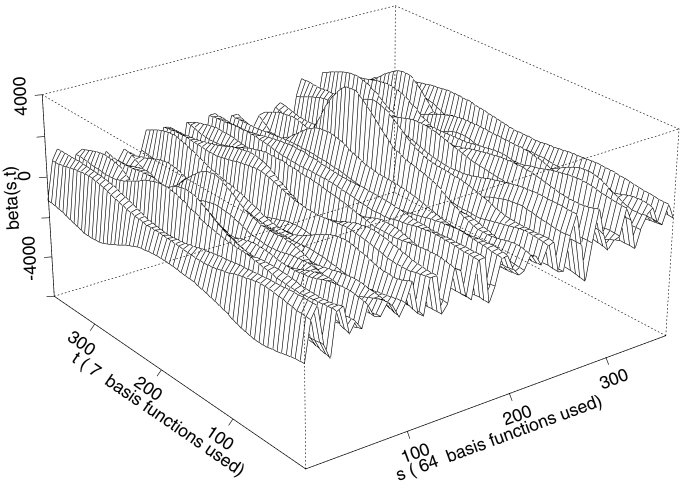

Figure 16.4. Perspective plot of estimated β function truncating the basis for the log precipitations to 7 terms.

the unsmoothed function shown in Figure 16.1, except that it is smoother as a function of t. However, it is excessively rough as a function of s. Thus, although the predictions are aesthetically attractive as smooth functions, they provide an optimistic assessment of the quality of the prediction, and an implausible mechanism by which the prediction takes place.

16.2.3 Restricting both bases

Sections 16.2.1 and 16.2.2 illustrated advantages in truncating both the $\boldsymbol{\eta}$ basis of the predictors and the $\boldsymbol{\theta}$ basis of the responses to obtain useful and sensible estimates. It should be stressed that the reason for doing this is not the same in both cases. Truncating the $\boldsymbol{\eta}$ basis for the covariates is essential to avoid over-fitting, while the $\boldsymbol{\theta}$ basis is truncated to ensure that the predictions are smooth.

Let us combine these different reasons for truncating the bases, and truncate both the predictor basis $\boldsymbol{\eta}$ and the response basis $\boldsymbol{\theta}$. Figures 16.6, 16.7 and 16.8 show the effects of truncating both bases to seven terms. We can discern several aspects of the effect of temperature on log precipitation. Temperature in February is negatively associated with precipitation throughout the year. Temperature around May is positively associated with precipitation in the summer months. Temperature in September has a strong negative association with precipitation in the autumn and

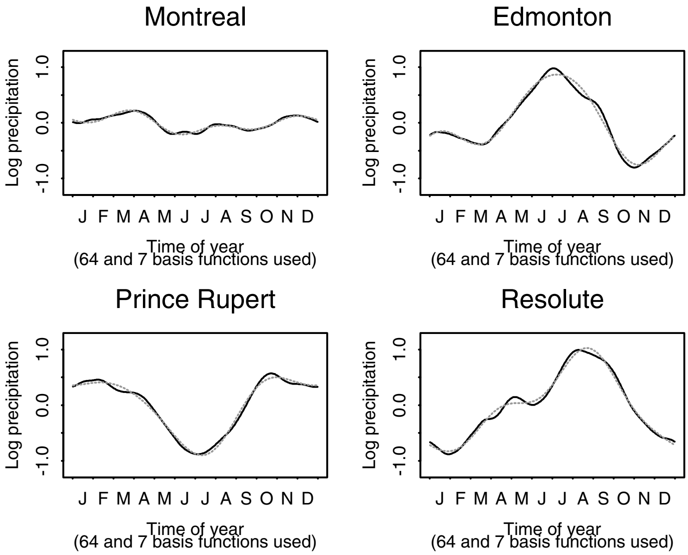

Figure 16.5. Original data (solid) and predictions (dashed) of log precipitation relative to annual mean for each of four weather stations. The prediction is carried out using an estimated β function with the basis for the log precipitations truncated to 7 terms.

winter, and finally, temperature in December associates positively with precipitation throughout the year, particularly with winter precipitation.

16.3 Assessing goodness of fit

There are various ways of assessing the fit of a functional linear model as estimated in Section 16.2. An approach borrowed from the conventional linear model is to consider the squared correlation function

$$R^2(t) = 1 - \sum_i \{\hat{y}_i(t) - y_i(t)\}^2 \Big/ \sum_i \{y_i(t) - \bar{y}(t)\}^2.$$

If we require a single numerical measure of fit, then the average of R^2 over t is useful, but using the entire function R^2 offers more detailed information about the fit. Figure 16.9 plots the R^2 function for the fit to the log precipitation data in Figure 16.6. The fit is generally reasonable, and is particularly good in the first five months of the year.

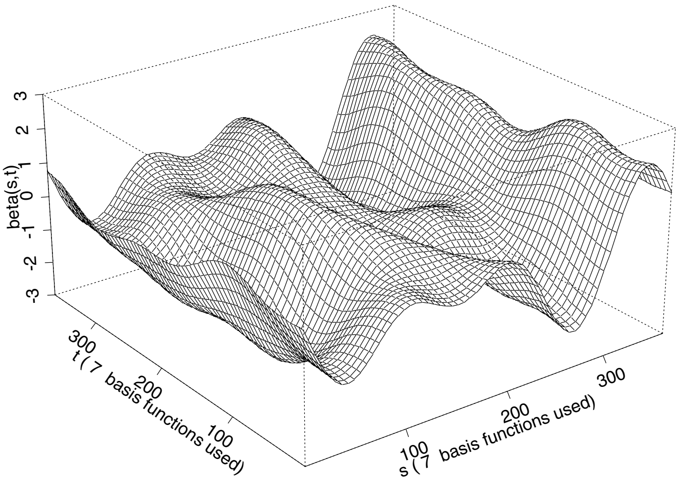

Figure 16.6. Perspective plot of estimated β function truncating both bases to seven terms.

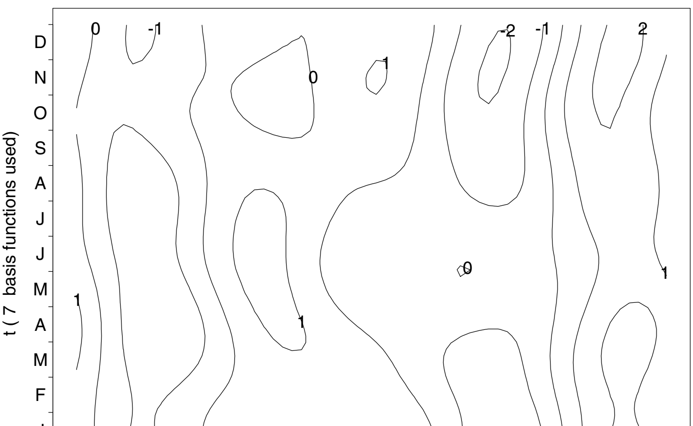

Figure 16.7. Contour plot of estimated β function truncating both bases to seven terms.

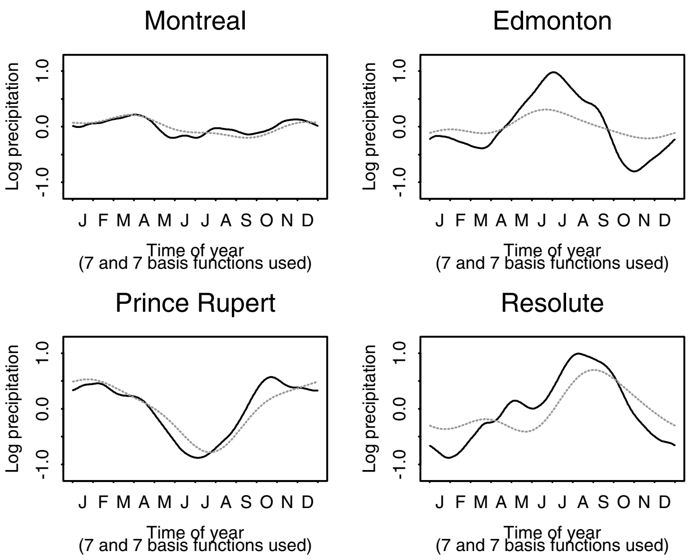

Figure 16.8. Original data (solid) and predictions (dashed) of log precipitation relative to annual mean for each of four weather stations. The prediction is carried out using an estimated β function with the both bases truncated to seven terms.

A complementary approach to goodness of fit is to consider an overall R^2 measure for each individual functional datum, defined by

$$R_i^2 = 1 - \int \{\hat{y}_i(t) - y_i(t)\}^2 \, dt \Big/ \int \{y_i(t) - \bar{y}(t)\}^2 \, dt.$$

For the four particular stations plotted in Figure 16.8, for instance, the values of R_i^2 are 0.96, 0.67, 0.63 and 0.81 respectively, illustrating that Montreal and Resolute are places whose precipitations fit closely to those predicted by the model on the basis of their observed temperature profiles; for Edmonton and Prince Rupert the fit is of course still quite good in that the temperature pattern accounts for over 60% of the variation of the log precipitation from the overall population mean. However, Figure 16.8 demonstrates that the pattern of precipitation, judged by comparing the predictions with the original data after subtracting the annual mean for the individual places, is predicted only moderately well for Resolute and is not well predicted for Edmonton. Figure 16.10 displays a histogram of all 35 R_i^2 values. At most of the stations, the R_i^2 value indicates reasonable or excellent prediction, but for a small proportion the precipitation pattern is not at all well predicted. Indeed, four stations (Dawson, Schefferville, Toronto and Prince George) have negative R_i^2 values, indicating that for

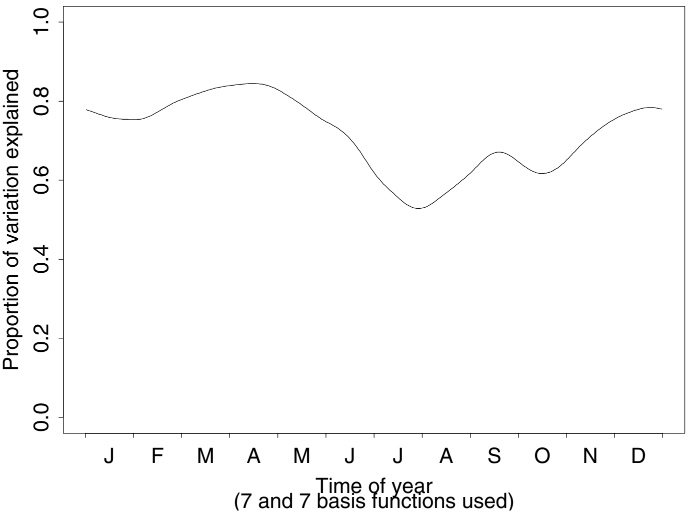

Figure 16.9. Proportion of variance of log precipitation explained by a linear model based on daily temperature records. The prediction is carried out using an estimated β function with both bases truncated to seven terms.

these places the population mean $\bar{y}$ actually gives a better fit to y_i than does the predictor $\hat{y}_i$.

To investigate this effect further, we use Figure 16.11 to show a plot of the residual mean square prediction error $\int(\hat{y}_i - y_i)^2$ against the mean square variation from the overall mean, $\int(y_i - \bar{y})^2$. The four places with negative values of R_i^2 are indicated by 0's on the plot. Each of the four places plotted in Figure 16.8 is indicated by the initial letter of its name. For most places the predictor has about one quarter the mean squared error of the overall population mean, and for many places the predictor is even better. The four places that yielded a negative value of R_i^2 did so because they were close (in three cases very close) to the overall population mean, not because the predictor did not work well for them. To judge accuracy of prediction for an individual place, it is clear that one needs to look a little further than just at the statistic R_i^2.

It is possible to conceive of an F-ratio function for the fit. We have

$$\hat{y}_i(t) - \bar{y}(t) = \sum_{j=1}^{J_0} C_{ij} \left(\sum_{k=1}^{K_0} B_{jk}\theta_k(t) \right) = \sum_{j=1}^{J_0} C_{ij}\theta_j(t).$$

By analogy with the standard linear model, we can ascribe $K_0 - 1$ degrees of freedom to the point-wise sum of squares $\sum_i \{\hat{y}_i(t) - \bar{y}(t)\}^2$ and $n - K_0$

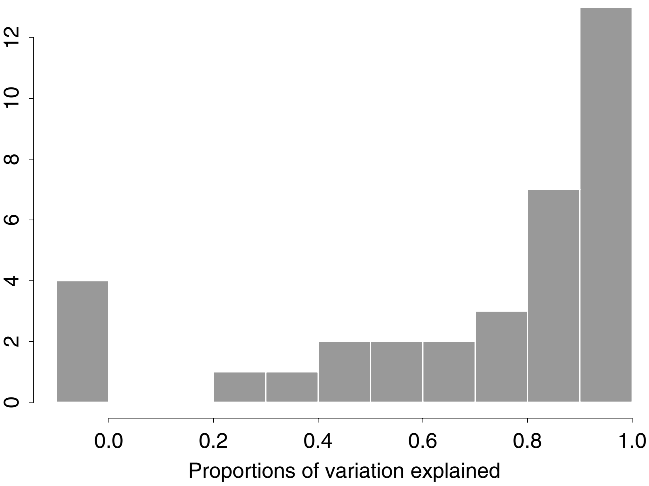

Figure 16.10. Histogram of individual proportions of variance R_i^2 in log precipitation explained by a linear model based on daily temperature records. The prediction is carried out using an estimated β function with both bases truncated to seven terms. The left-hand cell of the histogram includes all cases with negative R_i^2 values.

degrees of freedom to the residual sum of squares $\sum_i \{y_i(t) - \hat{y}_i(t)\}^2$. An F-ratio plot would be constructed by plotting

$$\texttt{FRATIO}(t) = \frac{\sum_i \{\hat{y}_i(t) - \bar{y}(t)\}^2/(K_0 - 1)}{\sum_i \{y_i(t) - \hat{y}_i(t)\}^2/(n - K_0)}.$$

However, the parameters $\theta_j(t)$ are not directly chosen to give the best fit of $\hat{y}_i(t)$ to the observed $y_i(t)$, and so the classical distribution theory of the F-ratio could be used only as an approximation to the distribution of $\texttt{FRATIO}(t)$ for each t.

Figure 16.12 plots the F-ratio for the fit to the log precipitation data. The upper 5% and 1% points of the $F_{6:28}$ distribution are given; within this model, this indicates that the effect of daily temperature on precipitation is highly significant overall.

We have not given much attention to the method by which the truncation parameters J_0 and K_0 could be chosen in practice. For many smoothing and regularization problems, the appropriate method of choice is probably subjective. The different roles of J_0 and K_0 lead to different ways of considering their automatic choice, if one is desired. The variable J_0 corresponds to a number of terms in a regression model, and so we could use a vari-

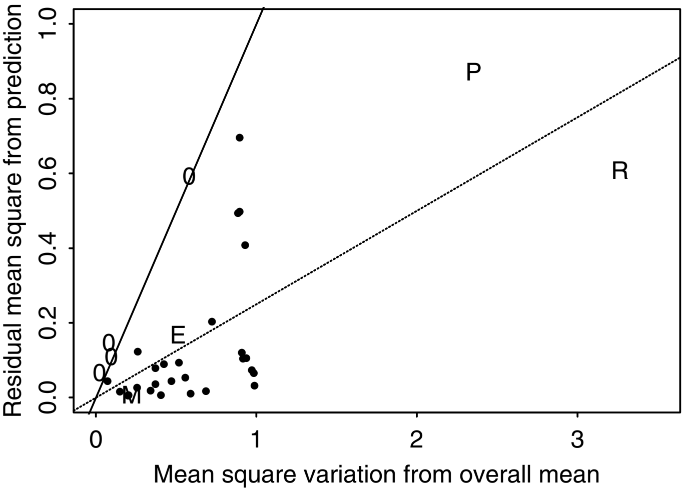

Figure 16.11. Comparison, for log precipitation, between mean square prediction errors and mean square variation from overall mean of log precipitation. The prediction is carried out using an estimated β function with both bases truncated to seven terms. The points for Montreal, Edmonton, Prince Rupert and Resolute are marked as M, E, P and R respectively. The points marked 0 yield negative R_i^2 values. The lines $y = x$ and $y = 0.25x$ are drawn on the plot as solid and dotted, respectively.

able selection technique from conventional regression, possibly adapted to give a functional rather than a numerical criterion, to indicate a possible value. On the other hand, K_0 is more akin to a smoothing parameter in a smoothing method, and so a method such as cross-validation might be a more appropriate choice. These questions are interesting topics for future investigation and research.

16.4 Computational details

Here we indicate how the fits discussed in Section 16.1 are computed. First, we have a look at the simpler case, used in that section, where the only regularization principle used is restricting the number of basis functions.

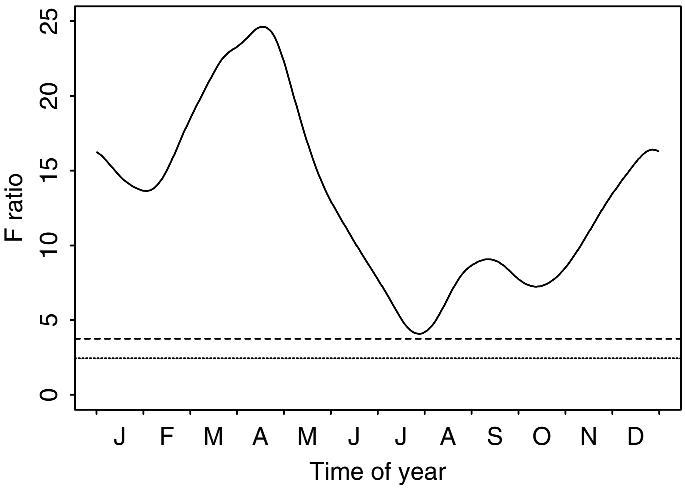

Figure 16.12. A plot of the F-ratio function for the prediction of log precipitation from daily temperature data. The prediction is carried out using an estimated β function with both bases truncated to seven terms. The horizontal lines show the upper 5% and 1% points of the $F_{6:28}$ distribution.

16.4.1 *Fitting the model without regularization*

Using more general matrix notation, our model is

$$\mathbf{y}^*(t) = \int \mathbf{z}^*(s)\beta(s,t)\,ds + \boldsymbol{\epsilon}(t)\ . \tag{16.6}$$

Recall that the bivariate regression function β has the expansion

$$\beta(s,t) = \boldsymbol{\theta}'(s)\mathbf{B}\boldsymbol{\eta}(t)\ ,$$

where basis system $\boldsymbol{\theta}$ has K_1 functions and $\boldsymbol{\eta}$ has K_2 functions. By substituting this expansion, the model becomes

$$\begin{aligned}\mathbf{y}^*(t) &= \int \mathbf{z}^*(s)\boldsymbol{\theta}'(s)\mathbf{B}\boldsymbol{\eta}(t)\,ds + \boldsymbol{\epsilon}(t)\\ &= \mathbf{Z}^*\mathbf{B}\boldsymbol{\eta}(t) + \epsilon(t),\end{aligned} \tag{16.7}$$

where the N by K_1 matrix $\mathbf{Z}^*$ is

$$\mathbf{Z}^* = \int \mathbf{z}^*(s)\boldsymbol{\theta}'(s)\,ds\ . \tag{16.8}$$

The second equation in (16.7), in effect, brings us back to the situation in Chapter 13, and in addition has the further simplification that the basis $\boldsymbol{\eta}$

is used for all the regression functions. Thus, we can use the computational details in that chapter to derive the following set of normal equations that must be solved for the regression function coefficient matrix $\mathbf{B}$:

$$\mathbf{Z}^{*'}\mathbf{Z}^{*}\mathbf{B}\int \boldsymbol{\eta}(t)\boldsymbol{\eta}'(t)\,dt = \mathbf{Z}^{*'}\int y(t)\boldsymbol{\eta}'(t)\,dt\ . \tag{16.9}$$

As before, we can re-express this equation in Kronecker product notation as

$$[\mathbf{J}_{\eta\eta}\otimes(\mathbf{Z}^{*'}\mathbf{Z}^{*})]\mathrm{vec}(\mathbf{B}) = \mathrm{vec}(\mathbf{Z}^{*'}\int \mathbf{y}(t)\boldsymbol{\eta}'(t)\,dt)\ , \tag{16.10}$$

where

$$\mathbf{J}_{\eta\eta} = \int \boldsymbol{\eta}(t)\boldsymbol{\eta}'(t)\,dt\ .$$

We note again that the numerical integration involved in the right side $\mathbf{Z}^{*'}\int \mathbf{y}\boldsymbol{\eta}'$ is, in effect, the set of inner products of the basis function vector $\boldsymbol{\eta}$ with the unit function $\mathbf{1}$ using the K_1 weighting functions $\mathbf{Z}^{*'}y$.

16.4.2 Fitting the model with regularization

Now let us consider the alternative strategy of endowing β with a generous number of basis functions for both $\boldsymbol{\theta}$ and $\boldsymbol{\eta}$. We need to define two roughness penalties: one for β's variation with respect to s, and another for its variation with respect to t.

Consider the s situation first. Let linear differential operator L_s be an appropriate operator for "curvature" in the larger sense to be applied to β as a function of s only. Our penalty is

$$\begin{aligned}
\mathtt{PEN}_s(\beta) &= \int\int [L_s\beta(s,t)]^2\,ds\,dt \\
&= \int\int [L_s\boldsymbol{\theta}'(s)\mathbf{B}\boldsymbol{\eta}(t)][L_s\boldsymbol{\theta}'(s)\mathbf{B}\boldsymbol{\eta}(t)]'\,ds\,dt \\
&= \int\int [L_s\boldsymbol{\theta}'(s)]\mathbf{B}\boldsymbol{\eta}(t)\boldsymbol{\eta}'(t)\mathbf{B}'[L_s\boldsymbol{\theta}(s)]\,ds\,dt \\
&= \int \mathrm{trace}[\mathbf{B}\boldsymbol{\eta}(t)\boldsymbol{\eta}'(t)\mathbf{B}'\mathbf{R}]\,dt \\
&= \mathrm{trace}[\mathbf{B}'\mathbf{R}\mathbf{B}\mathbf{J}_{\eta\eta}]\ ,
\end{aligned} \tag{16.11}$$

where order K_1 symmetric matrix $\mathbf{R}$ is

$$\mathbf{R} = \int [L_s\boldsymbol{\theta}(s)][L_s\boldsymbol{\theta}'(s)]\,ds\ .$$

Penalization of β with respect to t requires an analogous linear differential operator L_t to be applied to β as a function of t only. Following through the derivation above, we arrive at the following for the roughness penalty for t:

$$\begin{aligned}\mathtt{PEN}_t(\beta) &= \int\int [L_t\beta(s,t)]^2\,ds\,dt \\ &= \mathrm{trace}[\mathbf{B}'\mathbf{J}_{\theta\theta}\mathbf{S}\mathbf{B}]\ ,\end{aligned} \tag{16.12}$$

where order K_2 symmetric matrix $\mathbf{S}$ is

$$\mathbf{S} = \int [L_t\boldsymbol{\eta}(t)][L_t\boldsymbol{\eta}'(t)]\,dt\ ,$$

where

$$\mathbf{J}_{\theta\theta} = \int \boldsymbol{\theta}(t)\boldsymbol{\theta}'(t)\,dt\ .$$

When we add these two penalties, each multiplied by their respective smoothing parameters, the equations for $\mathbf{B}$ become

$$\mathbf{Z}^{*'}\mathbf{Z}^*\mathbf{B}\mathbf{J}_{\eta\eta} + \lambda_s\mathbf{R}\mathbf{B}\mathbf{J}_{\eta\eta} + \lambda_t\mathbf{J}_{\theta\theta}\mathbf{B}\mathbf{S} = \mathbf{Z}^{*'}\int \mathbf{y}\boldsymbol{\eta}'\ , \tag{16.13}$$

with the Kronecker product equivalent

$$[\mathbf{J}_{\eta\eta}\otimes(\mathbf{Z}^{*'}\mathbf{Z}^*) + \lambda_s\mathbf{J}_{\eta\eta}\otimes\mathbf{R} + \lambda_t\mathbf{S}\otimes\mathbf{J}_{\theta\theta}]\mathrm{vec}(\mathbf{B}) = \mathrm{vec}(\mathbf{Z}^{*'}\int \mathbf{y}\boldsymbol{\eta}')\ . \tag{16.14}$$

Finally, in order to compute standard errors, we will need to specify that $\mathbf{y} = \mathbf{C}\boldsymbol{\phi}$ where $\boldsymbol{\phi}$ is a system of K_y basis functions and $\mathbf{C}$ is the associated N by K_y coefficient matrix. In this case, we can express the estimate for $\mathbf{B}$ as

$$\mathrm{vec}(\hat{\mathbf{B}}) = [\mathbf{J}_{\eta\eta}\otimes(\mathbf{Z}^{*'}\mathbf{Z}^*) + \lambda_s\mathbf{J}_{\eta\eta}\otimes\mathbf{R} + \lambda_t\mathbf{S}\otimes\mathbf{J}_{\theta\theta}]^{-1}(\mathbf{J}_{\theta\eta}\otimes\mathbf{Z}^{*'})\mathrm{vec}(\mathbf{C})\ , \tag{16.15}$$

where

$$\mathbf{J}_{\theta\eta} = \int \boldsymbol{\theta}(t)\boldsymbol{\eta}'(t)\,dt\ .$$

16.5 The general case

The functional linear model (16.6) has been useful for exploring the implications of regularizing $\beta(s,t)$ with respect to each of its arguments. But it is, nevertheless, a model that is restricted in three important ways:

- The covariate function z that we used was a function of s alone. In fact, there is no reason why it might not also vary as a function of t, that is, take values $z(s,t)$. In fact, we assumed variation over t for the point-wise linear model in Chapter 14, and we may do so here, too.

- That we were able to integrate the influence of z over the entire year was made possible by the covariate being periodic, as we already observed. In many situations, and especially when both s and t index time, the case $s > t$ is inadmissible since causality does not operate backwards in time. In this case, we would have the model

$$y_i(t) = \int_0^t z_i(s)\beta(s,t)\,ds + \epsilon_i(t)$$

  and the matrix $\mathbf{Z}^*$ defined in (16.8) would now be a function of t rather than being constant.

- The expansion of β in (16.3), called a *tensor product* expansion, is highly specialized. In fact, it tends only be suitable when argument s can be integrated over its entire range, that is, including when the covariate is periodic.

Consequently, we need a formulation that can encompass not only the models that we have used up to now, but a range of others as well that may be important in other applications.

To get away from tensor-product expansions, let us now propose the general expansion

$$\beta(s,t) = \sum_k^{K_\beta} b_k\theta_k(s,t) \ = \boldsymbol{\theta}'(s,t)\mathbf{b}\ . \tag{16.16}$$

Moreover, let us assume that the covariate z takes values $z(s,t)$ varying over both arguments. Finally, let the interval of integration for argument s be allowed to vary over t, and we can use the notation Ω_t to indicate the interval associated with t. Our model now becomes

$$\begin{aligned} y_i(t) &= \int_{\Omega_t} z_i(s,t)\beta(s,t)\,ds + \epsilon_i(t) \\ &= \int_{\Omega_t} z_i(s,t)\boldsymbol{\theta}'(s,t)\mathbf{b}\,ds + \epsilon_i(t). \end{aligned} \tag{16.17}$$

Each of the previous models is contained within this one. For example, the multivariate covariate z_i in Chapter 13 does not vary with either argument, and no integration is involved, so that we can simply drop argument s from the model. The point-wise model in Chapter 14 is similar in that, since $\Omega_t = t$, argument s is again irrelevant in the sense that it can be folded into t itself. But in this case z_i does vary over t.

In this general case, we can still make an important simplification. As we did for the earlier specialized model, we can integrate out s by defining

$$z^*_{ik}(t) = \int_{\Omega_t} z_i(s,t)\theta'_k(s,t)\,ds \tag{16.18}$$

and then re-express this general model in a matrix version that no longer has an explicit role for s:

$$\mathbf{y}(t) = \mathbf{Z}^*(t)\mathbf{b} + \boldsymbol{\epsilon}(t), \tag{16.19}$$

where $\mathbf{Z}^*$ is an N by K_β matrix function. Of course, the removal of s is actually an illusion, since any evaluation of $\mathbf{Z}^*$ will automatically involve an integration over s.

Nevertheless, formally we now have a concurrent or point-wise functional linear model, but where the regression coefficient functions are all constant. In short, the most general case finally reduces to the point-wise linear model, but of course at the expense of replacing a single covariate $z_i(s,t)$ by a vector of K computed covariates $z^*_{ik}(t)$.

The estimated parameter vector $\hat{\mathbf{b}}$ satisfies

$$[\int \mathbf{Z}^{*'}\mathbf{Z}^* + \lambda_s \mathbf{R} + \lambda_t \mathbf{S}]\hat{\mathbf{b}} = \int \mathbf{Z}^{*'}\mathbf{y}, \tag{16.20}$$

where

$$\mathbf{R} = \int \int_{\Omega_t} [L_s \boldsymbol{\theta}(s,t)][L_s \boldsymbol{\theta}(s,t)]' \, ds \, dt$$

and

$$\mathbf{S} = \int \int_{\Omega_t} [L_t \boldsymbol{\theta}(s,t)][L_t \boldsymbol{\theta}(s,t)]' \, ds \, dt \ .$$

What kind of bivariate basis functions might we propose? The topic of smoothing data over higher numbers of dimensions has generated several examples. In *thin-plate spline smoothing*, for example, *radial basis functions* are used of the form

$$\theta_k(s,t) = \zeta_k(s^2 + t^2),$$

where the functions ζ_k are a *univariate* basis system.

An especially convenient and powerful class of bivariate basis functions are associated with *finite element* methods developed for numerical methods for solving partial differential equations over complex regions. These typically involve the approximate coverage of the two-dimensional domain $\Omega_t, t \in \mathcal{T}$ by a system of triangles, and the basis functions are piecewise linear "hat" functions having value one at a vertex shared by six triangles, and decreasing to zero on each of the distal edges. See Ramsay and Silverman (2002, Chapter 10) for an example.

16.6 Further reading and notes

Many important issues need further attention, but they would carry us beyond the objectives of this volume and would require technical resources rather out of line with what we are assuming.

There has been a considerable amount of work on reformulating time series methods for functional data, and especially in France and Spain. Bosq (2000) has a rather technical but broad overview. Besse, Cardot and Stephenson (2000) and Aguilera, Ocaña and Valderrama, M. J. (1999) developed autoregressive forecasting models for climatic variations.

From the perspective of classical frequentist statistics, just what is being estimated here and in Chapter 15 where the covariate structure is potentially of infinite dimension? As sample size increases, what conditions are required to assure the convergence of estimates $\hat{\boldsymbol{\beta}}$ to their population values? A good coverage of this issue along with some interesting consistency results can be found in Cuevas, Febrero and Fraiman (2002), as well as in earlier papers by Cardot, Ferraty and Sarda (1999) and Ferraty and Vieu (2001). These latter researchers are part of a group called STAPH that meets regularly to exchange findings on both foundational and application issues. Frank and Friedman (1993) also raised some of these issues in their survey on the use of regression in chemometrics.

If we adopt a Bayesian perspective, how can we propose prior distributions for the functional parameters like β that fulfill certain regularity conditions essential for coherent estimation? We have tended to adopt the perspective that the basis function expansion that we use is chosen essentially as a matter of both capturing certain known features of the problem and of computational convenience. This would imply that we wouldn't want to assume that the coefficients in the matrix $\mathbf{B}$ were the parameters, since we would freely admit that other investigators might use different basis expansions for perfectly good reasons.