10

Principal components analysis of mixed data

10.1 Introduction

It is a characteristic of statistical methodology that problems do not always fall into neat categories. In the context of the methods discussed in this book, we often have *both* a vector of data *and* an observed function on each individual of interest. In this chapter, we consider some ways of approaching such mixed data, extending the ideas of PCA that we have already developed.

In Chapter 7 we have discussed one way in which mixed data can arise. Consider the Canadian temperature data as a specific example. The registration process finds, for each weather station, a suitable phase shift to apply to the raw observed curve; the phase shifts are chosen to make the shifted records fit together as well as possible. The *vector part* of the record is in this case just the single number giving the size of the shift. The *functional part* of the record is the shifted curve.

The method we will develop in this chapter produces principal component weights that have the same structure as the mixed data themselves. So the variability accounted for by each principal component can itself be split into two parts, the part corresponding to variability in the phase shifts and the part corresponding to variability in the registered functions. The first four principal components for the Canadian temperature data are shown in Figure 10.1. Let $\hat{\mu}(s)$ be the mean of all the *registered* curves, in other words the mean of the functional parts of all the observations. We assume that the mean of all the phase shifts is zero.

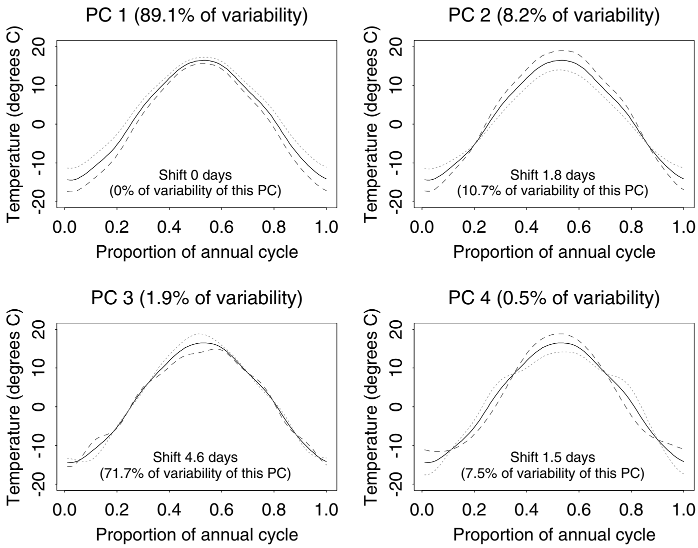

Figure 10.1. The mean Canadian temperature curve and the effects of adding and subtracting a suitable multiple of each PC curve, with the shift considered as a separate parameter.

The effect of each principal component is specified by a pair $(\xi_i(s), v_i)$, where $\xi(s)$ is the effect of variation in that component on the functional part, and v_i is the effect on the shift. Suppose, just for example, that the score at a particular weather station is 2.5 on the ith principal component and zero on all others. Then the functional part of the observation would be $\hat{\mu}(s) + 2.5\xi_i(s)$ and the phase shift would be $2.5v_i$. Note that the two effects go together, and the multiple of $\xi_i(s)$ is the same as that of v_i. In each case, the sign of the principal component has been taken to make the shift positive; this is by no means essential, but it leads to some simplicity of interpretation.

In the figure, the functional part of each principal component is illustrated by showing the effect on the overall mean $\hat{\mu}$ of adding and subtracting a suitable multiple of the relevant ξ_j. The fine dotted curve corresponds to adding the ξ_j and the dashed curve to subtracting. The shift part v_i is given numerically, for example 1.8 days for the second component. The figure also states what percentage of variability of each PC is accounted for by shift variability as opposed to the variability of the functional part.

Now consider the figure in detail. The first principal component—accounting for 89.1% of the variability in the original observations—entirely concerns the functional part, with 0% of the variability being in the shift component. A high score on this component goes along with a weather station that is warmer than average all the year round, but with a larger variation in the winter months.

The second component has 10.7% of its variability accounted for by a shift component, of size 1.8 days. The functional part of this component corresponds to a change in amplitude of the annual temperature variation. High positive scores on this component would indicate lower-than-average temperature variation over the year (cool summers and relatively warm winters) *together with* a positive shift value. The third component is very largely shift variation (71.7% of the variability). Associated with a positive shift is an increase in temperature at the high point of the summer, with very little effect elsewhere.

A comparison between Figure 10.1 and Figure 8.2 is instructive. Because the shift component has been explicitly separated out, less skill is needed to interpret the principal components in Figure 10.1. The percentage of variation explained by each of the first four principal components is very similar, but not quite identical, in the two analyses, for a reason discussed further in Section 10.4.2.

Of course, there are many other situations where we have numerical observations as well as functional observations on the individuals of interest, and the PCA methodology we set out can be easily generalized to deal with them.

10.2 General approaches to mixed data

We now consider mixed data in a more general context, bearing in mind the Canadian temperature data as a specific example. To be precise about notation, suppose that our observations consist of pairs $(x_i, \mathbf{y}_i)$, where x_i is a function on the interval $\mathcal{T}$ and $\mathbf{y}_i$ is a vector of length M. How might we use PCA to analyze such data?

There are three different ways of viewing the $\mathbf{y}_i$. First, it may be that the $\mathbf{y}_i$ are simply nuisance parameters, of no real interest to us in the analysis, for example corresponding to the time at which a recording instrument is activated. In this case we would quite simply ignore them. The $\mathbf{y}_i$ can be thought of as one of the features of almost all real data sets that we choose not to include in the analysis.

On the other hand, as in the temperature data example, both the functions x_i and the observations $\mathbf{y}_i$ may be of primary importance. The PCA of such *hybrid* data $(x_i, \mathbf{y}_i)$ is the case to which we give the most attention, from Section 10.3 onwards. There is some connection with the methodology described in Section 8.5 for bivariate curve data with values $(x_i(t), y_i(t))$, though in our case the second component is a scalar or vector rather than a function.

As a third and somewhat intermediate possibility, the $\mathbf{y}_i$ may be of marginal importance, our central interest being in the functions x_i. In this case, we could ignore the $\mathbf{y}_i$ initially and carry out a PCA of the curves $x_i(t)$ alone. Having done this, we could investigate the connection between the scores on the principal component scores and the variable(s) $\mathbf{y}_i$. We could calculate the sample correlations between the principal component scores and the components of the $\mathbf{y}_i$. Alternatively or additionally, we could plot the $\mathbf{y}_i$ against the principal component scores or use other methods for investigating dependence. In this general approach, the $\mathbf{y}_i$ would not have been used in the first part of the analysis itself; however, they would have played a key part in interpreting the analysis. It would be interesting, for example, to notice that a particular principal component of the x_i was highly correlated with $\mathbf{y}_i$. We develop this approach further in Section 10.5.2.

10.3 The PCA of hybrid data

10.3.1 Combining function and vector spaces

A typical principal component weight function would consist of a *pair* $(\xi, \mathbf{v})$, where $\mathbf{v}$ is an M-vector, and the principal component score of a particular observation would then be the sum

$$\eta_i = \int x_i(s)\xi(s)\, ds + \mathbf{y}_i'\mathbf{v}. \tag{10.1}$$

Another way of saying this is that the principal component would be made up of a functional part ξ and a vector part $\mathbf{v}$, corresponding to the functional and vector (or numerical) parts of the original data. A typical observation from the distribution of the data would be modelled as

$$\begin{pmatrix} x_i \\ \mathbf{y}_i \end{pmatrix} = \sum_j \eta_{ij} \begin{pmatrix} \xi_j \\ \mathbf{v}_j \end{pmatrix}, \tag{10.2}$$

where $(\xi_j, \mathbf{v}_j)$ is the jth principal component weight and, as j varies, the vectors of principal component scores $\eta_{ij} = \int x_i\xi_j + \mathbf{y}_i'\mathbf{v}_j$ are uncorrelated variables with mean zero.

This kind of hybrid data PCA can very easily be dealt with in our general functional framework. Define $\mathcal{Z}$ to the space of pairs $z = (x, \mathbf{y})$, where x is a smooth function and $\mathbf{y}$ is a vector of length M. Given any two elements $z_{(1)} = (x_{(1)}, \mathbf{y}_{(1)})$ and $z_{(2)} = (x_{(2)}, \mathbf{y}_{(2)})$ of $\mathcal{Z}$, define the inner product

$$\langle z_{(1)}, z_{(2)} \rangle = \int x_{(1)}x_{(2)} + \mathbf{y}_{(1)}'\mathbf{y}_{(2)}. \tag{10.3}$$

From (10.3) we can define the norm $\|z\|^2 = \langle z, z \rangle$ of any z in $\mathcal{Z}$.

Now that we have defined an inner product and norm on $\mathcal{Z}$, write z_i for the data pair $(x_i, \mathbf{y}_i)$. To find the leading principal component, we wish to find $\zeta = (\xi, \mathbf{v})$ in $\mathcal{Z}$ to maximize the sample variance of the $\langle \zeta, z_i \rangle$ subject to $\|\zeta\|^2 = 1$. The $\langle \zeta, z_i \rangle$ are of course exactly the same as the quantities $\eta_i = \int x_i(s)\xi(s)\, ds + \mathbf{y}_i'\mathbf{v}$ specified in equation (10.1).

Subsequent principal components maximize the same sample variance subject to the additional condition of orthogonality to the principal components already found, orthogonality being defined by the hybrid inner product (10.3). Principal components found in this way yield principal component scores that are uncorrelated, just as for conventional multivariate PCA.

The PCA of hybrid data is thus very easily specified in principle. However, there are several important issues raised by this idea, and we discuss these in the following sections.

10.3.2 Finding the principal components in practice

How do we carry out the constrained maximization of the sample variance of the $\langle \zeta, z_i \rangle$ in practice? Suppose that ϕ_k is a basis of K functions in which the functional parts x_i of the hybrid data z_i can be well approximated. Given any element $z = (x, \mathbf{y})$ of $\mathcal{Z}$, define the K-vector $\mathbf{c}$ to be the coefficients of x relative to the basis $\boldsymbol{\phi}$. Now let $p = K + M$, and let $\mathbf{w}$ be the p-vector

$$\mathbf{w} = \begin{bmatrix} \mathbf{c} \\ \mathbf{y} \end{bmatrix}.$$

Suppose that the basis $\boldsymbol{\phi}$ is an orthonormal basis, the Fourier functions, for example. Then the inner product (10.3) of any two elements $z_{(1)}$ and $z_{(2)}$ of $\mathcal{Z}$ is precisely equal to the ordinary vector inner product $\mathbf{w}_{(1)}'\mathbf{w}_{(2)}$ of the corresponding p-vectors of coefficients. Thus, if we use this method of representing members of $\mathcal{Z}$ by vectors, we have a representation in which the vectors behave exactly as if they were p-dimensional multivariate observations, with the usual Euclidean inner product and norm. It follows that we can use standard multivariate methods to find the PCA.

In summary, we can proceed as follows to carry out a PCA:

1. For each i, let $\mathbf{c}_i$ be the vector of the first K Fourier coefficients of x_i.

2. Augment each $\mathbf{c}_i$ by $\mathbf{y}_i$ to form the p-vector $\mathbf{w}_i$.

3. Carry out a standard PCA of the $\mathbf{w}_i$, by finding the eigenvalues and eigenvectors of the matrix $N^{-1}\sum_i \mathbf{w}_i\mathbf{w}_i'$.

4. If $\mathbf{u}$ is any resulting eigenvector, the first K elements of $\mathbf{u}$ are the Fourier coefficients of the functional part of the principal component, and the remaining elements are the vector part.

Since the procedure we have set out is a generalization of ordinary functional PCA, we may wish to incorporate some smoothing, and this is discussed in the next section.

10.3.3 Incorporating smoothing

To incorporate smoothing into our procedure, we can easily generalize the smoothing methods discussed in Chapter 9. The key step in the method is to define the roughness of an element $z = (x, \mathbf{y})$ of $\mathcal{Z}$. Let us take the roughness of z to be that of the functional part x of z, without any reference to the vector part $\mathbf{y}$. To do this, define $D^2 z$ to be equal to the element $(D^2 x, 0)$ of $\mathcal{Z}$ so that the roughness of z can then be written $\|D^2 z\|^2$, just as in the ordinary functional case. The norm is taken in $\mathcal{Z}$, but since the vector part of $D^2 z$ is defined to be zero, $\|D^2 z\|^2 = \|D^2 x\|^2$ as required.

Once we have defined the roughness of z, we can proceed to carry out a smoothed PCA using exactly the same ideas as in Chapter 9. As far as algorithms are concerned, the Fourier transform algorithm for the periodic case requires slight modification. Let z^* be the vector representation of an element z, of length $K + p$. The first K elements of z^* are the Fourier coefficients of the functional part x and the last p elements simply the vector part $\mathbf{y}$. The roughness of z is $\sum_{k=0}^{K-1} \omega_k^4 z_k^{*2}$ so the matrix $\mathbf{S}$ used in the algorithm described in Section 9.4.1 must be modified to have diagonal elements $(1 + \lambda\omega_k^4)^{-1/2}$ for $k < K$, and 1 for $K \le k < p$.

Apart from this modification, and of course the modified procedures for mapping between the function/vector and basis representations of elements of $\mathcal{Z}$, the algorithm is exactly the same as in Section 9.4.1. Furthermore, the way in which we can apply cross-validation to choose the smoothing parameter is the same as in Section 9.3.3.

To deal with the nonperiodic case, we modify the algorithm of Section 9.4.2 in the same way. The matrix $\mathbf{J}$ is a block diagonal matrix where the first K rows and columns have elements $\int \phi_j \phi_k$ and the last M rows and columns are the identity matrix of order M. The matrix $\mathbf{K}$ has elements $\int (D^2\phi_j)(D^2\phi_k)$ in its first K rows and columns, and zeroes elsewhere.

10.3.4 Balance between functional and vector variation

Readers who are familiar with PCA may have noted one potential difficulty with the methodology set out above. The variations in the functional and vector parts of a hybrid observation z are really like chalk and cheese: they are measured in units which are almost inevitably not comparable, and therefore it may well not be appropriate to weight them as we have. In the registration example, the functional part consists of the difference between the pattern of temperature on the transformed time scale and its population mean; the vector part is made up of the parameters of the time transformation. Clearly, these are not measured in directly compatible units!

One way of noticing the effect of noncomparability is to consider the construction of the inner product (10.3) on $\mathcal{Z}$, which we defined by adding the inner product of the two functional parts and that of the two vector parts. In many problems, there is no intrinsic reason to give these two inner products equal weight in the sum, and a more general inner product we could consider is

$$\langle z_{(1)}, z_{(2)} \rangle = \int x_{(1)} x_{(2)} + C^2 y'_{(1)} y_{(2)} \tag{10.4}$$

for some suitably chosen constant C. Often, the choice of C (for example $C = 1$) is somewhat arbitrary, but we can make some remarks that may guide its choice.

First, if the interval $\mathcal{T}$ is of length $|\mathcal{T}|$, then setting $C^2 = |\mathcal{T}|$ gives the same weight to overall differences between $x_{(1)}$ and $x_{(2)}$ as to differences of similar size in a single component of the vector part y. If the measurements are of cognate or comparable quantities, this may well be a good method of choosing C. On the other hand, setting $C^2 = |\mathcal{T}|/M$ tends to weight differences in functional parts the same as differences in all vector components.

Another approach, corresponding to the standard method of PCA relative to correlation matrices, is to ensure that the overall variability in the functional parts is given weight equal to that in the vector part. To do this, we would set

$$C^2 = \frac{\sum_i \|x_i - \bar{x}\|^2}{\sum_i \|\mathbf{y}_i - \bar{y}\|^2},$$

taking the norm in the functional sense in the numerator, and in the usual vector sense in the denominator.

Finally, in specific problems, there may be a particular rationale for some other choice of constant C^2, an example of which is discussed in Section 10.4.

Whatever the choice of C^2, the most straightforward algorithmic approach is to construct the vector representation z of any element $z = (x, y)$ of $\mathcal{Z}$ to have last M elements Cy, rather than just y. The first K elements are the coefficients of the representation of x in an appropriate basis, as before. With this modification, we can use the algorithms set out above. Some care must be taken in interpreting the results, however, because any particular principal component weight function has to be combined with the data values using the inner product (10.4) to get the corresponding principal component scores.

10.4 Combining registration and PCA

10.4.1 Expressing the observations as mixed data

We now return to the special case of mixed data obtained by registering a set of observed curves. For the moment, concentrate on data that may be assumed to be periodic on $[0, 1]$. We suppose that an observation can be modelled as

$$x(t+\tau) = \mu(t) + \sum_j \eta_j \xi_j(t) \tag{10.5}$$

for a suitable sequence of orthonormal functions ξ_j, and where η_j are uncorrelated random variables with mean zero and variances σ_j^2. The model (10.5) differs from the usual PCA model in allowing for a shift in time τ as well as for the addition of multiples of the principal component functions. Because of the periodicity, the shifted function $x(t + \tau)$ may still be considered as a function on $[0, 1]$.

Given a data set $x_1, \ldots, x_n$, we can use the Procrustes approach set out in Chapter 7 to obtain an estimate $\hat{\mu}$ of μ and to give values of the shifts $\tau_1, \ldots, \tau_n$ appropriate to each observation. Then we can regard the data as pairs $z_i = (\tilde{x}_i, \tau_i)$, where the τ_i are the estimated values of the shift parameter and the $\tilde{x}_i$ are the shifted mean-corrected temperature curves with values $x_i(t + \tau_i) - \hat{\mu}(t)$. Recall that a consequence of the Procrustes fitting is that the $\tilde{x}_i$ satisfy the orthogonality property

$$\int \tilde{x}_i D\hat{\mu} = 0. \tag{10.6}$$

10.4.2 Balancing temperature and time shift effects

We can now consider the effect of the methodology of Section 10.3 to the mixed data z_i obtained in the registration context. We seek principal components (ξ, v) that have two effects within the model (10.5): the addition of the function ξ to the overall mean $\hat{\mu}$, together with a contribution of v to the time shift τ.

In the special case of the registration data, there is a natural way of choosing the constant C^2 that controls the balance between the functional and shift components in the inner product (10.4). Suppose that x is a function in the original data function space, and that $z = (\tilde{x}, \tau)$ is the corresponding pair in $\mathcal{Z}$, so that

$$x(t) = \hat{\mu}(t - \tau) + \tilde{x}(t - \tau).$$

Because of the orthogonality property (10.6), we can confine attention to $\tilde{x}$ that are orthogonal to $\hat{\mu}$.

To define a norm on $\mathcal{Z}$, a requirement is that, at least to first order,

$$\|z\|^2 \approx \|x - \hat{\mu}\|^2 = \int [x(s) - \hat{\mu}(s)]^2 \, ds, \tag{10.7}$$

the standard squared function norm for $x - \hat{\mu}$. This means that the norm of any small perturbation of the mean function $\hat{\mu}$ must the same, whether it is specified in the usual function space setting as $x - \hat{\mu}$, or expressed as a pair z in $\mathcal{Z}$, consisting of a perturbation $\tilde{x}$ orthogonal to $\hat{\mu}$ and a time shift.

Suppose $\|\tilde{x}\|$ and τ are small. If we let

$$C^2 = \|D\hat{\mu}\|^2, \tag{10.8}$$

then, to first order in $\|\tilde{x}\|$ and τ,

$$x(t) - \hat{\mu}(t) \approx -\tau D\hat{\mu}(t - \tau) + \tilde{x}(t - \tau).$$

By the orthogonality of $\tilde{x}$ and $D\hat{\mu}$,

$$\|z - \hat{\mu}\|^2 \approx \int \tilde{x}^2(s) + C^2\tau^2(s) \, ds = \|\tilde{x}\|^2 + C^2\|\tau\|^2, \tag{10.9}$$

as required.

With this calculation in mind, we perform our PCA of the pairs $(\tilde{x}_i, \tau_i)$ relative to the inner product (10.4) with $C^2 = \|D\hat{\mu}\|^2$, and this was the way that C was chosen in Section 10.1. The percentage of variability of each principal component due to the shift was then worked out as $100C^2v_j^2$.

The use of this value of C provides approximate compatibility between the quantification of variation caused simply by the addition of a curve to the overall mean, and variation that also involves a time shift. It therefore accounts for the similarity of the percentages of variation explained by the various components in Figures 8.2 and 10.1.

10.5 The temperature data reconsidered

10.5.1 Taking account of effects beyond phase shift

In the temperature example, the shift effect is not necessarily the only effect that can be extracted explicitly and dealt with separately in the functional principal components analysis. We can also take account of the overall annual average temperature for each weather station, and we do this by extending the model (10.5) to a model of the form

$$x(t + \tau) - \theta = \alpha + \mu(t) + \sum_j \eta_j \xi_j(t), \tag{10.10}$$

where θ is an annual temperature effect with zero population mean. The η_j are assumed to be uncorrelated random variables with mean zero. The parameter α is the overall average temperature (averaged both over time and over the population). For identifiability we assume that $\int \mu(s)\,ds = 0$.

The data we would use to fit such a model consist of triples $(\breve{x}_i, \tau_i, \theta_i)$, where $\breve{x}_i$ are the observed temperature curves registered to one another by shifts τ_i, and with each curve modified by subtracting its overall annual average $\hat{\alpha}+\theta_i$. Here the number $\hat{\alpha}$ is the time average of all the temperatures observed at all weather stations, and the individual θ_i therefore sum to zero. Because the annual average $\hat{\alpha}+\theta_i$ has been subtracted from each curve $\breve{x}_i$, the curves $\breve{x}_i$ each integrate to zero as well as satisfying the orthogonality condition (10.6). The mean curve $\hat{\mu}$ is then an estimate of the mean of the registered curves $\breve{x}_i$, most straightforwardly the sample mean. In the hybrid data terms we have set up, the functional part of each data point is the curve $\breve{x}$, whereas the vector part is the 2-vector $(\tau_i, \theta_i)'$.

To complete the specification of (10.10) as a hybrid data principal components model, we regard τ and θ as random variables which can be expanded for the same η_j, as

$$\tau = \sum_j \eta_j v_j \text{ and } \theta = \sum_j \eta_j u_j,$$

where the v_j and u_j are fixed quantities. Thus, the jth principal component is characterized by a triple (ξ_j, v_j, u_j), constituting a distortion of the mean curve by the addition of a multiple of ξ_j, together with shifts in time and in overall temperature by the same multiples of v_j and u_j, respectively.

Just as before, we carry out a PCA of the hybrid data $\{(\breve{x}_i, \tau_i, \theta_i)\}$ with respect to a suitably chosen norm. To define the norm of a triple $(\breve{x}, \tau, \theta)$, consider the corresponding unregistered and uncorrected curve x, defined by

$$x(t+\tau) = \hat{\alpha} + \theta + \hat{\mu}(t) + \breve{x}(t).$$

Define $C_1 = \|D\hat{\mu}\|^2$ and $C_2 = |\mathcal{T}|$. Assume that $\breve{x}$ integrates to zero and satisfies (10.6).

By arguments similar to those used previously, using the standard square integral norm for $\breve{x}$,

$$\|x - \hat{\mu}\|^2 \approx \|\breve{x}\|^2 + C_1^2\tau^2 + C_2^2\theta^2.$$

Thus an appropriate definition of the norm of the triple is given by

$$\|(\breve{x}, \tau, \theta)\|^2 = \|\breve{x}\|^2 + C_1^2\tau^2 + C_2^2\theta^2.$$

In practice, a PCA with respect to this norm is carried out by the same general approach as before. For each i, the function $\breve{x}_i$ is represented by a vector $\breve{c}_i$ of its first K Fourier coefficients. The vector is augmented by the two values $C_1\tau_i$ and $C_2\theta_i$ to form the vector z_i. We then carry out a standard PCA on the augmented vectors z_i. The resulting principal component weight vectors are then unpacked into the parts corresponding to ξ_j, v_j and u_j, and the appropriate inverse transforms applied—just dividing by

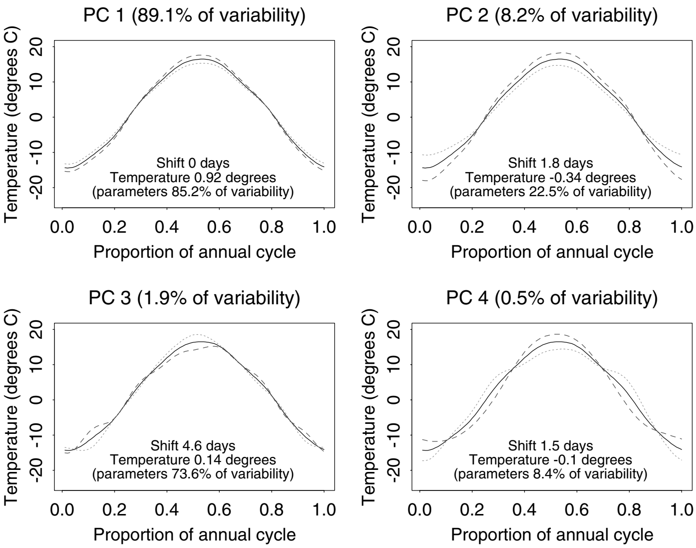

Figure 10.2. The mean Canadian temperature curve and the effect of adding and subtracting a suitable multiple of each PC curve, with the shift and annual average temperature considered as separate parameters.

C_1 and C_2 respectively in the case of the shift and overall temperature effects, and applying an inverse Fourier transform to the first K components of the vector to find ξ_j.

Figure 10.2 shows the effect of this approach applied to the Canadian temperature data. Notice that a component that was entirely variation in overall temperature would have a temperature effect of ± 1 degree, because time is scaled to make the cycle of unit length (with time measured in years) so that $C_2 = 1$. Because each principal component is scaled to have unit norm, the maximum possible value of $(C_2 u_i)^2$ is 1, with equality if and only if the other components are zero. Similarly, since $C_1 = 365/5.4$, a component that was entirely a time shift would have $v_i = \pm 5.4/365$ years, i.e., ± 5.4 days.

In each case in the figure, the proportions of variability due to the two parametric effects, shift and overall average temperature, are combined to give the percentage of variability due to the vector parameters. Principal component 1 is almost entirely due to the variation in overall temperature, with a small effect corresponding to a decrease in range between summer and winter. (Recall that the dotted curve corresponds to a positive multiple of the principal component curve ξ_i, and the dashed curve to a negative multiple.) Principal component 2 has some shift component, a moderate negative temperature effect, and mainly comprises the effect

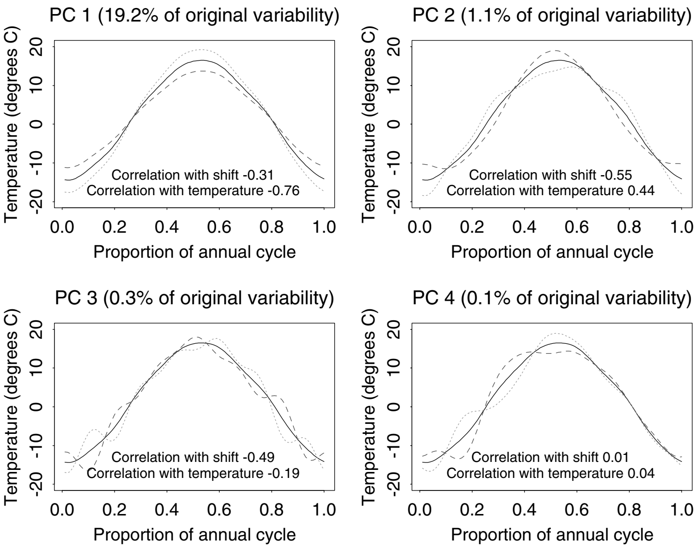

Figure 10.3. Principal component analysis carried out on the Canadian temperature curves adjusted for time shift and for annual average temperature.

of a decreased annual temperature range. Within this component, overall average temperature is positively associated with increased range, whereas in component 1 the association was negative. Principal component 1 accounts for a much larger proportion of the variability in the original data, and a slightly different approach in Section 10.5.2 shows that within the data as a whole, increased overall temperature is negatively correlated with higher range between summer and winter—colder places have more extreme temperatures.

Neither principal component 3 nor 4 contains much of an effect due to overall temperature. As before, component 3 is very largely shift, whereas component 4 corresponds to an effect unconnected to shift or overall temperature.

10.5.2 Separating out the vector component

This section demonstrates the other procedure suggested in Section 10.2. We carry out a principal components analysis on the *registered* curves $\breve{x}_i$ and then investigate the relationship between the resulting principal component scores and the parameters τ_i and θ_i arising in the registration process. Thus we analyze only the functional part of the mixed data, and the vector part is only considered later.

The effect of doing this is demonstrated in Figure 10.3. Removing the temperature and shift effects accounts for 79.2% of the variability in the original data, and the percentages of variability explained by the various principal components have been multiplied by 0.208, to make them express parts of the variability of the original data, rather than the adjusted data. For each weather station, we have a shift and annual average temperature as well as the principal component scores. Figure 10.3 shows the correlations between the score on the relevant principal component and the two parameters estimated in the registration.

We see that the components 3 and 4 in this analysis account for very little of the original variability and have no clear interpretation. Component 1 corresponds to an increase in range between winter and summer—the effect highlighted by component 2 in the previous analysis. We see that this effect is strongly negatively correlated with annual average temperature, and mildly negatively correlated with shift. Component 2 corresponds approximately to component 4 in the previous analysis, and is the effect whereby the length of summer is lengthened relative to that of winter. This effect is positively correlated with average temperature and negatively correlated with shift.