Estimating the Mean and Covariance Structure Nonparametrically When the Data are Curves

Author(s): John A. Rice and B. W. Silverman

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, Vol. 53, No. 1 (1991), pp. 233-243
Published by: Oxford University Press for the Royal Statistical Society
Stable URL: https://www.jstor.org/stable/2345738
Accessed: 06-07-2025 20:12 UTC

J. R. Statist. Soc. B (1991)
53, *No.* 1, *pp.* 233–243

Estimating the Mean and Covariance Structure Nonparametrically when the Data are Curves

By JOHN A. RICE and B. W. SILVERMAN†

University of California, San Diego, USA *University of Bath, UK*

[Received February 1989. Final revision November 1989]

SUMMARY

We develop methods for the analysis of a collection of curves which are stochastically modelled as independent realizations of a random function with an unknown mean and covariance structure. We propose a method of estimating the mean function nonparametrically under the assumption that it is smooth. We suggest a variant on the usual form of cross-validation for choosing the degree of smoothing to be employed. This method of cross-validation, which consists of deleting entire sample curves, has the advantage that it does not require that the covariance structure be known or estimated. In the estimation of the covariance structure, we are primarily concerned with models in which the first few eigenfunctions are smooth and the eigenvalues decay rapidly, so that the variability is predominantly of large scale. We propose smooth nonparametric estimates of the eigenfunctions and a suitable method of cross-validation to determine the amount of smoothing. Our methods are applied to data on the gaits of a group of 5-year-old children.

Keywords: CROSS-VALIDATION; CURVE ESTIMATION; EIGENFUNCTION EXPANSIONS; GAIT ANALYSIS; PRINCIPAL COMPONENTS; SMOOTHING

1. INTRODUCTION

In many experiments the basic observed responses are curves, rather than single data points, or, as in the classical multivariate case, vectors the elements of which are not tied together in any particular order. For instance, curves arise naturally as observations in the investigation of growth, where a separate growth curve is observed for each individual in a sample; Ramsay (1982) discusses other examples. Our own work is motivated by studies of walking where the basic observational unit is a curve that is the motion of a particular joint; see, for example, Sutherland *et al.* (1988). In this paper, we consider the extension to curve data of the existing methodology for nonparametric smoothing of standard regression functions as discussed, for example, by Silverman (1985), Eubank (1988) and Müller (1988).

As an illustrative example we use part of an extensive set of data collected by the Motion Analysis Laboratory at Children's Hospital, San Diego, California; see Olshen *et al.* (1989) for full details. The data are records of the angular rotations (in the saggital, or side, plane) of the knee and hip of 39 normal 5-year-old children; we shall concentrate mainly on the hip data. The observations are taken over a gait cycle consisting of one (double) step taken by each child. Fig. 1 gives a superposition of the 39 observed hip curves. It can be seen that the curves for different subjects exhibit similar shapes, showing a slow downswing followed by a more rapid upswing, but

†*Address for correspondence*: School of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK.

 0035–9246/91/53233 $2.00

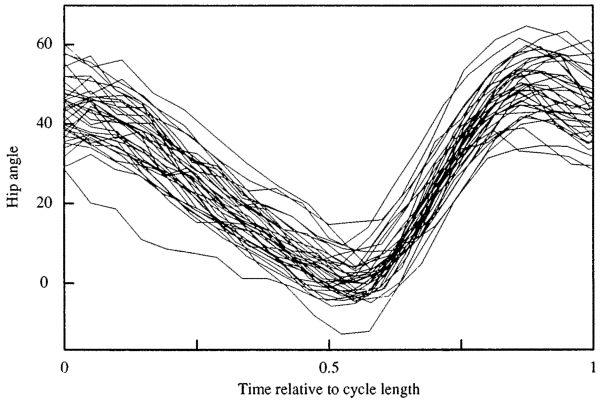

Fig. 1. Observed records of angular rotations of the hip in the saggital plane over one gait cycle in each of 39 subjects

there is considerable variability. In Section 2 we shall discuss the estimation of an overall mean function while the question of the analysis of the variability about this mean will be considered in Section 3. The full data have the interesting aspect that there are several curves per subject, corresponding to the motions of different joints; the extension of our ideas to deal with such multicurve data is discussed in Section 4.

We shall model the sample curves as being independent realizations of a stochastic process $X(t)$ that has mean $E\{X(t)\} = \mu(t)$ and covariance function $\text{cov}\{X(s), X(t)\} = \gamma(s, t)$. We assume that there is an orthogonal expansion (in the L^2 sense) of γ in terms of eigenfunctions:

$$\gamma(s, t) = \sum_{\nu} \gamma_\nu \phi_\nu(s) \phi_\nu(t).$$

A random curve from the population may then be expressed as

$$X(t) = \mu(t) + \sum_{\nu} \xi_\nu \phi_\nu(t)$$

where the ξ_ν are uncorrelated random variables with zero means and variances $E\xi_\nu^2 = \gamma_\nu$. The deviation of each sample function from the mean is thus a sum of orthogonal curves with uncorrelated random amplitudes. We shall suppose that the mean curve and the first few eigenfunctions are smooth and the eigenvalues γ_ν tend to zero rather rapidly so that the variability is predominantly of large scale and is well described as a sum of a small number of smooth curves with random amplitudes. We note that a model in which $X(t)$ is composed of a smooth sample function plus additive white noise can be incorporated into this framework by observing that the effect of white noise is to add a constant to each eigenvalue.

For statistical analysis, we assume a collection of n sample curves, each observed at times $t_1, t_2, \ldots, t_p$. The jth point on the ith curve will be denoted by $X_i(t_j) = X_{ij}$. Let Γ be the matrix with elements $\gamma(t_j, t_k)$, the covariance matrix of the observations on each curve. In our example the observations on each subject are taken at equally spaced intervals over the gait cycle, but the number of observations per subject ranges from 16 to 22 observations per cycle. Linear interpolation was used to produce 20

equally spaced observations per sample curve. We discuss the effects of this and other possible interpolation procedures, and their connection with numerical quadrature, in Section 5.

2. ESTIMATING THE MEAN

In this section we shall consider the estimation of the overall mean function $\mu(t)$ of a family of observed curves, a problem also discussed by Hart and Wehrly (1986). Let us first ignore the time sequence t_j and consider the estimation of the vector $\mu = \mu(t_j)$ as a standard problem in multivariate analysis. We have, for $i, k = 1, \ldots, n$ and $j, l = 1, \ldots, p$, $EX_{ij} = \mu_j$ and $\operatorname{cov}(X_{ij}, X_{kl}) = \delta_{ik}\Gamma_{jl}$. The standard Gauss–Markov theory of the general linear model shows that, regardless of the variance matrix Γ, the generalized least squares estimate of μ is obtained by averaging the data values separately at each time point t_j. If, as in many contexts, it is appropriate to assume that μ is a smooth curve, then the accuracy of estimation may be improved further by smoothing in some way.

One natural smoothing approach is to use penalized least squares. Let $\mathbf{X}_i$ be the vector $(X_{i1}, \ldots, X_{ip})^{\mathrm{T}}$. We estimate μ by the curve $\hat{\mu}$ that minimizes

$$n^{-1} \sum_i \|\mathbf{X}_i - \mu\|^2 + \alpha \int \mu''(t)^2 \, dt. \tag{2.1}$$

Here α is a positive smoothing parameter, and the integrated squared second derivative term quantifies the 'roughness' of μ. Minimizing expression (2.1) is equivalent to minimizing $\Sigma_j \{\bar{X}_j - \mu(t_j)\}^2 + \alpha \int \mu''(t)^2 \, dt$. This is the well-known spline smoothing approach (see, for example, Silverman (1985)) applied to the pointwise averages. The estimator $\hat{\mu}$ is a particular cubic spline with knots at the data points t_j and can be calculated in a number of operations linear in the number of data values.

In practice it is often satisfactory to choose the smoothing parameter α subjectively, but there has been considerable interest in automatic methods of choosing the smoothing parameter. A popular method is cross-validation, which usually works by leaving out data points one at a time, and choosing the value of α that works best for repredicting the data points omitted. In the context that we are considering, it is much more natural to leave out an entire observed curve at a time. For each k, let $\bar{\mathbf{X}}^{-k}$ be the vector of pointwise means calculated from the data leaving out curve k,

$$\bar{X}_j^{-k} = (n-1)^{-1} \sum_{i \neq k} X_{ij}.$$

Let $\hat{\mu}^{-k}$ be the estimate of μ obtained by minimizing

$$\sum_j \{\bar{X}_j^{-k} - \mu(t_j)\}^2 + \alpha \int \mu''(t)^2 \, dt. \tag{2.2}$$

Then the cross-validation score

$$S(\alpha) = \sum_{k=1}^{n} \sum_{j=1}^{p} \{X_{kj} - \hat{\mu}^{-k}(t_j)\}^2 \tag{2.3}$$

quantifies the predictive efficacy of estimates based on α. Any convenient numerical

search procedure can be used to minimize $S(\alpha)$ and hence to give an automatic choice of α.

Some computational economy is available by re-expressing formula (2.2). Let $A(\alpha)$ be the so-called *hat matrix* that maps the vector with components $\bar{X}_j$ to the vector of values $\hat{\mu}(t_j)$. Let $\hat{\mathbf{X}}^{(k)}$ be the vector obtained by smoothing just the kth curve, obtained by using the same algorithm as that used to find $\hat{\mu}$, but using the values X_{kj} instead of the values $\bar{X}_j$. For each k and j, we then have

$$\begin{aligned} X_{kj} - \hat{\mu}_j^{-k} &= X_{kj} - \sum_l A(\alpha)_{jl} \bar{X}_l^{-k} = X_{kj} - \sum_l A(\alpha)_{jl} \{(n-1)^{-1} n \bar{X}_l - (n-1)^{-1} X_{kl}\} \\ &= X_{kj} - \hat{\mu}(t_j) + (n-1)^{-1} \{\hat{X}_j^{(k)} - \hat{\mu}(t_j)\}. \end{aligned} \tag{2.4}$$

Use of the formula (2.4) in equation (2.3) makes the computation slightly simpler, and the computation of $S(\alpha)$ will then require $O(np)$ computer operations for each α.

What is the advantage of leaving out whole curves rather than individual data points when constructing a cross-validation score? A partial answer is given by studying the expectation properties of $S(\alpha)$. Define $M_n(\alpha)$ to be the summed mean-square error $\mathbf{E}\Sigma_j \{\hat{\mu}(t_j) - \mu(t_j)\}^2$ where $\hat{\mu}$ is obtained from a sample of n independent curves X_i. Then, since the individual curves each have the same distribution, we have

$$\begin{aligned} \mathbf{E} n^{-1} S(\alpha) &= \mathbf{E} n^{-1} \sum_{k=1}^{n} \sum_{j=1}^{p} (X_{kj} - \hat{\mu}_j^{-k})^2 = \sum_{j=1}^{p} \mathbf{E}(X_{nj} - \hat{\mu}_j^{-n})^2 \\ &= \sum_{j=1}^{p} \mathbf{E}(X_{nj} - \mu_j)^2 + \sum_{j=1}^{p} \mathbf{E}(\hat{\mu}_j^{-n} - \mu_j)^2 - 2 \sum_{j=1}^{p} \mathbf{E}(X_{nj} - \mu_j)(\hat{\mu}^{-n} - \mu_j) \\ &= \operatorname{tr}(\Gamma) + M_{n-1}(\alpha) \end{aligned} \tag{2.5}$$

since X_{nj} is independent of $\hat{\mu}_j^{-n}$ and $\mathbf{E}(X_{nj} - \mu_j) = 0$. Equation (2.5) shows that, regardless of the covariance function γ, the α that minimizes $n^{-1} ES(\alpha)$ will be the minimizer of $M_{n-1}(\alpha)$. Provided that n is at least moderate in size, the α that minimizes $M_{n-1}(\alpha)$ will be close to achieving a minimum of $M_n(\alpha)$. Thus, regardless of γ, minimizing the cross-validation score $S(\alpha)$ should give a good value of the smoothing parameter. Because of the presence of the covariance γ, the more common cross-validation score constructed by leaving out *individual* data points will not yield approximately an unbiased estimate of summed mean-square error. A cross-validation procedure based on leaving out values $\bar{X}_j$ will fail for the same reason. Our approach of deriving a procedure that does not depend on the covariance is in contrast with that of Hart and Wehrly (1986), who construct an estimate of $M_n(\alpha)$ based on a prior estimate of the covariance, and Diggle and Hutchison (1989) who assume a parametric form for the covariance. In our particular example, the automatic smoothing procedure has little effect, presumably both because the number of curves is relatively large and because the random components in each sample curve have little high frequency content.

3. ANALYSIS OF VARIABILITY: UNIVARIATE CASE

In this section we consider estimating the eigenfunctions of the covariance kernel γ. There is a variety of possible reasons for interest in the eigenfunctions and eigen-

values. They can often be interpreted physically, giving insight into, and a simple description of, the large-scale features of variability; for specific applications see Winant *et al.* (1975) and Servain and Legler (1986). The first few eigenfunctions and eigenvalues can be used for data reduction, as a part of feature extraction, for example. The scores of individual curves on the leading eigenfunctions can be used for description, clustering, classification and prediction; see Aubrey *et al.* (1980). Individual curves corrupted by white noise can be smoothed by expanding them in orthogonal eigenfunctions rather than in terms of arbitrarily chosen classical orthogonal functions with the expectation that the former expansion will converge more rapidly. We note that some smoothing techniques, such as spline smoothing and kriging, amount to using implicitly or explicitly chosen parametric covariance structures and their associated eigenfunctions for smoothing.

To describe our nonparametric procedure we first discuss the estimation of the eigenfunction corresponding to the largest eigenvalue. We assume throughout that the mean curve has been subtracted from the individual curves. It will be assumed without loss of generality that eigenvectors are normalized to have unit length. Let $\mathbf{G}$ denote the sample covariance matrix. A class of estimates is obtained by bounding the roughness of the eigenfunction, maximizing $\mathbf{u}^T\mathbf{G}\mathbf{u}$ subject to $\|\mathbf{u}\| = 1$ and $\mathbf{u}^T\mathbf{D}\mathbf{u} \leqslant \beta$, where β is a smoothing parameter and $\mathbf{D}$ is a roughening matrix for example of the form $\mathbf{F}^T\mathbf{F}$ where $\mathbf{F}$ is a second-differencing operator. An equivalent formulation, where the eigenfunction is penalized for roughness, is to maximize $\mathbf{u}^T(\mathbf{G} - \lambda\mathbf{D})\mathbf{u}$ subject to $\|\mathbf{u}\| = 1$. The solution to this problem, $\mathbf{u}_1(\lambda)$, is the eigenvector of $\mathbf{G} - \lambda\mathbf{D}$ with largest eigenvalue. To understand the effect of the smoothing parameter λ it is useful to note that the smallest eigenvalues of $\mathbf{D}$ have associated eigenfunctions which are smooth; for example, if $\mathbf{D}$ is constructed from a second-difference operator as above, it has two zero eigenvalues corresponding to constant and linear functions, and the eigenfunction corresponding to the kth smallest eigenvalue is roughly a polynomial of degree $k-1$. Thus, as λ increases, the solution $\mathbf{u}_1(\lambda)$ is pulled towards smooth functions, because it will tend to lie in the part of the space spanned by eigenfunctions of $\mathbf{D}$ with small eigenvalues. The largest eigenvalue of $\mathbf{G}$, $g_1(0) = \mathbf{u}_1(0)^T\mathbf{G}\mathbf{u}_1(0)$, is well known to be positively biased, a consequence of Jensen's inequality. The eigenvalue associated with the smoothed eigenvector $g_1(\lambda) = \mathbf{u}_1(\lambda)^T\mathbf{G}\mathbf{u}_1(\lambda)$ is necessarily smaller than $g_1(0)$, so that smoothing by an appropriate amount will reduce this bias. The estimation of smooth eigenfunctions has also been addressed by Besse and Ramsay (1986) from a rather different point of view.

An alternative characterization of the first eigenfunction $\mathbf{u}_1(0)$ of $\mathbf{G}$ is that it minimizes

$$\sum_{i=1}^{n} \|\mathbf{x}_i - (\mathbf{x}_i^T\mathbf{u})\mathbf{u}\|^2$$

and thus functions as a best predictor, in that the individual realizations $\mathbf{x}$ obtained from the population of curves can each be well modelled as a multiple of $\mathbf{u}_1(0)$. Playing this role, $\mathbf{u}_1(0)$ will be overly sensitive to individual data points and will tend to track high frequency components which are not present in the population eigenfunction $\phi_1(t)$. Constraining the solution to be smooth will counteract this tendency and will hopefully result in a predictor with smaller expected mean-square error.

The smoothing parameter λ can be chosen subjectively by plotting smoothed

vectors against the unsmoothed vector. We have also used cross-validation, carried out in the following way, analogous to the procedure suggested for the estimation of the mean curve: Let $\mathbf{G}^{(-i)}$ be the covariance matrix of the data leaving the ith curve out and let $\mathbf{u}_1^{(-i)}(\lambda)$ be the leading eigenvector of $\mathbf{G}^{(-i)} - \lambda\mathbf{D}$. The cross-validation score of λ is then

$$\mathrm{CV}(\lambda) = \sum_{i=1}^{n} \|\mathbf{x}_i - \{\mathbf{x}_i^{\mathrm{T}}\mathbf{u}_1^{(-i)}(\lambda)\}\mathbf{u}_1^{(-i)}(\lambda)\|^2. \tag{3.1}$$

We then choose the value of λ that minimizes this estimate of expected prediction error. As a guide to the range of values of λ to consider, we have found it effective to use the largest eigenvalue of $\mathbf{G}$ divided by a small positive eigenvalue of $\mathbf{D}$. The computations can be greatly speeded up by noting that, if ω_1 is the largest eigenvalue of $\mathbf{G} - \lambda\mathbf{D}$, the largest eigenvalue of $G^{(-i)} - \lambda\mathbf{D}$ for each i is necessarily in the interval $[\omega_1 - \|\mathbf{x}_i\|^2/n, \omega_1]$; standard techniques of numerical linear algebra can then be used to locate the eigenvalue precisely.

An analogue of result (2.5) can be developed in the present context. Let $\hat{\mathbf{u}}$ be an estimate of the leading eigenvector of a covariance matrix Γ based on a sample of size n. Let $\mathbf{X}$ be a random vector with mean zero and covariance matrix Γ, independent of the sample from which $\hat{\mathbf{u}}$ is constructed. Then a measure of the efficacy of $\hat{\mathbf{u}}$ as an estimate is given by the mean-square prediction error defined by

$$L_n(\hat{\mathbf{u}}) = E_{\mathbf{X}}\|\mathbf{X} - (\mathbf{X}^{\mathrm{T}}\hat{\mathbf{u}})\hat{\mathbf{u}}\|^2 = \mathrm{tr}(\Gamma) - \hat{\mathbf{u}}^{\mathrm{T}}\Gamma\hat{\mathbf{u}},$$

by the assumption that $\|\hat{\mathbf{u}}\| = 1$. Now define $R_n(\lambda) = E[L_n\{\mathbf{u}_1(\lambda)\}]$. Our aim is to choose a value of λ which yields a small value of $R_n(\lambda)$. An argument similar to that yielding result (2.5) shows that $n^{-1}\,\mathrm{CV}(\lambda)$ is an unbiased estimate of $R_{n-1}(\lambda)$. Thus the minimizer of $\mathrm{CV}(\lambda)$ is hopefully close to the minimizer of $R_n(\lambda)$.

We now turn to computation of the higher order eigenfunctions. Suppose that smooth orthogonal eigenfunctions $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_k$ have been found. Let $\mathbf{Q} = [\mathbf{Q}_1 \quad \mathbf{Q}_2]$ be an orthogonal matrix where the columns of $\mathbf{Q}_1$ consist of the k eigenfunctions already found and $\mathbf{Q}_2$ is an $n \times (n-k)$ matrix. We now wish to find a vector $\mathbf{u}_{k+1}(\lambda)$ to maximize $\mathbf{u}^{\mathrm{T}}(\mathbf{G} - \lambda\mathbf{D})\mathbf{u}$ subject to $\|\mathbf{u}\| = 1$ and the additional constraint $\mathbf{Q}_1^{\mathrm{T}}\mathbf{u} = \mathbf{0}$. This amounts to maximizing a quadratic form over a hyperplane and so can be accomplished by solving a lower dimensional eigenvalue problem. If $\mathbf{v}(\lambda)$ is the leading eigenvector of $\mathbf{Q}_2^{\mathrm{T}}(\mathbf{G} - \lambda\mathbf{D})\mathbf{Q}_2$, the solution to the constrained problem is $\mathbf{u}_{k+1}(\lambda) = \mathbf{Q}_2\,\mathbf{v}(\lambda)$.

To construct a cross-validation procedure for the higher order eigenfunctions, note that $\mathbf{v}(\lambda)$ is an estimate of the leading eigenvector of $\mathbf{Q}_2^{\mathrm{T}}\Gamma\mathbf{Q}_2$, the covariance matrix of the projected data $\mathbf{Q}_2^{\mathrm{T}}\mathbf{x}_i$. Exactly the same arguments as before then yield the cross-validation score

$$\mathrm{CV}_{k+1}(\lambda) = \sum_{i=1}^{n} \|\mathbf{Q}_2^{\mathrm{T}}\mathbf{x}_i - \{(\mathbf{Q}_2^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\,\mathbf{v}^{(-i)}(\lambda)\}\mathbf{v}^{(-i)}(\lambda)\|^2 \tag{3.2}$$

where $\mathbf{v}^{-i}(\lambda)$ is the leading eigenvector of $\mathbf{Q}_2^{\mathrm{T}}(\mathbf{G}^{(-i)} - \lambda\mathbf{D})\mathbf{Q}_2$. The form (3.2) is the most convenient for calculation because the length of the vectors, and hence the burden of the calculation, decreases as the program proceeds through higher eigenvalues. However, it is interesting that, substituting $\mathbf{u}_{k+1}^{-i}(\lambda) = \mathbf{Q}_2\,\mathbf{v}(\lambda)$, we

obtain from equation (3.2) after some easy manipulations exactly the same form as equation (3.1),

$$CV_{k+1}(\lambda) = \sum_{i=1}^{n} \|\mathbf{x}_i - \{\mathbf{x}_i^T \mathbf{u}_{k+1}^{(-i)}(\lambda)\}\mathbf{u}_{k+1}^{(-i)}(\lambda)\|^2. \tag{3.3}$$

As an example, we consider the curves of saggital hip angles displayed in Fig. 1. A plot of the empirical covariance function shows that the variance is higher at the ends of the cycle than in mid-cycle, that the covariance is everywhere positive and that the correlation between observations is smallest when they are taken about half a cycle apart. The first four eigenvalues of the covariance function are 652, 105, 77 and 33. These together account for 95% of the variability. Smoothing parameters were chosen by cross-validation as described earlier and the eigenvalues corresponding to the smoothed vectors were 631, 102, 75 and 32. Note the reduction over the unsmoothed values, especially for the first eigenvector. The roughness matrix was constructed to penalize second differences, periodically extended. The raw and smoothed eigenvectors are shown in Fig. 2. The first smoothed eigenfunction is nearly symmetric about the mid-cycle point and, although it resembles a cosine in shape, is of the same sign throughout the cycle, reflecting the positive nature of the covariance surface. Deviation from the mean for a single random curve is described by a random multiple of this eigenfunction, i.e. an overall positive or negative shift, the shift being slightly greater at the beginning and end of the cycle.

A useful way of presenting variability in the data and suggesting interpretations of the eigenfunctions is to examine individual curves that have high and low scores on particular eigenfunctions. For example, Fig. 3 shows the sample mean curve and the curves of the two individuals who had the highest and lowest scores on the second eigenfunction. Examination of these curves together with the eigenfunction plotted in Fig. 2 shows that the mode of variability captured by this eigenfunction is one in which a greater than average angle at the beginning of the gait cycle is associated with a delay in the second phase of the cycle. Similar figures confirm the remarks made earlier about the first eigenfunction and indicate that the third eigenfunction is associated with flatness or springiness of the cycle; large scores on this eigenfunction are associated with relatively high peaks and a low trough.

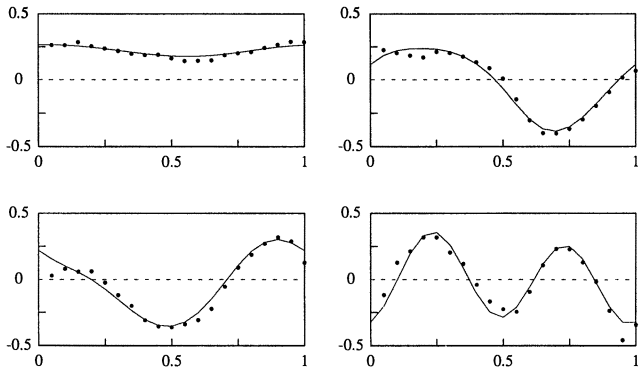

Fig. 2. First four raw eigenvectors (●) and smoothed eigenfunctions (———) of the covariance of the hip angle population

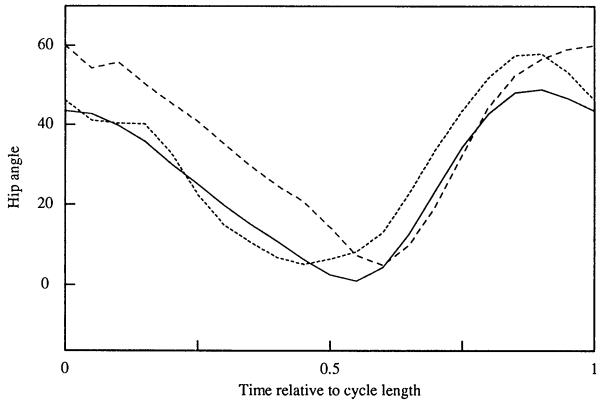

Fig. 3. Mean curve for the hip population (———) and the most extreme sample curves (----, ------) with respect to the second eigenfunction

To explore the connections with spectral analysis, we compared truncated expansions of data curves in terms of the estimated eigenfunctions with those in the same number of Fourier components. The latter choice would be most appropriate if the variability were stationary, since these would be the eigenfunctions. The expansions were all of the form $\hat{\mu} + \{(\mathbf{x} - \hat{\mu})^{\mathrm{T}}\mathbf{u}_1\}\mathbf{u}_1 + \{(\mathbf{x} - \hat{\mu})^{\mathrm{T}}\mathbf{u}_2\}\mathbf{u}_2 + \{(\mathbf{x} - \hat{\mu})^{\mathrm{T}}\mathbf{u}_3\}\mathbf{u}_3$, where $\mathbf{x}$ is a vector of data points, $\hat{\mu}$ the estimate of the overall mean vector and $\mathbf{u}_j$ the jth smooth eigenvector or normalized Fourier function. As we would perhaps expect, the approximations based on smooth eigenfunctions were usually somewhat closer to the data curves, achieving a root-mean-square error typically around two-thirds of those yielded by the Fourier approximation. Fourier analysis has been used to depress high frequency noise introduced by the digitization process (Sutherland *et al.*, 1988). As we remarked earlier, an expansion in smooth eigenfunctions could also be used for smoothing.

4. ANALYSIS OF VARIABILITY: MULTIVARIATE CASE

An unusual, and very interesting, aspect of this set of data is that there are several curves per subject, and we therefore wish to explore methods for describing and analysing the relationships between the curves—how the knee and hip move jointly, for example. As a subject progresses through a gait cycle, the hip and knee angles can be thought of as realizations $h(t)$ and $k(t)$ of random curves from populations with mean curves $\mu_H(t)$ and $\mu_K(t)$. An attractive way of thinking about these data, in a way that takes into account the dependence between them, is to consider the closed curve $(h(t), k(t))$ in the plane traced out as the cycle takes place. Each subject's curve can be viewed as a random dilation of the mean curve $(\mu_H(t), \mu_K(t))$. To explore the variation of these dilations, we carry out our smooth eigenfunction analysis on the n vectors $(h_i(t_1), \ldots, h_i(t_p), k_i(t_1), \ldots, k_i(t_p))$. The penalty matrix is of block diagonal form so that elements of h and k are not linked, but the covariance matrix allows correlation between hip and knee angles.

Fig. 4 displays the first two modes of dilation found in this way. The smoothing parameters were chosen by cross-validation. In the left-hand display, the dots show

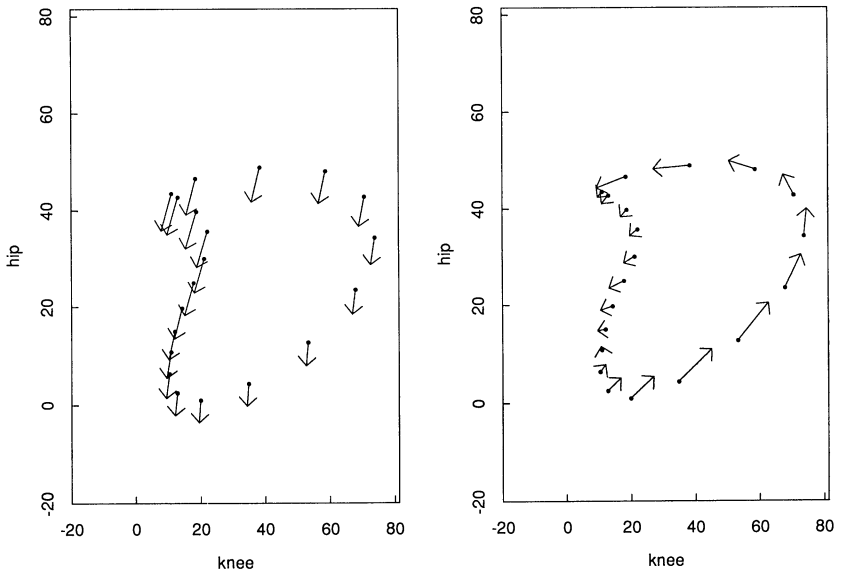

Fig. 4. Mean hip–knee gait cycle curve and its first two modes of random dilation

the mean curve and the arrows show the direction (modulo reflection) and the relative magnitudes of the first mode of random dilation. The pattern corresponds to a differential shift in the level of the hip curves, basically that seen and discussed earlier as the first hip eigenfunction, and shows that this shift is related to a similar variation of smaller magnitude in the knee. In the second mode of dilation, the greatest variation is in the latter part of the cycle—the right-hand section of the plot. The tendency of the tip of each vector during this phase of the cycle to point to the tail of the following vector suggests a kind of phase shift of this part of the gait cycle that takes place in both the hip and the knee curves if it takes place at all. It would clearly be possible to extend this form of analysis to higher dimensions (e.g. hip, thigh, knee and ankle), but presentation and thus interpretation of the results would be more difficult.

5. A MORE GENERAL FRAMEWORK

In this section we shall discuss briefly the extension of our methodology to the continuous time case, both for its own sake and for the insight that it gives into more general discrete time problems. Some related work is presented by Bouhaddou *et al.* (1987). Suppose that we are working on a fixed time interval, without loss of generality the unit interval, and that $X_1(t), \ldots, X_n(t)$ are observed sample curves. Let $\langle u, v \rangle$ be the inner product $\int u(t)\, v(t)\, dt$ and write $\|u\|^2 = \langle u, u \rangle$. Let $\Phi(u)$ be the roughness penalty $\int u''^2$. Our approach would use as an estimate of the mean function the curve μ that minimized $n^{-1}\Sigma\|X_i - \mu\|^2 + \alpha\, \Phi(\mu)$. An appropriate cross-validation score $S(\alpha)$ would be $n^{-1}\Sigma\|X_j - \hat{\mu}^{(-j)}\|^2$, where $\hat{\mu}^{(-j)}$ is the estimate obtained with smoothing parameter α leaving out curve X_j. It is trivial to prove that $E\, S(\alpha) = M_{n-1}(\alpha)$ where $M_n(\alpha)$ is the mean integrated square error $\mathbf{E}\|\hat{\mu} - \mu\|^2$. This holds regardless of the covariance function $\gamma(s, t)$.

Turning to the eigenfunctions of the variance structure, assume that the mean

function has been subtracted off. Our estimate of the leading eigenfunction is found by maximizing $\Sigma\langle u, X_i\rangle^2 - \lambda\,\Phi(u)$ subject to $\|u\| = 1$. A cross-validation choice of λ can be made by minimizing $n^{-1}\Sigma\|X_i - \langle u_1^{(-i)}(\lambda), X_i\rangle u_1^{(-i)}(\lambda)\|^2$ or equivalently by maximizing $n^{-1}\Sigma\langle u_1^{(-i)}(\lambda), X_i\rangle^2$. Again it is trivial to show that the cross-validation score is an unbiased estimator of the risk for an estimate based on $n-1$ sample curves. Higher order eigenfunctions are estimated in turn by performing the same calculations with the additional constraint that, for the jth eigenfunction u_j, $\langle u_j, \hat{u}_l\rangle = 0$ for $l = 1, \ldots, j-1$.

To apply the continuous formulation in practice, some discretization is inevitable. One natural approach is to characterize curves u in some discrete way, e.g. by specifying values on a fine fixed regular grid $t_1, \ldots, t_p$, as a vector $\mathbf{u}$. (Any computable, and hence finite dimensional, approximation to u, such as a spline, can be parameterized through its values on some finite set of points.) It is then necessary to use quadrature approximations to the inner product $\langle\cdot,\cdot\rangle$ and the roughness penalty Φ. The roughness penalty is most naturally approximated by a quadratic form such as $\mathbf{u}^T\mathbf{G}\mathbf{u}$ as used in Section 3. The simplest approximation for the inner product is to use $\langle u, v\rangle = p^{-1}\Sigma u(t_j)\, v(t_j)$. If this is done then all the calculations to be performed are exactly the same as those which we have set out earlier in the paper.

When values of the curves X_i are not necessarily available at the points t_j, some other quadrature approximation is needed. For example, suppose that the curve X_i is observed at discrete points $\tau_1, \tau_2, \ldots, \tau_q$ (not necessarily the same for each i). To implement our procedure we then need to be able to approximate $\langle u, x\rangle$ where u is known at the t_j and x is known at possibly different τ_k. A simple quadrature approximation is $\Sigma w_{jk}\, u(t_j)\, x(\tau_k)$ for some system of weights w_{jk}, e.g.

$$
\begin{aligned}
w_{jk} &= p^{-1}(\tau_{k+1} - t_j)/(\tau_{k+1} - \tau_k) && \text{if } \tau_k \leqslant t_j \leqslant \tau_{k+1},\\
w_{jk} &= p^{-1}(t_j - \tau_{k-1})/(\tau_k - \tau_{k-1}) && \text{if } \tau_{k-1} \leqslant t_j \leqslant \tau_k,\\
w_{jk} &= 0 \text{ otherwise.}
\end{aligned}
$$

This weight system has the property that if the t_j and τ_k are identical then $w_{jk} = p^{-1}\delta_{jk}$, and thus provides a generalization of the procedure already used where the $x(t_j)$ are directly available. In other cases it corresponds to linearly interpolating the available points on the curves x to give values $x(t_j)$ and then proceeding in exactly the way that we have described in detail. This is precisely what was done for the gait data. An interesting topic for further investigation would be the use of alternative approaches to the discrete representation of u and the quadrature approximation used; it is perhaps worth pointing out that every linear quadrature scheme $\Sigma w_{jk}\, u(t_j)\, x(\tau_k)$ is in a sense equivalent to an interpolation of the x to give values at the t_j followed by exact integration of an interpolation to the $u(t_j)\, x(t_j)$.

6. FURTHER ISSUES

In geophysical and atmospheric sciences, the sample curves are often recorded as functions of both space and time. Our analysis has corresponded to viewing curves taken at different points in time as replicate samples, but principal components analysis in these areas is usually regarded as a decomposition of the data into uncorrelated functions of both space and time, and the first few principal components often vary smoothly in time, reflecting seasonal variation, as well as in space. Our analysis

can be interpreted as focusing on only one of these dimensions.

Although we believe that our computations demonstrate the practical utility of our approach, a theoretical investigation of the sampling properties of the curve estimation procedures remains to be done. As an example of the sort of result which might be obtainable, we conjecture that for some range of values of the smoothing parameter the solution to the penalized criterion for eigenfunction estimation has smaller risk in the sense defined in Section 3 than the standard estimate and that the cross-validation procedure produces a consistent estimate of the optimal smoothing parameter. The literature on density estimation and single-curve nonparametric regression, e.g. Stone (1984), Rice (1984) and Härdle *et al.* (1988), strongly suggests that relevant theorems could be proved under suitable conditions, but at the cost of complicated technical arguments.

ACKNOWLEDGEMENTS

The support of the Science and Engineering Research Council and the National Science Foundation (grant DMS-8733306) is gratefully acknowledged.

REFERENCES

Aubrey, D. G., Inman, D. L. and Winant, C. D. (1980) The statistical prediction of beach changes in Southern California. *J. Geophys. Res.*, **85**, 3264–3276.

Besse, P. and Ramsay, J. O. (1986) Principal components analysis of sampled functions. *Psychometrika*, **51**, 285–311.

Bouhaddou, O., Obled, C. H. and Dinh, T. P. (1987) Principal component analysis and interpolation of stochastic processes: methods and simulation. *J. Appl. Statist.*, **14**, 251–267.

Diggle, P. J. and Hutchison, M. F. (1989) On spline smoothing with autocorrelated errors. *Aust. J. Statist.*, **31**, 166–182.

Eubank, R. L. (1988) *Spline Smoothing and Nonparametric Regression*. New York: Dekker.

Härdle, W., Hall, P. and Marron, J. S. (1988) How far are automatically chosen regression smoothing parameters from their optimum? *J. Am. Statist. Ass.*, **83**, 86–95.

Hart, J. D. and Wehrly, T. E. (1986) Kernel regression estimation using repeated measurements data. *J. Am. Statist. Ass.*, **81**, 1080–1088.

Müller, H.-G. (1988) Nonparametric regression analysis of longitudinal data. *Lect. Notes Statist.*, **46**.

Olshen, R. A., Biden, E. N., Wyatt, M. P. and Sutherland, D. H. (1989) Gait analysis and the bootstrap. *Ann. Statist.*, **17**, 1419–1440.

Ramsay, J. O. (1982) When the data are functions. *Psychometrika*, **47**, 379–396.

Rice, J. A. (1984) Bandwidth choice for nonparametric regression. *Ann. Statist.*, **12**, 1215–1230.

Servain, J. and Legler, D. M. (1986) Empirical orthogonal function analyses of tropical atlantic sea surface temperature and wind stress: 1964–1979. *J. Geophys. Res.*, **91**, 14181–14191.

Silverman, B. W. (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. R. Statist. Soc.* B, **47**, 1–52.

Stone, C. (1984) An asymptotically optimal window selection rule for kernel density estimates. *Ann. Statist.*, **12**, 1285–1297.

Sutherland, D. H., Olshen, R. A., Biden, E. N. and Wyatt, M. P. (1988) *The Development of Mature Walking*. Oxford: Blackwell Scientific.

Winant, C. D., Inman, D. L. and Nordstrom, C. E. (1975) Description of seasonal beach changes using empirical eigenfunctions. *J. Geophys. Res.*, **80**, 1979–1986.