Some Tools for Functional Data Analysis

Author(s): J. O. Ramsay and C. J. Dalzell

Source: *Journal of the Royal Statistical Society. Series B (Methodological)*, 1991, Vol. 53, No. 3 (1991), pp. 539-572

Published by: Wiley for the Royal Statistical Society

Stable URL: https://www.jstor.org/stable/2345586

J. R. Statist. Soc. B (1991)
53, *No.* 3, *pp.* 539–572

Some Tools for Functional Data Analysis

By J. O. RAMSAY† and C. J. DALZELL

McGill University, Montreal, Canada *Memorial University, St Johns, Canada*

[*Read before* The Royal Statistical Society *at a meeting organized by the* Research Section *on Wednesday, January 16th, 1991,* Dr F. Critchley *in the Chair*]

SUMMARY

Multivariate data analysis permits the study of observations which are finite sets of numbers, but modern data collection situations can involve data, or the processes giving rise to them, which are functions. Functional data analysis involves infinite dimensional processes and/or data. The paper shows how the theory of L-splines can support generalizations of linear modelling and principal components analysis to samples drawn from random functions. Spline smoothing rests on a partition of a function space into two orthogonal subspaces, one of which contains the obvious or structural components of variation among a set of observed functions, and the other of which contains residual components. This partitioning is achieved through the use of a linear differential operator, and we show how the theory of polynomial splines can be applied more generally with an arbitrary operator and associated boundary constraints. These data analysis tools are illustrated by a study of variation in temperature–precipitation patterns among some Canadian weather-stations.

Keywords: DIFFERENTIAL OPERATOR; FUNCTIONAL LINEAR MODEL; FUNCTIONAL PRINCIPAL COMPONENTS; L-SPLINES; NONPARAMETRIC REGRESSION; SMOOTHING

1. OBJECTIVES OF FUNCTIONAL DATA ANALYSIS

Functional data analysis (FDA) concerns data for which the ith observation is a real function, $x_i(t)$, $i=1, \ldots, n$, $t \in T$ where T is a real interval, and therefore each x_i is a point in some function space H. We wish to apply concepts such as linear modelling and principal components analysis familiar in multivariate data analysis (MDA) in this infinite dimensional domain.

The practical reasons for considering data analysis from a functional perspective are fourfold.

(a) Functional observations present themselves in applied contexts increasingly more frequently as automated on-line data collection facilities become accessible to more researchers. Moreover, smoothing and interpolation procedures can yield functional representations of finite sets of observations.
(b) Some modelling problems are more natural to think through in functional terms even though only finite numbers of observations are available.
(c) The objectives of an analysis can be functional in nature, as would be the case if finite data are used to estimate an entire function, its derivatives or the values of other functionals.
(d) Taking considerations such as smoothness into account for multivariate data arising from functional processes can have important implications for their analyses.

† *Address for correspondence*: Department of Psychology, McGill University, 1205 Dr Penfield Avenue, Montreal, Quebec, H3A 1B1, Canada.

 0035-9246/91/53539 $2.00

Furthermore, FDA is the natural sequel to MDA, where a datum x_i is a finite dimensional vector, just as the latter follows from the analysis of scalars. Thus the generalization of MDA to FDA has both an applied and an intrinsic interest, and may even lead to further development of MDA itself.

The most common type of data suitable for analysis by FDA are non-stationary time series with possibly unequally spaced observations. We shall sketch the application of FDA to multivariate or vector-valued functions, although the principal focus will be on univariate data. The data that will provide an applied context for our investigation of FDA are average monthly temperatures and precipitation levels at 35 Canadian weather-stations (Canadian Climate Program, 1982). Although there are only 12 mean values for each of the 35 stations, these can be assumed to arise from smooth functions, although certainly in the presence of some observational error or noise. The full curves in Fig. 1(a) display the harmonic smoothing spline temperature representations for Montreal and Resolute stations, which are located in the St Lawrence River valley and high Arctic respectively.

We shall need to recognize that much of the temperature variation for a given station can be accounted for by a shifted sine function with a period of 12 months, where

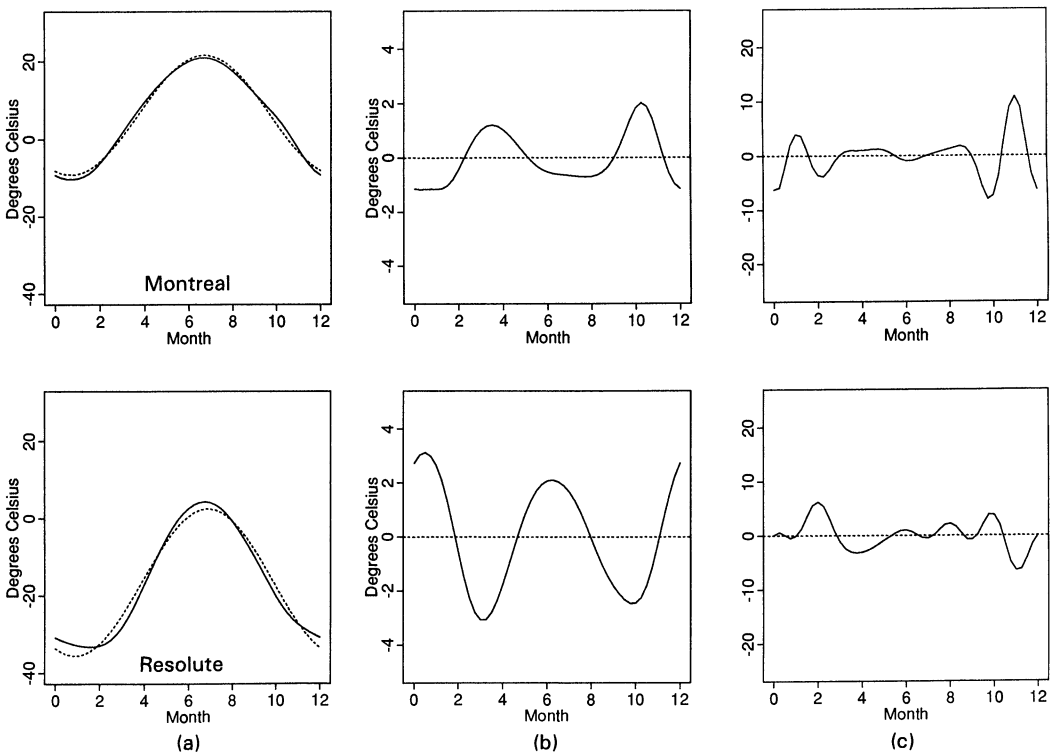

Fig. 1. (a) Mean monthly temperatures at two Canadian weather-stations (——, smoothing spline fit to these points using piecewise linear combinations of the functions (1, t, $\sin(\pi t/6)$, $\cos(\pi t/6)$, $t\sin(\pi t/6)$, $t\cos(\pi t/6)$); --------, best representation among linear combinations of (1, $\sin(\pi t/6)$, $\cos(\pi t/6)$)); (b) difference between the smooth fit and the purely sinusoidal representation; (c) result of applying the differential operator $L=(\pi/6)^2D+D^3$ to the spline smoothing curves (since linear combinations of (1, $\sin(\pi t/6)$, $\cos(\pi t/6)$) are in the null space of this operator, it represents 'harmonic acceleration' in the smoothing spline fits; all curves are constrained to be periodic)

phase and shift must be estimated separately for each station. But the atmosphere can be viewed as an extremely complex non-linear system only partially driven by a harmonic forcing function, and consequently this basic harmonic behaviour will be of secondary interest. Instead we shall be concerned to detect the effects of causal agents which will perturb this base-line behaviour. For example, although Fig. 1 indicates that temperatures at Montreal and Resolute are well approximated by sinusoidal variation (the broken line in Fig. 1(a)), each has a particular residual signature indicated by the full curve in Fig. 1(b) varying around 0 that invites further analyses.

This partitioning of functional variation into the structural and residual components in a meaningful way is the main goal of this paper. Fig. 1(c) indicates the functions $(\pi/6)^2 Dx = D^3 x$ for the same two weather-stations, where $x(t)$ is the mean temperature at t months and $D^j x$ denotes the jth derivative of x. For reasons that we shall develop later, these functions can be termed 'harmonic acceleration' and will yield further insights. Finally, functional versions of linear modelling and principal components analysis will prove helpful in identifying non-harmonic variation in the temperature functions.

Although the generalization of familiar statistical procedures such as regression analysis to observations in function space is hardly new (see for example Parzen (1961)), the topic has tended in the past to be treated in the context of stationary continuous time stochastic processes. Besse (1979) and Besse and Ramsay (1986) considered the functional principal components analysis of functions resulting from polynomial interpolation of observed values. A variety of recent developments have set the stage for a wider treatment of functional models, and this paper brings together some of these elements while at the same time tackling some challenging technical problems.

Spline smoothing and interpolation techniques represent data in function space by using a functional rationale (Eubank, 1988; Wegman and Wright, 1983; Silverman, 1985; Wahba, 1990). Although most of the current literature on splines in statistics is concerned with polynomial splines, this paper takes a more general approach involving splines which are piecewise elements of an essentially arbitrary finite dimensional function space, and is inspired by the treatment of L-splines by Kimeldorf and Wahba (1970, 1971), with extensions by Peele and Kimeldorf (1977). The monographs by Wahba (1990) and Schumaker (1981) can be consulted for more details concerning the mathematical basis for splines.

What happens when functional data analytic procedures are applied to spline and other types of smoothing functions computed from finite dimensional data? We consider the equivalences between functionally oriented analyses of these representations and corresponding analyses carried out on the original data.

In summary the definition of the FDA of a set of data which are functional either in fact or in principle usually implies the following four tasks:

(a) choice of function space in which the analysis is to take place—we comment further in this section;
(b) specification of the analysis in functional analytic terms—we do this for linear modelling and principal components analysis in Section 2;
(c) determination of how a finite dimensional observation vector is to be mapped into function space—there is a vast literature on this topic, and Section 3 does no more than to offer some notation;

(d) description of what the FDA of the functional representations of the finite observations means in terms of analysing the observations themselves—this is done in Section 4, and examples of such analyses are provided in Section 5.

In the interests of coherence and because of limitations of space, our approach to these tasks will be more geometrical than probabilistic, although probability considerations can be introduced along the lines of Wahba (1978).

1.1. *Choice of Function Space*

The analyses of functional data, $x_1, \ldots, x_n$, that are taken up in this paper presume an inner product or Hilbert space structure, i.e. $x_i \in H$ where H is a vector space of functions defined on a real interval T, and where H is complete and has an inner product $\langle , \rangle$. Of course, appropriate choices of inner product are also basic to MDA of observations $x_i \in \mathbf{R}^p$ (Eaton, 1983), with the most important example aside from $\langle x, y\rangle = x'y$ being $\langle x, y\rangle = x'\Sigma^{-1}y$, where Σ is the population variance matrix.

We begin by asking about the characteristics of Hilbert space H. If H is such that the evaluation function $\rho_s = x(s)$ is continuous for all $s \in T$, then a Hilbert function space defined on H is a *reproducing kernel Hilbert space* (RKHS) (Aronszajn, 1950). One way in which to assure the continuity of ρ_s is to impose a condition such as the existence of $m-1$ absolutely continuous derivatives and a square-integrable mth derivative, $m \geqslant 1$. Spaces of such functions will be indicated by H^m.

Associated with an RKHS is a symmetric positive bivariate function $k(s, t)$, $s, t \in T$, defined by the relation

$$\langle k(s, \ ), x\rangle = \rho_s(x) = x(s) \text{ and } k(s, \ ) \in H \text{ for all } s \text{ and } x. \tag{1}$$

In effect, $k(s, \ )$ is the representing function for function ρ_s whose existence is assured by the continuity of ρ_s and the Riesz representation theorem. In the finite dimensional case with the inner product $\langle x, y\rangle = x'\Sigma^{-1}y$ the reproducing kernel is Σ. We use this inner product for multivariate observations with covariance Σ, and correspondingly it can be shown that the functional inner product for which the reproducing kernel is k is appropriate for the analysis of certain stochastic processes with covariance function k (Kimeldorf and Wahba, 1970; Parzen, 1961; Wahba, 1978).

1.1.1. *Partitioning principle*

Does the function space H have some internal structure that will play a role in the analyses? This question is fundamental to effective data analysis, and we assume that $x = u + e$ where

(a) the *structural* component u is that which is predictable and obvious and
(b) the *residual* component e is the remainder, which may or may not yield significant insights but in any case requires a more exploratory approach.

This *partitioning principle* in the context of FDA implies that $H = H_1 \oplus H_2$. Subspace H_1 of H contains the structural components and is usually of finite dimensionality. Subspace H_2 contains the residual components and is usually infinite dimensional. Which of these two components will be of greater importance will depend on the objectives of the FDA.

The partitioning explored in this paper is achieved through the use of two linear operators, L and B. The first is a linear differential operator,

$$Lx = \sum_{j}^{m} \alpha_j D^j x, \tag{2}$$

where $D^{m-j}\alpha_j$ is an absolutely continuous function of t and $\alpha_m \neq 0$. By an appropriate choice of the differential operator L we can annihilate or filter out certain components regarded as structural by analysing Lx rather than x. The subspace H_1 is defined as $\ker(L)$ and therefore contains functions u such that $Lu = 0$ and is of dimension m.

The second linear operator B, such that $Bx \in \mathbf{R}^m$, is defined as a set of m boundary conditions or constraint functionals b_j having the property $\ker(L) \cap \ker(B) = 0$, i.e. the only function x satisfying $Lx = Bx = 0$ is $x = 0$. Subspace H_2 is defined as $\ker(B)$ and is of codimension m.

Structural component u and residual component e are associated with the subspaces $H_1 = \ker(L)$ and $H_2 = \ker(B)$ of elements of H mapped into 0 by the two linear operators. The complementarity condition $\ker(L) \cap \ker(B) = 0$ is required of L and B to achieve the partitioning.

The inner product of residual functions e and f in subspace H_2 can now be defined as

$$\langle e, f \rangle_2 = \int_T (Le)(t)(Lf)(t)\,dt, \tag{3}$$

and the corresponding inner product of u and v in subspace H_1 is

$$\langle u, v \rangle_1 = \sum_{j} b_j(u)\, b_j(v) = B(u)'\, B(v). \tag{4}$$

These induce the inner product of x and y in the entire space $H = H_1 \oplus H_2$

$$\langle x, y \rangle_0 = \langle x, y \rangle_1 + \langle x, y \rangle_2. \tag{5}$$

We shall designate the Hilbert space H equipped with this inner product as H_0. The associated norms will be indicated by $\|x\|_0$, $\|u\|_1$ and $\|e\|_2$. We shall use these inner products and norms in the next section to define various data analyses within each subspace.

Let us now turn to some examples of such structures.

1.1.2. *Example: linear trend, initial value constraints*

Let $H = H^2$ defined on $[0, T]$ and $L(x) = D^2x$, for which $\ker(L) = \text{span}\{1, t\}$. As boundary functionals we use the initial values $B(u) = (u(0), (Du)(0))'$. H_1 then contains functions of the form $c_1 + c_2 t$, with a natural inner product $\langle u, v \rangle_1 = u(0)\, v(0) + (Du)(0)(Dv)(0)$. Subspace H_2 contains functions without initial linear trend in the sense that $e(0) = (De)(0) = 0$ for all $e \in H_2$, and $\langle e, f \rangle_2 = \int (D^2e)(t)(D^2f)(t)\,dt$. One might use such a partitioning where one wished residual functions to have no linear trend at the origin.

1.1.3. *Example: linear trend, integral constraints*

This example contrasts with example 1.1.2 in the sense that, although L remains D^2, the boundary functionals are now $B(u) = (\int u(t)\,dt, \int t\, u(t)\,dt')$. By setting $B(u) = 0$

we are still defining H_1 as span $\{1, t\}$, but now with the inner product

$$\langle u, v\rangle_1 = (\int u(t)\,dt)(\int v(t)\,dt) + (\int t\,u(t)\,dt)(\int t\,v(t)\,dt).$$

This inner product is uniform in that it depends on the behaviour of u and v at all values of t rather than just on $t=0$. H_2 will have the same inner product as in the previous example, but will now consist of functions with linear trend eliminated by averaging across all values of t.

1.1.4. *Example: sinusoidal trend, integral constraints*

Now consider the analysis of functions $x \in H^3$ defined on $[0, T]$ with the operator $L = \theta^2 D + D^3$. The kernel of this operator is span$\{1, \sin(\theta t), \cos(\theta t)\}$, and thus is suitable for removing sinusoidal components of a specific frequency which vary around variable midlines. In the spirit of example 1.1.3 the boundary functionals are taken to be

$$B(u) = (\int u(t)\,dt,\ \int \sin(\theta t)\,u(t)\,dt,\ \int \cos(\theta t)\,u(t)\,dt)'.$$

This operator will be used in Section 4 to analyse average monthly temperatures at various weather-stations, where we shall be interested in removing strictly harmonic variation averaged across the entire year.

1.2. *Reproducing Kernels and Green's Functions*

The reproducing kernel k_1 for $H_1 = \ker(L)$ is constructed as follows. Let $u_1, \ldots, u_m$ be functions spanning H_1 and let $\mathbf{u} = (u_1, \ldots, u_m)'$. Defining the matrix U_1 to be the order m symmetric matrix with values $\langle u_i, u_j\rangle_1$, $i, j = 1, \ldots, m$, we have

$$k_1(s, t) = \mathbf{u}(s)'\, U_1^{-1}\mathbf{u}(t). \tag{6}$$

Thus, for example 1.1.2–1.1.3 we have that

$$k_1(s, t) = 4\{13 - 24(s+t) + 45st\}.$$

Computing the reproducing kernels k_2 for H_2 may not be a simple matter, and describing techniques for working these out for a specific choice of operators L and B is beyond the scope of this paper. We shall defer this problem to another paper (Dalzell and Ramsay, 1990). For example 1.1.2–1.1.3,

$$k_2(s, t) = \{4 - 22(s+t) + 156st - 210st^2 + 70t^3 - 70(t^4 + s^4) + 105st(s^3 + t^3) + 21(s^5 + t^5) - 42st(s^4 + t^4)\}/420, \qquad s \leqslant t.$$

Given that reproducing kernels k_1 and k_2 are available, the reproducing kernel for H_0 is simply $k_0 = k_1 + k_2$.

Green's function $g(t; w)$ associated with L and the boundary constraints $B(e) = 0$ satisfies

$$e(t) = \int g(t; w)(Le)(w)\,dw \qquad \text{for } B(e) = 0. \tag{7}$$

This Green's function plays an important role in that

(a) it provides a convenient means of computing the reproducing kernel,
(b) it is useful when computing Lx and

(c) it links the theory of RKHSs with the theory of ordinary differential equations.

Green's function can be derived simply from k_2 since it follows from equation (7) that $Lk_2(s, \ ) = g(s; \ )$, where L is applied to $k_2(s, t)$ as a function of t for fixed s. Conversely, the reproducing kernel k_2 can also be computed from g as

$$\begin{aligned} k_2(s, t) &= \langle k_2(s, \ ), k_2(t, \ ) \rangle_2 \\ &= \int (Lk_2)(s, w)(Lk_2)(t, w)\, dw \\ &= \int g(s; w), g(t; w)\, dw. \end{aligned} \tag{8}$$

2. DEFINITION OF FUNCTIONAL DATA ANALYSIS IN $H_0 = H_1 \oplus H_2$

In this section we apply data analyses familiar for observations $x_i \in \mathbf{R}^p$, $i = 1, \ldots, n$, to observations $x_i \in H$ that are elements of Hilbert space H_0, which is partitioned as described in the previous section. Illustrations of each of these analyses are deferred to Section 4, where they will be applied to smoothing spline representations of finite observations. The sample of n independent and identically distributed functions x_i is therefore an element of the Cartesian product $H_0^n = \otimes_1^n H_0$. We can designate the entire sample as the n-vector-valued function $\mathbf{x} = (x_1, \ldots, x_n)'$. Let ω_i denote a weight associated with observation i such that $\Sigma_i \omega_i = 1$. The order n diagonal matrix Ω contains these weights.

2.1. *The Linear Model*

A linear model for a dependent variable observation x_i implies a linear relationship between H and some other space G, i.e. $x_i \approx Z_i\beta$ where the continuous linear mapping $Z_i: G \to H$ is known, $i = 1, \ldots, n$, and $\beta \in G$. There is no need to assume that G is a Hilbert space, but it is assumed that Z_i is injective and that its image $Z_i(G)$ is closed. Collectively, the mappings Z_i define a mapping $\mathbf{Z}: G \to H_0^n$ such that $\mathbf{Z} = (Z_1, \ldots, Z_n)'$ and $\mathbf{x} \approx \mathbf{Z}\beta$. Thus, mapping $\mathbf{Z}$ corresponds to the design matrix in the usual MDA linear model situation.

Given a specific inner product for H_0^n, the least squares estimate of $\beta \in G$ is that element $\hat{\beta}$ which minimizes $\|\mathbf{x} - \mathbf{Z}\beta\|^2$. The approximation $\hat{\mathbf{x}} = \mathbf{Z}\hat{\beta}$ is the projection of $\mathbf{x}$ on the closed image $\mathbf{Z}(G)$ of G in H_0^n under mapping $\mathbf{Z}$. The fact that $\mathbf{Z}$ is injective and has a closed image ensures that the minimizing element $\hat{\beta}$ exists and is unique.

Associated with the operator $\mathbf{Z}$ is the left inverse operator satisfying the equation $\mathbf{Z}^-\mathbf{Z}\beta = \beta$ for all $\beta \in G$. Mapping $\mathbf{Z}\mathbf{Z}^-: H_0^n \to H_0^n$ is the projection operator P_H on the closed image $\mathbf{Z}(G)$ of G in H, and $\mathbf{Z}^-\mathbf{Z} = P_G$ is similarly the projection in G on the domain of $\mathbf{Z}$ or the range of $\mathbf{Z}^-$. The importance of $\mathbf{Z}^-$ is derived from the fact that $\hat{\beta} = \mathbf{Z}^-\mathbf{x}$ and that $\hat{\mathbf{x}} = \mathbf{Z}\mathbf{Z}^-\mathbf{x}$.

Here are some applications.

2.1.1. *Functional location and scale estimation*

Assume that each x_i is equal to some fixed function $\mu \in H$ plus some residual component $x_i - \mu$. In this case $G = H$ and the mapping Z_i is the identity mapping $\mathbf{1}_H$ for all i. One can quickly verify that the mapping $\mathbf{Z}^-$ defined by

$$\bar{x} = \mathbf{Z}^-\mathbf{x} = \sum_i \omega_i x_i, \tag{9}$$

where the ω_i are weights for the observations, is the required orthogonal left inverse of $\mathbf{Z}$. The k-sample situation in which the jth sample consists of functional observations x_{ij}, $i=1, \ldots, n_j$, can be treated in the same manner. The scalar variance estimate corresponding to the location estimate $\bar{x}$ is

$$\|\mathbf{x}-\mathbf{Z}\bar{x}\|^2=\sum_i \omega_i\|x_i-\bar{x}\|_0^2.$$

Covariation among the sampled functional observations is summarized by the bivariate covariance function

$$v(s, t)=\sum_i^n \omega_i\{x_i(s)-\bar{x}(s)\}\{x_i(t)-\bar{x}(t)\} \tag{10}$$

and correlation by the function

$$r(s, t)=v(s, t)/\{v(s, s)\, v(t, t)\}^{1/2}. \tag{11}$$

When x is Q-vector valued, we must also take into account the cross-covariances and cross-correlations among components of x for any pair of argument values s and t. Covariance function v then takes values which are order Q symmetric positive definite matrices.

2.1.2. *Finite dimensional regression*

We now propose that G is $\otimes_s^q H$, and that element β is the vector of functions $(\beta_1, \ldots, \beta_q)'$. Element $Z_i\beta$ is defined by

$$(Z_i\beta)(t)=\sum_s^q z_{is}\beta_s(t), \tag{12}$$

where the real coefficients z_{is} are known. In this case $\mathbf{Z}$ can be represented by an $n\times q$ matrix Z, just as in the multivariate case. Again, it falls out at once that Z^- is a minimum norm g-inverse of Z, and therefore that $Z^-=(Z'\Omega Z)^{-1}Z'\Omega$. Thus in the functional analysis-of-variance case $\mathbf{Z}$ can also be represented by a design matrix containing 0s and 1s just as it would be for MDA.

2.1.3. *Infinite dimensional regression*

When the linear modelling problem is functional all the way, G will be some space of bivariate functions $\beta(s, t)$ such as the product $\otimes_{s\in S}H$, where the index s takes values in a compact set S. The linear operator Z_i may then be represented by

$$(Z_i\beta)(t)=\int_S z_i(s)\,\beta(s, t)\,\mathrm{d}s. \tag{13}$$

More generally, let us assume that the functions z_i and $\beta(\ , t)$ are elements of some Hilbert space H_z with inner product $\langle\ ,\ \rangle_z$. Then we can state the linear model as

$$\hat{x}_i(t)=(Z_i\beta)(t)=\langle z_i, \beta(\ , t)\rangle_z. \tag{14}$$

The analogue to the cross-product matrix $Z'\Omega Z$ in the finite dimensional case is now the bivariate function

$$\Gamma(r, s)=\sum_i \omega_i z_i(r)\, z_i(s), \qquad r, s\in S,$$

and if we can give a precise meaning to the notion 'Γ^{-1}' corresponding to $(Z'\Omega Z)^{-1}$ in the finite regression case the operator $\mathbf{Z}^-$ defined by its image

$$\hat{\beta}(s, t) = (\mathbf{Z}^- x)(s, t) = \sum_i \omega_i x_i(t) \langle \Gamma^{-1}(s,) z_i( ) \rangle_z \tag{15}$$

will be what is required. It will be shown in Section 4 that the problem of approximating the mapping $\mathbf{Z}^-$ is much simplified if z_i and $\beta(, t)$ are elements of a reproducing kernel Hilbert space.

This by no means exhausts the range of linear models possible. For example, we might consider

$$(Z_i\beta)(t) = \sum_j z_{ij}(t)\beta_j$$

or

$$(Z_i\beta)(t) = \langle z_i(t,) \beta( ) \rangle_z,$$

where again $z_{ij}(t)$ or $z_i(t, s)$ are known functions.

2.2. *Principal Components Analysis*

The objective in principal components analysis is the orthogonal decomposition of the variance function v

$$v(s, t) = \sum_j \lambda_j \xi_j(t) \xi_j(s) \tag{16}$$

to isolate the dominant components of functional variation. The principal component functions or harmonics ξ_j can be visually inspected or subjected to further analyses, and the eigenvalues λ_j indicate the amount of variance attributable to each component. Moreover, we can choose to consider variation restricted to any of the three subspaces H_h, $h = 0$, 1 or 2, and in particular it may be of great interest to study the harmonics in the residual space H_2. Principal components analysis in L^2 was discused by Dauxois *et al.* (1982) and in H^m by Besse and Ramsay (1986) and Besse (1988).

We therefore define the principal components analysis of $\mathbf{x}$ to be the optimization problem

$$\max_{\xi_j} \langle v(s,), \xi_j \rangle_h \tag{17}$$

subject to the conditions

$$\| \xi_j \|_h = 1 \text{ and } \langle \xi_j, \xi_i \rangle_h = 0, \qquad i = 1, \ldots, j-1, \quad h = 0, 1, 2. \tag{18}$$

It follows from the Cauchy–Schwarz inequality that this problem is equivalent to the solution of the eigenequation

$$\langle v(s,), \xi_j \rangle_h = \lambda_j \xi_j(s). \tag{19}$$

With the covariance kernel v defined as in equation (10) in terms of a sample mean there can be at most $n-1$ non-zero eigenvalues λ_j and eigenfunctions ξ_j.

The decomposition (16) defined by eigenequation (19) then permits a reduced rank least squares approximation to v. Thus, the leading eigenfunctions ξ define the principal

components of variation among the sample functions x_i. The vector-valued case for X does not introduce any complications in principle, and the harmonics ξ_j will then also be vector valued (Ramsay, 1989).

Once the harmonics ξ_j are in hand, it will usually be worth displaying the component scores $f_{ij}=\langle x_i, \xi_j\rangle$ since these indicate the relative amount of component j in observation i and can therefore be useful in attaching some interpretation to components. We are also free to examine any linear transformation FT of the component scores as well as looking at the same linear combinations of harmonics. The transformation matrix T in this case will often result from a singular value decomposition of F.

3. REPRESENTING FINITE OBSERVATIONS IN H

Although modern realtime data acquisition technology often produces data that can be regarded as functional, it is usual to obtain a finite and therefore necessarily incomplete amount of information about a function. For example, temperature information at a weather-station may be compiled on a daily or monthly basis even though it is in principle available continuously, i.e. the datum for a specific observation is a vector $\mathbf{y}\in\mathbf{R}^p$ which is assumed to have the decomposition

$$\mathbf{y}=\boldsymbol{\tau}+\boldsymbol{\epsilon} \tag{20}$$

where the p-vector $\boldsymbol{\tau}$ is the errorless or systematic component in $\mathbf{y}$ and the noise or unsystematic component $\boldsymbol{\epsilon}$ is assumed to be an observation of a random $\mathbf{p}$-vector for which $E(\boldsymbol{\epsilon})=\mathbf{0}$ and $\mathrm{var}(\boldsymbol{\epsilon})=\Sigma$.

Assume that $\boldsymbol{\tau}$ is in turn the result of applying a linear evaluation operator

$$A=(\rho_{t_1}, \ldots, \rho_{t_p})'$$

to a function $x\in H_0$, so that $\boldsymbol{\tau}=A(x)$. More generally when $\mathbf{x}$ is a Q-vector-valued function denoted by

$$\mathbf{x}=(x_1, \ldots, x_Q)'\in\otimes H_0,$$

$A(\mathbf{x})$ will indicate the $p\times Q$ matrix of function values resulting from each component being evaluated at each sampling point. Assuming now that H_0 is decomposed into $H_1\oplus H_2$ as described in Section 1, we have the corresponding decomposition $\boldsymbol{\tau}=A(x)=A(u+e)=A(u)+A(e)$. The objective in smoothing is in effect to reverse the mapping A to pass from finite dimensional $\mathbf{y}$ to a representation $\hat{x}\in H_0$ which is in a specified sense close to the function x assumed to generate $\mathbf{y}$. Let $\hat{u}$ and $\hat{e}$ indicate the H_1 and H_2 components of $\hat{x}$ respectively.

The process of representing the data $\mathbf{y}\in\mathbf{R}^p$ by $\hat{x}\in H_0$ by smoothing splines involves minimizing with respect to $\hat{x}$ the criterion

$$W(\hat{x}\,|\,\mathbf{y}, \lambda)=\|\mathbf{y}-A(\hat{x})\|_N^2+\lambda\|\hat{x}\|_2^2, \tag{21}$$

where $\|\mathbf{y}\|_N^2=\mathbf{y}'N\mathbf{y}$ for some symmetric positive definite matrix N and $\lambda>0$. For example, if $\mathrm{var}(\boldsymbol{\epsilon})=\Sigma$ is known, it is reasonable to set $N=\Sigma^{-1}$. As $\lambda\to 0$, $\hat{x}$ approaches the minimum H_0-norm interpolant of $\mathbf{y}$.

Let $\mathbf{k}_h$ be the p-vector-valued function with elements $k_h(t_j, \ )$, and let $K_h=A(\mathbf{k}_h)$ be the order p symmetric positive semidefinite matrix with entries $k_h(t_i, t_j)$, $i, j=1, \ldots, p$ for $h=0, 1, 2$. It can be shown (Wahba, 1990) that $\hat{x}$ is of the form

$$\hat{x} = \hat{u} + \hat{e} = \mathbf{k}_1' \mathbf{c}_1 + \mathbf{k}_2' \mathbf{c}_2,$$

or in other words that component $\hat{u} \in \text{span}\{\mathbf{k}_1\} \subset H_1$ and component $\hat{e} \in \text{span}\{\mathbf{k}_2\} \subset H_2$. We also have that $L\mathbf{k}_2 = \mathbf{g}$, where $\mathbf{g}$ is the p-vector-valued function with elements $g(t_j;\ )$, and $g(t;\ w)$ is Green's function defined by equation (7). It then follows that $L\hat{x} = \mathbf{g}'\mathbf{c}_2$. The evaluation of $\hat{x}$ at the sampling points t_j is of the form

$$A(\hat{x}) = K_0 \mathbf{c}_0 = A(\hat{u}) + A(\hat{e}) = K_1 \mathbf{c}_1 + K_2 \mathbf{c}_2$$

since

$$\begin{aligned} \rho_{t_j}(\hat{x}) &= \langle k_0(t_j,\ ),\ \hat{x} \rangle \\ &= \langle k_0(t_j,\ ),\ \mathbf{k}_1' \mathbf{c}_1 + \mathbf{k}_2' \mathbf{c}_2 \rangle \\ &= \sum_i \{ c_{i1} \langle k_0(t_j,\ ),\ k_1(t_i,\ ) \rangle + c_{i2} \langle k_0(t_j,\ ),\ k_2(t_i,\ ) \rangle \} \\ &= \sum_i c_{i1} k_{ji}^1 + c_{i2} k_{ji}^2, \end{aligned} \tag{22}$$

where k_{ji}^h is the jith element of K_h, $h = 0$, 1 or 2.

The computation of the smoothing splines is therefore a matter of computing the appropriate coefficient vectors $\mathbf{c}_1$ and $\mathbf{c}_2$. This is much simplified if we assume that the sampling points t_j have been positioned so that $K_0 = K_1 + K_2$ is positive definite. For problems in which $H_1 = \ker(L)$ it will be assumed for convenience that the rank of $K_1 = m$ and that $p - m \leqslant \text{rank}(K_2) \leqslant p$. These assumptions can be relaxed at the expense of considerable additional complexity relative to the following discussion, but in most practical situations the sampling points will be chosen by design so that these assumptions hold.

Given these conditions on K_1 and K_2, a representation of $\hat{x}$ minimizing loss function W is

$$\hat{x} = \mathbf{k}_0' C_\lambda \mathbf{y},$$

where

$$C_\lambda = (K_0 K_0 + \lambda K_2)^{-1} K_0. \tag{23}$$

In the interpolation case equation (23) reduces to

$$C_\lambda = K_0^{-1}. \tag{24}$$

Smoothing parameter λ can either be chosen by inspection of the smooths, or by an automatic procedure such as generalized cross-validation (GCV) (Craven and Wahba, 1978; Bates *et al.*, 1987). We used GCV for the weather data, smoothing each variable separately, and choosing that value of λ which provided the approximate minimum of the GCV coefficient summed across the 35 replications. These values were 0.001 and 0.1 for temperature and log(precipitation) respectively, reflecting the fact that there is considerably more variability in rainfall data than in temperature. Fig. 2 shows the joint variation in temperature and precipitation for Edmonton in the prairies and the Pacific coastal station at Prince Rupert plotted as closed curves. (Any analyses of these data must also take into account their periodicity, so that a representation of these data in the interval $T = [0, 12]$ must also require that derivatives

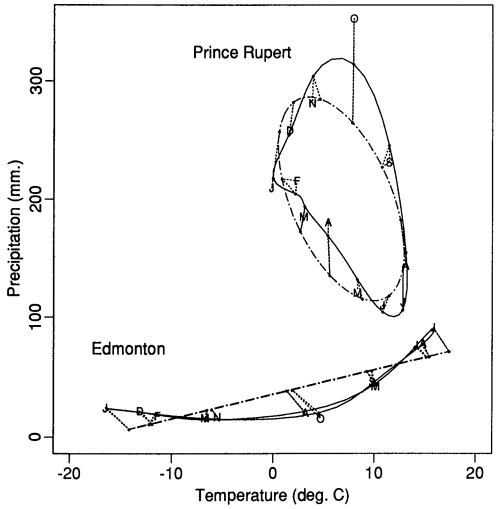

Fig. 2. Joint variation in average monthly temperature and precipitation for Edmonton and Prince Rupert displayed as closed curves on the temperature–precipitation plane (letters indicate actual observations for each month): ———, smoothing spline representation; ·····, purely sinusoidal or harmonic component in H_1 (broken lines join corresponding points)

at the end points match to order 2; this can either by achieved by the imposition of linear constraints on the parameter vectors or equivalently and more conveniently by adjoining copies of the data to the beginning and end of the original, so that observations extend over 36 sampling points, but results are considered only for the central 12.) The full curves indicate the harmonic smoothing splines in H_0 while the chain curves show the purely sinusoidal H_1 components which plot as ellipses. The letters indicate the actual data.

For even modest values of p the matrix K_0 can be poorly conditioned, and the use of K_0K_0 in equation (23) can lead to very serious rounding error problems. Bates *et al.* (1987) and Wahba (1990) contain further discussion of the computation of $\hat{x}$ and its components in H_1 and H_2.

4. RELATION BETWEEN FUNCTIONAL DATA ANALYSIS AND MULTIVARIATE DATA ANALYSIS OF FINITE OBSERVATIONS

In this section we study the equivalence between the conventional MDA of a set of observation vectors $\mathbf{y}_i$ and the functional data analysis of the corresponding functional representations $\hat{x}_i$, $i=1, \ldots, n$. In the previous section we have seen that $\hat{x}_i$ has the form

$$\hat{x}_i = \mathbf{k}_0' \mathbf{c}_{i0} = \hat{u}_i + \hat{e}_i = \mathbf{k}_1' \mathbf{c}_{i1} + \mathbf{k}_2' \mathbf{c}_{i2}$$

where $\mathbf{k}_h$ is the vector-valued function with jth element $k_h(t_j,)$, $h=0, 1, 2$. Each representing function $\hat{x}_i$ has values at $t_1, \ldots, t_p$ given by

$$A(\hat{x}_i) = A(\mathbf{k}_0' C \mathbf{y}_i) = A(\mathbf{k}_0') C \mathbf{y}_i = K_0 C \mathbf{y}_i. \tag{25}$$

We assume for simplicity that the same interpolation or smoothing process is used for each $\mathbf{y}_i$, and therefore that C is a matrix which does not depend on i. Let $\hat{\mathbf{x}}$ refer

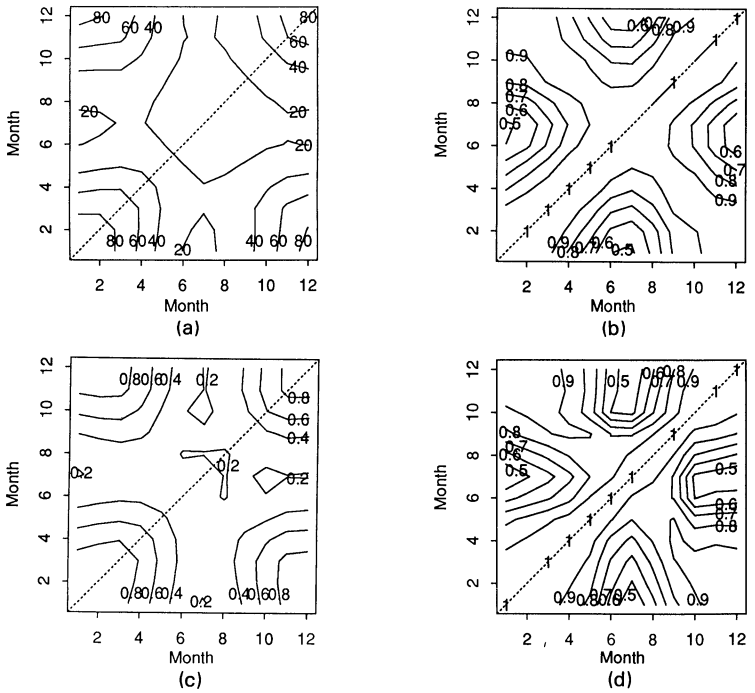

Fig. 3. Covariances among the mean temperatures and log(precipitation) taken across the 35 weather-stations represented as contour-plotted surfaces ((a), (c)) and the correlations ((b), (d)) (the ridges on the diagonal indicate a stronger covariation between temperatures in the spring and autumn than between those in the summer and winter): (a) temperature covariance; (b) temperature correlation; (c) log(precipitation) covariance; (d) log(precipitation) correlation

to the n-vector-valued function $(\hat{x}_1, \ldots, \hat{x}_n)'$, Y to the $p \times n$ matrix of observation vectors $\mathbf{y}_i$ and $A(\hat{\mathbf{x}})$ to the $p \times n$ matrix of approximating function values $\hat{x}_i(t_j)$. It follows that $\hat{\mathbf{x}} = Y'C'\mathbf{k}$, with order p matrix C being determined by the nature of the smoothing process.

4.1. *The Linear Model*

It will be assumed that each $\hat{x}_i$ is possibly already a residual after removing certain functional effects by linear modelling. The approximate covariance kernel $\hat{v}$ is consequently

$$\begin{aligned}\hat{v}(s, t) &= \hat{\mathbf{x}}(s)'\,\Omega\hat{\mathbf{x}}(t) \\ &= \mathbf{k}(s)'\,CY\Omega\,Y'C'\mathbf{k}(t). \end{aligned} \tag{26}$$

Fig. 3 displays the estimated covariance and correlation functions as surfaces by contour plotting. The diagonal ridges perpendicular to the broken lines indicate a much stronger correlation between spring and autumn than between winter and summer temperatures. This is a preliminary indication that factors are at work that influence weather patterns differently in various seasons, and that there are strong departures from covariance stationarity in these data.

For the finite dimensional linear model

$$(Z_i\beta)(t) = \sum_s^q z_{is}\beta_s(t),$$

the linear mapping $\mathbf{Z}$ is represented by the $n \times q$ coefficient matrix Z. The function

$$\hat{\beta} = \mathbf{Z}^{-}\hat{\mathbf{x}} = (Z'\Omega Z)^{-1}Z'\Omega Y'C'\mathbf{k}_0 \tag{27}$$

is the least squares estimate of the regression q-vector function β based on the approximating functions $\hat{\mathbf{x}}$. The values of these estimated regression functions at $t_1, \ldots, t_p$ are $A(\hat{\beta}) = (Z'\Omega Z)^{-1}Z'\Omega Y'C'K_0$ while the values of the least squares approximations to the functions $\hat{x}$ are

$$A(\hat{x}) = Z(Z'\Omega Z)^{-1}Z'\Omega Y'C'K_0.$$

Thus, this finite dimensional regression situation is equivalent to the MDA of the data matrix $Y'C'K_0$.

In situation (14) in which the regression functions are $\beta(s, \ )$, where index s takes values on a real interval S, the continuous analogue of the cross-product matrix $Z'\Omega Z$ is the bivariate function $\Gamma(r, s) = \Sigma_i \omega_i z_i(r) z_i(s) = \mathbf{z}(r)'\Omega\mathbf{z}(s)$. Its 'inverse' $\Gamma^{-1}(r, s)$ is defined in principle but can only be approximated in practice.

Assuming that z_i is an element of an RKHS with reproducing kernel k_z, we can propose that the function z_i be approximated by $\hat{z}_i$ which either interpolates or smooths values $z_i(s_1), \ldots, z_i(s_q)$, i.e.

$$\hat{z}_i = \sum_j d_{ij} k_z(s_j, \ ). \tag{28}$$

The $n \times q$ matrix $\hat{Z}$ of approximating function values $A(\hat{z}_i)$, $i = 1, \ldots, n$, has the form $A(\hat{\mathbf{z}}) = DK_z$, where D is an $n \times q$ matrix of coefficients and K_z is the symmetric order q matrix of reproducing function values.

The cross-product function $\hat{\Gamma}(r, s)$ for the approximations $\hat{z}_i$ can now be expressed as

$$\hat{\Gamma}(r, s) = \mathbf{k}_z(r)'D'\Omega D\mathbf{k}_z(s), \tag{29}$$

where $\mathbf{k}_z$ denotes the q-vector with elements $k_z(s_j, \ )$, $j = 1, \ldots, q$. The inverse function $\hat{\Gamma}^{-1}$ is then

$$\hat{\Gamma}^{-1}(r, s) = \mathbf{k}_z'(r)K_z^{-1}(D'\Omega D)^{-1}K_z^{-1}\mathbf{k}_z(s) \tag{30}$$

since this satisfies the equation $\hat{z}(r) = \langle \hat{\Gamma}^{-1}(r, \ ), \langle \hat{\Gamma}(\ , \ ), \hat{z}\rangle_z\rangle_z$. In applications, it may well be that $p > n$ and therefore that $(D'\Omega D)^{-1}$ may need to be replaced by $(D'\Omega D)^{+}$.

Using equation (30), we can now define the approximation to the least squares regression function β as

$$\hat{\beta}(s, t) = \hat{\mathbf{Z}}^{-}\mathbf{x} = \mathbf{k}_z'(s)K_z^{-1}(D'\Omega D)^{-1}D'\Omega\mathbf{x}(t). \tag{31}$$

Now, when $\hat{z}_i$ is approximated by interpolating the values of z_i at sampling points $s_1, \ldots, s_q$, we have that $D = ZK_z^{-1}$, where Z is the $n \times q$ matrix containing these values to be interpolated. In this case, equation (31) reduces to

$$\hat{\beta}(s, t) = \mathbf{k}_z'(s)(Z'\Omega Z)^{-1}Z'\Omega\mathbf{x}(t). \tag{32}$$

Replacing $\mathbf{x}$ by its smoothing approximation $Y'C'\mathbf{k}$ and evaluating $\hat{\beta}$ gives

$$A(\hat{\beta}) = K_z(Z'\Omega Z)^{-1}Z'\Omega Y'C'K_0. \tag{33}$$

The least squares approximation to the approximate observation function $\hat{\mathbf{x}}$ is then $\hat{\mathbf{x}} = \hat{Z}'(\hat{Z}\Omega\hat{Z}')^{-1}\hat{Z}\Omega Y'C'\mathbf{k}_0$ with values

$$A(\hat{\mathbf{x}}) = \hat{Z}'(\hat{Z}\Omega\hat{Z}')^{-1}\hat{Z}\Omega Y'C'K_0. \tag{34}$$

Thus, in the special case of interpolated regressor values the functional linear modelling of Y reduces to the MDA linear modelling of $Y'C'K_0 = Y'S'$. But we now have much more at our disposal than merely $A(\hat{\beta})$. We can also examine the H_1, H_2 and L components of $\hat{\mathbf{x}}$ at arbitrary evaluation points by replacing K_0 in equation (34) by the appropriate reducing kernel matrices.

4.1.1. *Functional regression of log(precipitation) on temperature*

What is the relation between precipitation and temperature, and how strong is it? Plots such as Fig. 2 suggest that temperature and precipitation covary to some extent, and that in general there is a tendency for more precipitation in the summer months. However, this relationship obviously varies from station to station. A closer study of the variation in the temperature–precipitation relationship is aided by first fitting a model which does not allow for such variation, and then examining the residuals. We propose here a functional regression of smoothed log(precipitation) on smoothed temperature by using model (14).

Expanding the infinite dimensional linear model, we have

$$\begin{aligned}\hat{x}_i(t) &= \int z_i(s)\,ds \int \beta(s,\,t)\,ds + \int \sin(\pi s/6)\, z_i(s)\,ds \int \sin(\pi s/6)\,\beta(s,\,t)\,ds \\ &\quad + \int \cos(\pi s/6)\, z_i(s)\,ds \int \cos(\pi s/6)\,\beta(s,\,t)\,ds + \int (Lz_i)(s)(L\beta)(s,\,t)\,ds \\ &= z_{i1}B_1(t) + z_{iS}B_S(t) + z_{iC}B_C(t) + \int (Lz_i)(s)(L\beta)(s,\,t)\,ds, \qquad (35)\end{aligned}$$

where x_i and z_i are now understood to be smoothing spline representations of the precipitation and temperature data for station i respectively. The purpose of the first three terms on the right-hand side of equation (35) is to account for the strictly harmonic H_1 variation in log(precipitation) as a function of temperature. The functions

$$B_1(t) = \int \beta(s,\,t)\,ds, \qquad B_S(t) = \int \sin(\pi s/6)\,\beta(s,\,t)\,ds$$

and $B_C(t) = \int \cos(\pi s/6)\,\beta(s,\,t)\,ds$ act as weights for the inner products of the constant, sine and cosine functions with the temperature functions, indicated in equation (35) by z_{i1}, z_{iS} and z_{iC} respectively. The final term, in contrast, accounts for strictly anharmonic variation in log(precipitation), and the bivariate function $(L\beta)(s,\,t)$ operates as a kernel function in a kernel smoothing process applied to Lz. It thus defines the optimal local average of L-temperature to predict anharmonic log(precipitation).

Fig. 4 gives the estimated regression functions $\hat{\beta}(\ ,\,t_j)$ for each month of the year. However, there is only a limited amount to be gained by direct inspection of these curves, and it is useful to decompose them into their harmonic H_1 and anharmonic H_2 components. Fig. 5 shows the weighting functions, $B_1(t)$, $B_S(t)$ and $B_C(t)$ for predicting the strictly harmonic variation in log(precipitation) as a function of temperature and Fig. 6 displays the functions $(L\hat{\beta})(\ ,\,t_j)$. From Fig. 5 we can conclude that most of the harmonic variation is accounted for by the component

$$z_{iS} = \int \sin(\pi s/6)\, z_i(s)\,ds$$

since only $B_S(t)$ differs substantially from 0. We also note that $B_S(t)$ is close to being a linear function of $\cos(\pi t/3)$. A slight shift of origin for time would probably improve the fit of this simple two-parameter approximation even further. As for the anharmonic component, Fig. 6 shows that it is much stronger for winter months, where it primarily

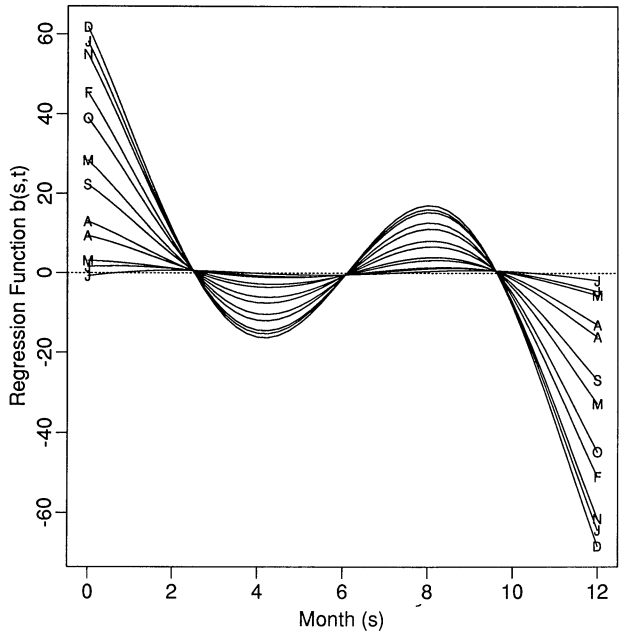

Fig. 4. Regression functions $\beta(s, t)$ plotted as functions of s for the regression of smoothed precipitation on smoothed temperature with the functional linear model $\hat{x}_i(t) = \langle z_i, \beta(\ , t)\rangle$ (letters correspond to months)

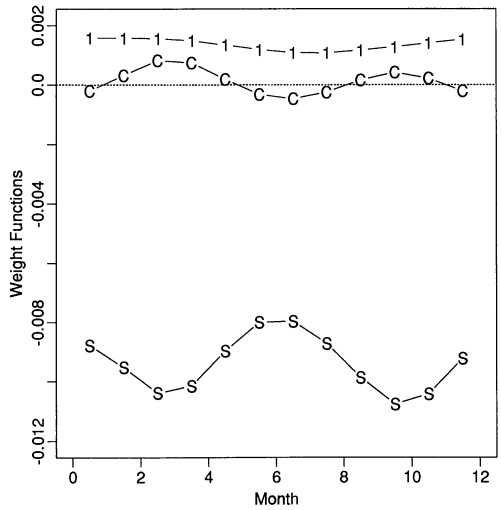

Fig. 5. Weight functions $\int \beta(s, t)\, ds$, $\int \sin(\pi s/6)\, \beta(s, t)\, ds$ and $\int \cos(\pi s/6)\, \beta(s, t)\, ds$ determining the harmonic H_1 component of $\hat{x}_i$ for the regression functions in Fig. 4

determines variation in summer precipitation levels. That the anharmonic component for the mid-winter months is exactly 0 follows from the end effects well known in spline smoothing.

In general, log(precipitation) is well approximated by temperature for almost all weather-stations, with the median multiple correlation being 0.964. The worst cases

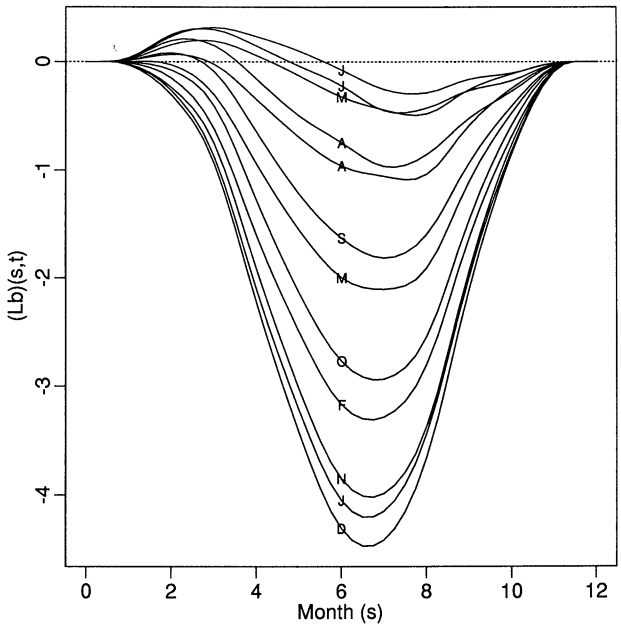

Fig. 6. Functions $(L\beta)(\ ,t)$ computed from the regression functions in Fig. 4: these determine the anharmonic H_2 component of $\hat{x}_i$ via the model $\hat{x}(t)_{2i} = \int (Lz_i)(s)(L\beta)(s,t)\,ds$ (letters correspond to months)

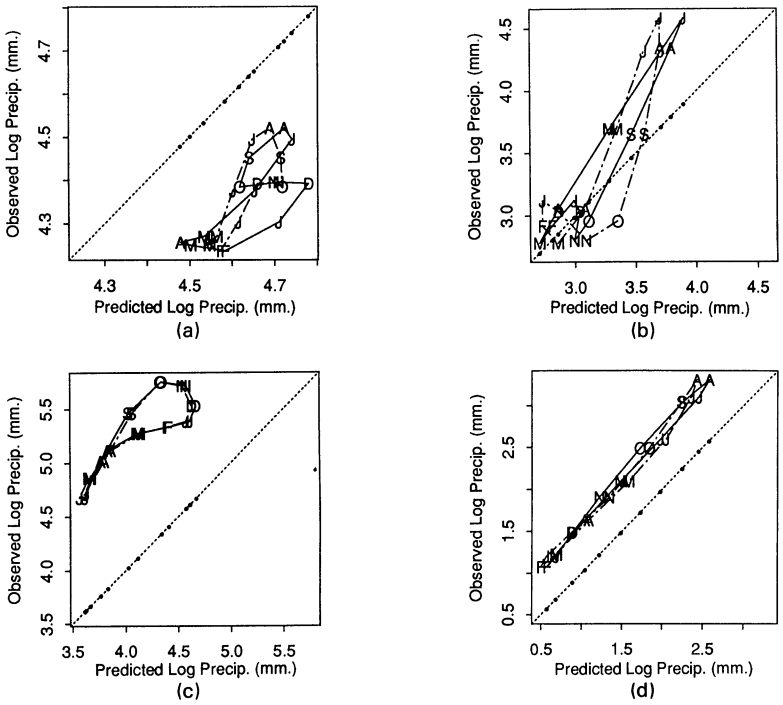

Fig. 7. Comparison of actual and predicted log(average monthly precipitation) for four selected weather-stations (the value of R is the product moment correlation, and therefore not affected by failure of the model to capture the correct slope or intercept; ——, prediction based on the complete model; --------, harmonic H_1 component of the prediction): (a) Montreal ($R = 0.737$); (b) Edmonton ($R = 0.964$); (c) Prince Rupert ($R = 0.862$); (d) Resolute ($R = 0.997$)

are Toronto (−0.119), Montreal (0.737) and Ottawa (0.773), possibly reflecting the microclimactic factors at work in large urban areas. There are clear patterns in the residuals, and these are evident in the plots of fitted log(precipitation) *versus* actual values for Edmonton, Montreal, Prince Rupert and Resolute shown in Fig. 7 (full curves). Also shown are the strictly harmonic components of fit (broken curves), and a comparison of the two fits shows that the H_2 contribution is important for Edmonton, but small and probably uninteresting for the other three weather-stations.

4.2. *Principal Components Analysis*

We now examine the consequences of applying functional principal components analysis to decompose the covariation among representations $\hat{x}_i$. The eigenequation (19) defining a principal functional component or harmonic ξ becomes

$$\langle \mathbf{k}(s)' CY\Omega Y' C' \mathbf{k}(\ ),\ \xi \rangle_h = \lambda \xi(s), \qquad \| \xi_h \| = 1,\ \ h = 0,\ 1,\ 2. \tag{36}$$

Since ξ will of necessity lie within span$\{\mathbf{k}\}$, we can use the representation $\xi = \mathbf{k}'\boldsymbol{\alpha}$ for some coefficient vector $\boldsymbol{\alpha} \in \mathbf{R}^p$. Defining $\boldsymbol{\eta} = A(\xi) = K\boldsymbol{\alpha} \in \mathbf{R}^p$, so that $\boldsymbol{\alpha} = K^{-1}\boldsymbol{\eta}$ and $\| \xi \|_h^2 = \boldsymbol{\alpha}' K_h \boldsymbol{\alpha} = \boldsymbol{\eta}' K^{-1} K_h K^{-1} \boldsymbol{\eta}$, leads to the re-expression of eigenequation (36) as

$$\mathbf{k}(s)' CY\Omega Y' C' K_h \boldsymbol{\alpha} = \lambda \mathbf{k}_h(s)' \boldsymbol{\alpha}, \qquad \boldsymbol{\alpha}' K_h \boldsymbol{\alpha} = 1. \tag{37}$$

Applying the evaluation operator A to eigenequation (36) and expressing it in terms of $\boldsymbol{\eta}$, we have the following eigenequation in $\mathbf{R}^p$

$$KCY\Omega Y' C' K_h K^{-1} \boldsymbol{\eta} = \lambda \boldsymbol{\eta}, \qquad \boldsymbol{\eta}' K^{-1} K_h K^{-1} \boldsymbol{\eta} = 1. \tag{38}$$

Let the order p symmetric positive semidefinite matrix M be defined as $K^{-1}K_hK^{-1}$. Then

$$KCY\Omega Y' C' KM\boldsymbol{\eta} = \lambda \boldsymbol{\eta}, \qquad \boldsymbol{\eta}' M \boldsymbol{\eta} = 1. \tag{39}$$

Thus, the principal components analysis of the approximate covariance operator $\hat{v}$ is equivalent to the eigenanalysis of the order p matrix $KCY\Omega Y' C' K$ in the metric M. Finally, this in turn is equivalent to the ordinary eigenanalysis

$$M^{1/2} KCY\Omega Y' C' K M^{1/2\prime} \boldsymbol{\nu} = \lambda \boldsymbol{\nu}, \qquad \boldsymbol{\nu}' \boldsymbol{\nu} = 1, \tag{40}$$

where $M^{1/2}$ is a symmetric square root of M or a Choleski factor. Then the corresponding harmonic function is $\xi = \mathbf{k}' K^{-1} M^{-1/2} \boldsymbol{\nu}$, and this is also the interpolant of the vector $\boldsymbol{\eta}$.

The principal component scores matrix F now becomes

$$F = \{\langle x_i,\ \xi_j \rangle\} = \{Y' C' K_h K^{-1} M^{-1/2} \boldsymbol{\nu}_j\}. \tag{41}$$

When the data are interpolated, so that $C = K^{-1}$, eigenequation (40) simplifies to $M^{1/2} Y\Omega Y' M^{-1/2\prime} \boldsymbol{\nu} = \lambda \boldsymbol{\nu}$ and $F = Y'\Lambda$ where $\Lambda = \{M^{1/2}\boldsymbol{\nu}_j\}$.

4.2.1. *Principal component analysis of temperature variation*

We turn now to the principal components analyses of the temperature data in subspace H_2, i.e. we wish to measure the size of residual components in temperature pattern x in terms of the harmonic acceleration Lx. Explicit account is taken of the observational noise in the data in that analyses are of the smoothing functions defined

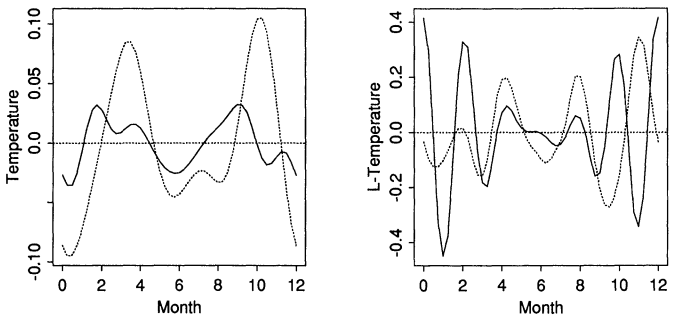

Fig. 8. (a) Rotated functional principal components or harmonics I (———) and II (--------) for variation in residual space H_2 and (b) corresponding components after applying the operator $L=(\pi/6)D+D^3$

Fig. 9. Rotated principal component scores for the principal components in Fig. 8 (a few weather-stations are unlabelled for clarity): -·-·-·-, unrotated harmonics

by equation (21) rather than the observations themselves. The first three eigenvalues account for 61.8%, 17.5% and 12.0% of the variance, and the remaining eigenvalues are relatively small. However, most of the variation among the principal component scores can be well represented in Euclidean space by two dimensions, and for simplicity we display in Fig. 8(a) the two orthonormal linear combinations of the three dominant eigenfunctions ξ which give the best two-dimensional representation of the principal

components scores. Fig. 8(b) shows the values of the same linear combinations $L\xi$, and thus their harmonic accelerations. Here it is evident that this first rotated harmonic has most of its acceleration in the winter months and is relatively symmetric about June. In contrast, the second rotated harmonic tends to increase in power from January to November. The transformed principal components scores are shown in Fig. 9 and permit us to see the relation between these harmonics and the temperature variation at each weather-station. A tight cluster of St Lawrence River valley stations faces a phalanx of Pacific, prairie and Arctic stations. They are separated primarily in terms of the amount of the first rotated harmonic, while the second rotated harmonic separates western and Arctic marine stations from continental stations. Since one of the forcing factors influencing weather is the effect of the high reflectance of snow cover during the winter season, we can conjecture that this second harmonic is associated with this effect. The three unrotated harmonics are represented in the plot by plotting the rows of $\Lambda^{-1}T$ as directions, and this indicates the proximity of each weather-station to each harmonic in this viewing plane.

4.2.2. *Principal component analysis of joint temperature–precipitation variation*

Something as complex as climate needs to be summarized in terms of many variables simultaneously, and the consideration of joint variation in temperature and precipitation is likely to be rather more useful than considering each variable separately. In particular, we wonder whether there are interesting components of variation involving interactions between these two variables that we might not see by separate principal components analyses.

The eigenanalysis (40) must now be based on the $12 \times 35 \times 2$ array resulting from combining the temperature and precipitation data. A general treatment of multiway principal components analysis can be made much more convenient by the use of tensor notation, and Ramsay (1989) should be consulted for such an approach. It can be shown, however, that this is essentially a matter of the eigenanalysis of $Y_{TP}\Omega Y_{TP}'$,

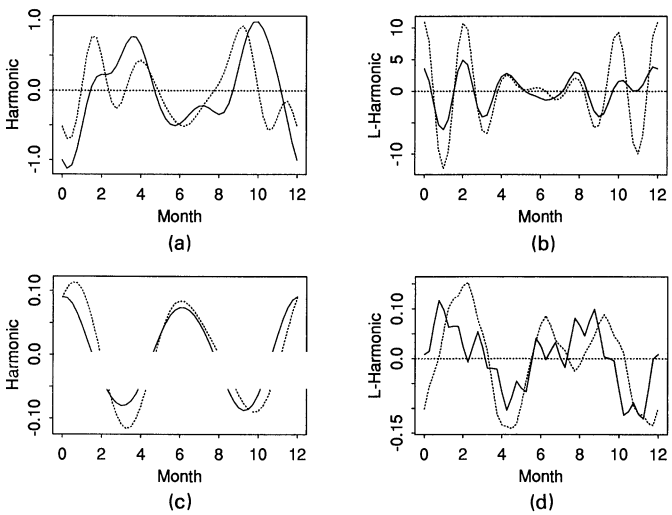

Fig. 10. Temperature ((a), (b)) and precipitation ((c), (d)) components of harmonics I (———) and II (· · · · ·) for a principal components analysis of joint variation in smoothed temperature and precipitation in H_2: (a) and (c), actual harmonics; (b) and (d), L-values

where Y_{TP} is the $2p \times n$ matrix resulting from the rowwise concatenation of the matrices $M^{1/2}KCY_T$ and $M^{1/2}KCY_P$ containing transformed temperature and log(precipitation) values respectively. The harmonics estimated from this process can then be separated into their temperature and precipitation components.

We again consider only components of variation in H_2. The first three eigenvalues accounted for 61.9%, 23.8% and 7.9% of the variance for a total of 93.6%, and the remaining eigenvalues decreased gradually in size. Fig. 10 shows the first two harmonics as well as the L-harmonics, which are split into their temperature and precipitation components. The distribution of principal components scores was roughly similar to that displayed in Fig. 9.

5. CONCLUSIONS AND DISCUSSION

What has been achieved by these functional data analyses over the outcome of the conventional analyses of residuals from the least squares fit of each set of observations by $(1, \sin(\pi t_j/6), \cos(\pi t_j/6))$, $j = 1, \ldots, 12$? In our view the following are distinctive features.

(a) The smoothing criterion (21) used to estimate the H_0 component is more faithful to the structure of the data in that observational error is separated from functional variation orthogonal to H_0, i.e. we have separated the multivariate distribution of observational error from the functional distribution of the latent regular mean temperature and precipitation functions. This was made possible by exploiting the presumed regularity of the functional variation in the smoothing process.

(b) Since the problem is functional in nature, it is valuable to estimate the functions involved directly, and this is especially so for the residual component $e(t)$. This permits us to look at many aspects of these functions besides their values at arbitrary values of t. Derivatives, integrals and many other functionals suggest themselves as potential quantifications of their behaviour.

(c) The functions Lx may be of direct interest. For example, if $x(t)$ is the trajectory of a moving body and $L = D^2$, then Lx is acceleration and the criterion $\|x\|^2 = \int (Lx)^2 \, dt$ measures power. Thus the penalty term in spline smoothing offers a potentially useful alternative to $\int e^2 \, dt$ as a measure of residual variation.

(d) The functional approach eliminates the need for equally spaced data or assumptions of stationarity underlying time—and frequency—domain time series analyses. We suspect that a large number of such analyses reported in applied journals are used on data which, even after possible deseasonalization and differencing, are unlikely to be either mean or covariance stationary.

(e) A functional perspective invites some graphical displays that might not readily occur to those familiar with MDA techniques. Fig. 2 displays simple harmonic variation as an ellipse on the temperature–precipitation plane, and this facilitates the viewing of anharmonic variation as departures from this pattern. Figs 4–6 provide various perspectives on the regression functions that help us to visualize the dependence of precipitation on temperature, and Fig. 7 plots residual functions even though a limited number of actual observations are involved.

We have emphasized throughout the *partitioning principle*, partly because we believe it to be an important part of most good data analyses, and partly because the vastness

of function space makes its use essential. Our use of the linear differential operator L along with boundary condition operator B as a device for achieving this partition is only intended to indicate one of many possible approaches. A projector can be combined with a differential operator, giving rise to partial splines (Cox *et al.*, 1988; Wahba, 1990), and yet other forms of partitioning are possible (Besse and Thomas-Agnan, 1989).

The scope of the paper has eliminated many possible topics and has also limited the perspective of the paper to a geometrical or data analytic approach. It would have been equally possible to adopt a stochastic process interpretation, following along the lines of Parzen (1961) and Wahba (1978), and we hope to pursue this line in future work. Inferential issues have not been touched at all, and there is much to do, although again we believe that tools available in MDA are both directly applicable for finite observations and can be extended without undue difficulty for FDAs.

Finally, the ultimate value of any set of statistical tools is determined by the value of their applications. Our aim in analysing weather data has not been to uncover new meteorological insights, but rather only to illustrate what FDA applications might look like. We believe that there is a wealth of interesting FDA applications waiting to be explored.

ACKNOWLEDGEMENTS

The authors wish to acknowledge the support of the Natural Sciences and Engineering Research Council of Canada through grants A320 and 3-31552 to the first and second author respectively. The first author would also like to acknowledge the support of the Educational Testing Service during his leave there.

REFERENCES

Aronszajn, N. (1950) Theory of reproducing kernels. *Trans. Am. Math. Soc.*, **68**, 337–404.

Bates, D., Lindstrom, M., Wahba, G. and Yandell, B. (1987) GCVPACK-routines for generalized cross validation. *Communs Statist. Simuln Computn*, **16**, 263–297.

Besse, P. (1979) Etude descriptive des processus: approximation et interpolation. *Thèse de Troisième Cycle*. Université Paul-Sabatier, Toulouse.

—— (1988) Spline functions and optimal metric in linear principal component analysis. In *Component and Correspondence Analysis: Dimensional Reduction by Functional Approximation* (eds J. L. A. van Rijkevorsel and J. de Leeuw). New York: Wiley.

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

Besse, P. and Thomas-Agnan, C. (1989) Le lissage par fonctions splines en statistique: revue bibliographique. *Report*. Université Paul-Sabatier, Toulouse.

Canadian Climate Program (1982) *Canadian Climate Normals, 1951–1980*. Ottawa: Environment Canada.

Cox, D. D., Koh, E., Wahba, G. and Yandell, B. S. (1988) Testing the (parametric) null model hypothesis in (semiparametric) partial and generalized spline models. *Ann. Statist.*, **16**, 113–119.

Craven, P. and Wahba, G. (1978) Smoothing noisy data with spline functions: estimating the correct degree of smoothing by the method of generalized cross-validation. *Numer. Math.*, **31**, 377–403.

Dalzell, C. J. and Ramsay, J. O. (1990) Computing reproducing kernels with arbitrary boundary constraints. Unpublished.

Dauxois, J., Pousse, A. and Romain, Y. (1982) Asymptotic theory for the principal component analysis of a vector random function: some applications to statistical inference. *J. Multiv. Anal.*, **12**, 136–154.

Eaton, M. L. (1983) *Multivariate Statistics: a Vector Space Approach*. New York: Wiley.

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

Kimeldorf, G. S. and Wahba, G. (1970) A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.*, **41**, 495–502.
——— (1971) Some results of Tchebycheffian spline functions. *J. Math. Anal. Applic.*, **33**, 82–94.
Parzen, E. (1961) An approach to time series analysis. *Ann. Math. Statist.*, **32**, 951–989.
Peele, L. and Kimeldorf, G. (1977) Prediction functions and mean-estimation functions for a time series. *Ann. Statist.*, **5**, 709–721.
Ramsay, J. O. (1989) The data analysis of vector-valued functions. In *Data Analysis, Learning Symbolic and Numeric Knowledge* (ed. E. Diday). Commack: Nova.
Schumaker, L. (1981) *Spline Functions: Basic Theory*. New York: Wiley.
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.
Wahba, G. (1978) Improper priors, spline smoothing and the problem of guarding against model errors in regression. *J. R. Statist. Soc.* B, **40**, 364–372.
——— (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.
Wegman, E. J. and Wright, I. W. (1983) Splines in statistics. *J. Am. Statist. Ass.*, **78**, 351–365.

DISCUSSION OF THE PAPER BY RAMSAY AND DALZELL

Bernard Silverman (University of Bath): I would like strongly to support the authors of tonight's paper in encouraging the consideration of data analysis from a functional perspective. I very much hope and expect that this paper and its discussion will be a focus for future methodological work in this area and for applications of techniques of functional data analysis (FDA) in a wide range of fields.

Recently I have been involved in work on FDA in collaboration with various colleagues. In Rice and Silverman (1991) we describe approaches to the estimation of the mean and covariance functions of a collection of observed curves. I shall describe this work fairly briefly and refer readers to the published paper for the diagrams and further details. Our motivation was a set of data on gait collected at Children's Hospital, San Diego; see Sutherland *et al.* (1988). Both the similarities to and the differences from tonight's paper may be of interest.

In our data—as in the Canadian weather data—the data curves are cyclical and are observed at a finite number of points, not necessarily the same for each curve. The approach of tonight's authors is first to smooth the data and then to use the resulting curves as the observations for the next part of the analysis. We chose instead to estimate the functional mean μ and the principal components of the variance function v (in the notation of Section 2.1.1) directly from the observed data. Because Professor Ramsay and Professor Dalzell have previously smoothed the observations they can estimate μ and v by equations (9) and (10) without any smoothing; in our case smoothing needs to be incorporated at the estimation stage. Our approach has the advantage that there is a more natural 'functional' form of cross-validation, where entire curves $x_i(\,)$ are omitted in turn, and that the smoothing parameter is chosen directly by reference to the particular quantity, mean or principal component, being estimated. A cross-validation score constructed in this way is, up to an irrelevant constant, almost exactly an unbiased estimator of the mean-square error of the estimate, *regardless of the covariance structure of the observations*. In the Ramsay–Dalzell approach, the generalized cross-validation step for the construction of the individual curves may give difficulties if there is autocorrelation in the observations on each curve.

The principal components of the covariance can be characterized as best predictors, in that, for example, individual realizations from the population of curves with covariance v can be well modelled as multiples of the leading principal function ξ_1. If the data are not previously smoothed, then, for reasons given in Rice and Silverman (1991), some smoothing will improve the estimation of this principal function. A natural approach is to bound the roughness of each estimated eigenfunction $\hat{\xi}_j$, by adding a constraint on the smoothness of $\hat{\xi}_j$ to conditions (18). The value of this constraint acts as a smoothing parameter, and an equivalent approach is to penalize the empirical covariance matrix by a multiple of a roughening matrix. Again, a natural functional cross-validation approach is obtained by leaving out entire observed curves in turn.

Our data set was also multivariate in that several angles were observed simultaneously through the gait cycle. The kind of diagram given by the authors in their Fig. 2 is most interesting, in providing a parametric cyclical fit to the data. This kind of curve is a valuable

complement to figures of the sort that we present in our Fig. 4, where the way in which the principal components of variability affect the cyclic mean curve is represented. Of course, our curve can only be drawn for the population as a whole, while the authors' curves can be drawn out for each individual weather-station.

Most of the detailed work of Rice and Silverman (1991) is presented for the case where all the curves are observed at the same time points. For the more general case where the time points are different on each curve, it is necessary to be able to approximate inner products of the form $\int u(t)\, x(t)\, dt$ where u is given at points t_i and x at points τ_j. In Rice and Silverman (1991) we used a simple quadrature idea based on linear interpolation, but I am glad to report that work done in collaboration with David Walter on a range of quadrature methods has demonstrated that the choice of quadrature method is in practice immaterial to the estimated mean and principal components, at least for the gait data.

An aspect of tonight's paper that might be worrying to the non-specialist reader is the very early introduction of concepts like reproducing kernel Hilbert space. Readers should be assured that an understanding of mathematical functional analysis is not in principle necessary for the application of FDA, but concepts of FDA will need careful development to make them accessible to the applied statistician. I fear that mathematical concepts may have got the better of the authors in their decision in Fig. 6 to provide plots of $L\beta$. A more easily interpretable quantity to plot might be $L^*L\beta$, where L^*L is the operator such that, in the notation of equation (3),

$$\langle e, f\rangle_2 = \int (L^*Le)(t)\, f(t)\, dt.$$

The operator L is crucial to the definition of the spaces H_1 and H_2, but once the residual space has been defined might it not be better, in many circumstances, to produce plots in which the role of L is hidden? A plot of $L^*L\beta$ can, I think, be interpreted without reference to L at all, but only depends on the decomposition into 'structural' and 'residual' spaces.

It is rare in statistical writing to distinguish between a finite dimensional space or a Hilbert space and its dual. For example in Section 2.1.3 $\beta(\ ,\ t)$ is really a member of the dual of H_z (which is, by a standard theorem, isometrically isomorphic to H_z). This distinction is not mere pedantry but needs to be borne in mind when interpreting results of FDA. Even in the finite dimensional case it is easy to be misled. For example the remark in Section 4.1.1 that 'most of the harmonic variation is accounted for by . . . z_{iS} . . . since only $B_S(t)$ differs substantially from 0' needs to be qualified in the light of knowledge of the relative variability of z_{i1}, z_{iS} and z_{iC} over i. I suspect that z_{i1} is much more variable and so the most important contribution overall will be that of B_1. A final point concerning interpretation is that the functions $B_S(t)$ and $B_C(t)$ in Fig. 5 display a strong component of variability at double the frequency of their corresponding harmonics. How does this tie in with the assertion that log(precipitation) is well approximated by temperature?

It is clear that FDA is an exciting and largely unexplored field of great potential. Professor Ramsay and Professor Dalzell have travelled all the way from Canada to give us a fascinating and thought-provoking contribution. For both answering and asking important questions they deserve congratulation, and it is with great pleasure that I propose the vote of thanks.

John T. Kent (University of Leeds): I like the idea of thinking of measurements collected through time as a *function* rather than as a multivariate *vector*. However, I question whether the methodology needs to be as complicated as the authors make out in this paper. What I would like to do in my discussion is to explore the possibility of a less elegant but simpler approach to functional data analysis.

As a first step I propose to use an off-the-shelf smoother such as a cubic spline (based on the differential operator $L = D^2$), rather than a smoother based on a general linear differential operator. There are at least two possible virtues in this approach. First, it is much more straightforward to compute a cubic spline, as the required software is already widely available. Second, it can be helpful to use the same smoother when comparing different examples.

Indeed it seems rather unsatisfactory to me to let the smoother depend on the number of terms in the regression equation. For instance, in the Canadian climate example, a glance at Fig. 1(b), which contains the residuals from fitting a regression on the first-order harmonics, suggests that most of the remaining variability lies in the second-order harmonics. If these second-order harmonics are also included in the regression equation, the differential operator for the smoother will involve fifth-order derivatives, an order which seems excessive.

Thus I would first fit a cubic smoothing spline to each set of time measurements in the data set. If a residual analysis is called for, I might fit the regression to the (continuous) cubic smoothing spline

rather than to the (discrete) raw data. Residuals can then be computed by subtracting the fitted regression curve from each cubic smoothing spline.

Next I would like to compare the cubic smoothing spline and other smoothing splines in more detail. I am pleased that Professor Ramsay and Professor Dalzell have chosen to illustrate their methodology on the simple example of regularly spaced periodic measurements. For in this case there is an elementary trigonometric approximation to smoothing splines which both illuminates the effect of smoothing and indeed forms a smoothing procedure in its own right. Wahba (1990), chapter 2, gives further information about this approximation. The key idea involves restricting attention to trigonometric polynomials with only as many terms as there are measurements. The details are as follows.

Given regularly spaced measurements $\{y_j, j=1, \ldots, p\}$ on the interval $[0, 1]$ with $p=2q$, let

$$f(t)=b_0+2^{1/2}\sum_{k=1}^{q-1}\{b_k\cos(2\pi kt)+a_k\sin(2\pi kt)\}+b_q\cos(2\pi qt),$$

$t\in[0, 1)$, be the interpolating trigonometric polynomial. To specify the type of smoothing, use a linear differential operator

$$L=\sum_{i=0}^{I}\alpha_i D^i$$

with constant coefficients, where the sum contains only odd or even values of i. Then it is not difficult to show that the trigonometric approximation to the smoothing spline with smoothing parameter $\lambda\geqslant 0$ is the trigonometric polynomial with damped coefficients $a_k^*=a_k/c_k$ and $b_k^*=b_k/c_k$, where

$$c_k=1+h_k\lambda\{\Sigma\alpha_i(-2\pi k)^i\}^2.$$

Here $h_k=1$ for $k\neq 0$, q, and $h_0=h_q=\frac{1}{2}$. In general the higher frequencies are damped more strongly than the low frequencies. Also the coefficients of any harmonics in the kernel of L (e.g. b_0 for $L=D^2$, and b_0, b_1 and a_1 for $L=(2\pi)^2D+D^3$) are unaltered by this smoothing.

Further, this trigonometric approximation offers a clarification of the effect of a functional principal component analysis. Consider a data set with n sets of measurements. Each set of measurements can be viewed as a p-vector of measurements in the time domain or rotated to a p-vector of trigonometric coefficients in the frequency domain. For both unsmoothed and smoothed data, principal component analysis (PCA) is effectively the same in both settings. However, the effect of smoothing can be seen most clearly in the frequency domain where a functional PCA has the following interpretation. First rescale the coefficients to damp down the higher order harmonics as described above. Then carry out a conventional multivariate PCA using the p rescaled coefficients as new variables.

As you can see I have found the proposals in this paper very stimulating and exciting. It therefore gives me great pleasure to second the vote of thanks.

The vote of thanks was passed by acclamation.

Bruce W. Turnbull (Cornell University, Ithaca): I was looking for an application in my own work. The following example is typical. In a clinical trial to assess the effects of dietary supplementation of selenium on the prevention of various forms of cancer, it is important to monitor the plasma level of the nutrient in each subject. The response curve is sigmoidal in shape; it rises from a base-line level at the start of treatment and then plateaus after a few weeks. Typically only a few plasma level readings are taken during this time, but on several hundred patients. Of course, there is large subject-to-subject variation. It is important to understand the underlying shape of the curves as this helps to determine how to lag the outcome variables in an analysis of efficacy. Departures from the curves are also important as these can identify groups of non-responding or non-complying patients. We have performed some preliminary analysis of these data by fitting regression splines with random effects; however, can the authors' methods be used to good advantage here?

Peter J. Diggle (Lancaster University): The paper discusses the increasingly important problem of what to do when confronted by a data set consisting of several curves. I would like to give two examples of a related class of problems in which the data themselves are numerical, but their most effective summary is in terms of several empirically derived curves.

The first example concerns replicated spatial point patterns in a designed experiment. In the simplest case of a completely randomized design, we are given a pattern of n_{ij} points from the jth replicate in

the ith treatment group. Replicates might be patterns consisting of the locations of cell nuclei in tissue sections from different subjects. A useful summary of the second-moment properties of each pattern is the estimated K-function $\hat{K}_{ij}(s)$ (Ripley, 1977). Suppose that we want to test for differences among the underlying group mean K-functions $K_i(s)$. One way is to construct standardized residual functions,

$$R_{ij}(s) = n_{ij}^{-1/2}\{\hat{K}_{ij}(s) - \hat{K}_{i.}(s)\},$$

where the dot denotes an appropriately weighted average over the relevant subscript, and from these to generate bootstrap re-samples of estimated K-functions *under the null hypothesis that* $K_i(s) = K(s)$ *for all* i,

$$\hat{K}^*_{ij}(s) = \hat{K}_{..}(s) + n_{ij}^{1/2} R^*_{ij}(s)$$

where the $R^*_{ij}(s)$ are bootstrap re-samples from the collection of $R_{ij}(s)$. As with Professor Silverman's cross-validation procedure for time series data, the bootstrap operates at the level of the complete *functions* $R_{ij}(s)$. Details, including choice of test statistic and an application to neuroanatomical data, are in Diggle and Lange (1991).

The second example concerns spectral analysis of replicated time series. Consider m replicates of a stationary process $\{Y_t\}$ with spectrum $f(\omega)$. Classical theory suggests that the individual periodogram ordinates should be (asymptotically) independent exponential variates. In particular, if $I_k(\omega_j)$ denotes the periodogram ordinate of the kth replicate at the jth Fourier frequency, then

$$I_k(\omega_j) \sim f(\omega_j)\chi^2_2/2. \tag{42}$$

In my experience, mostly with time series of hormonal concentrations in serial blood samples, the variability in the $I_k(\omega_j)$ over replicates at fixed ω_j is often much larger than would be consistent with equation (42). A plausible explanation is that the spectrum of $\{Y_t\}$ varies stochastically between replicates. This suggests an elaboration of equation (42) in which $f(\omega)$ is replaced by

$$f_k(\omega) = f(\omega)\ Z_k(\omega) \tag{43}$$

where the $\{Z_k(\omega)\}$ are independent realizations of a unit-mean stochastic process. For questions concerning the underlying biology of the system, as opposed to prediction for individual replicates, the relevant inferences concern $f(\omega)$ rather than $f_k(\omega)$. There may then be merit in treating the periodograms of the m replicates as the basic (functional) units of data, rather than the individual periodogram ordinates. Ibrahim Al Wasel and I are exploring procedures based on parametric or nonparametric versions of equation (43) and will report our results in due course.

A. C. Davison (University of Oxford): I should first like to say how much I appreciated the clear presentation of this stimulating paper.

Theoretical advantages obviously accrue from suitable formulation of statistical problems, but like some other discussants I wonder whether reproducing kernel Hilbert spaces are essential to developing the methods proposed in the paper. There are obvious didactic advantages if not.

For some time I have had data from the Physiological Flow Studies Unit at Imperial College on the velocity profiles of blood in arteries. The data are collected with an electronic probe and consist of successive velocity profiles across the diameter of an artery over the cycle of a heart-beat. The profile can be written $V(x, t)$, where x is the distance along the diameter of an artery and t time. The velocities are measured at points separated by a fixed distance δ, with $V(\ , t) = 0$ at the edges of the artery and in the surrounding tissue. The boundary of the artery dilates slightly over the period of a heart-beat, and it is not known where the boundary is except by observing where $V(x, t) = 0$. One purpose of an analysis would be to produce 'average' profiles for patients to discriminate between people with arteriosclerosis and healthy individuals. The methods described by the authors seem appropriate to the job. Have they any experience of this type of problem, or any advice to offer?

Frank Critchley (University of Warwick): In welcoming this innovative paper, I would like to comment on four aspects.

Philosophy

Its philosophy is close to that of van der Heijden *et al.* (1989), namely an exploratory data analysis of residuals from a model that is of secondary interest. Can fitting and/or diagnostics be introduced in situations where, *a priori*, there are several competing models and/or unknown model parameters? Clearly such model choice interacts strongly with residual analysis.

Generality

How general could functional data analysis (FDA) become? The methods presented can, in principle, generalize any multivariate data analysis (MDA) which, viewed geometrically, involves projection in a suitable inner product space. This includes both non-linear regression and projection pursuit. At another level, just as FDA extends MDA extends univariate data analysis by progressing from a scalar to a vector to a function so again, in principle, FDA can be extended by replacing the scalar argument t of a function x by a vector (or by a function!: the process can then be iterated!). The first of these embraces both spatial data with t containing the co-ordinates at which x occurs, and spatiotemporal data with t also containing a time co-ordinate. This provides a possible framework for Professor Diggle's contribution and for the example in the present paper respectively.

Modelling

The spatial nature of the data has modelling implications. The phases and shifts at the several weather-stations could be estimated jointly based on a suitable spatial process, rather than separately as in the paper. And it questions the present choice of λ which treats the stations as replicates. Also on the modelling aspect, do the authors have an explanation of the constant shift underfitting or overfitting evident in three of the diagrams in Fig. 7?

Why do it in function space?

FDA undoubtedly merits greater attention. However, certain cautions seem in order.

(a) You cannot get something for nothing. The switch from the finite to the infinite is more apparent than real. Only a finite dimensional part of function space is being considered. And even there *interval* estimates may well be wide.
(b) The present approach confounds the estimation of x from finite data with its partitioning into structural and residual components. In situations where, unlike the data in the present paper, there is appreciable observational error the methods proposed here can be expected to differ considerably from those discussed by Professor Silverman.
(c) Aspects of FDA can be seen as a roundabout way of reinventing the wheel when the end result is a straightforward MDA albeit in a metric inspired by the idea of filtering. However, why not go for that metric directly? This is the view taken by Kruskal and Winsberg in de Leeuw *et al.* (1986).

Actually, I believe there is ample scope for constructive interplay between MDA and FDA.

Randy Eubank (Texas A&M University, College Station): I would like to congratulate Professor Ramsay and Professor Dalzell on a fascinating and insightful exposition. They have presented a rich new set of data analysis tools while opening up many challenging problems for future research. My comments pertain to an inferential issue associated with their partitioning of functional variation into structural and residual components. One may often wonder whether all the systematic or explainable variation in x as a function of t resides in its structural component alone. For the data in the paper this corresponds to asking, for example, if all the systematic time variation in temperature is attributable to purely sinusoidal components.

Using the data $\mathbf{y}$ for a specific observation (as in Section 3), we might conduct such inference by examining $\mathbf{y}-\hat{u}$. A natural approach is to fit a smoothing spline to $\mathbf{y}-\hat{u}$ to detect any remaining patterns in t. The resulting fit is found to be $\hat{e}=K_2c_2$ and a corresponding test statistic is $T=\hat{e}'\hat{e}$. (This is recognized as an approximation to the rather obvious measure $p\int\hat{e}^2$ in the case of uniformly spaced t_j.) Under certain conditions (Eubank and Spiegelman, 1990; Jayasuriya, 1990), T may be recentred and rescaled in such a way that it has an approximate normal distribution. An alternative test statistic is provided by $\|\hat{x}\|^2$ whose limiting null distribution can often be handled by using techniques similar to those required for T. For multiple response vectors simultaneous assessments can be made using, for example, Bonferroni's inequality. It would have been interesting to see what these statistics told us about the data in Fig. 1.

The basic principle behind the goodness-of-fit tests described above is quite general. We fit a parametric model in t and then smooth the difference between the response and parametric fit to detect any remaining systematic variation. The magnitude of the smoothed difference function is then assessed by using an appropriate norm. This approach can also be used to determine the goodness of fit of some of the linear models discussed in this paper. I would be interested in the authors' views on this.

Philippe Besse (Université Paul Sabatier, Toulouse): Over the past two decades, spline functions have been successfully introduced in various areas of statistics for nonparametric estimation or research on optimal data transformations. This paper is a synthesis which illustrates what kinds of result can be obtained when these tools are applied in the framework of multivariate analyses of processes or functional data.

This approach focuses on the choice of a metric in linear modelling and principal components analysis (PCA), i.e. the choice of an inner product or a semi-inner product within a Hilbert functional space (although usual statistical packages impose the metric associated with the identity matrix). In this functional context, L-splines and reproducing kernels are very efficient tools, since they allow us to consider an infinite variety metric; thus, the problem now seems to be: how could we optimize this choice?

More generally, this technological transfer of numerical tools requires some statistical considerations. Cubic or 'natural' splines have a simple mechanistic interpretation: they minimize a quadratic quantity consistent with an 'energy' under constraints and with boundary conditions. What are the statistical interpretations of such properties?

This general problem can be illustrated on a special case. Spline functions and PCA can both be considered as smoothing tools. The first is driven by a smoothing parameter λ whose optimal value depends on data regularity. It is illustrated by the example dealing with temperature and precipitation data. In contrast, PCA of longitudinal data gives approximations of trajectories whose smoothness usually depends on the number q of components that are retained. A short study on artificial noisy data by Besse and Pousse (1991) shows that λ and q must be jointly optimized to obtain a better estimation of the original data. This is achieved by considering a mean-square risk proposed by Besse (1990) to study the stability of projection spaces in PCA and whose jackknife estimate is approached by perturbation theory.

These remarks agree with the conclusion of the paper where the authors hope to pursue their work to adopt a 'stochastic process interpretation' and to introduce some 'inferential issues'.

The following contributions were received in writing after the meeting.

Naomi S. Altman (Cornell University, Ithaca): Functional data analysis (FDA) provides a novel and flexible approach to repeated measures analysis. Unlike the classical approach to repeated measures analysis of designed experiments, FDA does not require that all experimental units be measured at the same time intervals, or estimation of the covariance structure.

Approaches to repeated measures analysis similar to FDA have been suggested in the past. When the data come from a stationary equally spaced time series, Brillinger (1973, 1982) suggested the use of multivariate data analysis (MDA) on the power spectrum. Another commonly used approach is to model each series as a parametric (or more recently nonparametric) curve and to analyse parameters or features of the curve, by using MDA. FDA offers advantages in relaxing some of the assumptions of these approaches. As well, analysing the entire curve, rather than a small set of features, should improve the flexibility of the approach.

One shortcoming of FDA, as presented here, is the lack of error estimates for the estimated parameter curves. Possibly Hotelling's method of tubes (Johansen and Johnstone, 1990) could be applied to this problem.

The authors have suggested the use of generalized smoothing splines to model the curves in terms of their structural and residual components. A simpler approach would be to model each curve semiparametrically, as

$$x_i(t) = h_i(t, \theta) + k_i(t)$$

where $h_i(t, \theta)$ has known parametric form and $k_i(t)$ is smooth. The components can readily be computed by using standard methodology, and can be analysed in the same method as the components of the spline decomposition.

The use of nonparametric regression to model the functional component of the observations requires the determination of smoothing parameters. Methods based on estimates of prediction error, such as cross-validation, are known to fail dramatically when the error correlations tend to be positive (Altman, 1990; Hart, 1990). Routine methods, such as the method of Gasser *et al.* (1989), may be less sensitive to error correlation. It is likely that optimal smoothing parameters will vary with the type of analysis intended, and that these parameter selection methods will need to be adjusted for FDA. It would be interesting to know how sensitive FDA methods are to selection of smoothing parameters and to choice of smoothing method.

D. R. Brillinger (University of California, Berkeley): After reading the paper, to obtain a feeling for what was proposed, I was led to try to think of a single-realization ordinary time series circumstance that was similar. To that end consider the model

$$y(t) = \mathbf{u}' \mathbf{A}_1(t) + \sum_{s=-\infty}^{\infty} e(s)\, A_2(t-s) + \epsilon(t)$$

$t = 0, \pm 1, \ldots$ with $\mathbf{A}_1(t)$ made up of things like t, t^2, $\cos(\omega t)$, $\sin(\omega t)$, with $A_2(\ )$ something like a realization of a stationary process and with $\epsilon(t)$ (stationary) noise. In practice we could have data $\{\mathbf{A}_1'(t), A_2(t), y(t)\}$, $t = 0, 1, \ldots, T-1$, available and wish to estimate $\mathbf{u}$, $e(\ )$ and characteristics of $\epsilon(\ )$. I do not remember having seen an investigation of this 'mixed transfer function' model. With $A_2(\ )$ absent we have the time series regression model, often studied under Grenander's conditions. With $A_1(\ )$ absent we have the usual linear time invariant system model.

A variety of estimation techniques can be considered for the parameters of the model: maximum likelihood, Gaussian, smoothness priors, penalized likelihood, the inverse problem, weighted least squares etc. Naïve estimates would be given by

$$\hat{\mathbf{u}} = \left\{ \sum_t \mathbf{A}_1(t)\, \mathbf{A}_1'(t) \right\}^{-1} \left\{ \sum_t \mathbf{A}_1(t) y(t) \right\}$$

$$\hat{E}(\lambda) = \hat{f}_{\hat{n}\hat{n}}(\lambda)^{-1}\, \hat{f}_{\hat{n}y}(\lambda)$$

where $E(\ )$ denotes the Fourier transform of $e(\ )$, where $\hat{n}(t) = y(t) - \hat{\mathbf{u}}' \mathbf{A}_1(t)$ and where $\hat{f}_{\hat{n}\hat{n}}(\ )$, $\hat{f}_{\hat{n}y}(\ )$ are power and cross-spectral statistics. If $\mathbf{A}_1'(t) = \{\cos(\omega t), \sin(\omega t)\}$ with ω unknown, we can envisage a further stage of minimization to estimate ω. When $y(\ )$ and $A_2(\ )$ are vector valued we could go on to a frequency side canonical analysis in the manner of Brillinger (1969).

In summary, what my reading of the approach of the paper leads to in ordinary (stationary) time series analysis is a pertinent mixed transfer function model. Analytic interest arises from the differing regularity conditions for $\mathbf{A}_1(\ )$ and $A_2(\ )$.

The paper of Professor Ramsay and Professor Dalzell is timely and stimulating. I congratulate them. I have to confess though to being somewhat aghast when I read the sentence 'Our aim in analysing weather data has not been to uncover new meteorological insights'. I take the aim of the analysis of scientific data as new scientific insight.

Chong Gu (Purdue University, West Lafayette): I am glad to have the opportunity to contribute to the discussion of Ramsay and Dalzell's interesting paper, and I take this opportunity to seek clarification about the motivation of functional data analysis (FDA) and the relation between FDA and multivariate data analysis (MDA), especially about the notion of principal component analysis.

My understanding of equation (12) is a model

$$y_i(t) = \sum_{s=1}^{q} z_{is}\, \beta_s(t) + \epsilon_i(t),$$

where the $y_i(t)$ are observed t-varying responses, the z_{is} form the usual design matrix, the $\beta_s(t)$ are t-varying linear coefficients, and $v(s, t)$ as in equation (10), which might be estimated by

$$\sum_{i=1}^{n} \omega_i \{y_i(s) - \hat{y}_i(s)\}\{y_i(t) - \hat{y}_i(t)\}$$

in model (12), is the covariance function of $\epsilon_i(t)$. So if we freeze t the model reduces to a standard linear model. Also t could be varying discretely, say on $\{1, \ldots, 12\}$ (R^{12} is a reproducing kernel Hilbert space). If my understanding is correct, then the principal component analysis of Section 2.2 does not have any statistical resemblance to the conventional principal component analysis in MDA, but rather is similar in spirit to spectral analysis in time series.

Now I try to understand the data analysis. To keep things simple I shall try to duplicate the authors' FDA, according to my understanding, using MDA (or an FDA on a discrete domain). Let $y_i(t)$ be the average temperature at station i during month t. After fitting harmonics in t for each i individually (note that this is *not* model (12)), we obtain residuals $\hat{x}_i(t)$. Now we assume that $\hat{x}_i(t)$, $i = 1, \ldots, 35$, behave like independent and identically distributed (IID) samples from a random distribution on R^{12} and estimate the covariance matrix of this distribution by equation (10). Here model (12) *is* used, but

only in a trivial form ($q=1$, $z=1$). Now an eigenvalue analysis of matrix $v(s, t)$, which is a principal component analysis by the authors' definition, provides the most variable projection of the residual vectors (the first principal component), say $\Sigma_{t=1}^{12} \alpha_t X(t)$, the most variable projection orthogonal to the first (the second principal component), etc. If this outline roughly describes a parallel to the authors' analysis, what do these principal components tell us, what is the rationale of such an analysis in the first place and what is the justification for treating the $\hat{x}_i(t)$ as IID samples?

Finally I note that in an analysis of joint temperature–precipitation variation (Section 4.2.2) $v(s, t)$ would be 2×2 matrix valued, and an eigenvalue analysis of $v(t, t)$ for fixed t looks more like a conventional principal component analysis.

M. C. Jones (The Open University, Milton Keynes): The trick that Professor Silverman mentioned in his spoken contribution of selecting a few curves for display out of a set of functional data, based on order statistics of their scores in principal component directions, has had its virtues extolled in Jones and Rice (1992).

As an example not in that paper but relevant to the work of Professor Ramsay, the splotch that is Fig. 1 of Besse and Ramsay (1986) and reproduced as Fig. 11 here—it comprises 42 superimposed curves

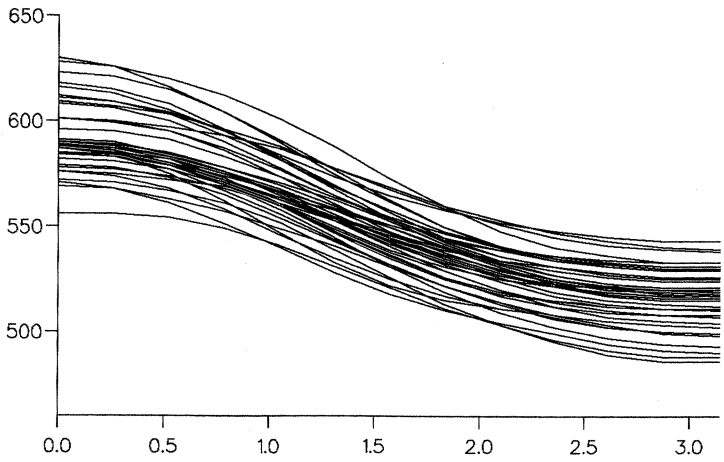

Fig. 11. 42 data curves of Besse and Ramsay (1986)

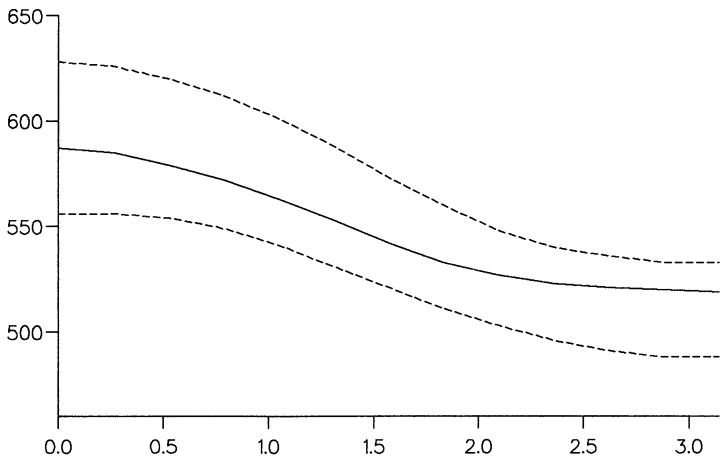

Fig. 12. Curves corresponding to first, 42nd (- - - -) and 22nd (———) ordered scores on the first principal component, extracted from Fig. 11

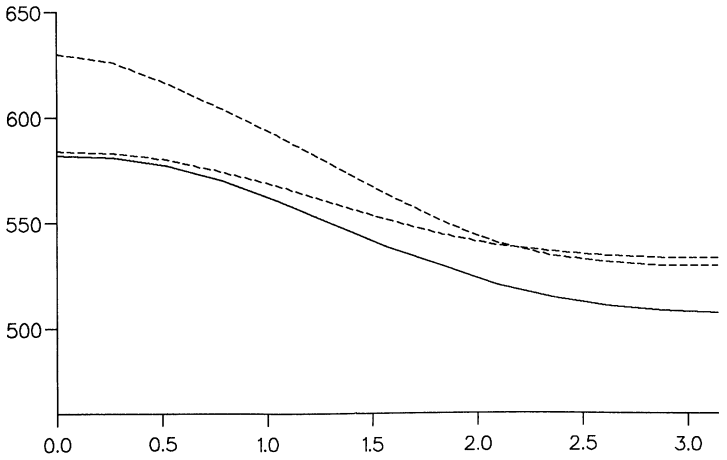

Fig. 13. As for Fig. 12, except that this uses the second principal component

representing tongue movement involved in speech production—has its main features more clearly displayed in Figs 12 and 13. In Fig. 12, the full curve is the data function corresponding to the median score (actually the 22nd) and the broken curves are the data curves associated with the most extreme scores on the first ('classical') principal component. This mode of variation, which accounts for some 90% of the variability in the functional data, clearly represents overall level. Similar curves with respect to the second principal component, worth a little over 8% of the overall variation, are given in Fig. 13; here, the interpretation is in terms of steepness or shallowness of the downward slope of the curves. (The principal components themselves are drawn in Fig. 5 of Besse and Ramsay (1986).) Aside from the aesthetic advantages of Figs 12 and 13 over Fig. 11, the relative sizes of the first two eigenvalues give a feel for how many of the curves are roughly parallel (most) and how many cross (not many), an impression not easy to be clear about from the naïve data plot.

Examples, of a more substantial nature, in Jones and Rice (1992), include the application of this functional data display technique to simulated sets of kernel density estimates and to raw curve data on ozone levels.

Edward J. Wegman (George Mason University, Fairfax): As the authors suggest, modern automated data collection allows data to be collected in the form of functions. Even with automated data collection, recording and processing of the data is more likely to take a digital form, so that analogue recording of functions is unlikely. But automated data collection allows a scope of data collection so that nonparametric function estimation, rather than simply parameter estimation, is feasible. Thus, the authors make the point that, even though we have a finite collection of data recorded, we are often interested in nonparametrically estimating functions on a continuum. Hence I agree strongly with motivations (b) and (c) but am somewhat sceptical of (a) (Section 1).

My earlier approaches to this problem (Wegman, 1984) were to conceive of an admissible subset of a normed linear function space $\mathcal{M}$ and a convex optimization functional $\mathcal{L}$. The idea was to construct the function space to have $\mathcal{M}$ contain only functions having the known properties of the function to be estimated. Smoothness and order preserving are typical such properties. The convex functional is chosen as desired but can be such functions as log-likelihood, penalized least squares or the like. Wegman (1984) formulates this framework and gives existence and uniqueness results but does not prescribe computational algorithms for calculating these functional estimators. The Wegman $\mathcal{M}$ space seems to be equivalent to the Ramsay–Dalzell H_1 space.

Ramsay and Dalzell offer a somewhat more restricted framework but in return obtain more manageable computational algorithms. They restrict themselves to Hilbert spaces H^m, with $m-1$ absolutely continuous derivatives and a square integrable mth derivative. This essentially makes the framework of this paper a Sobolev space of functions, an L^2-like space which makes the estimators found by the Ramsay–Dalzell methodology generalized L-splines. The only inner product function space of real-valued functions is of course L^2. Hence, the inference being discussed is squared error based.

Although Wegman (1984) discusses only existence and uniqueness, Le and Wegman (1990) have applied that methodology to the estimation and detection of transient (sonar) signals. Their methodology focuses on signal waveforms which are both time limited and frequency band limited in some specific senses. The application of the function estimation methodology gives rise to solutions which are generalized L-splines and Le and Wegman's results are probably special cases of the present paper. Although Sobolev spaces introduce a rich structure principally because the reproducing kernel Hilbert space property allows the comparatively easy formulation of basis functions spanning the space, other basis functions hold at least as much promise. In particular, wavelet bases (see Daubechies (1988), Daubechies *et al.* (1986) and Heil and Walnut (1989)) offer alternatives which are highly appropriate for both time series and nonparametric regression applications.

The **authors** replied later, in writing, as follows.

We are grateful to the discussants for adding so much to the paper. Several directions for further work are indicated, and some fascinating potential applications of functional data analysis (FDA) are offered.

We wonder whether some discussants might not have entirely appreciated that this was not designed to be a paper about one more smoothing procedure. If the only objective is to smooth the data, possibly before some analyses of the smooths, it seems likely that a wide variety of smoothers might do about the same job, as suggested by Buja *et al.* (1989).

Instead, we wanted to propose the use of a differential operator L as a substantively interesting technique for partitioning the functional variation into two components. One of these components consists of a simple 'carrier signal' u, such as a sinusoid plus a constant, and the other contains a residual function e. We may not only be interested in e in itself, but also in terms of the function Le, which in some sense may indicate the 'power' of the external influences perturbing the fundamental, u. For example, cubic splines can be thought of not only as an addition to being a useful smoothing tool, but also a technique permitting us to separate linear from non-linear behaviour, and studying the latter in terms of its 'acceleration', D^2e. Now, where such an interpretation of the functional data has appeal, the penalty functional $\|Le\|^2 = \int (Le)^2(t)\,dt$ also has a natural interpretation. It is exactly what we would want to use to be as conservative as possible about how much functional variation spills over into the residual space H_2, because it penalizes the power of external influences in this generalized sense.

We react to the excellent FDA example of Rice and Silverman by inviting them to go further by looking directly at

$$(Lx)(t) = \{(2\pi)^2 D + D^3\}(t),$$

since a shifted sinusoid is such a strong component of their data, and since we are hardly surprised to see a mechanical system like the skeletal system designed to minimize non-harmonic variation.

However, if the only role that L plays is to partition functional variation into orthogonal components, it may be that techniques like partial splines will serve as well. We also wonder whether, for large numbers of observation values t_j, it might be more convenient to work with kernel smoothers using kernels constructed from linear combinations of derivative estimation kernels.

Incidentally, we might add a criticism here of our own work. Although we analysed the climate data as if they were a sinusoid superimposed on possibly non-periodic components, partly to illustrate a technique useful for non-periodic data, it would have been preferable to use strictly periodic splines, or splines on the unit circle. A minor modification of the derivation described in Wahba (1990), p. 21, shows that the H_2-reproducing kernel for the above operator is

$$k_2(s, t) \propto \sum_{n=2}^{\infty} \frac{\cos\{2\pi n(t-s)\}}{(n^3-n)^2}, \qquad s \leqslant t. \tag{44}$$

When we analyse the data with this kernel, the results are very similar to those reported in the paper, except that $L\beta$ displayed in Fig. 6 is no longer required to be 0 at the extremes.

The arterial blood flow problem described by Dr Davison sounds absolutely fascinating. We would propose a toroidal co-ordinate system, θ, ϕ, r, where r measures the distance from the centre of the artery to the point of observation, θ is the angle of this point and ϕ is a value in $[0, 2\pi]$ indicating time within the heart-beat cycle. The harmonic functions satisfying Laplace's equation in this system are of the form

$$\exp(-ik\theta)\exp(-im\phi)\,J_n(c_n r), \qquad k, m = 1, \ldots, \quad n = 0, 1, \ldots, \tag{45}$$

where J_n is the Bessel function of the first kind and c_n is a known constant which positions the first zero of the nth Bessel function at the arterial wall, where blood flow will be 0. Since only J_0 is non-zero at the centre, it can be considered as measuring the average blood flow across the arterial disc for any fixed time t. The differential operator required would appear to be that which annihilates the components $\exp(-i\theta)\exp(-i\phi)\,J_0(c_0 r)$. Since not only is equation (45) an orthogonal system, but so is any order of its derivatives, it turns out to be easy to work out its reproducing kernel. The situation resembles that of splines on the sphere, which is also discussed in Wahba (1990).

The growth curve problem described by Professor Turnbull seems to call for the differential operator that annihilates span$\{1, \Psi\}$, where $\Psi(t) = 1/\{1+\exp(-a-bt)\}$. Assuming that constants a and b are known, the required differential operator is $L = D^2 - b(1-2\Psi)D$. We are investigating a similar problem in collaboration with R. Bock at the University of Chicago and have developed the reproducing kernel k_2 using the approach of Dalzell and Ramsay (1990). Of course, in many situations constants a and b will require some level of estimation, and here the exciting work mentioned by Professor Besse has much to offer.

As we announced in the paper, the perspective in the paper was constrained by space to be data analytic, as opposed to statistical. But we appreciate the enormous importance of interval estimation and inference in this domain. We thank Professor Altman and Professor Eubank for suggesting some useful results. And we certainly agree with Dr Jones that this is a fertile area for considering new graphical display techniques such as those which he has proposed.

Several discussants wonder whether the machinery of functional analysis in general, and reproducing kernel Hilbert spaces in particular, is really essential to FDA. In part they point out, along with Dr Critchley and Professor Kent, that in certain settings the implied results are quite familiar and have been around for a long time. To this we respond that we expect that more general tools should reduce to the familiar in certain simple situations, such as splines on the circle with equally spaced t_j. We would mistrust something that failed to do this. However, we are unrepentant about the use of functional analysis and cannot resist pointing out that it was commonplace only 25 years ago to argue that familiarity with linear algebra was indispensable to multivariate data analysis. Review papers such as Rao (1964) typically opened with basic results on eigenanalysis to help those along whose mathematical training was weak in this regard. The step from linear algebra to most of functional analysis is not large, and we expect that the need to model functional data will eventually force this topic into the training of data analysts and statisticians.

Finally, we want to take yet another opportunity to recognize the seminal contributions to FDA of Professor P. Besse, Professor J. Dauxois and Professor A. Pousse, associated now or formerly with Université Paul Sabatier, Toulouse, as well as those of several other colleagues in France.

REFERENCES IN THE DISCUSSION

Altman, N. S. (1990) Kernel regression estimation of data with correlated errors. *J. Am. Statist. Ass.*, **85**, 749–759.

Besse, Ph. (1990) P.C.A. stability and choice of dimensionality. To be published.

Besse, Ph. and Pousse, A. (1991) Extension des analyses factorielles. In *Modèles pour l'Analyse des Données Multidimensionelles* (eds J. J. Droesbeke, B. Fichet and Ph. Tassi). Paris: Economica.

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

Brillinger, D. R. (1969) The canonical analysis of stationary time series. In *Multivariate Analysis II* (ed. P. R. Krishnaiah). New York: Academic Press.

——— (1973) The analysis of time series collected in an experimental design. In *Multivariate Analysis III* (ed. P. R. Krishnaiah), pp. 241–256. New York: Academic Press.

——— (1981) *Time Series: Data Analysis and Theory*. San Francisco: Holden-Day.

Buja, A., Hastie, T. and Tibshirani, R. (1989) Linear smoothers and additive models (with discussion). *Ann. Statist.*, **17**, 453–555.

Dalzell, C. J. and Ramsay, J. O. (1990) Computing reproducing kernels with arbitrary boundary constraints. Unpublished.

Daubechies, I. (1988) Orthonormal basis of compactly supported wavelets. *Communs Pure Appl. Math.*, **41**, 909–996.

Daubechies, I., Grossmann, A. and Meyer, Y. (1986) Painless non-orthogonal expansions. *J. Math. Phys.*, **27**, 1271–1283.

Diggle, P. J. and Lange, N. (1991) Analysis of variance for replicate spatial point patterns in neuronatomy. *J. Am. Statist. Ass.*, **86**, in the press.

Eubank, R. L. and Spiegelman, C. H. (1990) Testing the goodness-of-fit of a linear model via nonparametric regression techniques. *J. Am. Statist. Ass.*, **85**, 387–392.

Gasser, T., Kneip, A. and Köhler (1989) A flexible and fast method for automatic smoothing and differentiation. Unpublished.
Hart, J. D. (1990) Kernel regression with time series errors. *J. R. Statist. Soc.* B, **53**, 173–187.
van der Heijden, P. G. M., de Falguerolles, A. and de Leeuw, J. (1989) A combined approach to contingency table analysis using correspondence analysis and log-linear analysis (with discussion). *Appl. Statist.*, **38**, 249–292.
Heil, C. and Walnut, D. (1989) Continuous and discrete wavelet transforms. *SIAM Rev.*, **31**, 628–666.
Jayasuriya, B. (1990) Testing for a polynomial model using nonparametric regression techniques. *PhD Dissertation*. Texas A&M University, College Station.
Johansen, S. and Johnstone, I. M. (1990) Hotelling's theorem on the volume of tubes: some illustrations in simultaneous inference and data analysis. *Ann. Statist.*, **18**, 652–684.
Jones, M. C. and Rice, J. A. (1992) Displaying the important features of large collections of similar curves. *Am. Statistn*, **46**, in the press.
Le, H. T. and Wegman, E. (1990) Nonparametric identification of linear system response. In *Signal Processing V* (eds L. Torres, E. Masgrau and M. Lagunas), pp. 229–232. Amsterdam: Elsevier.
de Leeuw, J., Heiser, W., Meulman, J. and Critchley, F. (eds) (1986) *Multidimensional Data Analysis*. Leiden: DSWO.
Rao, C. R. (1964) The use and interpretation of principal component analysis in applied research. *Sankhya* A, **26**, 329–358.
Rice, J. A. and Silverman, B. W. (1991) Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B, **53**, 233–243.
Ripley, B. D. (1977) Modelling spatial patterns (with discussion). *J. R. Statist. Soc.* B, **39**, 172–212.
Sutherland, D. H., Olshen, R. A., Biden, E. N. and Wyatt, M. P. (1988) *The Development of Mature Walking*. Oxford: Blackwell Scientific.
Wahba, G. (1990) *Spline Models for Observational Data*. Philadelphia: Society for Industrial and Applied Mathematics.
Wegman, E. (1984) Optimal nonparametric function estimation. *J. Statist. Planng Inf.*, **9**, 375–388.