Statistica Sinica **14**(2004), 631-647

FUNCTIONAL AND LONGITUDINAL DATA ANALYSIS: PERSPECTIVES ON SMOOTHING

John A. Rice

University of California, Berkeley

Abstract: The perspectives and methods of functional data analysis and longitudinal data analysis for smoothing are contrasted and compared. Topics include kernel methods and random effects models for smoothing, basis function methods, and examination of correlates of curve shapes. Some directions in which methodology might advance are identified.

Key words and phrases: Functional data analysis, longitudinal data analysis, nonparametric curve estimation.

1. Introduction

Until recently, functional data analysis (FDA) and longitudinal data analysis (LDA) have been rather disjoint enterprises. Both are concerned with the analysis of data consisting of repeated measurements of objects over time. Measurements treated in the FDA literature typically are recorded by high frequency automatic sensing equipment, whereas those treated in the LDA literature are more typically sparsely, and often irregularly, spaced measurements on human or other biological subjects. The aims of the analysis are often somewhat different, partly because of different scientific subject matter (it is interesting to read and compare the introductions of the two classic texts of Diggle, Liang, and Zeger (1994) and Ramsay and Silverman (1997)). Those of FDA tend to be exploratory − to represent and display data in order to highlight interesting characteristics, perhaps as input for further analysis − whereas those of LDA have a stronger inferential component. This contrast can be seen in the use of estimated time-correlation functions in the two areas; correlation functions are used in the FDA literature in a descriptive manner to characterize time dependencies in the curves, whereas an important aim in the LDA literature of estimating these correlation functions is to draw valid inferences. Another important difference is that the LDA literature has had to pay much more attention to data that are missing because of a variety of mechanisms.

Despite these differences in focus, there are many common aims, among them are the following:

1. Characterization of average or "typical" time course.
2. Estimation of individual curves from noisy and, often in LDA studies, sparse data. Functionals of these curves, such as derivatives and locations and values of extrema, are sometimes also of interest.
3. Characterizing homogeneity and patterns of variability among curves, and identifying unusual ones.
4. Assessing the relationships of shapes of curves to covariates.

Many of these objectives entail smoothing individual curves, either explicitly or implicitly. That is the primary focus of this paper.

I will attempt to compare and contrast LDA and FDA perspectives and methods of smoothing for this volume of *Statistica Sinica*, based on the stimulating conference, Emerging Issues in Longitudinal Data Analysis, which brought the two communities together. Themes I pursue include considering how strength is borrowed in FDA and LDA, how they can borrow strength from each other, the relationships between stochastic models and smoothing algorithms, and directions in which current methodology might advance. This paper thus constitutes a selective comparison and a view toward future directions, from my personal perspective, and is not intended to be a complete, all-encompassing review.

The remainder of the paper is organized as follows. In Section 2, kernel methods and their relation to random effects models are considered from both the FDA and LDA perspectives. Basis function methods are considered in Section 3. Linear and non-linear methods of interpolation and extrapolation are treated in Section 4, where I try to make the case for going beyond linear Gaussian models. Section 5 is concerned with assessing the relationships of correlates to curve shapes, touching briefly on joint modeling. Some final remarks are made in Section 6.

2. Local Kernel Smoothing and Random Effects Models

I first compare kernels that are typically used in FDA and those arising from an LDA perspective. I then discuss smoothing of multiple curves and illustrate how FDA can benefit from LDA practices.

2.1. Comparison of kernels

Kernel smoothing (linear averaging with smooth weights) can arise from different perspectives. As a simple example, consider the Nadaraya-Watson estimate

$$\hat{f}(t) = \frac{\sum_{i=1}^{n} Y_i w(t - t_i)}{\sum_{i=1}^{n} w(t - t_i)}. \tag{1}$$

From the FDA or non-parametric function estimation perspective, this estimate is appealing *prima facie* for purposes of estimating a broad variety of functions

$f(t)$ under weak assumptions. In the LDA literature, smoothing procedures more typically arise from stochastic models. For a fixed value of t, the estimate (1) would correspond to a model $Y_i = f(t) + \epsilon_i(t)$, $\mathrm{Var}(\epsilon_i(t)) \propto (w(t-t_i))^{-1}$. It has the property that as $|t-t_i|$ increases as t is moved away from the data, $\hat{f}(t)$ tends to a constant that depends upon the particular kernel used and the data (Y_i, t_i). For example, David Ruppert pointed out in a personal communication that if $t_1 < t_2 < \cdots < t_n$ and the kernel is Gaussian, then as t decreases, $\hat{f}(t) \to Y_1$.

This rather strange behavior could be avoided by formulating a Gaussian random effects model, $f(t) \sim N(0, \tau^2)$, leading to the estimate

$$E[f(t)|y_1, \cdots, y_n] = \frac{\sum_{i=1}^n y_i w(t-t_i)}{\tau^{-2} + \sum_{i=1}^n w(t-t_i)}. \tag{2}$$

The FDA community would recognize this as a ridge estimate, as proposed in Seifert and Gasser (1998) to ameliorate fluctuations of local polynomial estimates in regions of sparse data. This estimate extrapolates in a much more natural way than does (1), tending to 0 (or to the prior mean) as t moves away from the t_i.

In the LDA literature, smoothers and extrapolators typically arise from stochastic, or random-effects, models, in particular from classical time-series models, for example Jones (1993). An individual trajectory is modeled as a signal, a Gaussian process $Z(t)$ with mean $\mu(t)$ and covariance function $K(s,t)$, which is observed at times t_i with noise, $Y(t_i) = Z(t_i) + \epsilon_i$. The smoothed/extrapolated estimate of $Z(t)$ is then $\hat{Z}(t) = E(Z(t)|Y)$ which is a weighted linear average of the observed data. A kernel, like that in (2), is thus implicitly determined. The parameters of the kernel, most importantly the bandwidth, are determined as maximum likelihood estimates of the parameters of the covariance function. In practice, the covariance function is usually chosen for computational convenience rather than on principled grounds, and depends upon a parameter governing the correlation decay rate, i.e., a bandwidth parameter. Since according to the folk wisdom of kernel estimation, the shape of the kernel matters far less than the bandwidth, other aspects of the specification of the covariance function may not be of primary importance. Rather than estimating the bandwidth by maximum likelihood, FDA practioners are more likely to examine graphically the effects of a variety of bandwidths, or to use a data-driven selection procedure such as cross-validation.

The correspondence of smoothing and stochastic models was pointed out in the case of smoothing splines in Kimeldorf and Wahba (1970) and the implicitly determined kernel was studied in Silverman (1984). The relationship of smoothing splines to stochastic random effects models was exploited in Brumback and

Rice (1998) and Wang (1998) for the analysis of longitudinal data. There is often a useful formal correspondence between smoothing methods using a penalty function and random effects (Bayesian) models.

2.2. Smoothing multiple curves

With a view of learning how an LDA approach may be suggestive when viewed from an FDA perspective, let us now consider a simple model, involving no covariates other than time, a mean function $\mu(t)$, depending upon linear parameters, individual random curves $f_i(t)$ depending on random linear parameters, and a parametrized covariance function. The model is

$$Y(t_{ij}) = \mu(t_{ij}) + f_i(t_{ij}) + \epsilon_{ij}, \tag{3}$$

where t_{ij}, $i = 1, \ldots, n$, $j = 1, \ldots, n_i$ is the time of the jth measurement of the ith individual. Such models are common in the LDA literature. The EM algorithm (Laird and Ware (1982)) proceeds iteratively. Given estimates at stage k,

1. $f_i^{(k+1)}(t)$ = weighted linear combination of $\{y(t_{ij}) - \mu^{(k)}(t_{ij})\}$,
2. $\mu^{(k+1)}(t)$ = weighted linear combination of the averages of $\{y(t_{ij}) - f_i^{(k)}(t_{ij})\}$,
3. Parameters of the covariance function are estimated by maximum likelihood based on the expected values of the sufficient statistics.

This suggests an FDA analogue. As an illustration, I will use data on childrens' gaits (Olshen, Biden, Wyatt and Sutherland (1989) and Rice and Silverman (1991)), the angles formed by each child's ankle during a single phase of a gait cycle as shown in Figure 1.

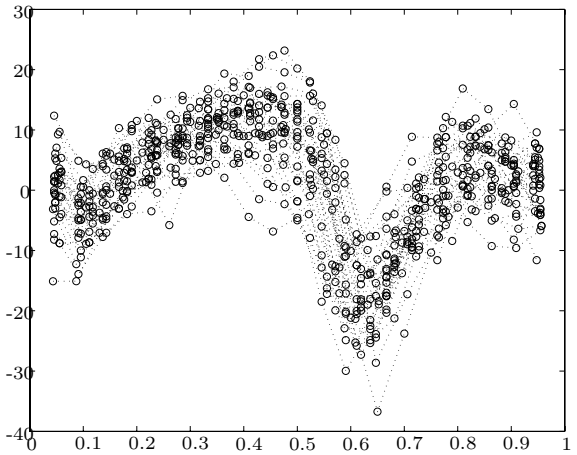

Figure 1.The angles formed by the ankles of 39 children during a single phase of a gait cycle.

Typical FDA practice would be to smooth each curve independently of the others. The result of using a local linear smoother with a bandwidth equal to 0.2 is shown as the dashed line in Figure 2. I refer to this as a "one-pass" smooth.

A "two-pass" procedure was then used, mimicking the first two EM steps above: (1) the individual curves were first smoothed, the average of the smooths was found; (2) this average was subtracted from the individual data points of each curve, and the results were smoothed again. (For a more elaborate related procedure see Wu and Zhang (2002).) Again, a local linear smoother with a fixed bandwidth equal to 0.2 was used. (No attempt to adaptively determine the bandwidth was made. An analogue would be to use cross validation in place of the maximum likelihood estimation above, but I have not investigated this further.) The resulting estimate for a single child is shown as the dotted line in Figure 2. Note how the traditional "one-pass" smoothing undershoots the peaks and valleys much more substantially than does the "two-pass" procedure derived from the LDA perspective. One usually thinks of borrowing strength to reduce variance, but here by using all the curves in conjunction, the bias in estimating individual curves is substantially decreased! It appears that rather than smoothing each curve in isolation, practitioners of FDA would gain by using all the curves. This point occurs again in the next section.

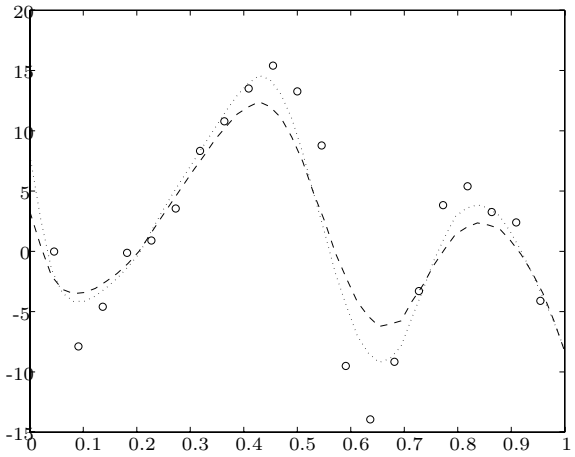

Figure 2. Comparison of one-pass smoothing (dashed line) and two-pass smoothing (dotted line) on ankle angle measurements (circles).

Other possibilities for smoothing functional data are suggested by representing the first two EM steps as

1. $\hat{f}_i^{(k+1)} = \mathsf{Smooth}\ (Y_i - \hat{\mu}^{(k)})$,
2. $\hat{\mu}^{(k+1)} = \mathsf{Smooth}\ (\mathsf{Average}\{Y_i - \hat{f}_i^{(k)}\})$.

The smoothing operator could be nonlinear, for example robust or constrained to be monotone. It could be based on local likelihoods. Similarly, averaging could be robust. The centering (subtraction of the mean) in the first step could be with respect to a registration template (see Ramsay and Silverman (1997) for a discussion of registration), with respect to a subset of the data such as a set of nearest neighbors, or with respect to the conditional expectation given a covariate. Looking ahead to the next section, the centering could be formed by a projection of an average onto a set of basis functions.

3. Basis Function Methods

Both the FDA and LDA literature have a history of using expansions in basis functions, but the points of view have been rather different. The typical FDA approach is to fit each curve individually by ordinary least squares to data on a regular grid as

$$f_i(t) = \sum_{k=1}^{K} U_{ki}\psi_k(t), \tag{4}$$

where the functions $\{\psi_k(t)\}$ form a basis and the U_{ki} are the (random) coefficients of curve i. As the dimension K of the subspace grows, increasingly oscillatory functions are admitted. The measurements of the individual random curves are thus represented as

$$Y_i(t_{ij}) = f_i(t_{ij}) + \epsilon_i(t_{ij}). \tag{5}$$

The $\epsilon_i(t)$ are composed of measurement errors and high frequency contributions from the $\psi_j(t)$, $j > K$. The choice of the truncation point is frequently made by some sort of cross-validation procedure. This finite dimensional representation is avowedly an approximation and the basis functions are chosen for good approximation properties.

In FDA, the covariance structure is of direct interest rather than viewed as a nuisance parameter complicating inferences. It is usually estimated directly from the observations $Y(t_{ij})$, if they are on a regular grid.

The basis formed by the eigenfunctions of the covariance function is stressed in FDA (Rice and Silverman (1991)). The random curves can be expanded in this basis and, because of well known extremal properties of the eigenfunctions and eigenvalues, a truncated expansion is optimal in a mean square sense among all expansions in a set of basis functions of the same order. The underlying notion is that although the curves are evaluated at a large number of points, there are not a large number of important "degrees of freedom," or modes of variation, in the ensemble. The eigenvalue-eigenfunction decomposition affords a compact representation in which the leading functions often provide an interpretable description of the major modes of variation in the collection of curves and by which

projections of individual curves onto the eigenfunctions assists in characterizing homogeneity and patterns of variability among curves, and identifying unusual ones.

In the LDA tradition the design points might well be sparse and perhaps irregularly spaced. The data vector of subject i might be modeled as

$$Y_i(t_{ij}) = \mu_i(t_{ij}) + \sum_{k=1}^{K} U_{ki}\psi_k(t_{ij}) + \epsilon_i(t_{ij}). \tag{6}$$

Here $\mu_i(t)$ is typically a parametric fixed-effects function, perhaps involving covariates (nonparametric estimates have been considered in Zhang, Lin, Raz and Sowers (1998), for example). These fixed effects are of primary interest. The functions $\psi_k(t)$ are typically polynomials and K is small, so that perhaps only linear or quadratic functions are included. The observation noise process $\epsilon_i(t)$ is usually either specified to be white noise or a convenient continuous time stationary process − the Ornstein-Uhlenbeck process is a popular choice. The covariance structure of the random trajectories $Y_i(t)$ is thus determined by the functions $\psi_k(t)$, the covariance of the coefficients U_{ki} and the covariance function of $\epsilon(t)$. This covariance structure is not of primary interest and is largely specified *a priori*. Contrasting this to the FDA-style expansion, we see that the high-frequency component, $\epsilon_i(t)$ in (6) corresponds to the higher order $\psi_k(t)$ and the white noise in (4) and (5).

The coefficients U_{ki} in (6) and their covariance structure are estimated by standard mixed-model technology. For linear, Gaussian models, the smoothed observations are

$$\hat{Y}_i(t) = E[(\mu_i(t) + \sum_{k=1}^{K} U_{ki}\psi_k(t_{ij}))|Y_i)] = W_i Y_i, \tag{7}$$

where W_i is a matrix of weights derived from the covariance structure and Y_i is the vector of observations on subject i. Dropping the Gaussian assumption, the estimate is the best linear approximation to the Bayes estimate, the BLUP; "best" given the model, of course.

It is interesting to contrast the estimates of the coefficients of individual curves with respect to a given eigenfunction as they would be constructed by FDA and LDA. The use of these coefficients is illustrated in Jones and Rice (1992). For simplicity of notation, suppose that the data are on a regular grid. Let Y be the observations of a curve, which are assumed to be corrupted by white noise with variance σ^2. The FDA-style estimate of the coefficient of the kth eigenfunction ψ_k with eigenvalue λ_k would be $Z_k = \langle Y, \psi_k \rangle$ whereas the LDA-style estimate would be $Z_k = (\lambda_k/(\lambda_k + \sigma^2))\langle Y, \psi_k \rangle$. The latter damps the

contributions from the eigenfunctions with small eigenvalues, which are typically rough, and hence smooths more. (Neither of these estimates takes explicitly into account that the eigenfunctions and eigenvalues are in fact estimated from the data.) This approach is explored in Yao, Müller, Clifford, Dueker, Follet, Lin, Buchholz and Vogel (2003). Generally, the random effects models used in LDA borrow strength across curves in estimating coefficients, resulting in shrunken estimates, whereas this practice is not followed in FDA, which estimates the coefficients corresponding to each curve in isolation.

A hybrid approach was proposed in Shi, Weiss and Taylor (1996) and Rice and Wu (2001). Basis functions were used as an approximating set in the spirit of FDA, but fit to the irregular and sparse data typical of LDA, using standard algorithms for Gaussian mixed models. Both papers used splines, but other basis functions could be used as well. Adaptive, data-driven methods were considered for choosing the truncation point K. A smooth estimate of the covariance function of the underlying random curves was thus determined by the estimated covariance function of the random coefficients U_{ki} and the basis functions ψ_k, from which eigenfuntions could be estimated if desired. (An alternative approach of estimating the eigenfunctions and eigenvalues directly from sparse irregular data is given in James, Hastie and Sugar (2001).)

4. Interpolation and Extrapolation

Specification of the mean and covariance function, as in framework of the previous section, allows the interpolation of data at unobserved time points via (7). The covariance function can be estimated as in the previous section, or, alternatively, can be estimated directly by smoothing scatterplots, Staniswalis and Lee (1998) and Diggle and Verbyla (1998). It appears that there has been little work on analyzing the consequences of the nature and degree of smoothing done in estimating the covariance function on the properties of the estimated "BLUP." (The presence of the B and the U in this context is not really appropriate, nor is the L, since the covariance function is estimated, leaving a rather uninformative acronymn.)

The apparent effectiveness of these linear methods is quite surprising given that the data are sparse and irregular and perhaps quite non-Gaussian. A typical task is to estimate $E(Y(t_0))$ given $Y(t_1), Y(t_2), Y(t_3)$, say, and there may be no instances of data for which Y is observed at these four points.

Multivariate normality, which conveniently only requires information on pairs to construct a full joint distribution, must be a quite dubious model in many longitudinal data sets. The presence of variable biological timing, for example coupled phase shifts, in growth curves can give rise to non-Gaussianity. One would think that in many applications the points on the random curves would

be highly constrained in a nonlinear way and that the distributions would be concentrated on non-linear manifolds. Indeed non-Gaussianity can be seen quite dramatically in the gait data. To empirically examine the joint distribution of the ankle trajectories, I interpolated the data linearly at 20 equispaced points and then used, xgobi (http://www.research.att.com/areas/stat/xgobi/), an interactive graphical tool for the exploration of multivariate data, to search for non-Gaussian projections of the 39 20-vectors. Two interesting projections are shown in Figure 3. (The precise effects of the linear interpolation are difficult to understand, but I would imagine that this process would make the results more, rather than less, Gaussian.)

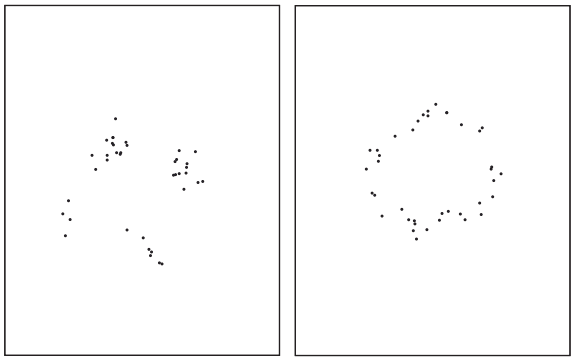

Figure 3. Two highly non-Gaussian projections of ankle-angle trajectories obtained by projection pursuit. Each point corresponds to an individual curve.

Thus one would expect that modern nonparametric regression procedures would be useful, but there has been little work on how to incorporate them in the context of functional data analysis (for example, Ramsay and Silverman (1997) is largely devoted to the extension of classical Gaussian multivariate analysis to the functional setting). Perhaps the key to the effectiveness of the "BLUP" is that the estimate is additive, i.e., that the estimate is a weighted average of simple univariate linear predictors, weights being given by the estimated covariance function. It would be worthwhile to examine the feasibility of suitable nonparametric additive models (Hastie and Tibshirani (1990)). More precisely, a family of smoothly related additive models would be required to handle the various time configurations of observed data and times at which interpolation was desired.

The results of a simple experiment illustrate the potential for using nonlinear methods for interpolation and extrapolation. Data points with $t < 0.3$ were deleted from an ankle-angle curve, a smooth curve was then fit to the remaining points, and the seven nearest neighbors among the individually smoothed curves

for the other subjects were found and averaged. This average alone would not be a suitable estimate of the missing curve, because there would be a discontinuity where it joined the observed curve. Continuity of the curve and its derivative was enforced in the following way: to extrapolate at a point $t < 0.3$, a weighted linear combination of the nearest neighbor average and a linear extrapolation from the observed curve was computed, with the weights decaying linearly from $w(0) = 1$ to $w(0.3) = 0$. Since the point is to illustrate the plausibility of nonlinear methods in general, and of a nearest-neighbor based method in particular, no attempt was made to fine-tune the procedure; for example a weighted average of nearest neighbors could be found, the number of neighbors varied, and the extrapolation weights varied. Figure 4 shows the results of performing this on four randomly selected curves, with varying degrees of success.

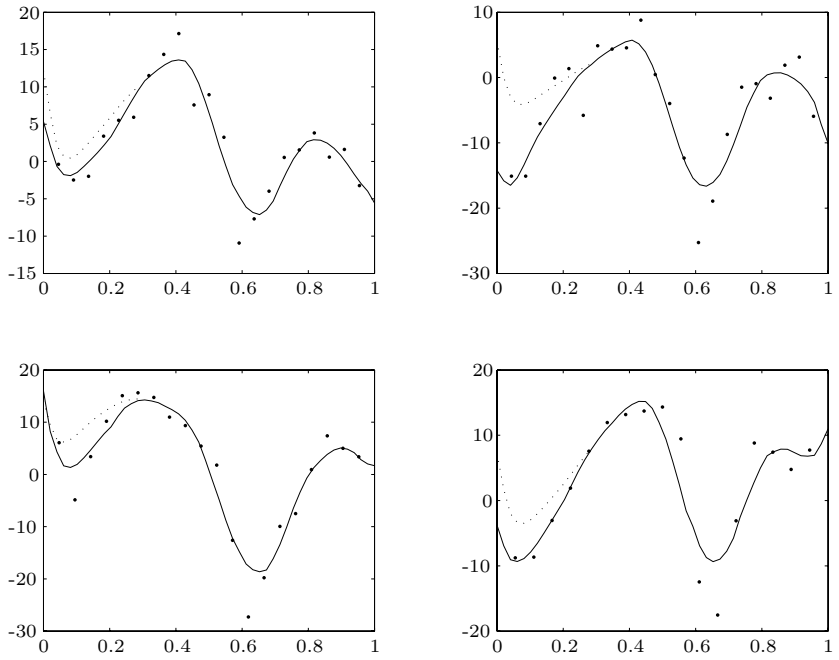

Figure 4. Four examples of using nearest neighbors to fill in missing data for $t < 0.3$. The original data are shown as points, the smooths using all the data are solid lines, and the extrapolated values are shown by dashed lines.

5. Correlates of Curve Shape

Linear models for functional and scalar covariates have been developed in the FDA literature, e.g., Ramsay and Silverman (1997). For the most part, though, this nonparametric point of view has not permeated the LDA literature, and has

not addressed common problems such as data sparsity, irregularity, censoring and truncation. Within the framework of basis function modeling the individual random curves are reduced to finite vectors of coefficients, making it possible to examine the dependence of the curve shapes on covariates, via classical linear or modern nonparametric techniques. Thus, if Z is a covariate,

$$E(Y_i(t)|Z) = \sum_{k=1}^{K} E(U_{ki}|Z)\psi_k(t), \tag{8}$$

and studying the dependence of curve shapes on Z reduces to studying its relationship to individual coefficients. Nonparametric methods of clustering and classification can be based on the coefficients, e.g., James and Sugar (2003). One of the main advantages of the basis function approach is that the reduction to a finite dimensional representation allows subsequent use of a variety of standard statistical tools.

To illustrate some possibilities for going beyond classical methods, and using nonparametric smoothing techniques for exploratory data analysis, I will use some data kindly provided by Hans Müller and Jane-Ling Wang on daily fertility of a cohort of 1,000 medflies (Carey, Liedo, Müller, Wang and Vaupel (1998a) and Carey, Liedo, Müller, Wang and Chiou (1998b)). Some examples of individual fertility curves are shown in Figure 5. There are large daily fluctuations for each fly and large variation from fly to fly.

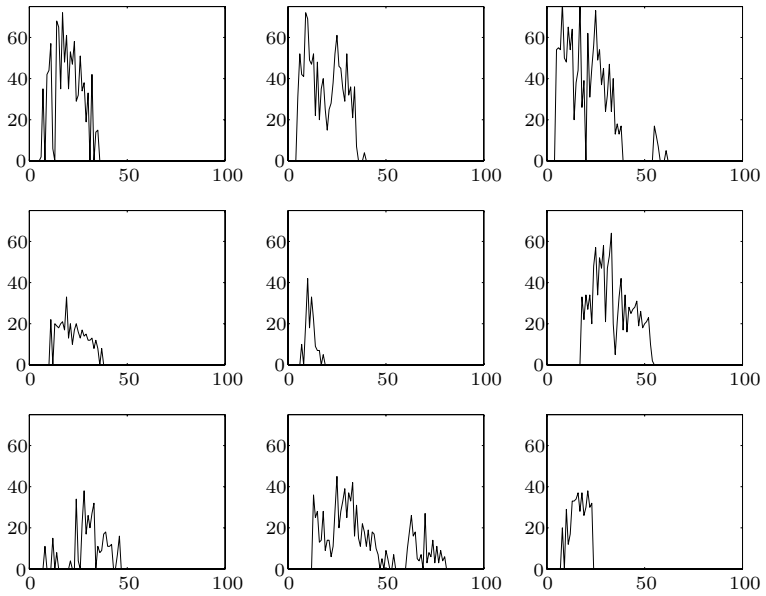

Figure 5. Fertility curves of nine medflies. Number of eggs produced is shown on the vertical axis and time in days is shown on the horizontal axis.

This fascinating data set has been used in a variety of studies on aging. One important aim in the study of aging is to investigate how patterns of reproductive activity relate to longevity. There is a large literature on the concept of a "cost of reproduction." The papers cited above contain further references and point to the influence of remaining reproductive potential on mortality.

In what follows, I will illustrate the potential for smoothing methods to elucidate relationships between longevity and reproduction curves in a model-free manner. Let $Y_i(t)$ denote the number of eggs laid on day t by fly i and let L_i denote its lifetime. We wish to examine the relationship between lifetime and the function $Y(t)$ (one noteworthy point is that $Y(t)$ is only defined for $1 \le t \le L$). This endeavor is a component of "joint modeling" in the LDA literature Wulfsohn and Tsiatis (1997), Wang and Taylor (2001), Henderson, Diggle and Dobson (2000) and Tsiatis and Davidian (2003)). Parametric survival models, such as Cox proportional hazard, play a central role in these approaches; in contrast, and in the FDA spirit, smoothing can be used in an exploratory way to help characterize the relationship between a longitudinal process and time to death.

The raw data are shown in Figure 6, in which the egglaying profiles are ordered by longevity (a similar plot appears in Carey, Liedo, Müller, Wang and Vaupel (1998a)). Even if the Figure is interactively viewed from many angles, and with the aid of color, the large variability makes it very hard to get a sense of underlying patterns.

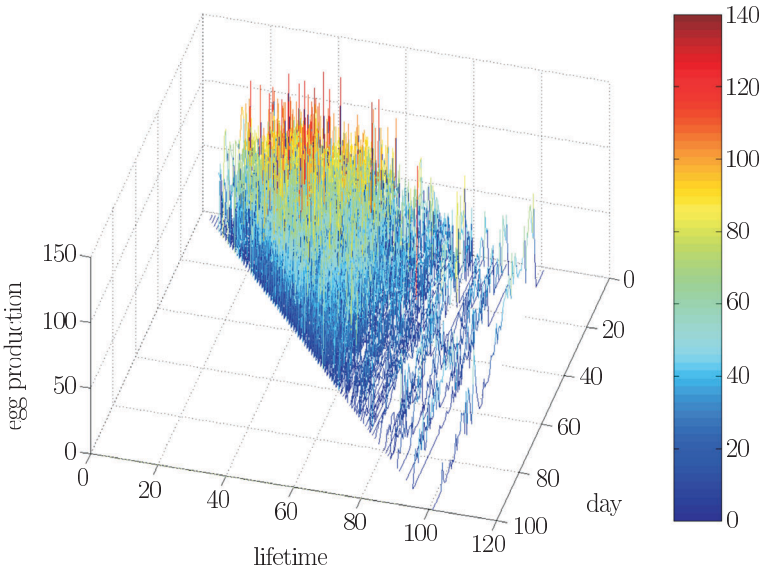

Figure 6. Egg laying profiles ordered by lifetime.

We thus consider smoothing, to form a nonparametric estimate of $E(Y(t)|L)$, for $t = 1, \ldots, L$. We construct this estimate in two stages. (1) An initial estimate of $E(Y(t)|L)$ is constructed as follows: for each fixed t, consider all the values $Y_i(t)$ for flies which have lived at least t days, ordered by lifetimes L_i, and smooth them with respect to L by a local polynomial smoother. The results of this are shown in the left panel of Figure 7. (2) Fixing $L = \ell$, the function $E(Y(t)|L = \ell)$ is estimated by smoothing the previous result with respect to t, yielding the results shown in the right panel of the figure. In the second stage smoothing, I preserved the initial run of 0's for small t that was produced in the first stage. There is an interesting tradeoff between smoothing with respect to L and with respect to t; in this case it seems advantageous to smooth with respect to L first, since the subsequent smoothing with respect to t, if needed, can be quite light.

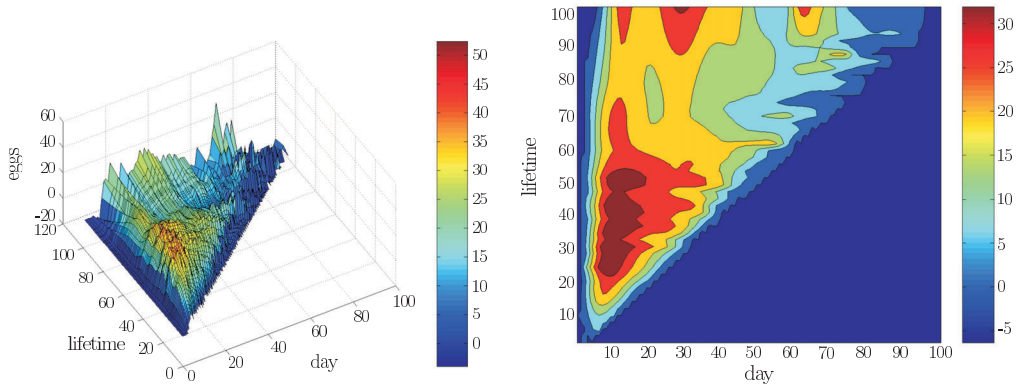

Figure 7. Left panel: smoothing with respect to lifetime. Right panel: smoothing the results shown in the left panel with respect to time.

The figures show that the initial period of latency, or immaturity, increases with the lifetime of the fly. The contours slope gradually away from the vertical axis and the time of peak egg laying (10−20 days) shifts later in time with increased lifetimes in the range of about 10 to 60 days, as if those flies were running on progressively slower "clocks." Reproductive activity after this peak tapers off more gradually with increased lifetime. The peak disappears, or broadens substantially, for flies living 60-90 days. The longest lived flies exhibit additional fertility peaks at about 35 and 65 days. Although these phenomena can be discerned by careful study of Figure 6, they are brought into vivid focus by this simple smoothing procedure.

The conditional expectation of $Y(t)$, $1 \le t \le 100$, given total egg production, $\sum_s Y(s)$, was estimated in a similar way. The biological interest is that total egg production can be viewed as a measure of "success," and there are various possible

strategies for how to optimize this by spacing production through time. The results are shown in Figure 8. In the left panel, there is little sign of the rotation of the contours seen in Figure 7, although a hint of second peak production period appears for the most prolific flies. The right panel shows the result of normalizing each egg laying profile by the total number of eggs laid. With the exception of some deviation for small times and some broadening as production increases, the curves exhibit little change with increasing total fertility. This is consistent with the multiplicative effects model of Chiou, Müller and Wang (2002), which used functional principal components and covariates.

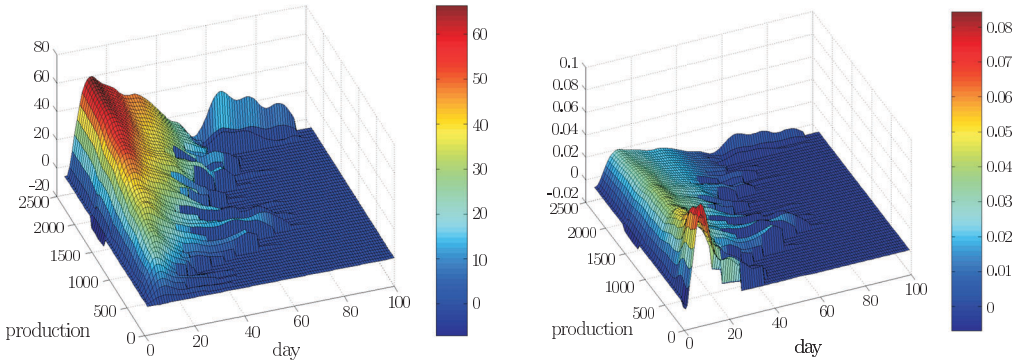

Figure 8. Left panel: smoothing with respect to total egg production. Right panel: normalization by total egg production.

Figure 7 suggests that lifetimes may be somewhat predictable from the egg production profiles, although the figure shows the conditional expectation of the latter given the former, rather than visa-versa. To explore this possibility, I conducted a simple illustrative experiment. The egg laying profiles for the first 30 days of medflies which lived that long were used to attempt to predict the corresponding lifetimes. There were 667 such. The individual profiles were smoothed with a local linear smoother, $k(x) \propto (1 - x^2)^2$ with a bandwidth of 15 days. The lifetime of an individual medfly was then predicted as the mean lifetime of its 25 nearest neighbors, using the l_1 norm as a distance between profiles. The resulting correlation was 0.27, so there was some, but not much, predictability by this method. (In general, correlations are small in these data, for example the correlation between lifetime and cumulative production for the first 15 days is 0.18 and the correlation between the total number of eggs in the first 30 days and the lifetime beyond 30 days is less than 0.1). Modest exploration of the effects of modifying the weights in the norm and the weighting of nearest neighbors made little appreciable difference. The point of this example is to

illustrate the possibility of using nonparametric regression techniques in such contexts, not to find the optimal such; for example, it would be natural to try tree-based regression methods based on features of the curves, such as the latency time noted above. An analysis of a very different spirit, based on a Cox model using curve features as predictors, is presented in Müller, Carey, Wu, Liedo and Vaupel (2001).

6. Final Remarks

One of the themes of this paper is that there are opportunities for a larger role to be played by FDA-style methods in LDA and that the approaches taken in LDA, in particular borrowing strength between curves as in mixed and random-effects modeling, could be profitably considered by the FDA community.

A second theme is that there may be interesting opportunities for developing methods that go beyond linear/Gaussian approaches (Zhang (1997) is an interesting example). I have illustrated possibilities by using nearest neighbor methods. These may not ultimately be the most effective, but they do have the advantage of applicability in non-linear, non-Gaussian situations; if the curves do tend to lie on low dimensional manifolds, this structure is tracked implicitly by nearest neighbor methods without the necessity of explicit construction.

Survival analysis, one of the key concerns of the LDA community, has been largely ignored by the FDA community, which to-date has had little to say about curves defined on variable domains (lifetimes), the problems posed by joint modeling, and the difficulties of missing data of various kinds (e.g., informative and non-informative drop-out). I would hope that the techniques and the exploratory spirit of FDA would have something to offer here. I offer a modest contribution in Section 5.

This paper has not addressed problems of inference, which are of central importance to the LDA community. Inference is often addressed in that literature by likelihood analysis of rather elaborate (and frequently dubious) models. Inference has not been a central interest in the FDA community, which typically prefers to avoid parametric models of stochastic processes and rely on simple techniques such as the bootstrap and permutation tests as needed. (However, see Fan and Lin (1998)).

Acknowledgements

This work was supported by a grant from the National Science Foundation. I also wish to thank Hans Müller and Jane-Ling Wang for sharing the medfly data, for several stimulating conversations, and for comments on an early draft of this manuscript. The helpful comments of two referees led to improvements in exposition. More generally, I would like to thank the participants in the conference for sharing many interesting ideas.

References

Brumback, B. and Rice, J. A. (1998). Smoothing spline models for the analsysis of nested and crossed samples of curve. *J. Amer. Statist. Assoc.* **93**, 944-961.

Carey, J. R., Liedo, P., Müller, H.-G., Wang, J.-L. and Chiou, J.-M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproductionin a large cohort of mediterranean fruit fly females. *J. Gerontology: Biol. Sci.* **53**A, B245-B251.

Carey, J. R., Liedo, P., Müller, H.-G., Wang, J.-L. and Vaupel, J. W. (1998). A simple graphical technique for displaying individual fertility data and cohort survival: case study of 1000 mediterranean fruit fly females. *Functional Ecology* **12**, 359-363.

Chiou, J. M., Müller, H.-G. and Wang, J.-L. (2002). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. Technical report, University of California, Davis.

Diggle, P., Liang, K.-Y. and Zeger, S. (1994). *Analysis of Longitudinal Data*. Oxford University Press, New York.

Diggle, P. and Verbyla, A. (1998). Nonparametric estimation of covariance structure in longitudinal data. *Biometric* **54**, 401-415.

Fan, J. and Lin, S.-K. (1998). Tests of significance when the data are curves. *J. Amer. Statist. Assoc.* **93**(443), 1007-1021.

Hastie, T. and Tibshirani, R. (1990). *Generalized Additive Models*. Chapman and Hall, New York.

Henderson, R., Diggle, P. and Dobson, A. (2000). Joint modelling of longitudinal measurements and event time data. *Biostatistics* **1**, 465-480.

James, G., Hastie, T. and Sugar, C. (2001). Principal component models for sparse functional data. *Biometrika* **87**, 587-602.

James, G. and Sugar, C. (2003). Clustering for sparsely sampled functional data. *J. Amer. Statist. Assoc.* **98**, 397-408.

Jones, M. C. and Rice, J. A. (1992). Displaying the important features of large collections of similar curves. *Amer. Statist.* **46**, 140-145.

Jones, R. H. (1993). *Longitudinal Data with Serial Correlation: a State Space Approach*. Chapman and Hall, New York.

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.

Laird, N. M. and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

Müller, H.-G., Carey, J. R., Wu, D., Liedo, R. and Vaupel, J. W. (2001). Reproductive potential predicts longevity of female mediterranean fruit flies. *Proc. Roy. Soc. Ser. B* **268**, 445-450.

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

Ramsay, J. O. and Silverman, B. W. (1997). *Functional Data Analysis*. Springer, New York.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. Roy. Statist. Soc. Ser. B.* **53**, 233-243.

Rice, J. A. and Wu, C. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57**, 253-259.

Seifert, B. and Gasser, T. (1998). Ridging methods in local polynomial regression. In *Dimension Reduction, Computational Complexity, and Information* **30** (Edited by S. Weisburg), 467–476. Fairfax Station, Virginia. Interface Foundation of North America.

Shi, M., Weiss, R. E. and Taylor, J. M. G. (1996). An analysis of pediatric aids cd4 counts using flexible random curves. *Appl. Statist.* **45**, 151-164.

Silverman, B. W. (1984). Spline smoothing: the equivalent variable kernel method. *Ann. Statist.* **12**, 898-916.

Staniswalis, J. G. and Lee, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Amer. Statist. Assoc.* **93**, 1403-1418.

Tsiatis, A. A. and Davidian, M. (2004). Joint modeling of longitudinal and time-to-event data: an overview. *Statist. Sinica* **14**, 809-834.

Wang, Y. (1998). Mixed-effects smoothing spline ANOVA. *J. Roy. Statist. Soc. Ser. B* **60**, 159–174.

Wang, Y. and Taylor, J. M. G. (2001). Jointly modelling longitudinal and event data, with application to AIDS studies. *J. Amer. Statist. Assoc.* **96**, 895-905.

Wu, H. and Zhang, J.-T. (2002). Local polynomial mixed-effects models for longitudinal data. *J. Amer. Statist. Assoc.* **97**, 883-897.

Wulfsohn, M. S. and Tsiatis, A. A. (1997). A joint model for survival and longitudinal data measured with error. *Biometrics* **53**, 330-339.

Yao, F., Müller, H.-G., Clifford, A. J., Dueker, S. R., Follet, J., Lin, Y., Buchholz, B. A. and Vogel, J. S. (2003). Shrinkage estimation for functional principal component scores, with application to the population kinetics of plasma folate. *Biometrics* **59**, 676-685.

Zhang, D., Lin, X., Raz, J. and Sowers, M. (1998). Semiparametric stochastic mixed models for longitudinal data. *J. Amer. Statist. Assoc.* **93**, 710-719.

Zhang, H. (1997). Multivariate adaptive splines for longitudinal data (MASAL). *J. Comput. Graph. Statist.* **6**, 74-91.

Department of Statistics, University of California, Berkeley, 367 Evans Hall #3860, Berkeley, California 94720-3860, U.S.A.

E-mail: rice@stat.berkeley.edu

(Received March 2003; accepted December 2003)