Statistica Sinica **14**(2004), 675-693

FUNCTIONAL RESPONSE MODELS

Jeng-Min Chiou[1], Hans-Georg Müller[2] and Jane-Ling Wang[2]

[1]*National Health Research Institutes, Taipei and* [2]*University of California, Davis*

Abstract: We review functional regression models and discuss in more detail the situation where the predictor is a vector or scalar, such as a dose, and the response is a random trajectory. These models incorporate the influence of the predictor either through the mean response function, through the random components of a Karhunen-Loève or functional principal components expansion, or by means of a combination of both. In a case study, we analyze dose-response data with functional responses from an experiment on the age-specific reproduction of medflies. Daily egg-laying was recorded for a sample of 874 medflies in response to dietary dose provided to the flies. We compare several functional response models for these data. A useful criterion to evaluate models is a model's ability to predict the response at a new dose. We quantify this notion by means of a conditional prediction error that is obtained through a leave-one-dose-out technique.

Key words and phrases: Dose-response, eigenfunctions, functional data analysis, functional regression, multiplicative modeling, principal components, smoothing.

1. Introduction

This study is motivated by a medfly experiment described in Carey, Liedo, Harshman, Zhang, Müller, Partridge and Wang (2002). Age-specific reproduction in terms of daily eggs laid was recorded for 1,200 female medflies that were fed one of 12 dietary doses, ranging from ad libitum (full diet) to 30% of ad libitum diet. Due to an abundance of early deaths, the two lowest doses were omitted from our analysis, which is based on data for 874 flies that laid eggs and were assigned to ten dose levels, with varying numbers of flies per dose level.

The shape changes in age-specific fecundity that occur when dietary dose varies are of interest. Carey et al. (2002) demonstrated that the total number of eggs produced by individual flies is monotone increasing as dietary dose increases, with a saturation characteristic. A biological question of interest is whether this increase in the total is due to a sustained increase over all ages, or whether flies start to produce eggs earlier, for example, under a richer diet. Figure 1 displays typical response curves for three doses, 50%, 75% and 100% of the ad lib diet. A general increase in the level of the egg-laying profiles as dietary levels increase is clearly visible, and it is of interest whether additional, perhaps more subtle, changes in the level of egg-laying occur.

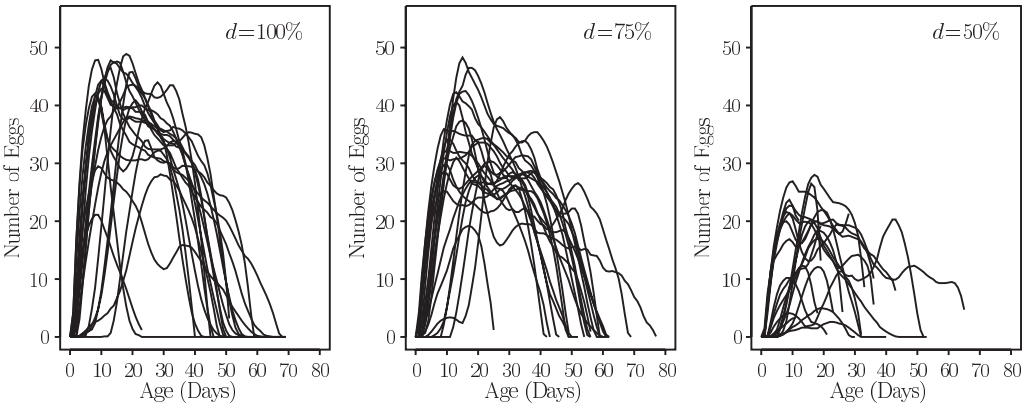

Figure 1. Smoothed egg-laying trajectories of twenty randomly selected individuals at dose levels 100%, 75% and 50%. The bandwidth is the same for all curves, and is obtained by minimizing the sum of all individual leave-one-point-out squared prediction errors.

The models we consider for our analysis fall within the general class of functional regression models (Ramsay and Silverman (1997)). These are regression models that include some components (predictors or responses) that may be viewed as random curves. Three subcategories of such models can be distinguished: both predictors and responses are functions; predictors are functions and responses are scalars; predictors are scalars and responses are functions. Due to the nature of our data, we are primarily concerned here with the latter class. We begin with a brief review of the two model classes with functional predictors and discuss the functional response model in Section 2.

Models where both predictors and responses are functions. These are usually referred to as functional regression models and were introduced in Ramsay and Dalzell (1991). These models are extensions of the multivariate linear regression model $E(Y|X) = BX$, employing a regression parameter matrix B, to the case of infinite-dimensional or functional data. The data are a sample of pairs of random functions $(X(t), Y(t))$, with X the predictor and Y the response functions.

The extension of the linear regression model to functional data is then

$$E(Y(t)|X) = \mu(t) + \int X(s)\beta(s,t)\,ds,$$

with a parameter function β and a mean response function $\mu(t)$. The estimation of the parameter function $\beta(\cdot,\cdot)$ is a main goal, and amounts to solving an inverse problem. One could attempt to generalize the least squares normal equation for $Y \in \mathbb{R}^p$, given by $\mathrm{Cov}\,(X,Y) = \mathrm{Cov}\,(X)B$ from multidimensional to functional data, leading to the "Functional Normal Equation"

$$r_{XY} = R_{XX}\beta, \quad \text{for} \quad \beta \in L^2,$$

where L^2 is the space of square integrable functions and $R_{XX} : L^2 \to L^2$ is the autocovariance operator of X, defined by

$$(R_{XX}\beta)(s,t) = \int r_{XX}(s,w)\beta(w,t)dw.$$

Here $r_{XX}(s,t) = \mathrm{Cov}[X(s), X(t)]$ and $r_{XY}(s,t) = \mathrm{Cov}[X(s), Y(t)]$ are the auto- and cross-covariance functions. Since R_{XX} is a compact operator in L^2, it is not invertible. A functional version of a generalized inverse is needed, and exists under some technical assumptions. According to He (1999) and He, Müller and Wang (2000, 2003), compare also Dauxois, Pousse and Romain (1982), if

$$X(t) = \sum_{j=1}^{\infty} \xi_j \varphi_j(t), \quad Y(t) = \sum_{j=1}^{\infty} \zeta_j \psi_j(t)$$

are the Karhunen-Loève expansions of random functions X and Y, assuming zero means and eigenfunctions φ_j, ψ_j of autocovariance operators R_{XX}, R_{YY}, a unique solution to the functional normal equation exists in the image space $\mathcal{I}(R_{XX})$, provided

$$\sum_{i,j} \left(\frac{E(\xi_i \zeta_j)}{\mathrm{Var}(\xi_i)} \right)^2 < \infty.$$

In this case, the solution of the functional normal equation can be written as

$$\beta^*(s,t) = \sum_{j=1}^{\infty} \rho_j u_j(s) R_{YY} v_j(t)$$

in terms of canonical correlations ρ_j and canonical weight functions u_j and v_j, or alternatively in terms of the eigenfunctions as

$$\beta^*(s,t) = \sum_{j,k=1}^{\infty} \frac{\mathrm{Cov}(\xi_j, \zeta_k)}{\mathrm{Var}(\xi_j)} \varphi_j(s)\psi_k(t).$$

A practical method to estimate β is to discretize the problem, first solving for a matrix B in a multivariate regression approach, and then adding a smoothing step to obtain the smooth regression parameter function β. Regularization is achieved by truncating the series at a finite number of summands K. For further asymptotic results, we refer to Fan and Zhang (1998) and Cuevas, Febrero and Fraiman (2002).

Models where predictors are functions and responses are scalars. Such models have been investigated recently by Cardot, Ferraty and Sarda (1999), Cardot, Ferraty, Mas and Sarda (2003), James (2002), Ratcliffe, Heller and Leader (2002) and Müller and Stadtmüller (2004). This type of data arises for example in

functional prediction problems, such as predicting adult height from a number of child growth measurements. Other applications are to discriminant and cluster analysis problems for functional data and random trajectories. Such problems are of interest in bioinformatics, where one goal is to classify genes according to their observed time-dependent expression profile.

A generalized functional model along these lines, allowing for generalized responses, has been proposed by James (2002), who models the principal component functions in a suitable B-spline basis and emphasizes applied aspects. Müller and Stadtmüller (2004) proposed a version of functional discriminant analysis via functional binomial regression and developed asymptotic distribution theory, using functional principal component scores. With a suitable link function g, centered random predictor functions $X(t) \in L^2(dw)$ and responses $Y \in \mathbb{R}$, the model is

$$E(Y|X) = g(\alpha + \int \beta(t)X_i(t)\,dw(t)).$$

Binomial functional regression could also be based on spline coefficient representations of the predictor functions, an approach that goes back to Shi, Weiss and Taylor (1996) and Rice and Wu (2001). Besides regression approaches, another method for curve clustering that has been investigated in recent years is based on modes in function space (Gasser, Hall and Presnell (1998), Hall and Heckman (2002) and Liu and Müller (2003)).

The remainder of the paper is organized as follows. Models for the case of a functional response and a scalar predictor are discussed in the following section and response models that include residual processes are proposed. The fitting of these models is the theme of Section 3, and special features of functional responses in dose-response analysis are discussed in Section 4. Section 5 is a case study of reproductive trajectories for medflies in response to dietary dose, and Section 6 is devoted to issues of model choice.

2. Modeling Functional Responses

In the data set that motivated this study, the predictor is a dose (amount of diet), while for each individual the response is a random trajectory (daily egg-laying). Similar situations were considered in Chiou, Müller, Wang (2003) and Chiou, Müller, Wang and Carey (2003) for different types of egg-laying data, and for the setting of motion capture data by Faraway (1997). An alternative approach models the influence of the covariate on the flow of time rather than on the amplitude of the response curves (Capra and Müller (1997)). In this paper, we introduce a general model which extends the approaches considered previously in Chiou, Müller, Wang (2003) and Chiou, Müller, Wang and Carey (2003). The

covariates are assumed to influence the shape of the response functions mainly through effects on the overall mean function and, in addition, on the conditional distributions of the functional principal component scores, if additional random components lead to better fits.

The random part of the model involves the Karhunen-Loève decomposition of the residual processes obtained after subtracting the fitted mean effects from the response functions. This means that residual processes are decomposed into orthonormal eigenfunctions ψ_k, $k = 1, 2, \ldots$, of the autocovariance operator R_{YY} defined above, with functional principal component scores as coefficients. This is an extension of the usual spectral decomposition of a covariance matrix, using eigenvectors, to the functional case. Then the eigenvectors become eigenfunctions and many new issues arise, such as convergence of the expansion to a limit, where different modes of convergence can be considered.

We assume that the eigenfunctions are smooth (say, twice differentiable) and that they do not change as the covariate level changes. An alternative approach would be a model where the eigenfunctions themselves change with the covariate; this increases the complexity but would be of interest for some cases, for example in applications where the covariate influences the size of the support of the response function.

The concept of functional principal components was developed in an early paper by Rao (1958) on growth curves, extending the multivariate technique of principal components analysis to the infinite-dimensional case. Practical and smoothing issues in estimating functional principal components were studied by Castro, Lawton and Sylvestre (1986) and Rice and Silverman (1991), with an interesting proposal to describe samples of random curves through principal components in Jones and Rice (1992).

Functional principal components are valued as a device for dimension reduction, which is essential for functional data. Infinite-dimensional trajectories can conveniently be reduced to a finite number of functional principal components that often have an applied interpretation as "modes of variation". Other approximation techniques such as orthogonal series, wavelets or B-splines may serve the same purpose and are preferable in some applications. While choosing an alternative base rather than the eigenbase has the advantage that it eliminates the need to estimate the eigenfunctions, easing the computational burden, such choices also encounter a number of disadvantages, e.g., a comparably larger number of basis functions will be necessary and there will be correlations between the functional principal component scores.

Assume that data consist of a sample of covariate vectors $\mathbf{Z}$ and associated response curves Y, $(\mathbf{Z}_i, Y_i(t))$, $i = 1, \ldots, n, t \in T$, where T is an interval and the covariate vector $\mathbf{Z}$ is in $\mathbb{R}^q$, $q \geq 1$. In developing the proposed functional response

model, we first approximate $E(Y(t)|\mathbf{Z} = \mathbf{z})$ by a product function $\mu_0(t)\theta(\mathbf{z})$, then introduce a "residual process" whose random effects are functions of the predictors $\mathbf{Z}$.

We assume that there exist twice continuously differentiable functions $\mu_0, \mu_1 : \mathbb{R} \to \mathbb{R}$ such that $E\{Y(t)\} = \mu_0(t)$ is the overall mean of the response processes $Y(t)$. Defining $\theta(\mathbf{z}) = \arg\min_\theta\{\int_T [E(Y(t) \mid \mathbf{Z} = \mathbf{z}) - \mu_0(t)\theta]^2\, dt\}$, we find by taking the derivative w.r. to θ that the minimizing function is

$$\theta(\mathbf{z}) = \frac{\int \mu_0(t)E(Y(t)|\mathbf{Z} = \mathbf{z})dt}{\int \mu_0^2(t)dt}, \tag{1}$$

provided $\int \mu_0^2(t)dt > 0$. Furthermore, since $E[E(Y(t)|\mathbf{Z})] = \mu_0(t)$, we find $E(\theta(\mathbf{Z})) = 1$, so that the least squares approximation $\mu_0(t)\theta(\mathbf{Z})$ to $E(Y(t)|\mathbf{Z})$ is well defined. In the case of multivariate covariates $\mathbf{Z}$ we make the single index assumption that there exists a parameter vector $\gamma \in \Re^q$, satisfying $\|\gamma\| = 1$, such that $\mu_1(\gamma'\mathbf{z}) = \theta(\mathbf{z})$. As a consequence, $E\{\mu_1(\gamma'\mathbf{Z})\} = 1$. The function $\mu_1(\cdot)$ thus provides an amplitude modifying factor dependent on the linear predictor $\gamma'\mathbf{Z}$.

We have seen that $\mu(t, \mathbf{z}) = \mu_0(t)\mu_1(\gamma'\mathbf{z})$ provides a least squares approximation to $E(Y(t)|\mathbf{Z} = \mathbf{z})$, and if indeed $E(Y(t)|\mathbf{Z} = \mathbf{z}) = \mu(t, \mathbf{z})$, this model corresponds to the product surface model described in detail in Chiou, Müller, Wang and Carey (2003). We note that other functional forms could be considered for $\mu(t, \mathbf{z})$ as well but, according to our experience with functional data, the product form seems to have the widest applicability in practice.

If $E(Y(t)|\mathbf{Z} = \mathbf{z}) \neq \mu(t, \mathbf{z})$, the residual process $R(t) = Y(t) - \mu_0(t)\mu_1(\gamma'\mathbf{Z})$ plays an additional role in the modeling of the response curves. Since $ER(t) = 0$, as $E\{\mu_1(\gamma'\mathbf{Z})\} = 1$, we can expand this process in terms of its eigenfunctions. Assuming that the first K eigenfunctions $\psi_k : T \to \mathbb{R}$ are sufficient for this expansion, we obtain $R(t) = \sum_{k=1}^{K} A_k\psi_k(t)$, where $E(A_k) = 0$, $\mathrm{Cov}\,(A_k, A_\ell) = \lambda_k\delta_{k\ell}$, using the Kronecker delta notation.

Furthermore, conditional on the covariates, we find

$$E\{R(t)|\mathbf{Z}\} = \sum_{k=1}^{K} E(A_k|\mathbf{Z})\psi_k(t) = \sum_{k=1}^{\infty} \eta_k(\mathbf{Z})\psi_k(t)$$

with functions $\eta_k : \mathbb{R}^q \to \mathbb{R}$, and for the conditional covariance (and with $s = t$, the conditional variance)

$$\mathrm{Cov}\,\{R(s), R(t)|\mathbf{Z}\} = \sum_{k,\ell} \mathrm{Cov}\,(A_k,\, A_\ell|\mathbf{Z})\,\psi_k(s)\psi_\ell(t).$$

Extending the single index dimension reduction scheme to random components of the model proves useful for cases where the covariate is a vector. Assume

that each of the conditional link functions associated with a random component, $\eta_k(\mathbf{z}) = E(A_k|\mathbf{Z} = \mathbf{z})$, is a smooth function of a single index formed from the data; i.e., there exist parameter vectors $\beta_k \in \mathbb{R}^p$, $\|\beta_k\| = 1$ and twice continuously differentiable functions $\alpha_k : \mathbb{R} \to \mathbb{R}$ such that

$$\alpha_k(\beta_k'\mathbf{z}) = \eta_k(\mathbf{z}) = E(A_k|\mathbf{Z} = \mathbf{z}), \quad k = 1, 2, \ldots. \tag{2}$$

Prediction of $Y(t)$, given a covariate level $\mathbf{z}$, may then be based on the conditional expectation

$$E\{Y(t)|\mathbf{Z} = \mathbf{z}\} = \mu(t, \mathbf{z}) + \sum_{k=1}^{K} \alpha_k(\beta_k'\mathbf{z})\psi_k(t). \tag{3}$$

We note that model (3) reduces to the functional quasi-likelihood regression model of Chiou, Müller, Wang (2003) when $\mu(t, \mathbf{z}) = \mu_0(t)$, and to the multiplicative effects model of Chiou, Müller, Wang and Carey (2003) when $\mu(t, \mathbf{z}) = \mu_0(t)\mu_1(\gamma'\mathbf{z})$ and $\alpha_k = 0$. The inclusion of the residual process $R(t)$ allows one to check the validity of this product surface model. Instead of the multiplicative assumption for the mean function, the conditional mean function $\mu(t, \mathbf{z})$ can be modeled as an arbitrary smooth surface in t and $\gamma'\mathbf{z}$ (Chiou, Müller, Wang and Carey (2003)).

If the residual process $R(t)$ is included in the model, the influence of the covariates on the shape of the response function Y is determined by the link functions and their single indices, the overall mean function and the eigenfunctions. To study the effect of covariates, one can look at predictions for various covariate settings, for example varying one covariate at a time. Approximate inference for covariate effects can be obtained by formulating tests for the single indices, assuming one can ignore the estimation of the eigen- and link functions. Confidence regions for the parameters can be translated into uniform confidence bands for the predicted trajectories, applying projection methods similar to Scheffé's method; compare Yao, Müller and Wang (2003) for an example in the functional setting. The link functions could be parameterized by low-order polynomials or B-splines, allowing one to apply crude significance tests for regression effect (non-constant link).

3. Fitting Functional Response Models

Fitting the models that include the residual process $R(t)$ is based on one-dimensional and two-dimensional scatterplot smoothers and the Quasi-Likelihood with Unknown Link and Variance Function Estimation (QLUE) algorithm, that is applicable to any regression data with vector predictors and scalar response (Chiou and Müller (1998)). The output of the smoothing steps are smooth

curve and surface estimates. We choose local weighted least squares to implement smoothing, noting that any of a number of alternative smoothing methods could be used. The output of QLUE consists of nonparametric link and variance functions, the parameter estimates for the single index parameter, and their asymptotic covariance matrix. The latter can be used for inference, pertaining to the asymptotic normality of the parameter estimates. Detailed descriptions of smoothing and QLUE algorithms can be found in Chiou, Müller and Wang (2003). We note that the QLUE steps are intended to provide dimension reduction through single index models, and may be replaced by other dimension reduction methods such as SIR (Li (1991)), (generalized) projection pursuit or additive modeling. The numerical implementation described in the following makes use of algorithms of Chiou, Müller and Wang (2003) and Chiou, Müller, Wang and Carey (2003).

We now make the following assumptions regarding the discrete sampling of the response trajectories $Y_i(t)$: there are m_i measurements available for the ith trajectory which are made at the levels $t_{i1}, \ldots, t_{im_i}$. We assume $\min\{m_i, i = 1, \ldots, n\} \to \infty$ as $n \to \infty$ and that the measurement locations $\{t_{ij}\}$ for each i are generated by a smooth design density (Müller (1984)); all these design densities are uniformly smooth and are uniformly bounded away from 0 on the compact domain on which the $Y(t)$ live. Explicit formulas embodying these requirements are easily derived and are useful for asymptotic derivations for the consistency of mean function and eigenfunction estimates, for example along the lines of Cardot, Ferraty and Sarda (1999), Cardot, Ferraty, Mas and Sarda (2003) and Yao, Müller and Wang (2003).

To fit the model, one first estimates the overall mean function $\mu_0(t)$ by aggregating the observations of all subjects into one scatterplot and applying a smoother. For simplicity, we ignore the dependence of repeated measurements coming from the same subject; it was demonstrated in Welsh, Lin and Carroll (2002) and Wang (2003) that adjusting to a given dependence structure can lead to substantial efficiency gains. Whether these gains hold up when the correlations need to be estimated remains an open question.

The next step is the estimation of μ_1 and γ_1: according to (1), a natural estimator for $\theta(\mathbf{Z_i})$ is given by

$$\theta(\mathbf{Z}_i) = \frac{\sum_{j=1}^m \hat{\mu}_0(t_j) Y_i(t_j)}{\sum_{j=1}^m \hat{\mu}_0^2(t_j)}.$$

To infer the single index components γ and μ_1, we then apply QLUE to the scatterplot $(\mathbf{Z_i}, \theta(\mathbf{Z_i}))$. This provides both the link function estimate $\hat{\mu}_1$ and the parameter estimate $\hat{\gamma}$, leading to the functional least squares estimate

$$\hat{\mu}(t, \mathbf{Z}) = \hat{\mu}_0(t)\, \hat{\mu}_1(\hat{\gamma}'\mathbf{Z}). \tag{4}$$

We note that in the case of a one-dimensional predictor, the QLUE step above will be replaced by a simple smoothing step for the scatterplot $(Z_i, \theta(Z_i))$, leading to just $\hat{\mu}_1$ and $\gamma = 1$.

Obtaining $R_i(t) = Y_i(t) - \hat{\mu}(t, \mathbf{Z_i})$, we calculate sample covariances between $R(t), R(s)$, from one or few subjects where $R_i(t), R_i(s)$ are observed simultaneously. These raw covariances are smoothed with a surface smoother and the smooth covariance surface is then discretized in order to obtain eigenvalues/eigenvectors of the resulting covariance matrix. The resulting eigenvectors are smoothed to obtain the eigenfunction estimates $\hat{\psi}_k$ (see Rice and Silverman (1991); Staniswalis and Lee (1998)) for residual processes $R(\cdot)$.

Next, given a choice K for the number of random components, the random effect functions $\eta_k(\cdot)$, $k = 1, \ldots, K$ in (2) are obtained via iterative procedures. Suppose that a current estimate $\hat{E}(R_i(t)|\mathbf{Z} = \mathbf{Z}_i)$ of

$$E(R_i(t)|\mathbf{Z} = \mathbf{Z}_i) = \sum_{k=1}^{K} \alpha_k(\beta_k' \mathbf{z})\psi_k(t)$$

is obtained by substituting current estimates for $\alpha_k(\cdot)$ and β_k. Let

$$\hat{A}_{ik} = \int_T \{R_i(t) - \hat{E}(R_i(t)|\mathbf{Z} = \mathbf{Z}_i)\}\hat{\psi}_k(t)\, dt, \quad k = 1, \ldots, K,$$

implemented by numerical integration. Applying QLUE (or a simple smoother in the one-dimensional case) to the scatterplot $(\mathbf{Z_i}, \hat{A}_{ik})$ then yields updates for parameter estimates $\hat{\beta}_k$ and link function estimates $\hat{\eta}_k(\cdot)$. After fitting the residual processes $R_i(t)$, the fitted response curve at covariate level $\mathbf{Z}$ is

$$\hat{Y}(t|\mathbf{Z}) = \hat{\mu}_0(t)\hat{\mu}_1(\hat{\gamma}'\mathbf{Z}) + \sum_{k=1}^{K} \hat{\alpha}_k(\hat{\beta}_k{}'\mathbf{Z})\hat{\psi}_k(t). \tag{5}$$

Basic consistency properties of the estimates can be derived analogously to Chiou, Müller and Wang (2003) and Chiou, Müller, Wang and Carey (2003). The estimation of mean, covariance and eigenfunctions works also for the case of noisy and irregular data. Numerical approximations of integrals $\hat{A}_{ik}$ however will encounter problems. These can be overcome by employing alternative shrinkage estimators for functional principal component scores (Yao et al. (2003)).

4. Dose-Response Analysis With Functional Responses

The univariate predictor "dose" d replaces here the generic multivariate predictor $\mathbf{Z}$ of the previous section. As the dose levels are experimentally determined, the requirements of Section 3 pertaining to the behavior of random covariates

Z need to be modified for this fixed design case, replacing, for example, constraints on expected values by analogous constraints on discrete sums. Since dose is a univariate predictor, dimension reduction is not needed in the covariate space. The iterative QLUE steps of the algorithm of Section 3 are replaced by one-dimensional regressions in this case. For implementing these regressions we consider both parametric and nonparametric approaches.

Initial nonparametric fits can guide parametric model choices. In our application, this points to quadratic regressions as reasonably close parametric fits, and least squares fits of such models indeed were found to work well. Compared to nonparametric smoothers, parametric fits have the general disadvantage of less flexibility and potential lack-of-fit problems, but on the positive side they succinctly summarize the regression relationship in terms of a few fitted parameters. They do not require choice of a smoothing parameter and allow testing for significance of the regression relationship, if one is willing to ignore potential perturbations stemming from the estimation of mean and eigenfunctions. The ability to test for regression effects is an attractive feature of parametric regressions. For example, one might want to test whether, for a given k, the k-th random component (ψ_k, A_k) should be included in the model, and to assess the overall relevance of dose level for the shape of the response curves.

For nonparametric fitting, the necessary choice of smoothing parameters may be based on selection criteria for individual scatterplots, for example cross-validation or related criteria, leading to a separate bandwidth choice for each scatterplot smoothing step. Or one can aim at minimizing an overall sum of individual prediction error estimates, such as the leave-one-curve-out prediction error of Rice and Silverman (1991),

$$PE = \sum_{i=1}^{n} \sum_{j=1}^{m_i} \{\hat{Y}_i^{(-i)}(t_{ij}) - Y_i(t_{ij})\}^2, \tag{6}$$

to be minimized in dependence on bandwidths and number of components K, assuming n subjects and m_i measurement times t_{ij} per subject. Here $\hat{Y}_i^{(-i)}$ is the prediction for trajectory Y_i, obtained by deleting the data for the i-th subject.

An alternative is to consider prediction of the mean response at each dose level, which is especially attractive for dose-response situations where typically several experimental units are assigned to the same dose, as in our example data. When using (6), one is trying to predict individual response trajectories. This is not very meaningful if individual trajectories display substantial random variation. Therefore we propose to predict the conditional mean response $E(Y(t)|Z = d)$ instead, by leaving out all data recorded at the dose level d, and then to compare this prediction with the observed empirical average response at the corresponding dose level.

Formally, assume the data are $(d_i, Y_i(t))$, $i = 1, \ldots, n$, d_i denoting the dose level administered to the ith subject, and that the dose levels d_i are grouped into L distinct levels $D_1, \ldots, D_L$, $L << n$. Define dose level average trajectories (which may be obtained from pointwise means coupled with interpolation or smoothing with a small bandwidth)

$$\bar{Y}(t|D_\ell) = \sum_{i=1}^{n} Y_i(t) 1_{\{d_i = D_\ell\}} / \sum_{i=1}^{n} 1_{\{d_i = D_\ell\}}, \quad \ell = 1, \ldots, L.$$

This empirical mean response curve at dose D_ℓ is compared with the predicted mean trajectory at the same dose level, based on a current model and independent of the actual observations made at that dose level. We refer to this device as leave-one-dose-out predictor. Noting that the single indices in (5) are simply replaced by dose levels,

$$\hat{Y}^{(-\ell)}(t|D_\ell) = \hat{\mu}_0^{(-\ell)}(t)\hat{\mu}_1^{(-\ell)}(D_\ell) + \sum_{k=1}^{K} \hat{\alpha}_k^{(-\ell)}(D_\ell)\hat{\psi}_k^{(-\ell)}(t),$$

where function estimates $\hat{\mu}_0^{(-\ell)}(\cdot)$, $\hat{\mu}_1^{(-\ell)}(\cdot)$, $\hat{\alpha}_k^{(-\ell)}(\cdot)$ and $\hat{\psi}_k^{(-\ell)}(\cdot)$ are computed from a reduced sample where all subjects i with dose $d_i = D_\ell$ have been omitted. We then obtain the conditional prediction error

$$CPE = \sum_{\ell=1}^{L} \frac{1}{L} \int_T \{\hat{Y}^{(-\ell)}(t|D_\ell) - \bar{Y}(t|D_\ell)\}^2 dt, \tag{7}$$

where the integrals are approximated by suitable sums. We note that computing leave-one-dose-out conditional prediction error (7) is considerably faster than computing leave-one-curve-out prediction error (6).

5. Analysis of Egg-Laying Profiles in Response to Dietary Dose

The data resulted from a dose-response experiment with functional responses, the trajectories of daily egg-laying for female medflies, recorded for ten different dietary dose levels. Egg-laying is a measure of reproductive success and its overall pattern reflects both physiological constraints and evolutionary strategy to cope with varying environments. Biological background and details of the experiment are described in Carey et al. (2002).

We include records of egg-laying from 0 to 40 days of age for flies that survive 10 days and produce eggs. Summary data are provided in Table 1. Median lifetime does not increase beyond dose levels of 0.70 and is short for the lowest dose group at level 0.50, with similar findings for the median number of eggs.

Table 1. Median lifetime and number of eggs at each dose level for medfly reproduction data.

Dose level	No. flies	Median lifetime	Median no. eggs
0.50	61	33.0	383.0
0.60	89	48.0	836.0
0.65	84	47.5	849.0
0.70	93	55.0	1116.0
0.75	88	56.0	1181.5
0.80	91	56.0	1187.0
0.85	89	49.0	1162.0
0.90	95	51.0	1197.0
0.95	87	53.0	1367.0
1.00	97	49.0	1264.0
Total	874	51.0	1120.0

Egg-production requires protein and a reduction in diet is consequently associated with fewer total eggs, as demonstrated in Table 1 and Carey et al. (2002). Underlying such a reduction in overall reproductive output can be a variety of physiological responses to food shortage: a delay in starting or raising the level of egg production, lowered egg-production proportionally at all ages, an earlier end to egg production, or other shape changes in egg-laying trajectories.

The left panel of Figure 2 demonstrates the estimated unconditional overall mean $\hat{\mu}_0(t)$ for $\mu_0(t)$, which depicts the basic shape of an egg-laying curve in

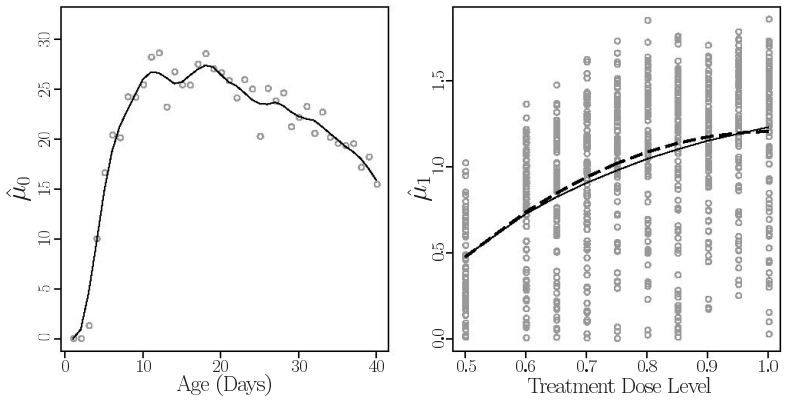

Figure 2. Function estimates of the mean function and multiplicative components, with overall mean function $\hat{\mu}_0(t)$ (left) and multiplicative effect function $\hat{\mu}_1(d)$ (right), for mean component $\hat{\mu}(t, d)$, with dietary dose level d as covariate. The dashed curve on the right panel is fitted by least squares quadratic regression and the solid curve by nonparametric regression with cross-validated bandwidth.

dependence of age. The estimated functional multiplicative effect function $\hat{\mu}_1(d)$ for $\mu_1(d)$, that reflects the effect of dose on mean response, is displayed in the right panel of Figure 2. This is a monotone increasing function, for which a quadratic regression appears to provide an adequate fit. The least squares regression coefficients for the quadratic fit are listed in Table 2 and are significant, if we are willing to assume that usual inference applies. Cross-sections at the dose levels for the functional multiplicative effect mean surface $\hat{\mu}(t,d) = \hat{\mu}_0(t)\hat{\mu}_1(d)$ are shown in Figure 6, left panel.

To describe the results for various models we highlight two special cases. One case that was investigated previously in Chiou, Müller and Wang (2003) includes the residual process $R(\cdot)$ but does not include a covariate effect on the conditional mean, stipulating $\mu_1(\cdot) \equiv 1$, and is referred to as Model I in the following. The model with general multiplicative mean function μ_1 and residual process $R(\cdot)$ is referred to as Model II.

Table 2. Estimated coefficients of the linear quadratic regression for $\mu_1(d)$.

Variable	Coeff	S.E.	p-value
Intercept	-1.6548	0.3587	0.0000
d	5.6685	0.9549	0.0000
d^2	-2.8079	0.6193	0.0000

The first three estimated eigenfunctions $\hat{\psi}_k(t)$ for Model I, corresponding to the first three largest eigenvalues, are shown in Figure 3, and they are very close to those obtained for Model II (not shown). The eigenfunctions essentially provide a decomposition of the peak in the egg-laying trajectory that occurs soon after egg production starts, in terms of its timing and shape. The link function estimates $\hat{\alpha}_k$, associated with the first three eigenfunctions, are shown in Figure 4 for Model I and in Figure 5 for Model II, obtained with both least squares quadratic regression fits (dashed) and local linear regression fits (solid). Quadratic regression is seen to provide a reasonably good approximation.

For Model I, the third link function appears almost flat, indicating that there is not much to be gained by including more than two random components. This is corroborated by the approximate inference provided in Table 3, where the estimated regression coefficients for the three quadratic regressions for these link functions are summarized. The regression coefficients for the first two components ($k = 1$ and $k = 2$) are significant, while those for the third component are not. This provides support for the choice of $K = 2$ random components, and indicates that the residual process $R(t)$ provides a necessary addition to this model.

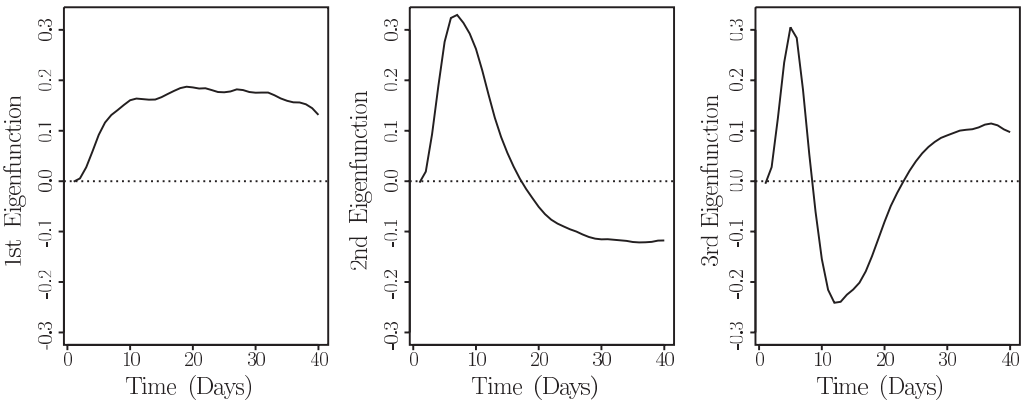

Figure 3. The first three estimated eigenfunctions $\{\hat{\psi}_k\}_{k=1,2,3}$ for Model I for the medfly dose-response data. The first eigenfunction explains 35.31%, the second additional 16.84%, and the third additional 8.82% of total variation. Bandwidths are selected by cross-validation.

For Model II, the approximate inference provided for the quadratic regression fits for the link functions in Table 4 indicates that only the second random component contributes significantly. The insignificance of the first random component link function is likely a consequence of the fact that the shape of the first estimated eigenfunction $\hat{\psi}_1$ resembles that of the estimated overall unconditional mean function $\hat{\mu}_0$. The contribution of the first random component then diminishes as a similar effect is already reflected in the multiplicative mean structure. Also for Model II, the residual process $R(t)$ seems to be a necessary additional component.

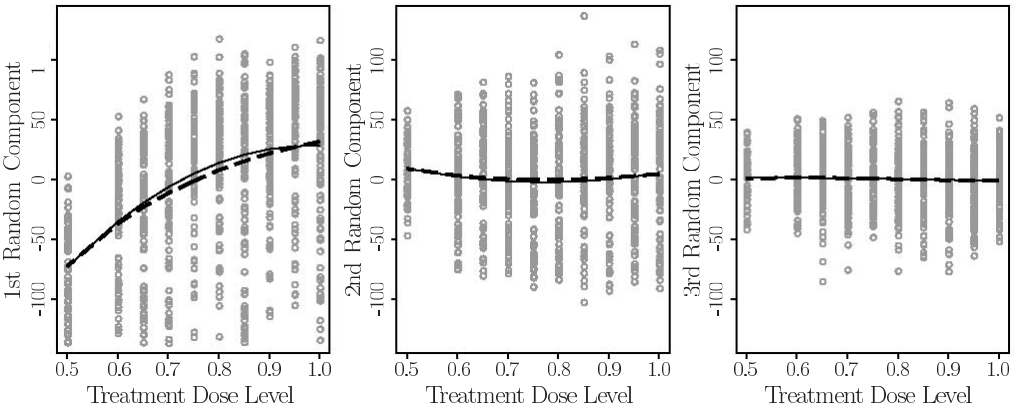

Figure 4. Estimated smooth link functions $\hat{\alpha}_k$ for Model I, $k = 1, 2, 3$, for the random components as functions of dose level, based on a sample of 874 medflies. The dashed curves are fits obtained from least squares quadratic regression, and the solid curves are nonparametric regression fits, using cross-validation bandwidths.

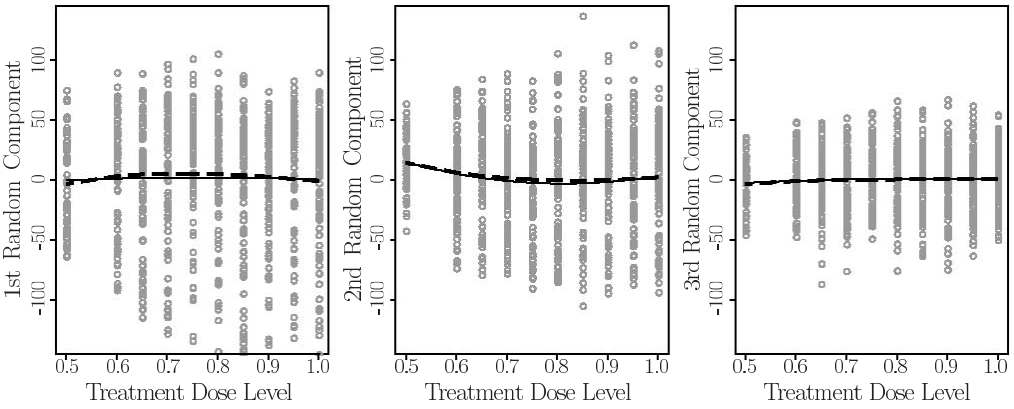

Figure 5. Same as Figure 4, but for Model II.

Table 3. Estimated least squares regression coefficients for Model I, fitting quadratic regressions for link functions $\alpha_k(d)$, $k = 1, 2, 3$.

	$k=1$			$k=2$			$k=3$		
Variable	Coeff	S.E.	p-value	Coeff	S.E.	p-value	Coeff	S.E.	p-value
Intercept	-387.52	49.38	0.000	86.99	37.87	0.022	4.23	23.78	0.859
d	843.08	131.48	0.000	-230.19	100.83	0.023	-3.32	63.32	0.958
d^2	-427.05	85.27	0.000	148.52	65.39	0.023	-1.93	41.06	0.963

Table 4. Same as Table 3, but for Model II.

	$k=1$			$k=2$			$k=3$		
Variable	Coeff	S.E.	p-value	Coeff	S.E.	p-value	Coeff	S.E.	p-value
Intercept	-15.60	49.28	0.752	115.36	37.91	0.002	-18.05	23.81	0.449
d	45.87	131.21	0.727	-291.29	100.92	0.004	44.09	63.40	0.487
d^2	-29.88	85.09	0.726	179.11	65.45	0.006	-25.31	41.11	0.538

Cross-sections at each of the experimental dose levels through the overall fitted surface $\hat{Y}(t, d)$, obtained by fitting Model II, including the multiplicative mean component and the first two random components of the residual process $R(t)$ with associated link functions estimated by quadratic regression, are in the right panel of Figure 6. We note that differences with the fitted multiplicative mean surface without residual process on the left panel of this figure are noticeable in the cross-sections for smaller doses. The question arises which model ultimately is preferable. As dose levels increase, the mean component of Model II is increasing, with a saturation characteristic, as seen from Figure 2.

Additional shape changes are represented by the second random component of the residual process, as determined by the second eigenfunction ψ_2 in Figure 3, middle panel, and by the second random components link function in Figure 4,

middle panel. An increase in this component signals an earlier and sharper peak in the egg-laying trajectory. The model predicts that such peaks are more likely to be found for small doses and for large doses, but less so for the middle doses. This is reflected in the earlier and sharper rise to the peak observed for small doses in the fitted model as compared to the fitted multiplicative mean surface (comparing the left and the right panels of Figure 6) without residual process. In contrast, Model II with residual process reflects shifts in the peak location, and therefore this more complex model adapts in a flexible way to functional response data.

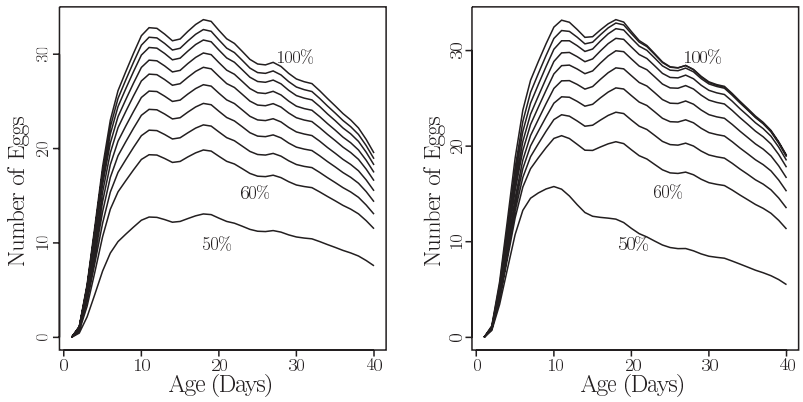

Figure 6. Cross-sections at each of the experiment's dose levels through the fitted multiplicative mean surface $\hat{\mu}(t,d)$, with the multiplicative effect function $\mu_1(d)$ fitted by nonparametric regression (left panel, corresponding to Model II with $K=0$), and of the fitted surface $\hat{Y}(t,d)$, with additional random components link functions $\hat{\eta}_k(d)$ fitted by linear quadratic regressions, corresponding to Model II with $K=2$ (right panel).

6. Model Comparisons via Conditional Prediction Error

In order to compare model choices for these data, we investigate the predictive capabilities of various model fits. The criterion that we adopt is the conditional prediction error CPE at (7), based on the leave-one-dose-out technique, comparing the model predictions with the mean over all response curves observed at the experiment's dose levels.

Table 5 provides comparisons between the more flexible nonparametric and the more interpretable quadratic fits for the link functions, for both Model I and Model II and with/without residual process components. We find that, with regard to the multiplicative function μ_1, the quadratic fit is clearly preferred for $K=0$, while it is sometimes slightly better and sometimes slightly worse for $K>1$; for the link functions α_k there is no clear preference. Given the advantages

of parametric fitting, such fits seem justified in view of these results. Model I, without the covariate effect in the mean function, clearly requires inclusion of the residual process, and adding $K = 2$ random components leads to the best fit among all such models. The differences with Model II, also including two random components from the residual process, at $K = 2$, are negligible, except for the nonparametric fitting of the α_k where Model I is worse.

Table 5. Conditional prediction errors with nonparametric (*Nonp*) and quadratic least squares (*Para*) fits for functions $\mu_1(\cdot)$ and $\alpha(\cdot)$ and with the number of random components (eigenfunctions) for the residual process varying between 0 and 3. Model I does not include a multiplicative mean function μ_1, while Model II does.

Model	$\mu_1(\cdot)$	$\alpha_k(\cdot)$	$K = 0$	$K = 1$	$K = 2$	$K = 3$
I	-	*Para*	1324.62	132.46	118.68	125.74
	-	*Nonp*	1324.62	150.21	139.68	144.23
II	*Nonp*	*Nonp*	161.86	133.92	115.18	119.26
	Nonp	*Para*	161.86	131.25	108.59	115.78
	Para	*Nonp*	136.60	135.06	114.90	118.77
	Para	*Para*	136.60	144.34	119.12	125.72

For $K = 0$, Model I does not include a residual process, just fits an overall mean function μ_0 irrespective of the dose level, and thus serves as a null model that does not include any dose effect on the response function. It is clear from the results that this model is not appropriate. Model II without residual process (i.e., $K = 0$), works considerably better. In terms of the CPE criterion, this multiplicative mean surface model is not as good as extended models that include random components from the residual process, but the differences are small for the parametric link functions for μ_1. We find that for $K = 3$, the CPE values start to increase, confirming that inclusion of $K = 2$ random components is the best choice, if a residual process is added. The overall best results are achieved for $K = 2$ and Model II.

Acknowledgements

We thank James Carey for sharing the medfly dose-response data. Special thanks are due to John Rice for stimulating discussions and for comments on an earlier version which led to major improvements and advances. We are also grateful to an associate editor and two referees for their helpful remarks, and to the consulting editor for stylistic improvements. This research was supported in part by National Health Research Institutes Grant (Taiwan), National Science Foundation Grants DMS-98-03267, DMS-00-79430, DMS-02-04869 and National Institutes of Health Grant 99-SC-NIH-1028.

References

Capra, W. B. and Müller, H. G. (1997). An accelerated time model for response curves. *J. Amer. Statist. Assoc.* **92**, 72-83.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statist. Probab. Lett.* **45**, 11-22.

Cardot, H., Ferraty, F., Mas, A. and Sarda, P. (2003). Testing hypotheses in the functional linear model. *Scand. J. Statist.* **30**, 241-255.

Carey, J. R., Liedo, P., Harshman, L., Zhang, Y., Müller, H. G., Partridge, L. and Wang, J. L. (2002). Life history response of Mediterranean fruit flies to dietary restriction. *Aging Cell* **1**, 140-148.

Castro, P. E., Lawton, W. H. and Sylvestre, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* **28**, 329-337.

Chiou, J. M. and Müller, H. G. (1998). Quasi-likelihood regression with unknown link and variance functions. *J. Amer. Statist. Assoc.* **92**, 72-83.

Chiou, J. M., Müller, H. G. and Wang, J. L. (2003). Functional quasi-likelihood regression models with smooth random effects. *J. Royal Statist. Soc. Ser. B* **65**, 405-423.

Chiou, J. M., Müller, H. G., Wang, J. L. and Carey, J. R. (2003). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statist. Sinica* **13**, 1119-1133.

Cuevas, A., Febrero, M. and Fraiman, R. (2002). Linear functional regression: the case of fixed design and functional response. *Canad. J. Statist.* **30**, 285-300.

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. Multivariate Anal.* **85**, 54-77.

Fan, J. and Zhang, J. T. (1998). Functional linear models for longitudinal data. *J. Roy. Statist. Soc. Ser. B* **39**, 254-261.

Faraway, J. J. (1997). Regression analysis for a functional response. *Technometrics* **39**, 254-261.

Gasser, T., Hall, P. and Presnell, B. (1998). Nonparametric estimation of the mode of a distribution of random curves. *J. Roy. Statist. Soc. Ser. B* **60**, 681-691.

Hall, P. and Heckman, N. E. (2002) Estimating and depicting the structure of a distribution of random functions. *Biometrika* **89**, 145-158.

He, G. (1999). Functional canonical analysis and linear modelling. Thesis, University of California, Davis.

He, G., Müller, H. G. and Wang, J. L. (2000). Extending correlation and regression from multivariate to functional data. In *Asymptotics in Statistics and Probability* (Edited by M. Puri), 197-210. VSP International Science Publishers, The Netherlands.

He, G., Müller, H. G. and Wang, J. L. (2003). Functional canonical analysis for square integrable stochastic processes. *J. Multivariate Anal.* **85**, 54-77.

James, G. M. (2002). Generalized linear models with functional predictors. *J. Roy. Statist. Soc. Ser. B* **64**, 411-432.

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

Li, K.-C. (1991). Sliced inverse regression for dimension reduction. *J. Amer. Statist. Assoc.* **86**, 316-327.

Liu, X. and Müller, H. G. (2003). Modes and clustering for time-warped gene expression profile data. *Bioinformatics* **19**, 1937-1944.

Müller, H. G. (1984). Optimal designs for nonparametric kernel regression. *Statist. Probab. Lett.* **2**, 285-290.

Müller, H. G. and Stadtmüller, U. (2004). Generalized functional linear models. (Preprint available at http://anson.ucdavis.edu/∼mueller/preprints.html). *Ann. Statist.* To appear.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *J. Roy. Statist. Soc. Ser. B* **53**, 539-572.

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

Rao, C. R. (1958). Some statistical models for comparison of growth curves. *Biometrics* **14**, 1-17.

Ratcliffe, S. J., Heller, G. Z. and Leader, L. R. (2002). Functional data analysis with application to periodically stimulated foetal heart rate data. II: functional logistic regression. *Statist. Medicine* **21**, 1115-1127.

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.

Shi, M. G., Weiss, R. E. and Taylor, J. M. G. (1996). An analysis of paediatric CD4 counts for acquired immune deficiency syndrome using flexible random curves. *Appl. Statist.* **45**, 151-163.

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

Wang, N. (2003). Marginal nonparametric kernel regression accounting for within-subject correlation. *Biometrika* **90**, 43-52.

Welsh, A., Lin, X. and Carroll, R. J. (2002). Marginal longitudinal nonparametric regression: locality and efficiency of spline and kernel estimates. *J. Amer. Statist. Assoc.* **97**, 482-493.

Yao, F., Müller, H. G., Clifford, A. J., Dueker, S. R., Follett, J., Lin, Y., Buchholz, B. 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.

Yao, F., Müller, H. G. and Wang, J. L. (2003). Functional data analysis for sparse longitudinal data. Preprint.

Division of Biostatistics and Bioinformatics, National Health Research Institutes, Taipei 115, Taiwan, R.O.C.

E-mail: jmchiou@nhri.org.tw

Department of Statistics, University of California, Davis, CA 95616, U.S.A.

E-mail: mueller@wald.ucdavis.edu

Department of Statistics, University of California, Davis, CA 95616, U.S.A.

E-mail: wang@wald.ucdavis.edu

(Received May 2003; accepted December 2003)