Biometrika (2010), , 4, *pp*. 791–805
doi: 10.1093/biomet/asq056
Advance Access publication 15 November 2010

ƽ

BY HANS-GEORG MÜLLER

Department of Statistics, University of California, Davis, One Shields Avenue, Davis, California 95616, U.S.A.
mueller@wald.ucdavis.edu

AND FANG YAO

Department of Statistics, University of Toronto, 100 Saint George Street, Toronto, Ontario M5S 3G3, Canada
fyao@utstat.toronto.edu

SUMMARY

We consider the problem of estimating functional derivatives and gradients in the framework of a regression setting where one observes functional predictors and scalar responses. Derivatives are then defined as functional directional derivatives that indicate how changes in the predictor function in a specified functional direction are associated with corresponding changes in the scalar response. For a model-free approach, navigating the curse of dimensionality requires the imposition of suitable structural constraints. Accordingly, we develop functional derivative estimation within an additive regression framework. Here, the additive components of functional derivatives correspond to derivatives of nonparametric one-dimensional regression functions with the functional principal components of predictor processes as arguments. This approach requires nothing more than estimating derivatives of one-dimensional nonparametric regressions, and thus is computationally very straightforward to implement, while it also provides substantial flexibility, fast computation and consistent estimation. We illustrate the consistent estimation and interpretation of the resulting functional derivatives and functional gradient fields in a study of the dependence of lifetime fertility of flies on early life reproductive trajectories.

Some key words: Derivative; Functional data analysis; Functional regression; Gradient field; Nonparametric differentiation; Principal component.

1. INTRODUCTION

Regression problems where the predictor is a smooth square integrable random function $X(t)$ defined on a domain $\mathcal{T}$ and the response is a scalar Y with mean $E(Y) = \mu_Y$ are found in many areas of science. For example, in the biological sciences, one may encounter predictors in the form of subject-specific longitudinal time-dynamic processes such as reproductive activity. For each such process, one observes a series of measurements and it is then of interest to model the dependence of the response on the predictor process (Cuevas et al., 2002; Rice, 2004; Ramsay & Silverman, 2005). Examples include studies of the dependence of remaining lifetime on fertility processes (Müller & Zhang, 2005), and a related analysis that we discuss in further detail in § 5 below. This concerns the dependence of total fertility on the dynamics of the early fertility process in a study of biodemographic characteristics of female medflies. Here, we observe trajectories of fertility over the first 20 days of life, measured by daily egg-laying for a sample

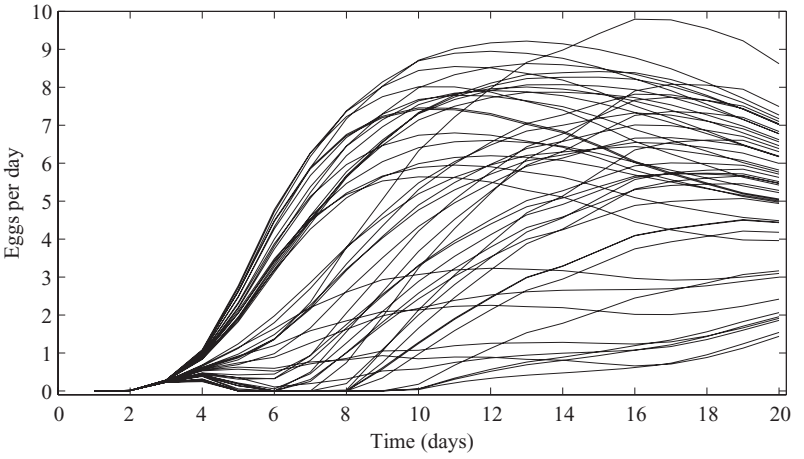


Fig. 1. Egg-laying trajectories (eggs per day) for 50 randomly selected flies, for the first 20 days of their lifespan.

of medflies, as illustrated for 50 randomly selected flies in Fig. 1, showing the realizations of the predictor process. The total number of eggs laid over the lifetime of a fly is also recorded and serves as scalar response.

In this and other functional regression settings, one would like to determine which predictor trajectories will lead to extreme responses, for example by identifying zeros of the functional gradient field, or to characterize the functional directions in which responses will increase or decrease the most, when taking a specific trajectory as a starting point. In some applications, these directions may give rise to specific interpretations, such as evolutionary gradients (Kirkpatrick & Heckman, 1989; Izem & Kingsolver, 2005). The advantage of functional over multivariate analysis for biological data in the form of trajectories was recently demonstrated in Griswold et al. (2008). The need to analyze the effects of changes in trajectories in the field of biological evolution and ecology and to address related questions in other fields motivates the development of statistical technology to obtain a functional gradient at a function-valued argument, e.g. a particular predictor function. It is thus of interest to develop efficient and consistent methods for the estimation of functional gradients.

In the framework of functional predictors and scalar responses, derivatives are defined as functional directional derivatives that indicate how changes in the predictor function in a specified functional direction are associated with corresponding changes in the scalar response. Similarly to the classical regression setting of a scalar predictor and scalar response, this problem can be easily solved for a functional linear regression relationship where the derivative corresponds to the slope parameter, respectively, regression parameter function, as we demonstrate below. The problem is harder and more interesting in the nonlinear situation where the classical analogue would be the estimation of derivatives of a nonparametric regression function (Gasser & Müller, 1984; Zhou & Wolfe, 2000).

When tackling this functional differentiation problem, one realizes that the space in which the predictor functions reside is infinite-dimensional and therefore sparsely populated, so that estimation techniques will be subject to a rather extreme form of the curse of dimensionality. This problem arises for a functional regression even before considering derivatives. The conventional approach to reducing the very high dimensionality of the functional regression problem is through the well-established functional linear model, which implies strong dimension reduction through structural assumptions. The structural constraints inherent in these models often prove to be too

restrictive, just as is the case for ordinary linear regression. A reasonably restrictive yet overall sufficiently flexible approach to dimension reduction in regression models with many predictors is additive modelling (Stone, 1985; Hastie & Tibshirani, 1986) and its extension to functional regression (Müller & Yao, 2008).

Postulating a regression relation that is additive in the functional principal components of predictor processes, but otherwise unspecified, provides a particularly useful set of constraints for the estimation of functional derivatives. The resulting functional derivative estimates are straightforward to implement, even for higher order derivatives, and require nothing more than obtaining a sequence of nonparametric estimators for the derivatives of one-dimensional smooth functions, with the principal components of the predictor processes as respective arguments. This approach is easily extended to the estimation of functional gradient fields and to the case of higher derivatives and is supported by consistency properties. Functional gradient fields emerge as a useful tool to aid in the interpretation of functional regression data.

2. Additive modelling of functional derivatives

2·1. *Preliminary considerations*

To motivate our procedures, first consider the case of a more conventional functional linear model. Assume the predictor process X has mean function $E\{X(t)\} = \mu_X(t)$ and covariance function $\mathrm{cov}\{X(t_1), X(t_2)\} = G(t_1, t_2)$, and the response is a scalar Y with mean $E(Y) = \mu_Y$. We denote centred predictor processes by $X^c(t) = X(t) - \mu_X(t)$. In the functional linear model, a scalar response Y is related to the functional predictor X via (Ramsay & Dalzell, 1991)

$$\Gamma_L(x) = E(Y \mid X = x) = \mu_Y + \int_{\mathcal{T}} \beta(t) x^c(t)\, ds. \tag{1}$$

Here, Γ_L is a linear operator on the Hilbert space $L^2(\mathcal{T})$, mapping square integrable functions defined on the finite interval $\mathcal{T}$ to the real line, and β is the regression parameter function, assumed to be smooth and square integrable. Recent work on this model includes Cai & Hall (2006), Cardot et al. (2007) and Li & Hsing (2007). Since the functional linear model at (1

Karhunen–Loève expansion

$$X(t) = \mu_X(t) + \sum_k \xi_{Xk}\phi_k(t), \qquad \xi_{Xk} = \int X^c(t)\phi_k(t)dt. \tag{2}$$

The random variables ξ_{Xk} are the functional principal components, also referred to as scores. These scores are uncorrelated and satisfy $E(\xi_{Xk}) = 0$ and $\mathrm{var}(\xi_{Xk}) = \lambda_k$ (Ash & Gardner, 1975).

The derivative of a Gâteaux differentiable operator Γ, mapping square integrable functions to real numbers, evaluated at $x = \sum_k \xi_{xk}\phi_k$, is an operator $\Gamma_x^{(1)}$ that depends on x and has the property that, for functions u and scalars δ,

$$\Gamma(x + \delta u) = \Gamma(x) + \delta\,\Gamma_x^{(1)}(u) + o(\delta) \tag{3}$$

as $\delta \to$

should adequately represent predictors X, and the component scores that correspond to the coordinate values that represent X are independent for Gaussian processes, which is particularly beneficial in the additive framework. The functional additive regression framework, where the response depends on predictor processes through smooth functions of the predictor functional principal components, embodies sensible structural constraints and dimension reduction and provides a structural compromise that is well suited for the estimation of functional gradients.

The functional additive framework, described in Müller & Yao (2008), revolves around an additive functional operator Ω,

$$\Omega(x) = E(Y \mid X = x) = \mu_Y + \sum_{k=1}^{\infty} f_k(\xi_{xk}),$$

subject to $E\{f_k(\xi_{Xk})\} = 0$ $(k = 1, \ldots)$, for the scores ξ_{Xk} as defined in (2). Applying (3) and (4) to Ω, for functions $x = \sum_k \xi_{xk}\phi_k$ and $u = \sum_k \xi_{uk}\phi_k$, we have

$$\Omega(x + \delta u) = \mu_Y + \sum_k f_k(\xi_{xk} + \delta\xi_{uk}) = \Omega(x) + \delta \sum_k f_k^{(1)}(\xi_{xk})\xi_{uk} + o(\delta), \tag{5}$$

which leads to

$$\Omega_x^{(1)}(u) = \sum_{k=1}^{\infty} f_k^{(1)}(\xi_{xk})\xi_{uk} = \sum_{k=1}^{\infty} \omega_{xk}\Phi_k(u), \quad \Omega_x^{(1)} = \sum_k f_k^{(1)}(\xi_{xk})\Phi_k \tag{6}$$

and $\omega_{xk} = f_k^{(1)}(\xi_{xk})$ for the functional additive model.

It is of interest to extend functional derivatives also to higher orders. This is done by iterating the process of taking derivatives in (3). Generally, the form of the pth derivative operator is rather unwieldy, as it depends not only on x, but also on $p - 1$ directions $u_1, \ldots, u_{p-1}$, which are used to define the lower order derivatives, and its general form will be $\Gamma^{(p)}_{x;u_1,\ldots,u_{p-1}} = \sum_k \gamma_{x;u_1,\ldots,u_{p-1};k}\Phi_k$. The situation however is much simpler for the additive operators Ω, where

$$\Omega(x + \delta u) = \mu_Y + \sum_k f_k(\xi_{xk} + \delta\xi_{uk}) = \Omega(x) + \sum_{j=1}^{p} \frac{1}{j!}\delta^j \left\{ \sum_k f_k^{(j)}(\xi_{xk})\xi_{uk}^j \right\} + o(\delta^p).$$

The separation of variables that is an inherent feature of the additive model implies that one does not need to deal with the unwieldy cross-terms that combine different $u_j s$, limiting the usefulness of functional derivatives of higher order in the general case. The straightforwardness of extending functional derivatives to higher orders is a unique feature of the additive approach, as pth order derivative operators

$$\Omega_x^{(p)}(u) = \sum_{k=1}^{\infty} f_k^{(p)}(\xi_{xk})\{\Phi_k(u)\}^p \tag{7}$$

can be easily obtained by estimating pth derivatives of the one-dimensional nonparametric functions f_k. As in ordinary multivariate calculus, higher order derivatives can be used to characterize extrema or domains with convex or concave functional regression relationships and also for diagnostics and visualization of nonlinear functional regression relations. As it enables such estimates, while retaining full flexibility in regard to the shape of the derivatives, the framework of additive models is particularly attractive for functional derivative estimation.

3. ESTIMATION AND ASYMPTOTICS

In order to obtain additive functional derivatives (6), we require estimates of the defining coefficients ω_{xk}, for which we use $\hat{\omega}_{xk} = \hat{f}_k^{(1)}(\xi_{xk})$. Thus, the task is to obtain consistent estimates of the derivatives $f_k^{(1)}$ for all $k \geqslant 1$. The data recorded for the ith subject or unit are typically of the form $\{(t_{ij}, U_{ij}, Y_i), i = 1, \ldots, n,\ j = 1, \ldots, n_i\}$, where predictor trajectories X_i are observed at times $t_{ij} \in \mathcal{T}$, yielding noisy measurements

$$U_{ij} = X_i(t_{ij}) + \epsilon_{ij} = \mu_X(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t_{ij}) + \epsilon_{ij}, \tag{8}$$

upon inserting representation (2), where ϵ_{ij} are independent and identically distributed measurement errors, independent of all other random variables, and the observed responses Y_i are related to the predictors according to $E(Y \mid X) = \mu_Y + \sum_k f_k(\xi_{Xk})$. A difficulty is that the ξ_{ik} are not directly observed and must be estimated. For this estimation step, one option is to use the principal analysis by conditional expectation procedure (Yao et al., 2005) to obtain estimates $\hat{\xi}_{Xi}$ in a preliminary step. Briefly, the key steps are the nonparametric estimation of the mean trajectory $\mu_X(t)$ and of the covariance surface $G(t_1, t_2)$ of predictor processes X, obtained by smoothing pooled scatter-plots. For the latter, one omits the diagonal elements of the empirical covariances, as these are contaminated by the measurement errors. From estimated mean and covariance functions, one then obtains eigenfunction and eigenvalue estimates (Rice & Silverman, 1991; Staniswalis & Lee, 1998; Boente & Fraiman, 2000).

We implement all necessary smoothing steps with local linear smoothing, using automatic data-based bandwidth choices. Additional regularization is achieved by truncating representations (2) and (8) at a suitable number of included components K, typically chosen data-adaptively by pseudo-BIC or similar selectors, or simply as the smallest number of components that explain a large enough fraction of the overall variance of predictor processes. We adopt the latter approach in our applications, requiring that 90% of the variation is explained. Given the observations made for the ith trajectory, best linear prediction leads to estimates of the functional principal components ξ_{ik}, by estimating $E(\xi_{ik} \mid U_i) = \lambda_k \phi_{ik}^{\mathrm{T}} \Sigma_{U_i}^{-1}(U_i - \mu_{X_i})$, where $U_i = (U_{i1}, \ldots, U_{in_i})^{\mathrm{T}}$, $\mu_{X_i} = \{\mu_X(t_{i1}), \ldots, \mu_X(t_{in_i})\}^{\mathrm{T}}$, $\phi_{ik} = \{\phi_k(t_{i1}), \ldots, \phi_k(t_{in_i})\}^{\mathrm{T}}$, and the (j, l) entry of the $n_i \times n_i$ matrix Σ_{U_i} is $(\Sigma_{U_i})_{j,l} = G_X(t_{ij}, t_{il}) + \sigma_X^2 \delta_{jl}$, with $\delta_{jl} = 1$, if $j = l$, and $\delta_{jl} = 0$, if $j \neq l$. One then arrives at the desired estimates $\hat{\xi}_{ik}$ by replacing the unknown components λ_k, ϕ_k, μ_X, G_X and σ^2 by their estimates. For densely observed data, a simpler approach is to insert the above estimates into (2), $\hat{\xi}_{ik} = \int \{\hat{X}_i(t) - \hat{\mu}(t)\}\hat{\phi}_k(t)dt$. These integral estimators require smoothed trajectories $\hat{X}_i$ and therefore dense measurements per sampled curve.

Once the estimates $\hat{\xi}_{ik}$ are in hand, we aim to obtain derivative estimates $\hat{f}_k^{(\nu)}$, the νth order derivatives of the component functions f_k, $k \geqslant 1$, with default value $\nu = 1$. Fitting a local polynomial of degree $p \geqslant \nu$ to the data $\{\hat{\xi}_{ik}, Y_i - \bar{Y}\}_{i=1,\ldots,n}$, obtaining a weighted local least squares fit for this local polynomial by minimizing

$$\sum_{i=1}^{n} \kappa\left(\frac{\hat{\xi}_{ik} - z}{h_k}\right)\left\{Y_i - \bar{Y} - \sum_{\ell=0}^{p} \beta_\ell (z - \hat{\xi}_{ik})^\ell\right\}^2 \tag{9}$$

with respect to $\beta = (\beta_0, \ldots, \beta_p)^{\mathrm{T}}$ for all z in the domain of interest, leads to suitable derivative estimates $\hat{f}_k^{(\nu)}(z) = \nu!\hat{\beta}_\nu(z)$. Here, κ is the kernel and h_k the bandwidth used for this smoothing step. Following Fan & Gijbels (1996), we choose $p = \nu + 1$ for practical implementation.

The following result provides asymptotic properties for this procedure and also consistency of the resulting estimator for the functional derivative operator (6), i.e.,

$$\hat{\Omega}_x^{(1)}(u) = \sum_{k=1}^{K} \hat{f}_k^{(1)}(\xi_{xk})\xi_{uk}, \tag{10}$$

when $K = K(n) \to \infty$ components are included in the estimate and the predictor scores are independent. Gaussianity of predictor processes is not needed.

THEOREM 1. *Under Assumptions* A1–A4 *in the Appendix, for all* $k \geqslant 1$ *for which* λ_j, $j \leqslant k$ *are eigenvalues of multiplicity* 1, *letting* $\tau_j(\kappa^\ell) = \int u^j \kappa^\ell(u)du$, *as* $n \to \infty$, *it holds that*

$$(nh_k^3)^{1/2} \left\{ \hat{f}_k^{(1)}(z) - f_k^{(1)}(z) - \frac{\tau_4(\kappa) f_k^{(3)}(z) h_k^2}{6\tau_2(\kappa)} \right\} \xrightarrow{D} N\left\{ 0, \frac{\tau_2(\kappa^2)\mathrm{var}(Y \mid \xi_{Xk} = z)}{\tau_2^2(\kappa) p_k(z)} \right\} \tag{11}$$

for estimates (9), *where* p_k *is the density of* ξ_{Xk}. *Under the additional Assumption* A5

$$\sup_{\|u\|=1} \left| \hat{\Omega}_x^{(1)}(u) - \Omega_x^{(1)}(u) \right| \longrightarrow 0, \tag{12}$$

in probability for estimates (10), *at any* $x \in L^2(\mathcal{T})$, *as* $n \to \infty$.

For further details about the rate of convergence of (12), we refer to (A1) in the Appendix. For higher order functional derivatives, obtained by replacing estimates of first order derivatives $f^{(1)}$ by estimates of higher order derivatives $f^{(p)}$ in (7), one can prove similar consistency results.

4. SIMULATION STUDIES

To demonstrate the use of the proposed additive modelling of functional gradients, we conducted simulation studies for Gaussian and non-Gaussian predictor processes with different underlying models and data designs. In particular, we compared our proposal with functional quadratic differentiation, suggested by a referee, where one obtains derivatives by approximating the regression relationship with a quadratic operator,

$$\Gamma_Q(x) = E(Y \mid X = x) = \mu_Y + \int_{\mathcal{T}} \alpha(t)x(t)dt + \int_{\mathcal{T}} \beta(t)x^2(t)dt. \tag{13}$$

While this model can be implemented with expansions in B-splines or other bases, for the reasons outlined above, we select the orthogonal functional coordinates that are defined by the eigenfunctions of X. Inserting $\alpha(t) = \sum_k \alpha_k \phi_k(t)$, $\beta(t) = \sum_k \beta_k \phi_k(t)$, the functional derivative operator for (13) is seen to be $\Gamma_{Q,x}^{(1)} = \sum_k (\alpha_k + 2\beta_k \xi_{xk})\Phi_k$.

Each of 400 simulation runs consisted of a sample of $n = 100$ predictor trajectories X_i, with mean function $\mu_X(s) = s + \sin(s)$ $(0 \leqslant s \leqslant 10)$, and a covariance function derived from two eigenfunctions, $\phi_1(s) = -\cos(\pi s/10)/\sqrt{5}$, and $\phi_2(s) = \sin(\pi s/10)/\sqrt{5}$ $(0 \leqslant s \leqslant 10)$. The corresponding eigenvalues were chosen as $\lambda_1 = 4$, $\lambda_2 = 1$, $\lambda_k = 0$, $k \geqslant 3$, and the measurement errors in (8) as $\varepsilon_{ij} \sim N(0, 0{\cdot}4^2)$ independent. To study the effect of Gaussianity of the predictor process, we considered two settings: (i) $\xi_{ik} \sim N(0, \lambda_k)$, Gaussian; (ii) ξ_{ik} are generated from the mixture of two normals, $N\{(\lambda_k/2)^{1/2}, \lambda_k/2\}$ with probability $1/2$ and $N\{-(\lambda_k/2)^{1/2}, \lambda_k/2\}$ with probability $1/2$, a mixture distribution. Each predictor trajectory was sampled at locations uniformly distributed over the domain $[0, 10]$, where the number of noisy measurements was chosen separately and randomly for each predictor trajectory. We considered both dense and sparse design cases. For the dense design case, the number of measurements per trajectory was

Table 1. *Monte Carlo estimates of relative squared prediction errors for functional gradients with standard error in parenthesis, for both dense and sparse designs, based on* 400 *Monte Carlo runs with sample size* $n = 100$. *The underlying functional regression model is quadratic or cubic and the functional principal components of the predictor process are generated from Gaussian or mixture distributions*

Design	True model	Method	Gaussian	Mixture
Dense	Quadratic	FAD	0·134 (0·043)	0·139 (0·038)
		FQD	0·133 (0·023)	0·141 (0·019)
	Cubic	FAD	0·189 (0·047)	0·183 (0·045)
		FQD	0·368 (0·053)	0·337 (0·051)
Sparse	Quadratic	FAD	0·141 (0·049)	0·139 (0·041)
		FQD	0·136 (0·033)	0·137 (0·026)
	Cubic	FAD	0·228 (0·055)	0·208 (0·051)
		FQD	0·373 (0·050)	0·349 (0·055)

FAD, functional additive differentiation; FQD, functional quadratic differentiation.

selected from $\{30, \ldots, 40\}$ with equal probability, while for the sparse case, the number of measurements was chosen from $\{5, \ldots, 10\}$ with equal probability. The response variables were generated as $Y_i = \sum_k m_k(\xi_{ik}) + \epsilon_i$, with independent errors $\epsilon_i \sim N(0, 0{\cdot}1)$.

We compared the performance of quadratic and additive functional differentiation for two scenarios: (a) a quadratic regression relation with $m_k(\xi_k) = (\xi_k^2 - \lambda_k)/5$; (b) a cubic relation with $m_k(\xi_k) = \xi_k^3/5$. Functional principal component analysis was implemented as described in § 3. The functional derivatives were estimated according to (10) for the proposed additive approach and by a quadratic least squares regression of $\{Y_i - \bar{Y}\}$ on the principal components of X, then using the relation $f_k^{(1)}(\xi_{xk}) = \alpha_k + 2\beta_k \xi_{xk}$ for the quadratic operator.

The results for the overall relative estimation error of the functional gradients $\sum_{k=1}^{2} \|\hat{f}_k^{(1)} - m_k^{(1)}\|^2 / \|m_k^{(1)}\|^2/2$ in Table 1 suggest that the functional additive derivatives lead to similar estimation errors as the quadratic model when the underlying regression is of quadratic form, while the additive modelling leads to substantially improved estimation in all scenarios when the underlying model is cubic. Comparisons of functional linear derivatives using the operator $\Gamma_{L,x}^{(1)}$ with those obtained for additive derivative operators led to analogous results.

5. APPLICATION TO TRAJECTORIES OF FERTILITY

To illustrate the application of functional additive derivatives, we analyze egg-laying data from a biodemographic study conducted for 1000 female medflies, as described in Carey et al. (1998). The goal is to determine shape gradients in early life fertility trajectories that are associated with increased lifetime fertility. The selected sample of 818 medflies includes flies that survived for at least 20 days. The trajectories corresponding to the number of daily eggs laid during the first 20 days of life constitute the functional predictors, while the total number of eggs laid throughout the entire lifetime of a fly is the response. As a pre-processing step, a square root transformation of egg counts was applied.

Daily egg counts during the first 20 days of age are the observed data and are assumed to be generated by smooth underlying fertility trajectories. For 50 randomly selected flies, fitted predictor trajectories, obtained by applying the algorithm described in § 3, are shown in Fig. 1. Most egg-laying trajectories display a steep rise towards a time of peak fertility, followed by a sustained more gradual decline. There is substantial variation in the steepness of the rise to the

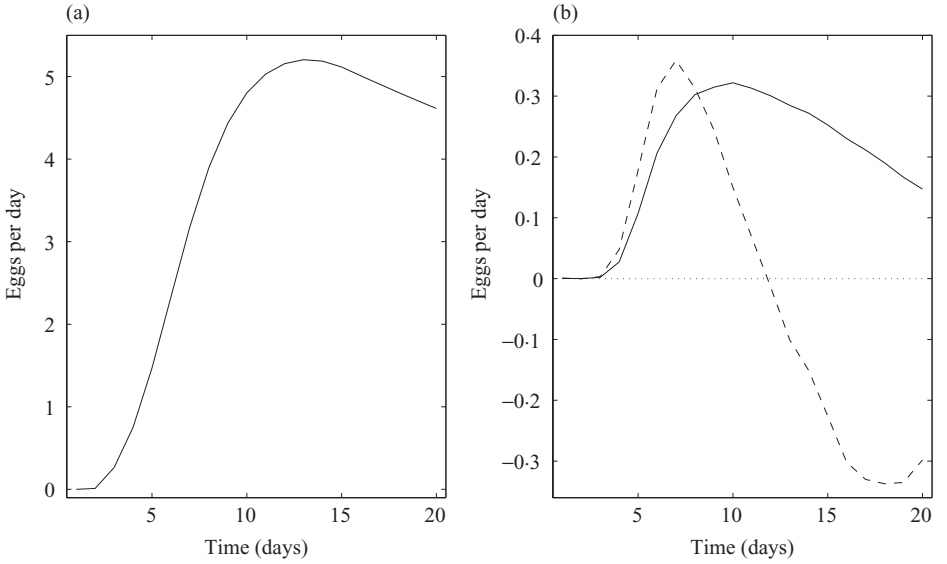


Fig. 2. Smooth estimates of mean function (a) and first (solid) and second (dashed) eigenfunction (b) of the predictor trajectories, explaining 72·1% and 18·6% of the total variation, respectively.

maximal level of egg-laying, and also in the timing of the peak and the rate of decline. Some trajectories rise too slowly to even reach the egg-laying peak within the first 20 days of life. Overall, the shape variation across trajectories is seen to be large.

The total egg count over the entire lifespan is a measure for reproductive success, an important endpoint for quantifying the evolutionary fitness of individual flies. It is of interest to identify shape characteristics of early life reproductive trajectories that are related to evolutionary fitness, i.e., reproductive success. Functional derivatives provide a natural approach to address this question. For the predictor processes, the smooth estimate of the mean fertility function is displayed in Fig. 2(a), while the estimates of the first two eigenfunctions are shown in Fig. 2(b), explaining 72·1% and 18·6% of the total variation of the trajectories, respectively. These eigenfunctions reflect the modes of variation (Castro et al., 1986) and the dynamics of predictor processes. Two components were chosen, accounting for more than 90% of the variation in the data.

We compared the 10-fold crossvalidated relative prediction errors for functional differentiation based on linear, quadratic and additive operators, with resulting error estimates of 0·163 for linear, 0·154 for quadratic and 0·120 for additive approaches. These results support the use of the additive differentiation scheme. For functional additive differentiation, nonparametric regressions of the responses on the first two functional predictor scores are shown in the upper panels of Fig. 3, overlaid with the scatter-plots of observed responses against the respective scores. The estimated first derivatives of these smooth regression functions are in the lower panels, obtained by local quadratic fitting, as suggested in Fan & Gijbels (1996). We find indications of nonlinear relationships. Both derivative estimates feature minima in the middle range and higher values near the ends of the range of the scores; their signs are relative to the definition of the eigenfunctions.

A natural perspective on functional derivatives is the functional gradient field, quantifying changes in responses against changes in the predictor scores. Functional gradients lend themselves to visualization if plotted against the principal components, which is useful if these components explain most of the variation present in the predictor trajectories. The functional gradient field for the eigenbase as functional coordinate system is illustrated in Fig. 4.

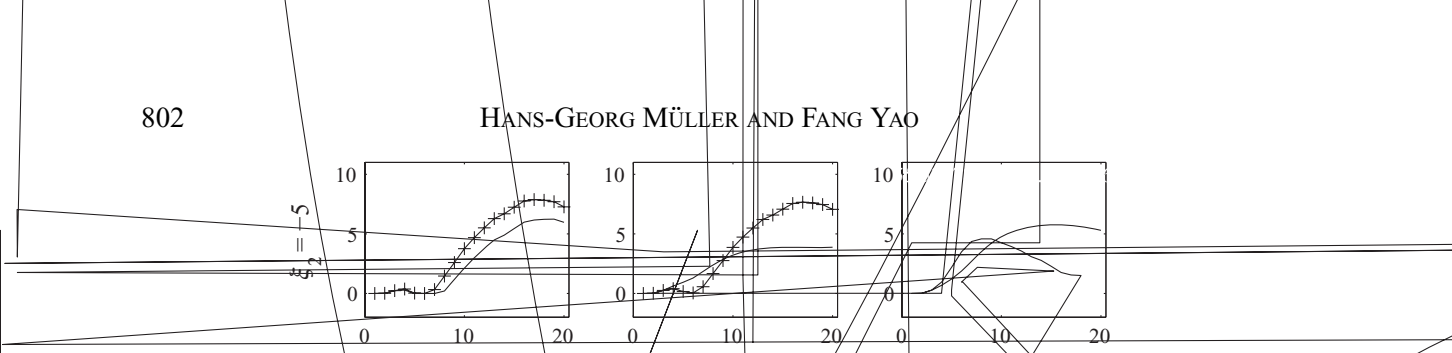
The base of each arrow corresponds to a test trajectory x at which the gradient $\{\Omega_x^{(1)}(\phi_1), \ldots, \Omega_x^{(1)}(\phi_K)\} = \{f_1^{(1)}(\xi_{x1}), \ldots, f_K^{(1)}(\xi_{xK})\}$ is determined, inserting estimates (10) for $K = 2$. The length of each arrow corresponds to the size of the gradient and its direction to the direction u of the functional gradient. If one moves a small unit length along the direction of each arrow, the resulting increase in the response is approximately proportional to the length of the arrow.

The functional gradient field is seen to be overall quite smooth in this application. Increases in total fertility occur when the first functional principal component score is increased and the second score is decreased; the size of the effect of such changes varies locally. Relatively larger increases in the fertility response occur for trajectories with particularly small values as well as large values of the first score, upon increasing this score. Increases of the second score generally lead to declines in reproductive success, and more so for trajectories that have mildly positive second scores. The gradient field also shows that there are no extrema in these data. It is thus likely that biological constraints prevent further increases of fertility by moulding the shapes of early fertility, specifically, in the direction of increasing first and decreasing second scores. The evolutionary force that will favourably select for flies with trajectories that are associated with overall increased fertility is thus likely in equilibrium with counteracting constraints.

Given that the most sustained increases in fertility are associated with increasing the first predictor score, it is of interest to relate this finding to the shape of the first eigenfunction. This eigenfunction is seen to approximately mimic the mean function, see Fig. 2, so that the increases in total fertility that result from increasing the first predictor score are obtained by increased egg-laying activity over the entire domain, paralleling the mean function. This can be viewed as multiplying the mean function by increasing factors; see Chiou et al. (2003) for a discussion of related multiplicative models for functional data. The second eigenfunction corresponds to a sharper early peak, followed by an equally sharp decline, so it is not surprising that the functional derivative in this direction is negative, indicating that a fast rise to peak egg-laying is detrimental to overall fertility, which is likely due to a high cost of early reproduction. We find that both changes in timing and levels of egg-laying are reflected in the functional gradient field, which delineates in compact graphical form the shape changes that are associated with increases in reproductive success.

It is instructive to compare given predictor trajectories with gradient-induced trajectories that are obtained when moving a certain distance, defined by the length of the arrow in the gradient field, along the functional gradient. The shape change from the starting trajectory to the gradient-induced trajectory then provides a visualization of the shape change represented by the functional gradient, corresponding to the shape change that induces the largest gain in lifetime fertility. For this analysis, we select nine test trajectories, which correspond to the bases of the corresponding arrows in the gradient field plot, representing subjects that have all possible combinations of the scores $\xi_1 = \{-7, 0, 7\}$ and $\xi_2 = \{-5, 0, -5\}$. The resulting trajectories are depicted in Fig. 5, arranged from left to right as ξ_1 increases, and from top to bottom as ξ_2 increases. The nine test trajectories, drawn as solid curves, are given by $x = \hat{\mu} + \sum_{k=1}^{2} \xi_{xk}\hat{\phi}_k$, with the values of ξ_{x1}, ξ_{x2} obtained by forming all combinations of the above values. The gradient-induced trajectories are $x^* = \hat{\mu} + \sum_{k=1}^{2}\{\xi_{xk} + \rho \hat{f}_k^{(1)}(\xi_{xk})\}\hat{\phi}_k$, where the scaling factor is $\rho = 10$ for enhanced visualization.

For all scenarios, the functional gradients point towards fertility trajectories that feature enhanced postpeak reproduction. For test trajectories with late timing of the initial rise in fertility, the gradients point towards somewhat earlier timing of the initial rise, as seen in the plots of the first column with $\xi_1 = -7$. These are also trajectories with relatively high peaks. For early steep rises, however, the gradients point towards delayed timing of the rise, seen for the combinations of $\xi_1 = 0, 7$ and $\xi_2 = -5, 0$. For the test trajectories with $\xi_1 = 0, 7$ and $\xi_2 = 5$, the timing of the


$\hat{\sigma}_{jj} = -5$
10
5
0
0
10
20

$f_k^{(1)}(\xi_{yk})| \to 0$. This property is implied by the Cauchy–Schwarz inequality and the following assumption.

Assumption A1. For all $k \geqslant 1$ and for all z_1, z_2 it holds that $|f_k^{(1)}(z_1) - f_k^{(1)}(z_2)| \leqslant L_k |z_1 - z_2|$ for a sequence of positive constants L_k such that $\sum_{k=1}^{\infty} L_k^2 < \infty$.

We consider integral estimators of the functional principal components and a fixed design with the t_{ij}s increasingly ordered. Write $\mathcal{T} = [a, b]$, $\Delta_i = \max\{t_{ij} - t_{i,j-1} : j = 1, \ldots, n_i + 1\}$, where $t_{i0} = a$ and $t_{i,n_i+1} = b$ for all subjects. We make the following assumptions for the design and the process X, denoting $\mathcal{T}^{\delta} = [a - \delta, b + \delta]$ for some $\delta > 0$ and $\min_i$ and $\max_i$ taken over $i = 1, \ldots, n$. Bandwidths $b_i = b_i(n)$ refer to the smoothing parameters used in the local linear least squares estimation steps for obtaining smoothed trajectories $\hat{X}_i$ and $\asymp$ denotes asymptotic equivalence.

Assumption A2. Assume that $X^{(2)}(t)$ is continuous on $\mathcal{T}^{\delta}$; $\int_{\mathcal{T}} E[\{X^{(k)}(t)\}^4]dt < \infty$, $k = 0, 2$; $E(\epsilon_{ij}^4) < \infty$; the functional principal components ξ_{xk} of X are independent.

Assumption A3. Assume that $\min_i n_i \geqslant m \geqslant Cn^{\alpha}$ for some constants $C > 0$ and $\alpha > 5/7$; $\max_i \Delta_i = O(m^{-1})$; there exists a sequence $b \asymp n^{-\alpha/5}$, such that $\max_i b_i \asymp \min_i b_i \asymp b$.

Assumption

001Tf1.0590TDbTj109.65F30TjF11Tf0.050.903TD1.66F3001Tf1.0590TDbTj13F153TF095000Tc0003.

the 3×1 unit vector with the second element equal to 1 and 0 otherwise. Define the hypothetical estimator $\tilde{f}_k^{(1)}(z) = e_2^{\mathrm{T}} S_n^{-1} T_n$.

To evaluate $|\hat{f}_k^{(1)}(z) - \tilde{f}_k^{(1)}(z)|$, one needs to bound the differences $D_{j,1} = \sum_i (\hat{w}_i \hat{\xi}_{ik}^j - w_i \xi_{ik}^j)$, $D_{\ell,2} = \sum_i (\hat{w}_i \hat{\xi}_{ik}^\ell - w_i \xi_{ik}^\ell) Y_i$ $(j = 0, \ldots, 4, \ell = 0, \ldots, 2)$, where $D_{j,1} = \sum_i \{(\hat{w}_i - w_i)\xi_{ik}^j + (\hat{w}_i - w_i)(\hat{\xi}_{ik}^j - \xi_{ik}^j) + w_i(\hat{\xi}_{ik}^j - \xi_{ik}^j)\} \equiv D_{j,11} + D_{j,12} + D_{j,13}$. Modifying the arguments in the proof of Theorem 1 in Müller & Yao (2008), without loss of generality considering $D_{0,1}$ and applying Lemma 1, for generic constants C_1, C_2,

$$h_k D_{0,1} \leqslant \frac{C_1}{n h_k} \sum_i |\hat{\xi}_{ik} - \xi_{ik}| \{I(|z - \xi_{ik}| \leqslant h_k) + I(|z - \hat{\xi}_{ik}| \leqslant h_k)\},$$

$$\leqslant \frac{C_2}{n h_k} \sum_i \|\hat{X}_i - X_i\| I(|z - \xi_{ik}| \leqslant h_k) + \frac{\|\hat{G} - G\|_S}{\delta_k} \frac{1}{n h_k} \sum_i \|X_i\| I(|z - \xi_{ik}| \leqslant h_k). \quad \text{(A4)}$$

Applying the law of large number for a random number of summands (Billingsley, 1995, p. 380) and the Cauchy–Schwarz inequality, the terms in (A4) are bounded in probability by

$$2 p_k(z) \big[\{E(\|\hat{X}_i - X_i\|^2)\}^{1/2} + \delta_k^{-1} \|\hat{G} - G\|_S \{E(\|X_i\|^2)\}^{1/2} \big].$$

Under Assumption 3, it is easy to see that $b^2 + (mb)^{-1/2} = o\{h_k^2 + (n h_k^3)^{-1/2}\}$ and $E\|\hat{G} - G\|_S = o\{h_k^2 + (n h_k^3)^{-1/2}\}$. Analogously one can evaluate the magnitudes of $D_{j,1}$ and $D_{\ell,2}$ for $j = 0, \ldots, 4$, $\ell = 0, 1, 2$, which leads to $|\hat{f}_k^{(1)}(z) - \tilde{f}_k^{(1)}(z)| = o_p\{|\tilde{f}_k^{(1)}(z) - f_k^{(1)}(z)|\}$. Combining this with standard asymptotic results (Fan & Gijbels, 1996) for $\tilde{f}_k^{(1)}(z)$ completes the proof of (11).

To show (12), observe $\int_{\mathcal{T}} \phi_k(t) u(t)\, dt \leqslant 1$ and $\int_{\mathcal{T}} \{\hat{\phi}_k(t) - \phi_k(t)\} u(t)\, dt \leqslant \|\hat{\phi}_k - \phi_k\|$ for $\|u\| = 1$ by the Cauchy–Schwarz inequality and the orthonormality constraints for the ϕ_k. Then

$$\sup_{\|u\|=1} \big|\hat{\Omega}_x^{(1)}(u) - \Omega_x^{(1)}(u)\big| \leqslant \sum_{k=1}^{K} \big\{ \big|\hat{f}_k^{(1)}(\xi_{xk}) - f_k^{(1)}(\xi_{xk})\big| + \big|\hat{f}_k^{(1)}(\xi_{xk}) - f_k^{(1)}(\xi_{xk})\big| \|\hat{\phi}_k - \phi_k\|$$

$$+ \big|f_k^{(1)}(\xi_{xk})\big| \|\hat{\phi}_k - \phi_k\| \big\} + \sum_{k=K+1}^{\infty} \big|f_k^{(1)}(\xi_{xk})\big|,$$

whence Lemma 1 and $E(\|\hat{\phi}_k - \phi_k\|) = o[\delta_k^{-1}\{h_k^2 + (n h_k^3)^{-1/2}\}]$ imply (12).

REFERENCES

ASH, R. B. & GARDNER, M. F. (1975). *Topics in Stochastic Processes*. New York: Academic Press [Harcourt Brace Jovanovich Publishers]. Probability and Mathematical Statistics, Vol. 27.

BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. New York: John Wiley & Sons Inc., 3rd ed. A Wiley-Interscience Publication.

BOENTE, G. & FRAIMAN, R. (2000). Kernel-based functional principal components. *Statist. Prob. Lett.* , 335–45.

CAI, T. & HALL, P. (2006). Prediction in functional linear regression. *Ann. Statist.* , 2159–79.

CARDOT, H., CRAMBES, C., KNEIP, A. & SARDA, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. *Comp. Statist. Data Anal.* , 4832–48.

CAREY, J. R., LIEDO, P., MÜLLER, H.-G., WANG, J.-L. & CHIOU, J.-M. (1998). Relationship of age patterns of fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females. *J. Gerontol. A.: Biol. Sci. Med. Sci.* , 245–51.

CASTRO, P. E., LAWTON, W. H. & SYLVESTRE, E. A. (1986). Principal modes of variation for processes with continuous sample curves. *Technometrics* , 329–37.

CHIOU, J.-M., MÜLLER, H.-G., WANG, J.-L. & CAREY, J. R. (2003). A functional multiplicative effects model for longitudinal data, with application to reproductive histories of female medflies. *Statist. Sinica* , 1119–33.

CUEVAS, A., FEBRERO, M. & FRAIMAN, R. (2002). Linear functional regression: The case of fixed design and functional response. *Can. J. Statist.* , 285–300.

FAN, J. & GIJBELS, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.

FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer.

GASSER, T. & MÜLLER, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method. *Scand. J. Statist.* , 171–85.
GRISWOLD, C., GOMULKIEWICZ, R. & HECKMAN, N. (2008). Hypothesis testing in comparative and experimental studies of function-valued traits. *Evolution* , 1229–42.
HALL, P., MÜLLER, H.-G. & YAO, F. (2009). Estimation of functional derivatives. *Ann. Statist.* , 3307–29.
HASTIE, T. & TIBSHIRANI, R. (1986). Generalized additive models (with discussion). *Statist. Sci.* , 297–318.
IZEM, R. & KINGSOLVER, J. (2005). Variation in continuous reaction norms: Quantifying directions of biological interest. *Am. Naturalist* , 277–89.
KIRKPATRICK, M. & HECKMAN, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and other infinite-dimensional characters. *J. Math. Biol.* , 429–50.
LI, Y. & HSING, T. (2007). On rates of convergence in functional linear regression. *J. Mult. Anal.* , 1782–804.
MÜLLER, H.-G. & YAO, F. (2008). Functional additive models. *J. Am. Statist. Assoc.* , 1534–44.
MÜLLER, H.-G. & ZHANG, Y. (2005). Time-varying functional regression for predicting remaining lifetime distributions from longitudinal trajectories. *Biometrics* , 1064–75.
RAMSAY, J. O. & DALZELL, C. J. (1991). Some tools for functional data analysis. *J. R. Statist. Soc.* B , 539–72.
RAMSAY, J. O. & SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. New York: Springer, 2nd ed.
RICE, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statist. Sinica* , 631–47.
RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when the data are curves. *J. R. Statist. Soc.* B , 233–43.
STANISWALIS, J. G. & LEE, J. J. (1998). Nonparametric regression analysis of longitudinal data. *J. Am. Statist. Assoc.* , 1403–18.
STONE, C. J. (1985). Additive regression and other nonparametric models. *Ann. Statist.* , 689–705.
YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Am. Statist. Assoc.* , 577–90.
ZHOU, S. & WOLFE, D. A. (2000). On derivative estimation in spline regression. *Statist. Sinica* , 93–108.

[*Received June* 2009. *Revised June* 2010]