

Additive Modeling of Functional Gradients

BY HANS-GEORG MÜLLER

*Department of Statistics, University of California, Davis, One Shields Avenue, Davis,
California, 95616, U.S.A.*

mueller@wald.ucdavis.edu

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 functional regression setting where one observes functional predictors and scalar responses. Derivatives are then defined as functional directional derivatives which indicate how changes in the predictor function in a specified functional direction are associated with corresponding changes in the scalar response. Aiming at 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 is consistent. We demonstrate the consistent

49 estimation and interpretation of the resulting functional derivatives and functional gradient fields
50 in a study of the dependence of lifetime fertility of flies on early life reproductive trajectories.

51 *Some key words:* Derivative; Functional Data Analysis; Functional Regression; Gradient Field; Nonparametric Dif-
52 ferentiation; Principal Component.

53

54

1. INTRODUCTION

55 Regression problems where the predictor is a smooth square integrable random function $X(t)$
56 defined on a domain \mathcal{T} and the response is a scalar Y with mean $EY = \mu_Y$ are found in many
57 areas of science. For example, in the biological sciences, one may encounter predictors in the
58 form of subject-specific longitudinal time-dynamic processes such as reproductive activity. For
59 each such process one observes a series of measurements and it is then of interest to model the
60 dependence of the response on the predictor process (Cuevas et al., 2002; Rice, 2004; Ramsay &
61 Silverman, 2005). Examples include studies of the dependence of remaining lifetime on fertility
62 processes (Müller & Zhang, 2005), and a related analysis which we discuss in further detail
63 in Section 5 below. This concerns the dependence of total fertility on the dynamics of the early
64 fertility process in a study of biodemographic characteristics of female medflies. Here we observe
65 trajectories of fertility over the first 20 days of life, measured by daily egg-laying for a sample
66 of medflies, as illustrated for 50 randomly selected flies in Figure 1, showing the realizations of
67 the predictor process. The total number of eggs laid over the lifetime of a fly is also recorded and
68 serves as scalar response.

69 In this and other functional regression settings examples, one would like to determine which
70 predictor trajectories will lead to extreme responses, for example by identifying zeros of the
71 functional gradient field, or to characterize the functional directions in which responses will
72 increase or decrease the most, when taking a specific trajectory as a starting point. In some appli-
73 cations, these directions may give rise to specific interpretations, such as evolutionary gradients
74 (Kirkpatrick & Heckman, 1989; Izem & Kingsolver, 2005). The advantage of functional over
75 multivariate analysis for biological data in the form of trajectories was recently demonstrated

76

77

78

79

80

97 in Griswold et al. (2008). The need to analyze the effects of changes in trajectories in the field
98 of biological evolution and ecology and to address related questions in other fields motivates
99 the development of statistical technology to obtain a functional gradient at a function-valued
100 argument, e.g., a particular predictor function. It is clearly of interest to develop efficient and
101 consistent methods for the estimation of functional gradients.

102 In the framework of functional predictors and scalar responses, derivatives are defined as func-
103 tional directional derivatives which indicate how changes in the predictor function in a specified
104 functional direction are associated with corresponding changes in the scalar response. Similarly
105 to the classical regression setting of a scalar predictor and scalar response, this problem can be
106 easily solved for a functional linear regression relationship where the derivative corresponds to
107 the slope parameter, respectively, regression parameter function, as we demonstrate below. The
108 problem is harder and more interesting in the nonlinear situation where the classical analogue
109 would be the estimation of derivatives of a nonparametric regression function (Gasser & Müller,
110 1984; Zhou & Wolfe, 2000).

111 When tackling this functional differentiation problem, one realizes that the space in which the
112 predictor functions reside is infinite-dimensional and therefore sparsely populated, so that esti-
113 mation techniques will be subject to a rather extreme form of the curse of dimensionality. This
114 problem already arises for the case of a functional regression, before even considering deriva-
115 tives. The conventional approach to reducing the very high dimensionality of the functional re-
116 gression problem is through the well-established functional linear model, which implies strong
117 dimension reduction through structural assumptions. The structural constraints inherent in these
118 model often prove to be too restrictive, just as is the case for ordinary linear regression. A rea-
119 sonably restrictive yet overall sufficiently flexible approach to dimension reduction in regression
120 models with many predictors is additive modeling (Stone, 1985; Hastie & Tibshirani, 1986) and
121 its extension to functional regression (Müller & Yao, 2008).

122 Postulating a regression relation that is additive in the functional principal components of
123 predictor processes, but otherwise unspecified, provides a particularly useful set of constraints
124
125
126
127
128

145 for the estimation of functional derivatives. The resulting functional derivative estimates are
 146 straightforward to implement, even for higher order derivatives, and require nothing more than
 147 obtaining a sequence of nonparametric estimators for the derivatives of one-dimensional smooth
 148 functions, with the principal components of the predictor processes as respective arguments.
 149 This approach is easily extended to the estimation of functional gradient fields and to the case of
 150 higher derivatives and is supported by consistency properties. Functional gradient fields emerge
 151 as a useful tool to aid in the interpretation of functional regression data.

152 2. ADDITIVE MODELING OF FUNCTIONAL DERIVATIVES

153 2.1. Preliminary considerations

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

$$159 \Gamma_L(x) = E(Y | X = x) = \mu_Y + \int_{\mathcal{T}} \beta(t)x^c(t) ds. \quad (1)$$

160 Here, Γ_L is a linear operator on the Hilbert space $L^2(\mathcal{T})$, mapping square integrable functions
 161 defined on the finite interval \mathcal{T} to the real line and β is the regression parameter function, assumed
 162 to be smooth and square integrable. Recent work on this model includes Cai & Hall (2006);
 163 Cardot et al. (2007); Li & Hsing (2007). Since the functional linear model at (1) is the functional
 164 extension of a simple linear regression model, one would intuitively expect that the functional
 165 derivative of the response $E(Y | X)$ with regard to X corresponds to the function β ; in the
 166 following we give a more precise account of this fact.

167 The necessary regularization for estimating the regression parameter function β in model (1),
 168 and thus for identifying the functional linear model, can be obtained through a truncated basis
 169 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 170 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 171 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 172 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 173 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 174 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 175 representation, for example via the eigenbasis of the predictor process X . While using the eigen-
 176 representation, for example via the eigenbasis of the predictor process X . While using the eigen-

193 basis representation of predictor processes is not inherently tied to the functional regression rela-
 194 tionship, it provides natural coordinates for the main directions needed for differentiation, which
 195 are provided by the eigenfunctions that capture the modes of variation of predictor processes.
 196 Introducing directions becomes necessary as the infinite-dimensional predictor space lacks nat-
 197 ural Cartesian coordinates, which are conventionally used for the familiar representation of the
 198 gradients of a function from \mathfrak{R}^p to \mathfrak{R} .

199 The sequence of orthonormal eigenfunctions ϕ_k associated with predictor processes X forms
 200 a basis of the function space, with eigenvalues λ_k , $k = 1, 2, \dots$, and satisfies the orthonormality
 201 relations $\int \phi_j(t)\phi_k(t) dt = \delta_{jk}$, where $\delta_{jk} = 1$ if $j = k$ and $= 0$ if $j \neq k$, and the eigenequa-
 202 tions $\int G(s, t)\phi_k(s)ds = \lambda_k\phi_k(t)$. The eigenfunctions ϕ_k are ordered according to the size of
 203 the corresponding eigenvalues, $\lambda_1 \geq \lambda_2 \geq \dots$. Predictor processes can then be represented by
 204 the Karhunen-Loève expansion

$$205 \quad X(t) = \mu_X(t) + \sum_k \xi_{Xk}\phi_k(t), \quad \text{where} \quad \xi_{Xk} = \int X^c(t)\phi_k(t)dt. \quad (2)$$

207 The random variables ξ_{Xk} are the functional principal components, also referred to as scores.
 208 These scores are uncorrelated and satisfy $E(\xi_{Xk}) = 0$ and $\text{var}(\xi_{Xk}) = \lambda_k$ (Ash & Gardner,
 209 1975).

210 The derivative of a Gâteaux differentiable operator Γ , mapping square integrable functions
 211 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
 212 property that, for functions u and scalars δ ,

$$213 \quad \Gamma(x + \delta u) = \Gamma(x) + \delta \Gamma_x^{(1)}(u) + o(\delta) \quad (3)$$

215 as $\delta \rightarrow 0$. The functional derivative operator at x is then characterized by a sequence of constants
 216 γ_{xk} corresponding to functional directional derivatives $\Gamma_x^{(1)}(\phi_k) = \gamma_{xk}$ in the directions of the
 217 basis functions ϕ_k , and accordingly can be represented as

$$218 \quad \Gamma_x^{(1)} = \sum_{k=1}^{\infty} \gamma_{xk} \Phi_k, \quad (4)$$

219
 220
 221
 222
 223
 224

where $\gamma_{xk} = \Gamma_x^{(1)}(\phi_k)$ is a scalar, and Φ_k denotes the linear operator with

$$\Phi_k(u) = \xi_{uk} = \int u(t)\phi_k(t)dt, \quad \text{for all } u \in L^2(\mathcal{T}).$$

To examine the functional derivative in the framework of the functional linear model (1), we use the representation of the regression parameter function β in the eigenbasis ϕ_k , $\beta(t) = \sum_k b_k \phi_k(t)$, $t \in \mathcal{T}$. This leads to an alternative representation of the operator in (1),

$$\Gamma_L(x) = \mu_Y + \sum_{k=1}^{\infty} b_k \xi_{xk} = \mu_Y + \sum_{k=1}^{\infty} b_k \Phi_k(x),$$

with the constraint $\mu_Y = \int_{\mathcal{T}} \beta(t)\mu_X(t)dt$. For any δ and arbitrary square integrable functions with representations $u = \sum_k \xi_{uk}\phi_k$ and $x = \sum_k \xi_{xk}\phi_k$, one then has

$$\Gamma_L(x + \delta u) = \mu_Y + \sum_k b_k(\xi_{xk} + \delta\xi_{uk}) = \Gamma_L(x) + \delta \sum_k b_k \xi_{uk}.$$

This immediately implies $\Gamma_{L,x}^{(1)} = \sum_{k=1}^{\infty} b_k \Phi_k$, and in this case $\gamma_{xk} = b_k$, so that the functional derivative does not depend on x , as expected. We may conclude from $\Gamma_{L,x}^{(1)}(\phi_k) = b_k$ that the functional derivative is characterized by the regression parameter function β in (1). Although the derivative of the functional linear operator Γ_L is therefore of limited interest, these considerations motivate the study of the derivative of a functional regression operator for the case of a more general nonlinear functional regression relation.

2.2. Derivatives for nonlinear functional regression

The functional linear model is too restrictive in many situations, and especially so when derivatives are of interest, while a completely nonparametric functional regression model is subject to the curse of dimensionality and faces practical difficulties (Ferraty & Vieu, 2006; Hall et al., 2009). At the same time, analogously to the multivariate situation, a coordinate system is needed on which to anchor directional derivatives. In the functional case, in the absence of Cartesian coordinates, a natural orthogonal coordinate system is provided by the eigenfunctions of predictors X . This is a privileged system in that relatively few components will adequately represent predictors X , and the component scores that correspond to the coordinate values that represent

289 X are independent for Gaussian processes, which is particularly beneficial in the additive frame-
 290 work. The functional additive regression framework, where the response depends on predictor
 291 processes through smooth functions of the predictor functional principal components, embodies
 292 sensible structural constraints and dimension reduction and provides a sensible structural com-
 293 promise that is well suited for the estimation of functional gradients.

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

296
$$\Omega(x) = E(Y | X = x) = \mu_Y + \sum_{k=1}^{\infty} f_k(\xi_{xk}),$$

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

300
$$\Omega(x + \delta u) = \mu_Y + \sum_k f_k(\xi_{xk} + \delta \xi_{uk}) = \Omega(x) + \delta \sum_k f_k^{(1)}(\xi_{xk}) \xi_{uk} + o(\delta), \quad (5)$$

301 which leads to

302
$$\Omega_x^{(1)}(u) = \sum_{k=1}^{\infty} f_k^{(1)}(\xi_{xk}) \xi_{uk} = \sum_{k=1}^{\infty} \omega_{xk} \Phi_k(u), \quad \Omega_x^{(1)} = \sum_k f_k^{(1)}(\xi_{xk}) \Phi_k \quad (6)$$

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

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

310
$$\Omega(x + \delta u) = \mu_Y + \sum_k f_k(\xi_{xk} + \delta \xi_{uk}) = \Omega(x) + \sum_{j=1}^p \frac{1}{j!} \delta^j \left\{ \sum_k f_k^{(j)}(\xi_{xk}) \xi_{uk}^j \right\} + o(\delta^p). \quad (7)$$

311 The separation of variables that is an inherent feature of the additive model implies that one
 312 does not need to deal with the unwieldy cross-terms which combine different u'_j s, limiting the
 313 usefulness of functional derivatives of higher order in the general case. The straightforwardness
 314

315
 316
 317
 318
 319
 320

337 of extending functional derivatives to higher orders is a unique feature of the additive approach,
 338 as p -th order derivative operators

$$339 \quad \Omega_x^{(p)}(u) = \sum_{k=1}^{\infty} f_k^{(p)}(\xi_{xk}) \{\Phi_k(u)\}^p \quad (8)$$

340 can be easily obtained by estimating p -th derivatives of the one-dimensional nonparametric func-
 341 tions f_k . As in ordinary multivariate calculus, higher order derivatives can be used to character-
 342 ize extrema or domains with convex or concave functional regression relationships and also for
 343 diagnostics and visualization of nonlinear functional regression relations. As it enables such es-
 344 timates, while retaining full flexibility in regard to the shape of the derivatives, the framework of
 345 additive models is particularly attractive for functional derivative estimation.
 346

347 3. ESTIMATION AND ASYMPTOTICS

348 In order to obtain additive functional derivatives (6), we require estimates of the defining coef-
 349 ficients ω_{xk} , for which we use $\hat{\omega}_{xk} = \hat{f}_k^{(1)}(\xi_{xk})$. Thus, the task is to obtain consistent estimates of
 350 the derivatives $f_k^{(1)}$ for all $k \geq 1$. The data recorded for the i -th subject or unit are typically of the
 351 form $\{(t_{ij}, U_{ij}, Y_i), i = 1, \dots, n, j = 1, \dots, n_i\}$, where predictor trajectories X_i are observed
 352 at times $t_{ij} \in \mathcal{T}$, yielding noisy measurements
 353

$$354 \quad U_{ij} = X_i(t_{ij}) + \epsilon_{ij} = \mu_X(t_{ij}) + \sum_{k=1}^{\infty} \xi_{ik} \phi_k(t_{ij}) + \epsilon_{ij}, \quad (9)$$

355 upon inserting representation (2), where ϵ_{ij} are independent and identically distributed measure-
 356 ment errors, independent of all other random variables, and the observed responses Y_i are related
 357 to the predictors according to $E(Y | X) = \mu_Y + \sum_k f_k(\xi_{Xk})$. A difficulty is that the ξ_{ik} are
 358 not directly observed and must be estimated. For this estimation step, one option is to use the
 359 Principal Analysis by Conditional Expectation (PACE) procedure (Yao et al., 2005) to obtain
 360 estimates $\hat{\xi}_{Xi}$ in a preliminary step. Briefly, the key steps are the nonparametric estimation of the
 361 mean trajectory $\mu_X(t)$ and of the covariance surface $G(t_1, t_2)$ of predictor processes X , obtained
 362 by smoothing pooled scatterplots. For the latter, one omits the diagonal elements of the empirical
 363 covariances, as these are contaminated by the measurement errors. From estimated mean and co-
 364

364

365

366

367

368

385 variance functions, one then obtains eigenfunction and eigenvalue estimates (Rice & Silverman,
386 1991; Staniswalis & Lee, 1998; Boente & Fraiman, 2000).

387 We implement all necessary smoothing steps with local linear smoothing, using automatic
388 data-based bandwidth choices. Additional regularization is achieved by truncating represen-
389 tations (2) and (9) at a suitable number of included components K , typically chosen data-
390 adaptively by pseudo-BIC or similar selectors, or simply as the smallest number of components
391 that explain a large enough fraction of the overall variance of predictor processes. We adopt the
392 latter approach in our applications, requiring that 90% of the variation is explained. Given the ob-
393 servations made for the i -th trajectory, best linear prediction leads to estimates of the functional
394 principal components ξ_{ik} , $E(\xi_{ik} | U_i) = \lambda_k \phi_{ik}^T \Sigma_{U_i}^{-1} (U_i - \mu_{X_i})$, where $U_i = (U_{i1}, \dots, U_{in_i})^T$,
395 $\mu_{X_i} = \{\mu_X(t_{i1}), \dots, \mu_X(t_{in_i})\}^T$, $\phi_{ik} = \{\phi_k(t_{i1}), \dots, \phi_k(t_{in_i})\}^T$, and the (j, l) entry of the
396 $n_i \times n_i$ 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
397 $j \neq l$. One then arrives at the desired estimates $\hat{\xi}_{ik}$ by substituting the unknown components
398 $\lambda_k, \phi_k, \mu_X, G_X$ and σ^2 by their estimates. For densely observed data, a simpler approach is to
399 plug in the above estimates into (2), $\hat{\xi}_{ik} = \int \{\hat{X}_i(t) - \hat{\mu}(t)\} \hat{\phi}_k(t) dt$. These integral estimators
400 require smoothed trajectories \hat{X}_i and therefore dense measurements per sampled curve.

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

$$405 \sum_{i=1}^n \kappa\left(\frac{\hat{\xi}_{ik} - z}{h_k}\right) \{Y_i - \bar{Y} - \sum_{\ell=0}^p \beta_\ell (z - \hat{\xi}_{ik})^\ell\}^2 \quad (10)$$

406 with respect to $\beta = (\beta_0, \dots, \beta_p)^T$ for all z in the domain of interest, leads to suitable derivative
407 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
408 step. Following Fan & Gijbels (1996), we choose $p = \nu + 1$ for practical implementation.

409
410
411
412
413
414
415
416

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

$$435 \quad \widehat{\Omega}_x^{(1)}(u) = \sum_{k=1}^K \widehat{f}_k^{(1)}(\xi_{Xk}) \xi_{uk}, \quad (11)$$

436 when $K = K(n) \rightarrow \infty$ components are included in the estimate and the predictor scores are
 437 independent. Gaussianity of predictor processes is not needed.
 438

439 **THEOREM 1.** *Under assumptions (A1), (A2) in the Appendix, for all $k \geq 1$ for which λ_j , $j \leq$
 440 k are eigenvalues of multiplicity 1, letting $\tau_j(\kappa^\ell) = \int u^j \kappa^\ell(u) du$, as $n \rightarrow \infty$, it holds that*

$$441 \quad (nh_k^3)^{1/2} \left\{ \widehat{f}_k^{(1)}(z) - f_k^{(1)}(z) - \frac{\tau_4(\kappa) f_k^{(3)}(z) h_k^2}{6\tau_2(\kappa)} \right\} \xrightarrow{D} N \left\{ 0, \frac{\tau_2(\kappa^2) \text{var}(Y | \xi_{Xk} = z)}{\tau_2^2(\kappa) p_k(z)} \right\} \quad (12)$$

442 for estimates (10), where p_k is the density of ξ_{Xk} . Under the additional assumption (A3),
 443

$$444 \quad \sup_{\|u\|=1} |\widehat{\Omega}_x^{(1)}(u) - \Omega_x^{(1)}(u)| \xrightarrow{p} 0, \quad (13)$$

445 for estimates (11), at any $x \in L^2(\mathcal{T})$.

446 For further details about the rate of convergence of (13) we refer to (15) in the Appendix. For
 447 higher order functional derivatives, obtained by replacing estimates of first order derivatives $f^{(1)}$
 448 by estimates of higher order derivatives $f^{(p)}$ in (8), one can prove similar consistency results.
 449

450 4. SIMULATION STUDIES

451 To demonstrate the use of the proposed additive modeling of functional gradients, we con-
 452 ducted simulation studies for Gaussian and non-Gaussian predictor processes with different
 453 underlying models and data designs. In particular, we compared our proposal with functional
 454 quadratic differentiation, suggested by a referee, where one obtains derivatives by approximat-
 455 ing the regression relationship with a quadratic operator,
 456

$$457 \quad \Gamma_Q(x) = E(Y | X = x) = \mu_Y + \int_{\mathcal{T}} \alpha(t) x(t) dt + \int_{\mathcal{T}} \beta(t) x^2(t) dt. \quad (14)$$

458 While this model can be implemented with expansions in B-splines or other bases, for the rea-
 459 sons outlined above, we select the orthogonal functional coordinates that are defined by the
 460

461
462
463
464

481 eigenfunctions of X . Inserting $\alpha(t) = \sum_k \alpha_k \phi_k(t)$, $\beta(t) = \sum_k \beta_k \phi_k(t)$, the functional deriva-
 482 tive operator for (14) is seen to be $\Gamma_{Q,x}^{(1)} = \sum_k (\alpha_k + 2\beta_k \xi_{xk}) \Phi_k$.

483 Each of 400 simulation runs consisted of a sample of $n = 100$ predictor trajectories X_i , with
 484 mean function $\mu_X(s) = s + \sin(s)$, $0 \leq s \leq 10$, and a covariance function derived from two
 485 eigenfunctions, $\phi_1(s) = -\cos(\pi s/10)/\sqrt{5}$, and $\phi_2(s) = \sin(\pi s/10)/\sqrt{5}$, $0 \leq s \leq 10$. The
 486 corresponding eigenvalues were chosen as $\lambda_1 = 4$, $\lambda_2 = 1$, $\lambda_k = 0$, $k \geq 3$, and the measurement
 487 errors in (9) as $\varepsilon_{ij} \sim N(0, 0.4^2)$, independent. To study the effect of Gaussianity of the predictor
 488 process, we considered two settings: (i) $\xi_{ik} \sim N(0, \lambda_k)$, Gaussian; (ii) ξ_{ik} are generated from the
 489 mixture of two normals, $N\{(\lambda_k/2)^{1/2}, \lambda_k/2\}$ with probability 1/2 and $N\{-(\lambda_k/2)^{1/2}, \lambda_k/2\}$
 490 with probability 1/2, a mixture distribution. Each predictor trajectory was sampled at locations
 491 uniformly distributed over the domain $[0, 10]$, where the number of noisy measurements was cho-
 492 sen separately and randomly for each predictor trajectory. We considered both dense and sparse
 493 design cases. For the dense design case, the number of measurements per trajectory was selected
 494 from $\{30, \dots, 40\}$ with equal probability, while for the sparse case, the number of measurements
 495 was chosen from $\{5, \dots, 10\}$ with equal probability. The response variables were generated as
 496 $Y_i = \sum_k m_k(\xi_{ik}) + \varepsilon_i$, with independent errors $\varepsilon_i \sim N(0, 0.1)$.

497 We compared the performance of quadratic and additive functional differentiation for two
 498 scenarios: (a) a quadratic regression relation with $m_k(\xi_k) = (\xi_k^2 - \lambda_k)/5$; (b) a cubic relation
 499 with $m_k(\xi_k) = \xi_k^3/5$. Functional principal component analysis was implemented as described in
 500 Section 3. The functional derivatives were estimated according to (11) for the proposed additive
 501 approach and by a quadratic least squares regression of $\{Y_i - \bar{Y}\}$ on the principal components
 502 of X , then using the relation $f_k^{(1)}(\xi_{xk}) = \alpha_k + 2\beta_k \xi_{xk}$ for the quadratic operator.

503 The results for the overall relative estimation error of the functional gradients
 504 $\sum_{k=1}^2 \|\hat{f}_k^{(1)} - m_k^{(1)}\|^2 / \|m_k^{(1)}\|^2 / 2$ in Table 1 suggest that the functional additive derivatives lead
 505 to similar estimation errors as the quadratic model when the underlying regression is of quadratic
 506 form, while the additive modeling leads to substantially improved estimation in all scenarios

507

508

509

510

511

512

529 when the underlying model is cubic. Comparisons of functional linear derivatives using operator
530 $\Gamma'_{L,x}$ with those obtained for additive derivative operators led to analogous results.

531

532 5. APPLICATION TO TRAJECTORIES OF FERTILITY

533 To illustrate the application of functional additive derivatives, we analyze egg-laying data from
534 a biodemographic study conducted for 1000 female medflies, as described in Carey et al. (1998).
535 The goal is to determine shape gradients in early life fertility trajectories that are associated with
536 increased lifetime fertility. The selected sample of 818 medflies includes flies which survived for
537 at least 20 days. The trajectories corresponding to the number of daily eggs laid during the first
538 20 days of life constitute the functional predictors, while the total number of eggs laid throughout
539 the entire lifetime of a fly is the response. As a pre-processing step, a square root transformation
540 of egg counts was applied.

541 Daily egg counts during the first 20 days of age are the observed data and are assumed to
542 be generated by smooth underlying fertility trajectories. For 50 randomly selected flies, fitted
543 predictor trajectories, obtained by applying the algorithm described in Section 3, are shown in
544 Figure 1. Most egg-laying trajectories display a steep rise towards a time of peak fertility, fol-
545 lowed by a sustained more gradual decline. There is substantial variation in the steepness of the
546 rise to the maximal level of egg-laying, and also in the timing of the peak and the rate of decline.
547 Some trajectories rise too slowly to even reach the egg-laying peak within the first 20 days of
548 life. Overall, the shape variation across trajectories is seen to be large.

549 The total egg count over the entire lifespan is a measure for reproductive success, an impor-
550 tant endpoint for quantifying the evolutionary fitness of individual flies. It is of interest to identify
551 shape characteristics of early life reproductive trajectories that are related to evolutionary fitness,
552 i.e., reproductive success. Functional derivatives provide a natural approach to address this ques-
553 tion. For the predictor processes, the smooth estimate of the mean fertility function is displayed
554 in the left panel of Figure 2, while the estimates of the first two eigenfunctions are shown in the
555 right panel, explaining 72.1% and 18.6% of the total variation of the trajectories, respectively.

556

557

558

559

560

577 These eigenfunctions reflect the modes of variation (Castro et al., 1986) and the dynamics of pre-
 578 dictor processes. Two components were chosen, accounting for more than 90% of the variation
 579 in the data.

580 We compared the 10-fold cross-validated relative prediction errors for functional differenti-
 581 ation based on linear, quadratic and additive operators, with resulting error estimates of 0.163
 582 for linear, 0.154 for quadratic and 0.120 for additive approaches. These results support the use of
 583 the additive differentiation scheme. For functional additive differentiation, nonparametric regres-
 584 sions of the responses on the first two functional predictor scores are shown in the upper panels
 585 of Figure 3, overlaid with the scatterplots of observed responses against the respective scores.
 586 The estimated first derivatives of these smooth regression functions are in the lower panels, ob-
 587 tained by local quadratic fitting, as suggested in Fan & Gijbels (1996). We find indications of
 588 nonlinear relationships. Both derivative estimates feature minima in the middle range and higher
 589 values near the ends of the range of the scores; their signs are relative to the definition of the
 590 eigenfunctions.

591 A natural perspective on functional derivatives is the functional gradient field, quantifying
 592 changes in responses against changes in the predictor scores. Functional gradients provide use-
 593 ful visualization if plotted against the principal components, which is useful if these compo-
 594 nents explain most of the variation present in the predictor trajectories. The functional gradient
 595 field for the eigenbase as functional coordinate system is illustrated in Figure 4. The base of
 596 each arrow corresponds to a test trajectory x at which the gradient $\{\Omega_x^{(1)}(\phi_1), \dots, \Omega_x^{(1)}(\phi_K)\} =$
 597 $\{f_1^{(1)}(\xi_{x1}), \dots, f_K^{(1)}(\xi_{xK})\}$ is determined, inserting estimates (11) for $K = 2$. The length of each
 598 arrow corresponds to the size of the gradient and its direction to the direction u of the functional
 599 gradient. If one moves a small unit length along the direction of each arrow, the resulting increase
 600 in the response is approximately proportional to the length of the arrow.

601 The functional gradient field is seen to be overall quite smooth in this application. Increases in
 602 total fertility result when increasing the first functional principal component score and decreasing
 603 the second score; the size of the effect of such changes varies locally. Relatively larger increases
 604

605

606

607

608

625 in the fertility response occur for trajectories with particularly small values as well as large
626 values of the first score, upon increasing this score. Increases of the second score generally
627 lead to declines in reproductive success, and more so for trajectories that have mildly positive
628 second scores. The gradient field also shows that there are no extrema in these data. It is thus
629 likely that biological constraints prevent further increases of fertility by moulding the shapes of
630 early fertility, specifically, in the direction of increasing first and decreasing second scores. The
631 evolutionary force that will favorably select for flies with trajectories that are associated with
632 overall increased fertility is thus likely in equilibrium with counter-acting constraints.

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

646 It is instructive to compare given predictor trajectories with gradient-induced trajectories that
647 are obtained when moving a certain distance, defined by the length of the arrow in the gradient
648 field, along the functional gradient. The shape change from the starting trajectory to the gradient-
649 induced trajectory then provides a visualization of the shape change represented by the functional
650 gradient, corresponding to the shape change that induces the largest gain in lifetime fertility. For
651 this analysis, we select nine test trajectories, which correspond to the bases of the corresponding

652

653

654

655

656

673 arrows in the gradient field plot, representing subjects that have all possible combinations of the
 674 scores $\xi_1 = \{-7, 0, 7\}$ and $\xi_2 = \{-5, 0, 5\}$. The resulting trajectories are depicted in Figure 5,
 675 arranged from left to right as ξ_1 increases, and from top to bottom as ξ_2 increases. The nine test
 676 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}
 677 obtained by forming all combinations of the above values. The gradient-induced trajectories are
 678 $x^* = \hat{\mu} + \sum_{k=1}^2 \{\xi_{xk} + \rho \hat{f}_k^{(1)}(\xi_{xk})\} \hat{\phi}_k$, where the scaling factor is $\rho = 10$ for enhanced visual-
 679 ization.

680 For all scenarios, the functional gradients point towards fertility trajectories that feature en-
 681 hanced post-peak reproduction. For test trajectories with late timing of the initial rise in fertility,
 682 the gradients point towards somewhat earlier timing of the initial rise, as seen in the plots of the
 683 first column with $\xi_1 = -7$. These are also trajectories with relatively high peaks. For early steep
 684 rises, however, the gradients point towards delayed timing of the rise, seen for the combinations
 685 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
 686 the rise is unaltered in the gradient but the height of the peak and post-peak fertility are sub-
 687 stantially increased. We thus find that the timing of the rise in the test trajectory and the size of
 688 its peak substantially influence the direction of the gradient. Large gradients and shape changes
 689 are associated with early and at the same time low rises; a typical example is the test trajectory
 690 with $\xi_1 = 0, \xi_2 = -5$. The gradients seem to point towards an optimal timing of the initial rise
 691 in fertility and in addition to high levels of post-peak reproduction. These features emerge as the
 692 crucial components for enhanced reproductive success.

693 Our analysis demonstrates that the study of fertility dynamics clearly benefits from employing
 694 functional derivatives and that additive operators provide an attractive implementation. Func-
 695 tional additive derivatives and the resulting functional gradients are expected to aid in the analysis
 696 of other complex time-dynamic data as well.

697

698

699

700

701

702

703

704

ACKNOWLEDGEMENTS

We are grateful to a reviewer for a detailed and constructive comments, which led to many improvements. This research was supported by the National Science Foundation and a NSERC Discovery grant.

APPENDIX

A1. Notations and assumptions

To guarantee expansions (5) and (6) and the differentiability of Ω , one needs to assume that, for any $x, y \in L^2(\mathcal{T})$ with $x = \sum_k \xi_{xk} \phi_k$ and $y = \sum_k \xi_{yk} \phi_k$, if $\|x - y\|_{L^2} \rightarrow 0$, then $\sum_k |f_k^{(1)}(\xi_{xk}) - f_k^{(1)}(\xi_{yk})| \rightarrow 0$. This property is implied by the Cauchy-Schwarz inequality and

(A1) For all $k \geq 1$ and for all z_1, z_2 it holds that $|f_k^{(1)}(z_1) - f_k^{(1)}(z_2)| \leq L_k |z_1 - z_2|$ 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 $\mathcal{T} = [a, b]$, $\Delta_i = \max\{t_{ij} - t_{i,j-1} : j = 1, \dots, n_i + 1\}$, 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 $i = 1, \dots, n$. 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. The following assumptions are referred to as (A2).

(A2.1) Assume that $X^{(2)}(t)$ is continuous on \mathcal{T}^δ ; $\int_{\mathcal{T}} E[\{X^{(k)}(t)\}^4] dt < \infty$, $k = 0, 2$; $E(\epsilon_{ij}^4) < \infty$; the functional principal components ξ_{xk} of X are independent.

(A2.2) Assume that $\min_i n_i \geq m \geq Cn^\alpha$ 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 $\max_i b_i \asymp \min_i b_i \asymp b$.

(A2.3) The kernel κ is a Lipschitz continuous symmetric density with compact support $[-1, 1]$.

To characterize the convergence rate of the functional derivative operator, define

$$\begin{aligned} \theta_k(z) &= \left\{ |f_k^{(3)}(z)| + \frac{|f_k^{(1)}(z)|}{\delta_k} \right\} h_k^2 + \left\{ \frac{\sigma_k(z)}{p_k^{1/2}(z)} + \frac{|f_k^{(1)}(z)|}{\delta_k} \right\} (nh_k^3)^{1/2}, \\ \theta^*(x) &= \sum_{k=1}^K \theta_k(\xi_{xk}) + \sum_{k=K+1}^{\infty} |f_k^{(1)}(\xi_{xk})|, \end{aligned} \tag{15}$$

where $\delta_1 = \lambda_1 - \lambda_2$ and $\delta_k = \min_{j \leq k} (\lambda_{j-1} - \lambda_j, \lambda_j - \lambda_{j+1})$ for $k \geq 2$. We also require

(A3) For any $x = \sum_{k=1}^{\infty} \xi_{xk} \phi_k \in L^2(\mathcal{T})$, $\sum_{k=1}^K \theta_k(\xi_{xk}) \rightarrow 0$, $\sum_{k=K+1}^{\infty} |f_k^{(1)}(\xi_{xk})| \rightarrow 0$, as $n \rightarrow \infty$.

A2. Proof of Theorem 1

We first state an important lemma that sets the stage for proving the main theorem. Define $\|F\|_S = \{\int \int F^2(s, t) ds dt\}^{1/2}$ for a symmetric bivariate function F .

LEMMA 1. Under (A2), if ϕ_k is of multiplicity 1 and $\hat{\phi}_k$ is chosen such that $\int \phi_k \hat{\phi}_k > 0$, then

$$E(\|\hat{X}_i - X_i\|^2) = O\{b^4 + (mb)^{-1}\}, \tag{16}$$

$$E(\|\hat{\mu} - \mu\|^2) \asymp E\{\|\hat{G} - G\|_S^2\} = O\{b^4 + (nmb)^{-1} + m^{-2} + n^{-1}\}, \tag{17}$$

$$|\hat{\lambda}_k - \lambda_k| \leq \|\hat{G} - G\|_S, \quad \|\hat{\phi}_k - \phi_k\| \leq 2\sqrt{2}\delta_k^{-1} \|\hat{G} - G\|_S, \tag{18}$$

$$|\hat{\xi}_{ik} - \xi_{ik}| \leq C(\|\hat{X}_i - X_i\| + \delta_k^{-1} \|X_i\| \|\hat{G} - G\|_S), \tag{19}$$

where (18) and (19) hold uniformly over i and k .

Proof of Theorem 1. We provide a brief sketch of the main steps. Denote $\sum_{i=1}^n$ by \sum_i and let $w_i = \kappa\{(z - \xi_{ik})/h_k\}/(nh_k)$, $\hat{w}_i = \kappa\{(z - \hat{\xi}_{ik})/h_k\}/(nh_k)$, $\theta_k = \theta_k(z)$, $S_n = (S_{n,j+l})_{0 \leq j, l \leq 2}$, with $S_{n,j} = \sum_i w_i (\xi_{ik} - z)^j$, $T_n = (T_{n,0}, T_{n,1}, T_{n,2})^T$, with $T_{n,j} = \sum_i w_i (\xi_{ik} - z)^j Y_i$, and $\hat{S}_n = (\hat{S}_{n,j+l})_{0 \leq j, l \leq 2}$, $\hat{T}_n = (\hat{T}_{n,0}, \dots, \hat{T}_{n,2})^T$ for the corresponding quantities with w_i and ξ_{ik} replaced by \hat{w}_i and $\hat{\xi}_{ik}$. From (10), the local quadratic estimator of the derivative function $f_k^{(1)}(z)$ can be written as $\hat{f}_k^{(1)}(z) = e_2^T \hat{S}_n^{-1} \hat{T}_n$, where e_2^T is the 3×1 unit vector with the 2nd element equal to 1 and 0 otherwise. Define the hypothetical estimator $\tilde{f}_k^{(1)}(z) = e_2^T S_n^{-1} T_n$.

817 To evaluate $|\hat{f}_k^{(1)}(z) - \tilde{f}_k^{(1)}(z)|$, one needs to bound the differences

818 $D_{j,1} = \sum_i (\hat{w}_i \hat{\xi}_{ik}^j - w_i \xi_{ik}^j), \quad D_{\ell,2} = \sum_i (\hat{w}_i \hat{\xi}_{ik}^\ell - w_i \xi_{ik}^\ell) Y_i, \quad j = 0, \dots, 4, \ell = 0, \dots, 2,$

819 where $D_{j,1} = \sum_i \{(\hat{w}_i - w_i) \xi_{ik}^j + (\hat{w}_i - w_i)(\hat{\xi}_{ik}^j - \xi_{ik}^j) + w_i(\hat{\xi}_{ik}^j - \xi_{ik}^j)\} \equiv D_{j,11} + D_{j,12} +$

820 $D_{j,13}$. Modifying the arguments in the proof of Theorem 1 in Müller & Yao (2008), without loss

821 of generality considering $D_{0,1}$ and applying Lemma 1, for generic constants C_1, C_2 ,

822
$$h_k D_{0,1} \leq \frac{C_1}{nh_k} \sum_i |\hat{\xi}_{ik} - \xi_{ik}| \{I(|z - \xi_{ik}| \leq h_k) + I(|z - \hat{\xi}_{ik}| \leq h_k)\},$$

823
$$\leq \frac{C_2}{nh_k} \sum_i \|\hat{X}_i - X_i\| I(|z - \xi_{ik}| \leq h_k) + \frac{\|\hat{G} - G\|_S}{\delta_k} \frac{1}{nh_k} \sum_i \|X_i\| I(|z - \xi_{ik}| \leq h_k) \quad (20)$$

825 Applying the law of large number for a random number of summands (Billingsley, 1995, page

826 380) and Cauchy-Schwarz inequality, the terms in (20) are bounded in probability by

827
$$2p_k(z) [\{E(\|\hat{X}_i - X_i\|^2)\}^{1/2} + \delta_k^{-1} \|\hat{G} - G\|_S \{E(\|X_i\|^2)\}^{1/2}].$$

829 Under (A2.2), it is easy to see that $b^2 + (mb)^{-1/2} = o\{h_k^2 + (nh_k^3)^{-1/2}\}$ and $E\|\hat{G} - G\|_S =$

830 $o\{h_k^2 + (nh_k^3)^{-1/2}\}$. Analogously one can evaluate the magnitudes of $D_{j,1}$ and $D_{\ell,2}$ for $j =$

831 $0, \dots, 4, \ell = 0, 1, 2$, which leads to $|\hat{f}_k^{(1)}(z) - \tilde{f}_k^{(1)}(z)| = o_p\{|\hat{f}_k^{(1)}(z) - f_k^{(1)}(z)|\}$. Combining

832 this with standard asymptotic results (Fan & Gijbels, 1996) for $\tilde{f}_k^{(1)}(z)$ completes the proof of

833 (12).

834 To show (13), observe $\int_{\mathcal{T}} \phi_k(t) u(t) dt \leq 1$ and $\int_{\mathcal{T}} \{\hat{\phi}_k(t) - \phi_k(t)\} u(t) dt \leq \|\hat{\phi}_k - \phi_k\|$ for

835 $\|u\| = 1$ by the Cauchy-Schwarz inequality and the orthonormality constraints for the ϕ_k . Then

836
$$\sup_{\|u\|=1} |\hat{\Omega}_x^{(1)}(u) - \Omega_x^{(1)}(u)| \leq \sum_{k=1}^K \left\{ |\hat{f}_k^{(1)}(\xi_{xk}) - f_k^{(1)}(\xi_{xk})| + |\hat{f}_k^{(1)}(\xi_{xk}) - f_k^{(1)}(\xi_{xk})| \|\hat{\phi}_k - \phi_k\| \right.$$

837
$$\left. + |f_k^{(1)}(\xi_{xk})| \|\hat{\phi}_k - \phi_k\| \right\} + \sum_{k=K+1}^{\infty} |f_k^{(1)}(\xi_{xk})|,$$

839 whence Lemma 1 and $E(\|\hat{\phi}_k - \phi_k\|) = o[\delta_k^{-1} \{h_k^2 + (nh_k^3)^{-1/2}\}]$ imply (13).

840

841

REFERENCES

842 ASH, R. B. & GARDNER, M. F. (1975). *Topics in Stochastic Processes*. New York: Academic Press [Harcourt Brace

843 Jovanovich Publishers]. Probability and Mathematical Statistics, Vol. 27.

844

845

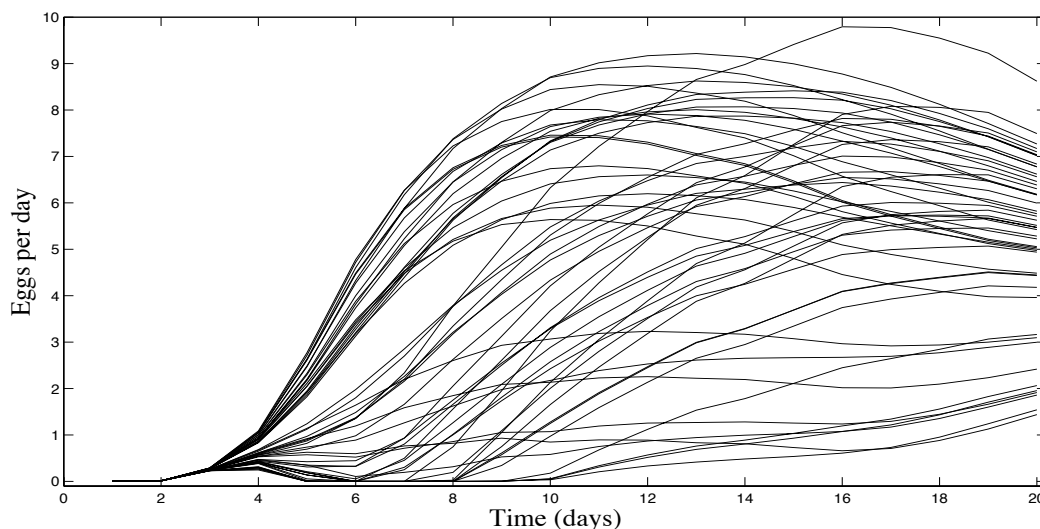
846

847

848

- 865 BILLINGSLEY, P. (1995). *Probability and Measure*. Wiley Series in Probability and Mathematical Statistics. New
866 York: John Wiley & Sons Inc., 3rd ed. A Wiley-Interscience Publication.
- 867 BOENTE, G. & FRAIMAN, R. (2000). Kernel-based functional principal components. *Statistics & Probability Letters*
868 **48**, 335–345.
- 869 CAI, T. & HALL, P. (2006). Prediction in functional linear regression. *The Annals of Statistics* **34**, 2159–2179.
- 870 CARDOT, H., CRAMBES, C., KNEIP, A. & SARDA, P. (2007). Smoothing splines estimators in functional linear
871 regression with errors-in-variables. *Computational Statistics & Data Analysis* **51**, 4832–4848.
- 872 CAREY, J. R., LIEDO, P., MÜLLER, H.-G., WANG, J.-L. & CHIOU, J.-M. (1998). Relationship of age patterns of
873 fecundity to mortality, longevity, and lifetime reproduction in a large cohort of Mediterranean fruit fly females.
874 *Journal of Gerontology - Biological Sciences and Medical Sciences* **53**, 245–251.
- 875 CASTRO, P. E., LAWTON, W. H. & SYLVESTRE, E. A. (1986). Principal modes of variation for processes with
876 continuous sample curves. *Technometrics* **28**, 329–337.
- 877 CHIOU, J.-M., MÜLLER, H.-G., WANG, J.-L. & CAREY, J. R. (2003). A functional multiplicative effects model for
878 longitudinal data, with application to reproductive histories of female medflies. *Statistica Sinica* **13**, 1119–1133.
- 879 CUEVAS, A., FEBRERO, M. & FRAIMAN, R. (2002). Linear functional regression: the case of fixed design and
880 functional response. *Canadian Journal of Statistics. La Revue Canadienne de Statistique* **30**, 285–300.
- 881 FAN, J. & GJIBELS, I. (1996). *Local Polynomial Modelling and its Applications*. London: Chapman & Hall.
- 882 FERRATY, F. & VIEU, P. (2006). *Nonparametric Functional Data Analysis*. New York: Springer, New York.
- 883 GASSER, T. & MÜLLER, H.-G. (1984). Estimating regression functions and their derivatives by the kernel method.
884 *Scandinavian Journal of Statistics. Theory and Applications* **11**, 171–185.
- 885 GRISWOLD, C., GOMULKIEWICZ, R. & HECKMAN, N. (2008). Hypothesis testing in comparative and experimental
886 studies of function-valued traits. *Evolution* **62**, 1229–1242.
- 887 HALL, P., MÜLLER, H.-G. & YAO, F. (2009). Estimation of functional derivatives. *The Annals of Statistics* **37**,
888 3307–C3329.
- 889 HASTIE, T. & TIBSHIRANI, R. (1986). Generalized additive models. *Statistical Science* **1**, 297–318. With discussion.
- 890 IZEM, R. & KINGSOLVER, J. (2005). Variation in continuous reaction norms: Quantifying directions of biological
891 interest. *American Naturalist* **166**, 277–289.
- 892 KIRKPATRICK, M. & HECKMAN, N. (1989). A quantitative genetic model for growth, shape, reaction norms, and
893 other infinite-dimensional characters. *Journal of Mathematical Biology* **27**, 429–450.
- 894 LI, Y. & HSING, T. (2007). On rates of convergence in functional linear regression. *Journal of Multivariate Analysis*
895 **98**, 1782–1804.
- 896 MÜLLER, H.-G. & YAO, F. (2008). Functional additive models. *Journal of the American Statistical Association*
103, 1534–1544.

- 913 MÜLLER, H.-G. & ZHANG, Y. (2005). Time-varying functional regression for predicting remaining lifetime distri-
914 butions from longitudinal trajectories. *Biometrics* **61**, 1064–1075.
- 915 RAMSAY, J. O. & DALZELL, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical*
Society: Series B (Statistical Methodology) **53**, 539–572.
- 916 RAMSAY, J. O. & SILVERMAN, B. W. (2005). *Functional Data Analysis*. Springer Series in Statistics. New York:
917 Springer, 2nd ed.
- 918 RICE, J. A. (2004). Functional and longitudinal data analysis: Perspectives on smoothing. *Statistica Sinica*, 631–647.
- 919 RICE, J. A. & SILVERMAN, B. W. (1991). Estimating the mean and covariance structure nonparametrically when
the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53**, 233–243.
- 920 STANISWALIS, J. G. & LEE, J. J. (1998). Nonparametric regression analysis of longitudinal data. *Journal of the*
921 *American Statistical Association* **93**, 1403–1418.
- 922 STONE, C. J. (1985). Additive regression and other nonparametric models. *The Annals of Statistics* **13**, 689–705.
- 923 YAO, F., MÜLLER, H.-G. & WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *Journal of*
the American Statistical Association **100**, 577–590.
- 924 ZHOU, S. & WOLFE, D. A. (2000). On derivative estimation in spline regression. *Statistica Sinica* **10**, 93–108.



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

937 [Received June 2009. Revised April 2010]

938

939

940

941

942

943

944

Table 1. Monte Carlo estimates of relative squared prediction errors for functional gradients with standard error in parenthesis, obtained for the proposed functional additive differentiation (FAD) and functional quadratic differentiation (FQD) methods 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)

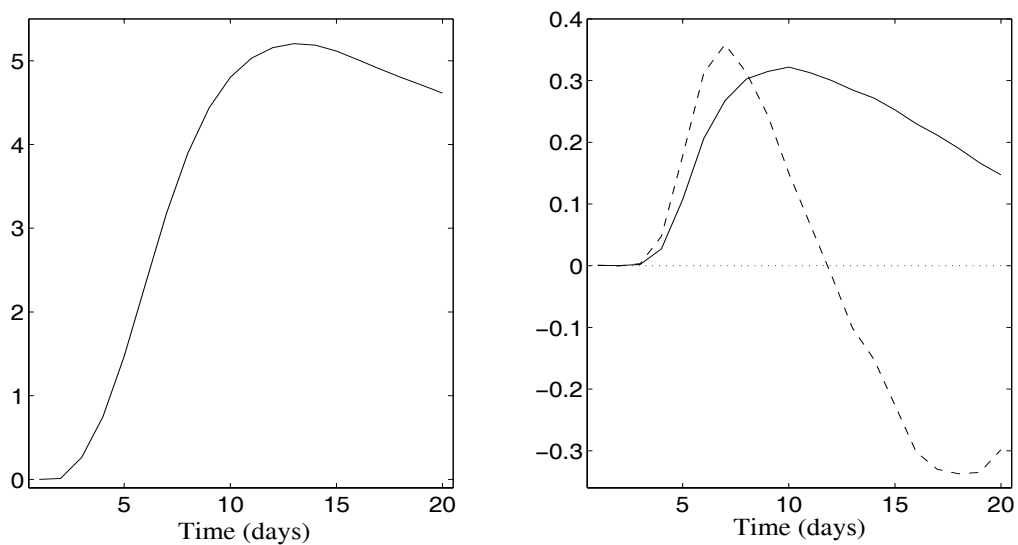


Fig. 2. Smooth estimates of mean function (left panel) and first (solid) and second (dashed) eigenfunction (right panel) of the predictor trajectories, explaining 72.1% and 18.6% of the total variation, respectively.

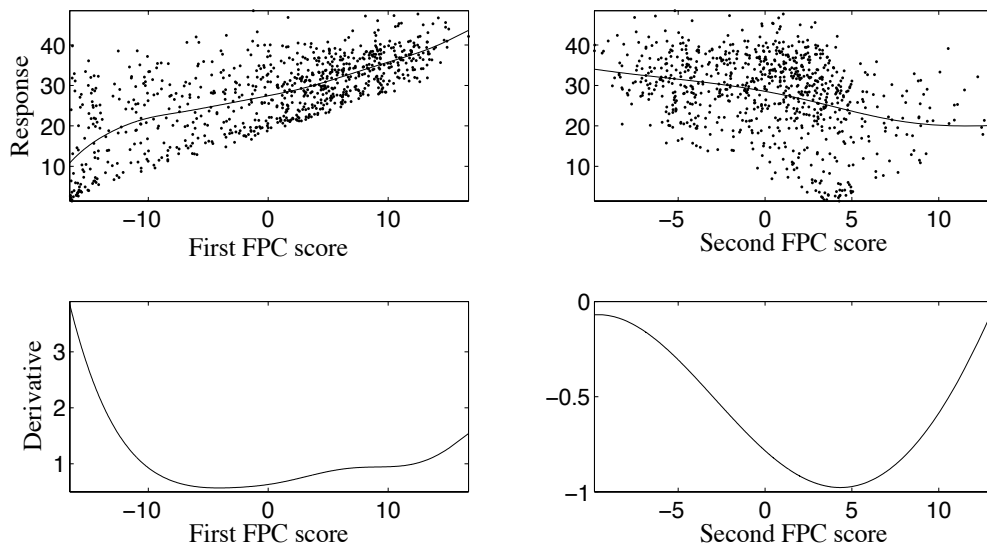


Fig. 3. Top panels: Nonparametric regression of the response (total fertility) on the first (left) and second (right) functional principal component (FPC) of predictor processes. Bottom panels: Estimated derivatives of the smooth regression functions in the top panels.

1057
 1058
 1059
 1060
 1061
 1062
 1063
 1064
 1065
 1066
 1067
 1068
 1069
 1070
 1071
 1072
 1073
 1074
 1075
 1076
 1077
 1078
 1079
 1080
 1081
 1082
 1083
 1084
 1085
 1086
 1087
 1088

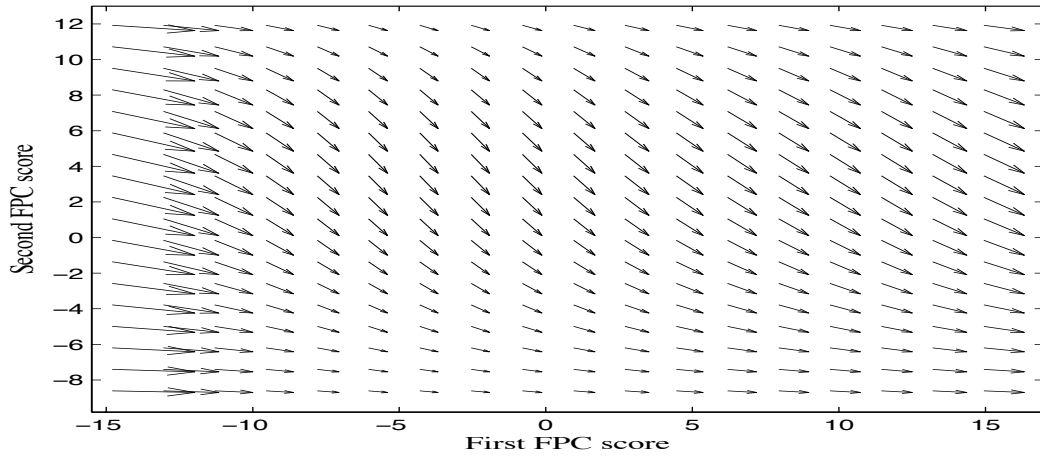


Fig. 4. Estimated functional gradient field for total fertility, differentiated against the predictor process, expressed in terms of gradients of the response with respect to first (abscissa) and second (ordinate) functional principal component (FPC). The arrows and their lengths indicate the direction and magnitude of the functional gradient at the predictor function that corresponds to the base of the arrow.

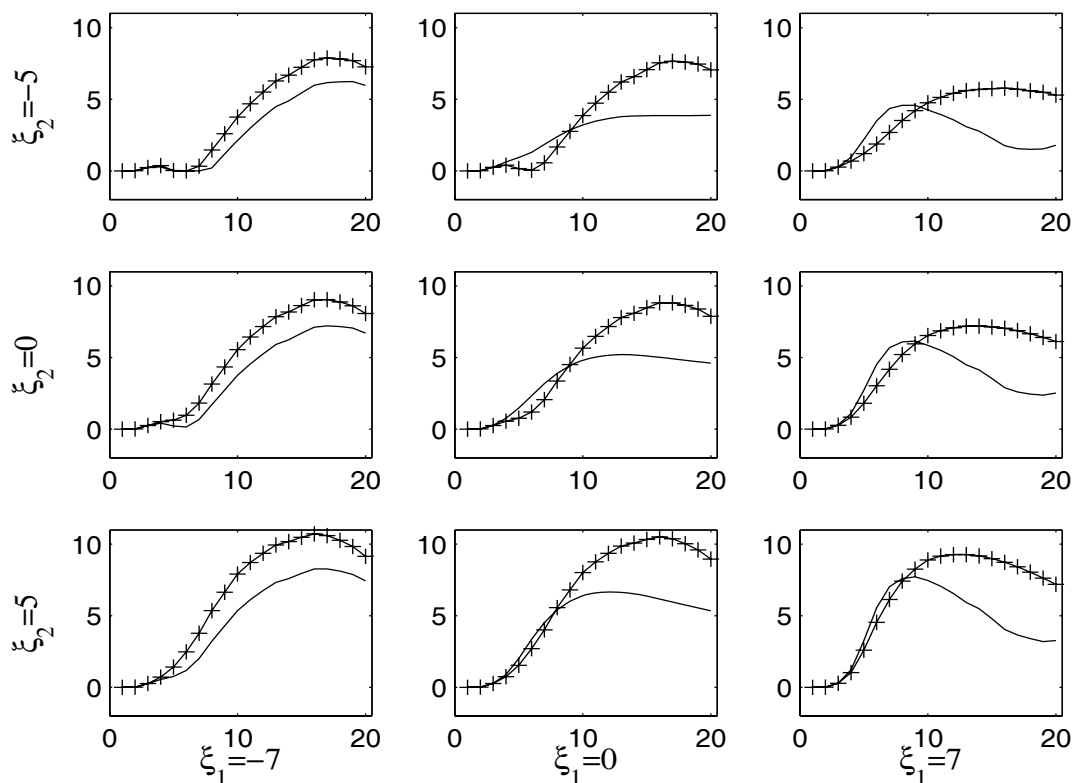


Fig. 5. Visualization of the shape changes in fertility trajectories induced by gradients: The test trajectories (solid) correspond to the nine combinations of the functional principal components obtained for $\xi_1 \in \{-7, 0, 7\}$ and $\xi_2 \in \{-5, 0, 5\}$ and lie at the base of the respective arrows in Figure 4. The gradient-induced trajectories (plus marked lines) are obtained after moving in the direction of the arrow by an amount that is proportional to the length of the arrow. For all panels, the abscissa indicates age (in days) and the ordinate fertility, measured as eggs per day.