Functional Additive Models

In commonly used functional regression models, the regression of a scalar or functional response on the functional predictor is assumed to be linear. This means that the response is a linear function of the functional principal component scores of the predictor process. We relax the linearity assumption and propose to replace it by an additive structure, leading to a more widely applicable and much more flexible framework for functional regression models. The proposed functional additive regression models are suitable for both scalar and functional responses. The regularization needed for effective estimation of the regression parameter function is implemented through a projection on the eigenbasis of the covariance operator of the functional components in the model. The use of functional principal components in an additive rather than linear way leads to substantial broadening of the scope of functional regression models and emerges as a natural approach, because the uncorrelatedness of the functional principal components is shown to lead to a straightforward implementation of the functional additive model, based solely on a sequence of one-dimensional smoothing steps and without the need for backfitting. This facilitates the theoretical analysis, and we establish the asymptotic consistency of the estimates of the components of the functional additive model. We illustrate the empirical performance of the proposed modeling framework and estimation methods through simulation studies and in applications to gene expression time course data.

KEY WORDS: Additive model; Asymptotics; Functional data analysis; Functional regression; Linear model; Principal components; Smoothing; Stochastic processes.

1.

A characteristic feature of functional regression models is that either the predictor or the response or both are functions, with the functional components typically assumed to be realizations of a stochastic process. The functional linear model is the commonly adopted functional regression model. It has been introduced in its most general form, where both predictor and response are functions, by Ramsay and Dalzell (1991). The case of a functional predictor and scalar response has been the focus of most research to date on the functional linear model, as well as somewhat artificial situations in which the functional data are assumed to be observed without noise and on a very dense and regular grid. For this case, Cardot, Ferraty, Mas, and Sarda (2003) provided consistency results and introduced a testing procedure. Theory for the case of fixed design and functional response was developed by Cuevas, Febrero, and Fraiman (2002). Ramsay and Silverman (2005) have summarized some of these developments.

Several extensions of the basic linear functional regression models have been proposed, often motivated by established analogous extensions of the classical multivariate regression models toward more general regression models. These include generalized functional linear models (James 2002; Escabias, Aguilera, and Valderrama 2004; Cardot and Sarda 2005; Müller and Stadtmüller 2005); modifications of functional regression for longitudinal, (i.e., sparse), irregular, and noisy data (Yao, Müller, and Wang 2005b); varying-coefficient functional models (Malfait and Ramsay 2003; Fan and Zhang 2000;

domain. Therefore, the predictor set is not countable, and an additive model in these predictors would require uncountably many additive components. We overcome this difficulty by taking as predictors the countable set of FPC scores, which can be truncated at an increasing sequence of finitely many predictors while representing the entire predictor function adequately.

For the case of a scalar response, the combination of FPC scores with an additive model, emphasizing applied modeling with readily available software tools, has been demonstrated in concurrent work with applied emphasis (Foutz and Jank 2008; Liu and Müller 2008; Sood, James and Tellis 2008). As we show, a key to both analysis and implementation of this combination is the uncorrelatedness of the functional predictor scores. The usual implementation of additive models, which must take into account dependencies among predictors, requires backfitting or similarly complex schemes. Because the functional predictor scores are uncorrelated, the additive fitting step can be greatly simplified and requires no more than one-dimensional smoothing steps, separately applied to each predictor score. This has important consequences: very fast and simple implementation and simplification, making it possible to study asymptotic properties of the resulting model, especially in the Gaussian case. The combination of FPC score predictors and additive models thus emerges as a particularly natural and flexible data-adaptive nonparametric framework for functional regression models. We refer to this approach as the *functional additive model* (FAM).

The article is organized as follows. Section 2 contains background from functional linear regression. The proposed FAMs are introduced in Section 3, and issues of fitting these models are the theme of Section 4. Asymptotic consistency properties are presented in Section 5, whereas Section 6 is devoted to a report on simulation results. An application to regression for gene expression time courses in the *Drosophila* life cycle is described in Section 7, and concluding remarks are given in Section 8.

for scalar responses Y and for each FPC score ζ_m of the response process. In practice, when fitting a functional regression model and estimating the regression parameter function β, we need to regularize by truncating these expansions at a finite number of components K and M.

3. A A

Suppose for the moment that the true FPC scores $_{ik}$ for predictor processes are known. Then the functional linear models (1) and (2) are reduced to a standard linear model with infinitely many of these FPC scores as predictors, as demonstrated in eqs (4) and (5). Moreover, the linear structure in the predictor scores and the uncorrelatedness of the FPC scores then imply that

$$E(Y - \ _Y|\ _k) = b_k\ _k \qquad \text{and} \qquad E(\zeta_m|\ _k) = b_{km}\ _k \tag{6}$$

for scalar and functional response models. Accordingly, the best predictor is the linear predictor, and the regressions on predictor scores are lines through the origin.

This observation provides a key motivation for the extension of the functional linear model to the FAM. It seems natural to generalize the well-known extension of (generalized) linear to (generalized) additive models (Hastie and Tibshirani 1990) to the functional case by replacing the linear terms $b_k\ _k$ and $b_{km}\ _k$ in (6) and (4)–(5) by more general relationships. We thus generalize the linear relationship with $_k$ to arbitrary functional relations $f_k(\ _k)$, where functions $f_k(\cdot), k = 1, 2, \ldots$ [resp. functions $f_{km}(\cdot)$, $k, m = 1, 2, \ldots$], are assumed to be smooth; beyond smoothness, nothing more is needed. This substitution transforms functional linear models (1) and (2) to FAMs with the underlying FPC scores $_k$ as predictors,

$$E(Y|X) = \ _Y + \sum_{k=1}^{\infty} f_k(\ _k) \tag{7}$$

and

$$E(Y(t)|X) = \ _Y(t) + \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} f_{km}(\ _k)\ \ _m(t), \tag{8}$$

for scalar and functional response cases. To ensure identifiability, we also require that

$$\begin{aligned} Ef_k(\ _k) &= 0, \qquad k = 1, 2, \ldots, \qquad \text{and} \\ Ef_{km}(\ _k) &= 0, \qquad k = 1, 2, \ldots, m = 1, 2, \ldots. \end{aligned} \tag{9}$$

In this model the linear relationship between the response Y and the predictor FPC scores $_k$ is replaced by an additive relation that gives rise to a far more flexible and essentially nonparametric model while avoiding the curse of dimension, which for infinite-dimensional functional data is unsurmountable if no structure is imposed. Beyond additivity, a second key assumption that we make from now on is that the predictor FPC scores $_k$ are independent. Because these scores are always uncorrelated, this assumption is, for example, satisfied for the case in which predictor processes are Gaussian. Then the basic FAM assumptions (7) and (8) imply that

$$\begin{aligned} E(Y - \ _Y|\ _k) &= E\{E(Y - \ _Y|X)|\ _k\} \\ &= E\left\{\sum_{j=1}^{\infty} f_j(\ _j)\Big|\ _k\right\} = f_k(\ _k), \end{aligned} \tag{10}$$

and for the functional response case, analogously,

$$\begin{aligned} E(\zeta_m|\ _k) &= E\{E(\zeta_m|X)|\ _k\} \\ &= E\left\{\sum_{j=1}^{\infty} f_{jm}(\ _j)\Big|\ _k\right\} = f_{km}(\ _k). \end{aligned} \tag{11}$$

The relations $f_k(\ _k) = E(Y - \ _Y|\ _k)$ and $f_{km}(\ _k) = E(\zeta_m|\ _k)$, for all $k, m = 1, 2, \ldots$, are straightforward generalizations of the decomposition of functional linear regression into simple linear regressions against the predictor FPC scores as in (6). They are key for surprisingly simple implementations of the FAM. Whereas complex iterative procedures are required to fit a regular additive model (backfitting and variants; see, e.g., Hastie and Tibshirani 1990; Mammen and Park 2005), representations (10) and (11) motivate a straightforward estimation scheme to recover the component functions f_k and f_{km}, through a series of one-dimensional smoothing steps. This not only leads to fast and easily diagnosed procedures for the underlying infinite-dimensional data, but also facilitates asymptotic analysis. The high degree of flexibility and the simplicity of model fitting makes FAM an especially attractive alternative to the special case of the standard functional linear models (1) and (2).

4. A A

We begin with an overview of the estimation procedures. In a first step, smooth estimates of the mean and covariance functions for the predictor processes are obtained by scatterplot smoothing. This is followed by FPC analysis, which yields estimates $\hat{}_k$ for the eigenfunctions, $\hat{}_k$ for the eigenvalues, and $\hat{}_{ik}$ for the FPC scores of individual predictor trajectories; some additional details are given in the Appendix. The estimation steps are implemented with the *principal analysis by conditional expectation* (PACE) approach, also regarding the choice of the number of included eigenfunctions K through a pseudo–Akaike information criterion (AIC) (Yao et al. 2005a), available in the PACE package. This has been shown to work for densely sampled trajectories and in the Gaussian case in addition to the case of sparse and irregular measurements, and also has been demonstrated to be fairly robust against violations of the Gaussian assumption.

Once these preliminary estimates are in hand, obtaining estimates $\hat{f}_k$ and $\hat{f}_{km}$ of the smooth component functions f_k and f_{km} is straightforward. We implement all smoothing steps with local polynomial fitting; other smoothing techniques can be used equally well. For the case of scalar responses, we estimate the functions f_k by fitting a local linear regression to the data $\{\hat{}_{ik}, Y_i\}_{i=1,\ldots,n}$, where $\hat{}_{ik}$ is obtained by (A.4) in the Appendix. Minimizing

$$\sum_{i=1}^{n} K_1\left(\frac{\hat{}_{ik} - x}{h_k}\right)\{Y_i - \beta_0 - \beta_1(x - \hat{}_{ik})\}^2 \tag{12}$$

with respect to β_0 and β_1, leads to $\hat{f}_k(x) = \hat{\beta}_0(x) - \bar{Y}$, where h_k is the bandwidth used for this smoothing step and K_1 is a symmetric probability density that serves as a kernel function. For the functional response case, the functions f_{mk} are analogously

estimated by passing a local linear smoother through the data $\{\hat{}_{ik}, \hat{\zeta}_{im}\}_{i=1,\ldots,n}$ that are obtained by (A.4), that is, minimizing

$$\sum_{i=1}^{n} K_1\left(\frac{\hat{}_{ik} - x}{h_{mk}}\right)\{\hat{\zeta}_{im} - \beta_0 - \beta_1(x - \hat{}_{ik})\}^2 \qquad (13)$$

with respect to β_0 and β_1, leading to $\hat{f}_{mk}(x) = \hat{\beta}_0(x)$, where h_{mk} is the bandwidth. Then the fitted version of the FAM (7) with scalar response is

$$\widehat{E}(Y|X) = \bar{Y} + \sum_{k=1}^{K} \hat{f}_k(\ _k). \qquad (14)$$

To quantify the strength of the regression relationship, we use a global measure similar to the coefficient of determination R^2 in standard linear regression (Draper and Smith 1998), with population and sample versions

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{i=1}^{n}\{Y_i - E(Y_i|X_i)\}^2}{\sum_{i=1}(Y_i - \ _Y)^2} \qquad \text{and} \\ \widehat{R}^2 &= 1 - \frac{\sum_{i=1}^{n}\{Y_i - \widehat{E}(Y_i|X_i)\}^2}{\sum_{i=1}(Y_i - \bar{Y})^2}, \end{aligned} \qquad (15)$$

where $E(Y_i|X_i)$ and $\widehat{E}(Y_i|X_i)$ for the ith subject are as in (7) and (14).

Analogously, the fitted FAM for functional responses, based on (8), is

$$\widehat{E}\{Y(t)|X\} = \hat{}_Y(t) + \sum_{m=1}^{M}\sum_{k=1}^{K} \hat{f}_{mk}(\ _k)\ \hat{}_m(t), \qquad t \in \mathcal{T}, \qquad (16)$$

and the strength of the regression for this case can be measured by

$$\begin{aligned} R^2 &= 1 - \frac{\sum_{i=1}^{n}\int [Y_i(t) - E\{Y_i(t)|X_i\}]^2\,dt}{\sum_{i=1}^{n}\int \{Y_i(t) - \ _Y(t)\}^2\,dt}, \\ \widehat{R}^2 &= 1 - \frac{\sum_{i=1}^{n}\sum_{l=2}^{m_i}[V_{il} - \widehat{E}\{Y_i(t_{il})|X_i\}]^2(t_{il} - t_{i,l-1})}{\sum_{i=1}^{n}\sum_{l=2}^{m_i}\{V_{il} - \ _Y(t_{il})\}^2(t_{il} - t_{i,l-1})} \end{aligned} \qquad (17)$$

for population and sample versions, where $E\{Y_i(t)|X_i\}$ and $\widehat{E}(Y_i(t)|X_i)$ for the ith subject are as in (8) and (16), and a dense grid of measurements t_{il} is assumed for each subject.

5. A

Establishing relevant asymptotic results requires studying the relationship between the true and estimated FPC scores $_{ik}$ and $\hat{}_{ik}$, ζ_{im} and $\hat{\zeta}_{im}$, $k = 1, \ldots, K$, $m = 1, \ldots, M$, because the estimates of the FAM component functions f_k and f_{km} must be based on the estimated scores. Starting with known convergence results for the estimated population components such as mean function, eigenfunction, and eigenvalue estimates in model (A.2) or (A.3) (see Yao et al. 2005a; Hall and Hosseini-Nasab 2006), a key step in the mathematical analysis is to establish exact upper bounds of $|\hat{}_{ik} - \ _{ik}|$ and $|\hat{\zeta}_{im} - \zeta_{im}|$, that are iid in terms of i or do not depend on i, $i = 1, \ldots, n$ (see the Supplement for details). The convergence properties of the estimated additive model components f_k in (7) or f_{mk} in (8) will follow from those upper bounds, because these estimates are obtained by applying a nonparametric smoothing method to $\{\hat{}_{ik}, Y_i\}$ or $\{\hat{}_{ik}, \hat{\zeta}_{im}\}$ for $i = 1, \ldots, n$.

Asymptotic results are obtained for $n \to \infty$, and the number of included components K and M needs to satisfy $K = K(n) \to \infty$ and $M = M(n) \to \infty$ for a genuinely functional (infinite-dimensional) approach. The results concern consistency of estimates (12) and (13) and of predicted responses (14) and (16) obtained for new predictor processes. Details on the regularity assumptions are given in the Appendix.

Theorem 1. Under assumptions (A1.1)–(A5), (C1.1), (C1.2), (C2.1), and (C2.3) (see the App.), in the scalar response case, for all $k \geq 1$ for which $\ _j$, $j \leq k$ are eigenvalues of multiplicity 1,

$$\hat{f}_k(x) - f_k(x) \xrightarrow{p} 0, \qquad (18)$$

for estimates (12). Under the additional assumptions (B1.1)–(B4) and (C2.2), in the functional response case, for all k and m for which $\ _j$, $j \leq k$ and $\ _l$, $l \leq m$ are eigenvalues of multiplicity 1,

$$\hat{f}_{km}(x) - f_{km}(x) \xrightarrow{p} 0, \qquad (19)$$

for estimates (13).

Additional results on the rates of convergence of $\tilde{}_k(x) = |\hat{f}_k(x) - f_k(x)|$ and $\tilde{\vartheta}_{mk}(x) = |\hat{f}_{mk}(x) - f_{mk}(x)|$ are given in the Supplement, eq. (41). Next, we consider consistency of the predictions obtained by applying FAM.

Theorem 2. Under (A1.1)–(A4), (A6), (A7), (C1.1), (C1.2), (C2.1), and (C2.3), for the scalar response case,

$$\widehat{E}(Y|X) - E(Y|X) \xrightarrow{p} 0, \qquad (20)$$

where $\widehat{E}(Y|X) = \bar{Y} + \sum_{k=1}^{K}\hat{f}_k(\ _k)$ as in (12). Under the additional assumptions (B1.1)–(B4), (B5), (B6), and (C2.2), it holds for the functional response case that for all $t \in \mathcal{T}$,

$$\widehat{E}\{Y(t)|X\} - E\{Y(t)|X\} \xrightarrow{p} 0, \qquad (21)$$

where $\widehat{E}\{Y(t)|X\} = \hat{}_Y(t) + \sum_{k=1}^{K}\sum_{m=1}^{M}\hat{f}_{mk}(\ _k)\ \hat{}_m(t)$, with $\hat{f}_{mk}(\ _k)$ as in (13).

Additional results on the rates of convergence of $\ ^*_n = |\widehat{E}(Y|X) - E(Y|X)|$ and $\vartheta^*_n = |\widehat{E}(Y(t)|X) - E(Y(t)|X)|$ are given in eq. (42) of the Supplement.

6. A

We conducted simulation studies to illustrate the empirical performance of the FAMs (7) with scalar and (8) with functional response. For both cases, we generated 200 simulation runs, each consisting of a sample of $n = 100$ predictor trajectories X_i, with mean function $\ _X(s) = s + \sin(s)$, $0 \leq s \leq 10$, and a covariance function derived from two eigenfunctions, $\ _1(s) = -\cos(\ s/10)/\sqrt{5}$ and $\ _2(s) = \sin(\ s/10)/\sqrt{5}$, $0 \leq s \leq 10$. The corresponding eigenvalues were chosen as $\ _1 = 4$, $\ _2 = 1$, and $\ _k = 0$, $k \geq 3$, with measurement errors in (A.2) as $\epsilon_{ij} \overset{\text{iid}}{\sim} \mathrm{N}(0, .5^2)$. We consider two different underlying distributions of the predictor FPC scores: (a) $\ _{ik} \sim \mathcal{N}(0, \ _k)$ and (b) $\ _{ik} = \sqrt{\ _k}(Z_{ik} - 4)/2$, where $Z_{ik} \overset{\text{iid}}{\sim} \text{gamma}(4, 1)$, which is a right-skewed distribution, $k = 1, 2$.

As for the design of the number and spacing of the measurement locations at which predictor trajectories are sampled, we considered both dense and sparse designs to check the robustness of our methods against violations of the dense design assumption. For the dense case, each predictor trajectory was sampled at locations distributed uniformly over the domain $[0, 10]$, where the number of measurements was chosen separately and randomly for each predictor trajectory, by selecting a number from $\{30, \ldots, 40\}$ with equal probability. For the more challenging sparse case, the number of measurements was chosen from $\{3, \ldots, 6\}$ with equal probability. For each simulation run, we generated 100 new predictor trajectories X_i^* with measurements U_{ij}^*, taken at the same time points U_{ij} as for the 100 observed predictor trajectories and 100 associated response variables (resp. response functions Y_i^*).

For the scalar response case, we generated responses $Y_i = \sum_{k=1}^{2} f_k(\ _{ik}) + \varepsilon_i$, where the f_k's are the true component functions that relate the FPC scores $_{ik}$ of the predictor trajectories X_i to the responses Y_i, with errors $\varepsilon_i \overset{\text{iid}}{\sim} \mathrm{N}(0, .1)$, so that $_Y = 0$ and $^2_Y = .1$. We compared the performance of fitting the FAM (7) and the functional linear regression model (1) under two situations: (a) the true functions $f_k(x) = x^2 - \ _k$ for $k = 1, 2$ are nonlinear, and (b) the true functions $f_1(x) = 2x$ and $f_2(x) = x/2$ are linear. As we demonstrated in (6), the functions f_k become lines through the origin for the case of the functional linear regression model. The representation (6) immediately suggests the estimates $\hat{f}_k(x) = \sum_{i=1}^{n}(Y_i - \bar{Y})(\hat{}_{ik} - \bar{\hat{}}_{\cdot k})\hat{}_k^{-1}x$, where $\bar{Y}_i = \sum_{i=1}^{n} Y_i/n$, $\bar{\hat{}}_{\cdot k} = \sum_{i=1}^{n} \hat{}_{ik}/n$, $\hat{}_k$ is the estimate of the kth eigenvalue of the predictor process X, and the $\hat{}_{ik}$'s are obtained by the PACE method in (A.4).

For both the scalar and functional response cases, we implemented the FAM as described in Section 4, including choosing the number of model components for the predictor processes by the AIC, and applied local polynomial smoothing to estimate the functions f_k, with leave-one-out cross-validation for automatic choice of the smoothing bandwidths. We fitted the functional linear model as described earlier and compared the quality of the prediction of responses for the new subjects.

For the functional response case, we also compared the predictive performance of FAM with that of functional linear regression. For the latter, we estimated the (in this case linear) component functions f_{km} by $\hat{f}_{km}(x) = \sum_{i=1}^{n}(\hat{\xi}_{im} - \bar{\xi}_{\cdot m})(\hat{}_{ik} - \bar{\hat{}}_{\cdot k})\hat{}_k^{-1}x$. The simulation settings were the same as those for the scalar response case. In addition, we generated functional response trajectories as $Y_i(t) = \ _Y(t) + \zeta_{i1}\ _1(t)$, where $_Y(t) = t + \sin(t)$, $\ _1(t) = -\cos(\ t/10)/\sqrt{5}$, and ζ_{i1} were the only nonzero FPC scores for the responses, $0 \le t \le 10$. To generate the scores ζ_{i1}, as before, we considered nonlinear and linear scenarios for the component functions $f_{km}, k = 1, 2, m = 1$: (a) the true functions $f_{k1}(x) = x^2 - \ _k$ were nonlinear, $k = 1, 2$, that is, $\zeta_{i1} = \sum_{k=1}^{2}(\ ^2_{ik} - \ _k)$, and (b) the true functions $f_{11}(x) = 2x$, $f_{21}(x) = x/2$ were linear, that is, $\zeta_{i1} = 2\ _{i1} + \ _{i2}/2$. The measurement errors ε_{il} in (A.3) were generated iid from $\mathcal{N}(0, .1)$. The designs for sampling the response trajectories were chosen in the same way as those for sampling the predictor trajectories, with both sparse and dense cases included.

To evaluate the prediction of new responses from future subjects, we generated 100 new predictor and response trajectories X_i^* and Y_i^*, with measurements U_{ij}^* and V_{il}^* taken at the same time points as U_{ij} and V_{il}. The quality of the responses was measured in terms of the relative prediction errors (RPEs),

$$\mathrm{RPE}_i = \frac{(Y_i^* - \hat{Y}_i^*)^2}{Y_i^{*2}} \qquad \text{and}$$
$$\mathrm{RPE}_{i,f} = \frac{\int\ (Y_i^*(t) - \hat{Y}_i^*(t))^2\,dt}{\int\ Y_i^{*2}(t)\,dt}, \tag{22}$$

for scalar and functional response cases.

The results for the RPE when the predictor FPC scores are normal, given in Table 1, suggest that FAM leads to similar prediction errors in the sparse and somewhat larger prediction errors in the dense case compared with the functional linear approach when the true functions f_k or f_{km} are linear, whereas FAM improves on functional linear regression when the underlying component functions are nonlinear. This holds equally for scalar and functional responses.

Similar results emerge for the case of right-skewed distributions (Table 2). Again, the median losses when using FAM for the case of an underlying linear model are small in the sparse case, whereas they are now more noticeable in the dense case. The improvements obtained when using FAM for the nonlinear case for both dense and sparse designs are found to persist for

Table 1. Simulation results for the comparison of predictions obtained by the FAM and functional linear regression (LIN), for models with scalar response [see (7) for the FAM version, (1) for the linear version] and with functional response [(8) for the FAM version, (2) for the linear version], for both dense and sparse designs

Design	Response	Model	True	25th	50th	75th	True	25th	50th	75th
Dense	Scalar	FAM	NLF	.0683	.2983	1.820	LF	.0075	.0431	.3024
		LIN		.6458	1.068	1.870		.0102	.0335	.1362
	Functional	FAM	NLF	.0025	.0086	.0279	LF	.0005	.0012	.0031
		LIN		.0109	.0363	.0705		.0004	.0009	.0019
Sparse	Scalar	FAM	NLF	.1124	.4437	2.884	LF	.0176	.1066	.7149
		LIN		.6614	1.071	1.822		.0270	.1133	.5378
	Functional	FAM	NLF	.0066	.0156	.0432	LF	.0023	.0042	.0090
		LIN		.0138	.0389	.0737		.0023	.0044	.0094

NOTE: The true component functions are linear (LF) and nonlinear (NLF), and the true FPC scores of the predictor process are generated from normal distributions, as described in Section 6. Simulations were based on 400 Monte Carlo runs with $n = 100$ predictors and responses per sample. Shown in the table are the Monte Carlo estimates of the 25th, 50th, and 75th percentiles of the relative prediction error, RPE (22).

Table 2. Simulation results obtained for the same settings as in Table 1, but with the true FPC scores of the predictor process generated from right-skewed distributions related to gamma(4, 1), as described in Section 6

Design	Response	Model	True	25th	50th	75th	True	25th	50th	75th
Dense	Scalar	FAM	NLF	.0163	.0885	1.035	LF	.0014	.0066	.0410
		LIN		.0396	.1827	3.383		.0008	.0030	.0135
	Functional	FAM	NLF	.0028	.0101	.0375	LF	.0005	.0013	.0034
		LIN		.0126	.0417	.0819		.0003	.0005	.0010
Sparse	Scalar	FAM	NLF	.0448	.2775	3.558	LF	.0291	.1511	1.031
		LIN		.1867	.7179	3.642		.0305	.1423	.9316
	Functional	FAM	NLF	.0072	.0146	.0485	LF	.0023	.0041	.0090
		LIN		.0156	.0422	.0803		.0021	.0042	.0091

the situation with skewed distributions. We note that the performance of FAM appears to be less stable compared with the linear model in the tails of the error distribution, as evidenced by occasional relatively large values in the 75th percentiles of RPEs. We conclude that in situations where the signal is not too weak, the losses when using FAM in the linear case are relatively small, whereas FAM performs better than the linear model in situations when the underlying regression relationship is nonlinear.

7. A

A

We applied our methods to gene expression profile data where both predictors and responses are functional. Arbeitman et al. (2002) obtained developmental gene expression profiles over the entire lifespan of *Drosophila*, and we applied functional regression to study the relation of the expression profiles for different developmental periods (see also Müller, Chiou, and Leng 2008). The genes in the biologically identified group of $n = 21$ "transiently expressed zygotic genes" show early peaks in expression during the embryonal phase and are active in the cellularization phase of the embryo. For this well-defined gene group, we studied how the expression in the pupa or metamorphosis phase depends on that in the embryo phase. The data consist of 31 measurements during the embryo phase (i.e., the predictor process) and 18 measurements during the pupa phase (i.e., the response process). For one of the genes, data were available only for the embryo phase, and thus we used the data for this gene only to carry out the FPC analysis for predictor processes. The linearly interpolated gene expression trajectories for both predictor and response processes, as well as their mean functions, are shown in Figure 1, confirming an early peak in the predictor trajectories and displaying quite a bit of variation between genes.

The AIC method selected three components for predictor processes and four components for response processes. The corresponding eigenfunctions are displayed in Figure 2; the legend explains the fraction of variation explained by the corresponding eigenvalues. Of interest are the pairwise scatterplots of all pairings of response FPC scores $\hat{\zeta}_{i1}, \hat{\zeta}_{i2}, \hat{\zeta}_{i3}$, and $\hat{\zeta}_{i4}$ versus the predictor FPC scores $\hat{}_{i1}$, $\hat{}_{i2}$, and $\hat{}_{i3}$, as shown in Figure 3. Judging from these scatterplots, there exist clear relationships between response and predictor scores; whereas several of these appear to be close to linear, for others a linear fit is not good, pointing to the presence of nonlinear relationships. To interpret these relations, we need to take the shape of both the predictor and response eigenfunctions into account; for example, the negative relationship between $\hat{\zeta}_{i1}$ and $\hat{}_{i1}$ implies that sharper

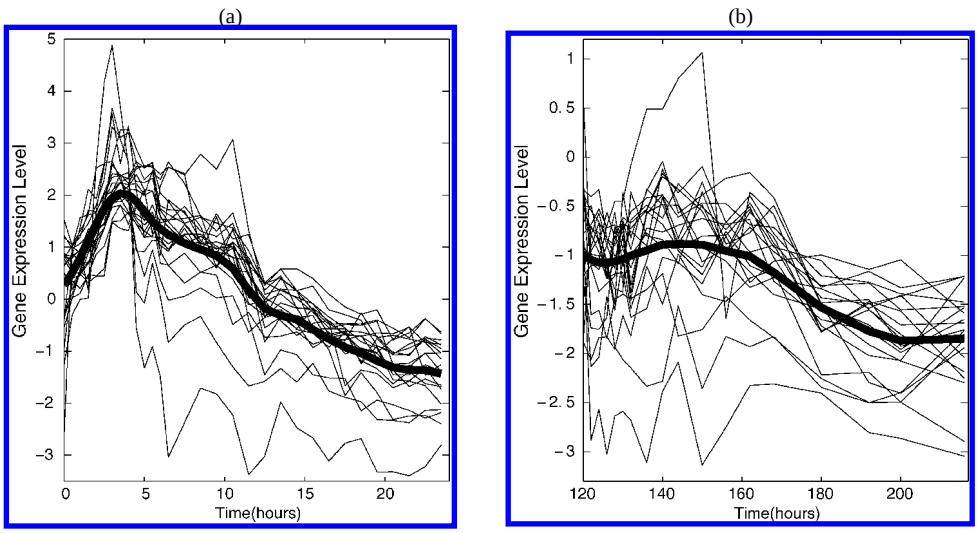

Figure 1. Observed (thin curves) gene expression levels and smoothed estimates of the mean functions (thick curves) for embryo phase (a) and pupa phase (b), for zygotic data.

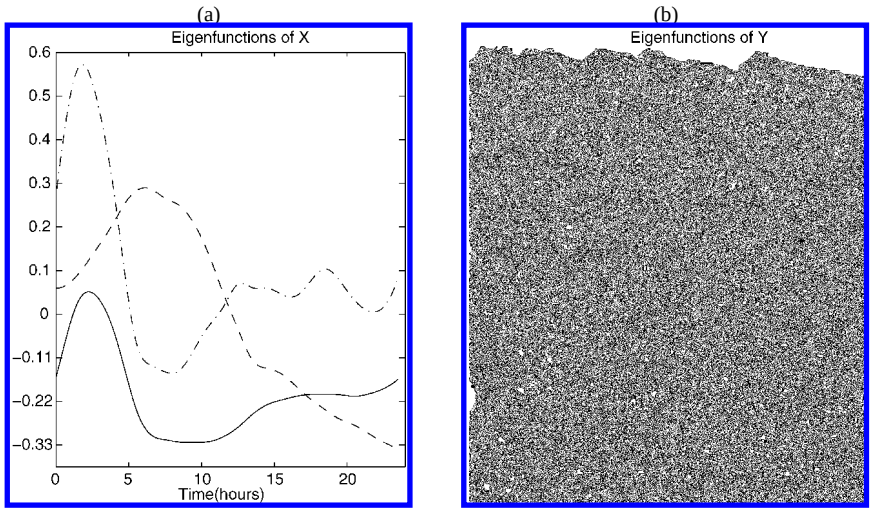

Figure 2. (a) Estimates of the first three eigenfunctions for predictor trajectories from the embryo phase (predictor process), $K = 3$ components selected by AIC, accounting for 80.69% (first component; —), 13.65% (second component; − −), and 2.71% (third component; − · − ·) of the total variation. (b) Estimates of the first 4 eigenfunctions of the pupa (metamorphosis) phase (response process), $M = 4$ components selected by AIC, accounting for 80.47% (first component; —), 8.89% (second component; − −), 4.2% (third component; − · − ·), and 1.55% (fourth component; · · · ·) of the total variation.

initial peaks and lower late embryonal gene expression is coupled with an overall lower pupa phase expression, in the sense that the contribution of the first eigenfunction in the response is reduced.

Interpretation of FAM can be aided by a "principal response plot," displayed in Figure 4, where one predictor FPC score is varied while the others are kept fixed at level 0. This can be visualized by looking at a set of predictor functions moving in a certain direction "away" from the mean function of the predictor process, where the direction depends on the chosen eigenfunction; each of these predictor functions, when plugged into FAM, then generates a corresponding response function.

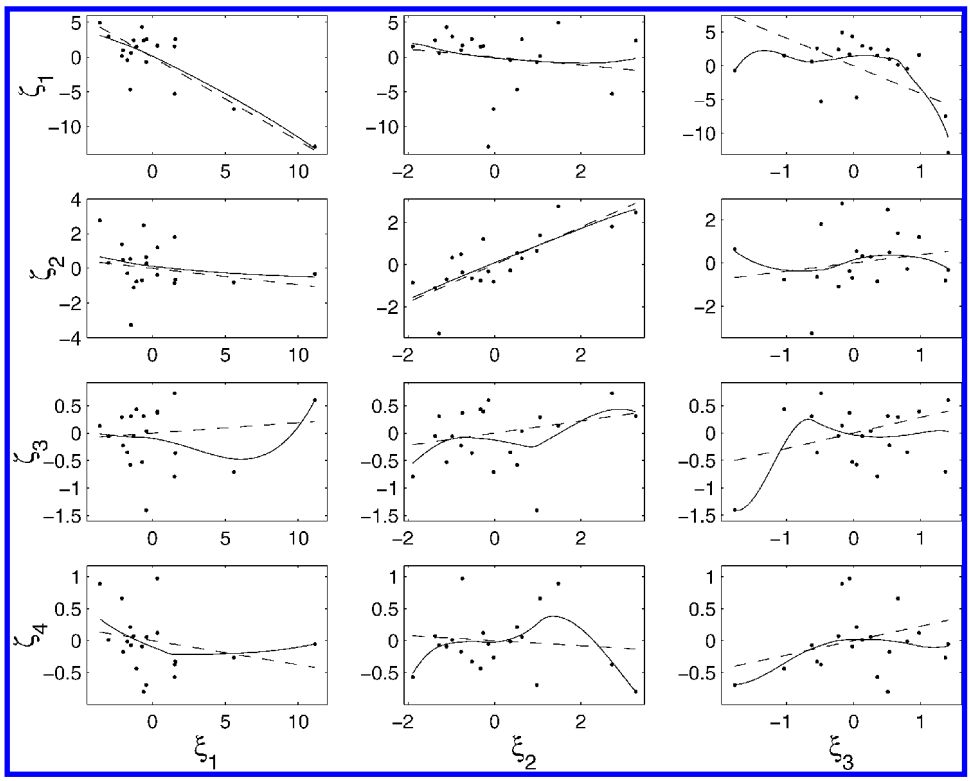

Figure 3. Scatterplots (·), local polynomial (—) and linear (− −) estimates for the regressions of estimated FPC scores of the pupa phase (responses, y-axis) versus those for the embryo phase (predictors, x-axis). The FPC scores of the embryo phase expressions are arranged from left to right ($_k, k = 1, 2, 3$), and the FPC scores of the pupa phase expressions are arranged from top to bottom ($\zeta_m, m = 1, 2, 3, 4$), for the zygotic data.

The set of these response functions is then jointly visualized with the set of predictor functions in such a way that the corresponding predictor/response pairs can be easily identified. This device can be used for each predictor score. Because these act independently from each other on the response, displaying a series of such plots for all relevant predictor scores then provides a graphical representation of the FAM.

The left panels of Figure 4 indicate that shifts in the levels on the right side of the peak of the predictor curves are associated with broad shifts up or down in responses, the middle panels indicate that the size of peak expression in predictors is associated with amplitude shifts in the right half of response curves, and the right panels indicate that combined time and amplitude shifts of predictor peaks are associated with strong amplitude shifts in responses in a nonlinear way. The principal response plots thus characterize the response changes induced by the modes of variation of the predictors. For proper interpretation, it is helpful to note that the actual sample variation of the predictor scores in the direction of each eigenfunction, as depicted in the top panels of Figure 4, depends on the size of the respective eigenvalue, which corresponds to the variance of the FPC scores. This variation accordingly is much smaller for the second or third eigenfunction, compared with the first eigenfunction, and thus the system's response function changes, as depicted in the lower panels, will be realized on increasingly smaller scales for real functional data as the order of the eigenfunction increases.

The increased bias when fitting functional linear regression to these data is evident in the observed leave-one-subject-out PREs, that is, the cross-validated observed version of (22), denoted by $\mathrm{RPE}_{(-i),f}$; mean and percentiles are listed in Table 3. The median and mean of the errors are larger for the functional linear model compared with the FAM. This finding is in line with the increase in functional R^2 (17) obtained for the FAM compared with the functional linear regression model.

Table 3. Functional R^2 (17), 25th, 50th and 75th percentiles and mean of the cross-validated observed relative prediction errors, $\mathrm{RPE}_{(-i),f}$ (22), comparing FAM and functional linear regression models for zygotic data

	25th	50th	75th	Mean	R^2
FAM	.0506	.0776	.1662	.1301	.19
LIN	.0479	.0891	.1727	.1374	.16

8. A <

The FAM strikes a fine balance between greatly enhancing flexibility, compared with the functional linear model, and preventing the curse of dimension incurred by a fully nonparametric approach, due to its sensible structural constraints. This is analogous to the situation in ordinary multiple regression, where additive models do not suffer from the curse of dimension in the way in which unstructured nonparametric models do. As a reviewer has pointed out, one also could imagine other additive regression models geared toward specific regression relations, where the predictors correspond to suitably cho-

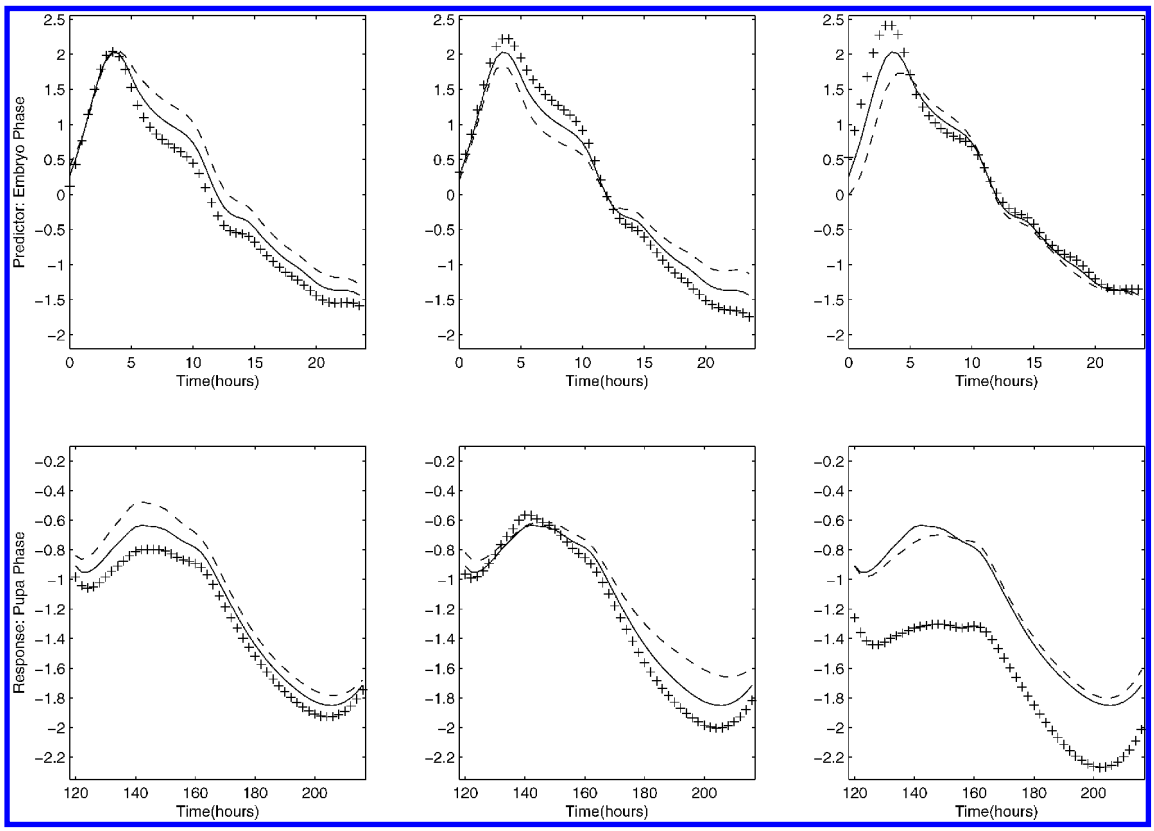

Figure 4. Top panels: Predictor trajectories $\hat{}_X(s) + \alpha\,\hat{}_k(s)$, for $\alpha = 0$ (—), $\alpha = 1$ (−−) and $\alpha = -1$ (++) for $k = 1$ (left panels), $k = 2$ (middle panels), and $k = 3$ (right panels). Bottom panels: Corresponding response trajectories $\hat{}_Y(t) + \sum_{m=1}^{M}\{\hat{f}_{km}(\alpha) + \sum_{\ell\neq k}\hat{f}_{\ell m}(0)\}\,\hat{}_m(t)$, for the zygotic data.

sen functionals of the predictor processes other than the FPC scores.

As we have demonstrated, specific advantages of using FPC scores in an additive regression model are that (a) such a model is a strict generalization of the functional linear model, which emerges as a special case; (b) asymptotic consistency results still can be derived under truly nonparametric (smoothness) assumptions for all component functions, including mean and covariance functions; and (c) implementing the resulting additive model requires no more than applying simple smoothing steps. The necessary smoothing parameters for the component functions can be easily selected in a data-adaptive way. The FAM does not impose a largely increased computational burden over the established functional linear regression model, and thus the added flexibility comes at a low additional computational cost. Judging from our simulations, the loss in efficiency against the functional linear regression model when the underlying functional regression is truly linear is quite limited. On the other hand, the increased flexibility of the FAM can lead to substantial gains.

A

A.1

The eigenfunctions $_k, k = 1, 2, \ldots$, for the representation of predictor processes X are the solutions of the equations $(A_{G_X} f)(s) = f(s)$ on the space of square-integrable functions $f \in L^2(\mathcal{S})$ under constraints, where the autocovariance operator $(A_{G_X} f)(t) = \int f(s) \times G_X(s,t)\, ds$ is a compact linear integral operator of Hilbert–Schmidt type (Ash and Gardner 1975). The constraints correspond to orthonormality of the eigenfunctions, that is, $\int _j(s)\ _k(s)\, ds = \delta_{jk}$, where $\delta_{jk} = 1$ if $j = k$ and $= 0$ if $j \neq k$. The eigenfunctions $_j$ are ordered according to the size of the corresponding eigenvalues, $_1 \geq _2 \geq \cdots$. Analogously, eigenfunctions and eigenvalues of the response process Y are denoted by $_m$ and $_m$. We assume that all functions $\{_X, _j\}$ and $\{_Y, _k\}$ are smooth (twice continuously differentiable).

We then have well-known representations $G_X(s_1, s_2) = \sum_k _k \times _k(s_1)\ _k(s_2)$ and $G_Y(t_1, t_2) = \sum_m _m\ _m(t_1)\ _m(t_2)$ of the covariance functions of X and Y, as well as the Karhunen–Loève expansions for processes X and Y,

$$\begin{aligned} X(s) &= _X(s) + \sum_{j=1}^{\infty} _j\ _j(s) \qquad \text{and} \\ Y(t) &= _Y(t) + \sum_{k=1}^{\infty} \zeta_k\ _k(t). \end{aligned} \tag{A.1}$$

Because $\{_k, k = 1, 2, \ldots\}$ and $\{_m, m = 1, 2, \ldots\}$ form orthonormal bases of the respective space of square-integrable functions, it follows that the regression parameter functions also can be represented on this basis; that is, there exist coefficients β_k and β_{km} such that $\beta(s) = \sum_{k=1}^{\infty} \beta_k\ _k(t)$ and $\beta(s,t) = \sum_{k=1}^{\infty}\sum_{m=1}^{\infty} \beta_{km}\ _k(s)\ _m(t)$.

According to (A.1), we may represent the measurements for predictor trajectories in (1) and both the predictor and response trajectories in (2) as

$$\begin{aligned} U_{ij} &= X_i(s_{ij}) + \epsilon_{ij} \\ &= _X(s_{ij}) + \sum_{k=1}^{\infty} _{ik}\ _k(s_{ij}) + \epsilon_{ij}, \\ & s_{ij} \in \mathcal{S}, 1 \leq i \leq n, 1 \leq j \leq n_i, \end{aligned} \tag{A.2}$$

$$\begin{aligned} V_{il} &= Y_i(t_{il}) + \varepsilon_{il} \\ &= _Y(t_{il}) + \sum_{m=1}^{\infty} \zeta_{im}\ _k(t_{il}) + \varepsilon_{il}, \\ & s_{il} \in \mathcal{T}, 1 \leq i \leq n, 1 \leq l \leq m_i. \end{aligned} \tag{A.3}$$

More specifically, writing $\mathbf{U}_i = (U_{i1}, \ldots, U_{in_i})^T$, $_{X_i} = (_X(s_{i1}), \ldots, _X(s_{in_i}))^T$, and $_{ik} = (_k(s_{i1}), \ldots, _k(s_{in_i}))^T$, the best linear prediction for $_{ik}$ is $_k\ _{ik}^T \mathbf{\Sigma}_{U_i}^{-1}(\mathbf{U}_i - _{X_i})$, where the (j,l) entry of the $n_i \times n_i$ matrix $\mathbf{\Sigma}_{U_i}$ is $(\mathbf{\Sigma}_{U_i})_{j,l} = G_X(s_{ij}, s_{il}) + _X^2 \delta_{jl}$, with $\delta_{jl} = 1$ if $j = l$ and $\delta_{jl} = 0$ if $j \neq l$. Then the estimates for the scores $_{ik}$ are obtained by substituting estimates for $_{X_i}$, $_k$ and $_{ik}$, $\mathbf{\Sigma}_{X_i}$ (see the Supplement) obtained from the entire data ensemble, leading to

$$\hat{}_{ik} = \hat{}_k\ \hat{}_{ik}^T \widehat{\mathbf{\Sigma}}_{U_i}^{-1}(\mathbf{U}_i - \hat{}_{X_i}), \tag{A.4}$$

where the (j,l) element of $\widehat{\mathbf{\Sigma}}_{U_i}$ is $(\widehat{\mathbf{\Sigma}}_{U_i})_{j,l} = \widehat{G}_X(s_{ij}, s_{il}) + \hat{}_Y^2(s_{ij})\delta_{jl}$. We note that it follows from results of Müller (2005) that as designs become dense, these best linear (PACE) estimates $\hat{}_{ik}, \hat{\zeta}_{im}$ (A.4) converge to those obtained by the more traditional integration-based estimates,

$$\begin{aligned} \hat{}_{ik}^I &= \sum_{j=2}^{n_i} (U_{ij} - \hat{}_X(t_{ij}))\ \hat{}_k(s_{ij})(s_{ij} - s_{i,j-1}), \\ \hat{\zeta}_{im}^I &= \sum_{j=2}^{n_i} (V_{ij} - \hat{}_Y(t_{ij}))\ \hat{}_k(t_{ij})(t_{ij} - t_{i,j-1}), \end{aligned} \tag{A.5}$$

which are motivated by the definition of the FPC scores as inner products, $_{ik} = \int \{X_i(s) - _X(s)\}\ _k(s)\, ds$ and $\zeta_{im} = \int \{Y_i(t) - _Y(t)\}\ _m(t)\, dt$. Thus, the PACE estimates and the estimates based on integral approximations can be considered equivalent in the dense design case that we consider here.

A.2

To obtain the FPC scores for predictor and response processes (in case of a functional response), we adopt the PACE methodology (Yao et al. 2005a). Estimating the predictor mean function $_X$ by local linear scatterplot smoothers, we minimize

$$\sum_{i=1}^{n}\sum_{j=1}^{n_i} K_1\left(\frac{s_{ij} - s}{b_X}\right)\{U_{ij} - \beta_0^X - \beta_1^X(s - s_{ij})\}^2 \tag{A.6}$$

with respect to β_0^X and β_1^X, to obtain $\hat{}_X(s) = \hat{\beta}_0^X(s)$. The kernel K_1 is assumed to be a smooth symmetric density function, and b_X is a bandwidth. Estimating the covariance function $G_X(s_1, s_2)$ of predictor processes X, we define $\mathrm{G}_i^X(s_{ij_1}, s_{ij_2}) = (U_{ij_1} - \hat{}_X(s_{ij_1}))(U_{ij_2} - \hat{}_X(s_{ij_2}))$ and the local linear surface smoother by minimizing

$$\begin{aligned} &\sum_{i=1}^{n}\sum_{1 \leq j_1 \neq j_2 \leq n_i} K_2\left(\frac{s_{ij_1} - s_1}{h_X}, \frac{s_{ij_2} - s_2}{h_X}\right) \\ &\qquad \times \left\{\mathrm{G}_i^X(s_{ij_1}, s_{ij_2}) - f\left(\beta^X, (s_1, s_2), (s_{ij_1}, s_{ij_2})\right)\right\}^2, \end{aligned} \tag{A.7}$$

where $f(\beta^X, (s_1, s_2), (s_{ij_1}, s_{ij_2})) = \beta_0^X + \beta_{11}^X(s_1 - s_{jl_1}) + \beta_{12}^X(s_2 - s_{ij_2})$ with respect to $_X = (\beta_0^X, \beta_{11}^X, \beta_{12}^X)^T$, yielding $\widehat{G}_X(s_1, s_2) = \hat{\beta}_0^X(s_1, s_2)$. Here the kernel K_2 is a two-dimensional smooth density with mean 0 and finite covariances, and h_X is a bandwidth. An essential feature is the omission of the diagonal elements $j_1 = j_2$, which are contaminated with the measurement errors.

To estimate the variance of the measurement error $_X^2$, we fit a local quadratic component orthogonal to the diagonal of G_X and a local linear component in the direction of the diagonal. We denote the diagonal of the resulting surface estimate by $\widetilde{G}_X(s)$ and a local linear smoother focusing on diagonal values $\{G_X(s,s) + _X^2\}$ by $\widehat{V}_X(s)$,

using bandwidth h_X^*. We let $a_0 = \inf\{s : s \in \mathcal{S}\}$, $b_0 = \sup\{s : s \in \mathcal{S}\}$, $|\mathcal{S}| = b_0 - a_0$, and $\mathcal{S}_1 = [a_0 + |\mathcal{S}|/4, b_0 - |\mathcal{S}|/4]$. The estimate of $\frac{2}{X}$ is

$$\hat{}_X^2 = 2 \int_1 \{\widehat{V}_X(s) - \widetilde{G}_X(s)\}\, ds/|\mathcal{S}| \tag{A.8}$$

if $\hat{}_X^2 > 0$, and $\hat{}_X^2 = 0$ otherwise. Estimates of eigenvalues/eigenfunctions $\{\ _k,\ \ _k\}_{k\geq 1}$ are obtained as numerical solutions $\{\hat{}_k, \hat{}$

To guarantee the consistency of predictions, we also require the following for fixed $_k$ and $t \in \mathcal{T}$:

(A7) $\sum_{k=1}^{K}[|f_k''(\ _k)|h_k^2 + n^{-1/2}\{\mathrm{var}(Y|\ _k)\}^{1/2}p_k^{-1/2}(\ _k) \times h_k^{-1/2}] \to 0$

(B6) $\sum_{k=1}^{K}\sum_{m=1}^{M}[|f_{mk}''(\ _k)\ _m(t)|h_{mk}^2 + n^{-1/2}\{\mathrm{var}(\zeta_m|\ _k)\}^{1/2} \times p_k^{-1/2}(\ _k)|\ _m(t)|h_{mk}^{-1/2}] \to 0.$

[Received August 2007. Revised June 2008.]

Arbeitman, M. N., Furlong, E. E. M., Imam, F., Johnson, E., Null, B. H., Baker, B. S., Krasnow, M. A., Scott, M. P., Davis, R. W., and White, K. P. (2002), "Gene Expression During the Life Cycle of *Drosophila melanogaster*," *Science*, 297, 2270–2275.

Ash, R. B., and Gardner, M. F. (1975), *Topics in Stochastic Processes*, New York: Academic Press.

Billingsley, P. (1995), *Probability and Measure* (2nd ed.), New York: Wiley.

Cai, T., and Hall, P. (2006), "Prediction in Functional Linear Regression," *The Annals of Statistics*, 34, 2159–2179.

Cardot, H., and Sarda, P. (2005), "Estimation in Generalized Linear Models for Functional Data via Penalized Likelihood," *Journal of Multivariate Analysis*, 92, 24–41.

Cardot, H., Ferraty, F., Mas, A., and Sarda, P. (2003), "Testing Hypotheses in the Functional Linear Model," *Scandinavian Journal of Statistics*, 30, 241–255.

Cuevas, A., Febrero, M., and Fraiman, R. (2002), "Linear Functional Regression: The Case of Fixed Design and Functional Response," *Canadian Journal of Statistics*