Published in final edited form as:

J R Stat Soc Ser C Appl Stat. 2020 January ; 69(1): 25–46. doi:10.1111/rssc.12376.

Longitudinal dynamic functional regression

Ana-Maria Staicu,

North Carolina State University, Raleigh, USA

Md Nazmul Islam,

North Carolina State University, Raleigh, USA

Raluca Dumitru,

University of North Florida, Jacksonville, USA

Eric van Heugten

North Carolina State University, Raleigh, USA

Summary.

The paper develops a parsimonious modelling framework to study the time-varying association between scalar outcomes and functional predictors observed at many instances, in longitudinal studies. The methods enable us to reconstruct the full trajectory of the response and are applicable to Gaussian and non-Gaussian responses. The idea is to model the time-varying functional predictors by using orthogonal basis functions and to expand the time-varying regression coefficient by using the same basis. Numerical investigation through simulation studies and data analysis show excellent performance in terms of accurate prediction and efficient computations, when compared with existing alternatives. The methods are inspired and applied to an animal science application, where of interest is to study the association between the feed intake of lactating sows and the minute-by-minute temperature throughout the 21 days of their lactation period. R code and an R illustration are provided.

Keywords

Functional data; Functional principal component analysis; Longitudinal study; Longitudinal functional regression; Penalization; Time-varying coefficient model

Address for correspondence: Ana-Maria Staicu, Department of Statistics, North Carolina State University, 2311 Stinson Drive, Raleigh, NC 27665-8203, USA. astaicu@ncsu.edu.

⁸Supplementary material

The on-line supplementary material contains technical arguments that are needed to justify the finite truncation that is used in the linear predictor as well as to calculate the conditional variance that is used in the construction of the prediction bands, additional numerical results with our simulation study and data analysis. Moreover, the R code for implementation of the proposed framework is posted publicly at <http://www4.stat.ncsu.edu/~staicu/software/LDFR.zip>. The fitting methodology is illustrated at http://www4.stat.ncsu.edu/~staicu/software/illustration_LDFR.html by using a generated data set.

Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article: ‘Supplementary material of “Longitudinal dynamic functional regression”’.

1. Introduction

Functional regression has attracted much interest in recent years; see Cai and Hall (2006), Cardot et al. (1999, 2003), Fan and Zhang (2000), Ivanescu et al. (2015), Morris and Carroll (2006), Müller (2005), Ramsay and Silverman (1997), Reiss and Ogden (2007) and Scheipl et al. (2015) to name a few. In this paper we consider longitudinal scalar-on-function regression for scalar outcomes and functional predictors observed repeatedly. This research is motivated by an animal science study of the effect of daily ambient air temperature on feed intake of sows during their lactation period. To be specific, a number of sows are observed for several days during their lactation period and, for each day, the total daily feed intake as well as the minute-by-minute daily temperature for the respective day are recorded. Fig. 1 illustrates the data for a randomly chosen sow: daily feed intake for each lactation day (Fig. 1(a)) and temperature profiles (Fig. 1(b)). Using the methods proposed we can predict the entire feed intake behaviour of a sow during her lactation period, based on the feed intake and the daily temperature profile recorded for only part of the duration of lactation.

Functional linear models for scalar-on-function regression are a popular regression model and assume that the effect of the functional predictor is captured by the integral of the predictor weighted by a smooth regression coefficient. Three estimation approaches are quite common: both the functional predictor and the smooth coefficient are expanded by using the empirical eigenbasis of the predictors covariance (Cardot et al., 1999); both the functional predictor and the smooth coefficient are expanded by using a B -spline basis and penalties are employed to control the smoothness of the parameter function (Ramsay and Silverman, 1997); or a mixture of these approaches, i.e. a predefined basis function is used to represent the smooth parameter, the empirical eigenbasis of the predictors covariance is used to expand the functional predictors and in addition penalties are used to control the smoothness of the parameter function (Cardot et al., 2003; Goldsmith et al., 2011). Extensions of these approaches to accommodate additional covariates or more flexible relationships have been discussed previously by Cardot and Sarda (2008), McLean et al. (2014), Morris and Carroll (2006), Müller and Yao (2012) and Ramsay and Silverman (2005).

The popular scalar-on-function regression model has been extended to analyse longitudinal functional data where the scalar response and the functional predictor are observed repeatedly for each subject. Goldsmith et al. (2012) introduced longitudinal penalized functional regression (LPFR), which assumes that the effect of the functional predictor is constant over time. They modelled the time invariant regression coefficient via the truncated polynomial basis and used a penalized-based approach for estimation. In a similar spirit, Gertheiss et al. (2013) introduced longitudinal functional principal component regression, where the response is regressed onto the functional principal components of the various structures that compose the functional predictors. The main limitation of these methods is the assumption that the effect of the predictor is invariant over time. Such an assumption may be viewed as strong and unrealistic in some situations. For example, in the motivating animal science study, a lactating sow's body can adjust to prolonged exposure to heat and, thus, the effect of heat on their feed intake behaviour is expected to change throughout their

lactation period. Therefore, assuming a time-varying relationship between temperature and feed intake is more appropriate.

Recently, Kundu et al. (2016) considered a time-varying functional coefficient in this longitudinal functional model framework (called ‘LPEER’) with the focus to capture a more complex time-varying effect. They focused on time-varying effects that exhibit smoothness properties on some subdomains and spiky behaviour on other subdomains; in contrast we focus on time-varying effects that are smooth across the entire domain. Kundu et al. (2016) modelled the bivariate regression coefficient as a linear combination of *user-specified* time-dependent basis functions with unknown time invariant coefficient functions. Whereas the time invariant functions are estimated via a penalty operator, which is informed by the scientific context of the problem and intended to recover spiky and smooth trends, the time-dependent basis functions are chosen arbitrarily and their number is selected by using the Akaike information criterion or pointwise confidence intervals for the time invariant coefficient functions. In their numerical study, Kundu et al. (2016) used two temporal functions $f_1(t) = 1$ and $f_2(t) = t$; the methodology is implemented solely for polynomial temporal functions. In general, it is not clear how to select an optimal temporal basis. In addition, the methodology is limited to Gaussian responses and it does not seem trivial to extend it to non-Gaussian cases. Our numerical experience with this method indicated that, when applicable and for smooth effects, the approach is quite computationally expensive. Furthermore, like LPFR and longitudinal functional principal component regression, this method also does not consider reconstruction of the full response trajectory, which is often a major goal in longitudinal studies involving a repeatedly measured response.

In this work, our goal is to develop a longitudinal dynamic functional regression (LDFR) framework to study time-varying association between Gaussian or non-Gaussian responses and functional covariates that are observed in a longitudinal design that enables us to recover the full response trajectory. There are three key novelties of this paper. The first is the model proposed for both Gaussian and non-Gaussian longitudinal responses. The second is the use of a combination of ideas from functional data analysis and longitudinal data analysis that has important advantages: it enables us to tackle a challenging problem that has not previously been solved in this generality, it takes into account the dependence between the functional predictors and it enables prediction of the full trajectory of the response. The third novelty is the way that the smoothness of the time-varying regression coefficient is modelled, by using a mixture of two common approaches in non-parametric regression. The main advantages of this approach are

- a. it enjoys a parsimonious modelling framework,
- b. it provides excellent numerical performance in terms of prediction accuracy,
- c. it is computationally efficient and
- d. it can be easily implemented by using well-developed freely available software.

In Section 2 we introduce the proposed methodology for responses in the exponential family. Section 3 describes the estimation procedure. Section 4 details the response trajectory

prediction. The methods are applied to the sow data in Section 5 and are investigated in simulation studies in Section 6. Finally, Section 7 concludes our paper.

2. Methodology proposed

Let the observed data be $[Y_{ij}, t_{ij}, \{(W_{ijr}, s_{ijr}): r=1, \dots, r_{ij}\}; i=1, \dots, I, j=1, 2, \dots, n_i]$, where i indexes the subject, j indexes the repeated observations, Y_{ij} is the response measured at time t_{ij} and W_{ijr} are the noisy functional covariates that are observed at points s_{ijr} . It is assumed that $t_{ij} \in \mathcal{T}$ and $s_{ijr} \in \mathcal{S}$ for closed and compact sets $\mathcal{T}$ and $\mathcal{S}$ respectively. We assume that, for each i and j , r_{ij} is large, and furthermore that the set $\{s_{ij1}, \dots, s_{ijr_{ij}}\}$ is dense in $\mathcal{S}$. Also we assume that, although n_i may be small for all i , the set $\{t_{ij}: j=1, \dots, n_i, i=1, \dots, I\}$ is dense in $\mathcal{T}$. To distinguish between the two sampling designs, we refer to the former by a *dense* functional design, and to the latter by a *sparse* longitudinal design. We consider that $W_{ijr} = \tau(s_{ijr}, t_{ij}) + X_i(s_{ijr}, t_{ij}) + \epsilon_{ij}^W(s_{ijr})$, where $\tau(\cdot, t_{ij})$ is the overall mean function at visit time t_{ij} , $X_i(\cdot, t_{ij})$ is the i th subject-specific latent signal corresponding to time t_{ij} and $\epsilon_{ij}^W(\cdot)$ is the j th visit-specific residual from the i th subject-specific deviation, $X_i(\cdot, t_{ij})$. It is assumed that $X_i(\cdot, \cdot)$ s are zero-mean smooth processes over $\mathcal{S} \times \mathcal{T}$ and independent and identically distributed (IID) over i ; also that $\epsilon_{ij}^W(\cdot)$ s are zero-mean processes, IID over both i and j . The structure of $\epsilon_{ij}^W(\cdot)$ is discussed in more detail in Section 2.2. For identifiability, it is assumed that the two processes $X_i(\cdot, \cdot)$ and $\epsilon_{ij}^W(\cdot)$ are mutually independent. For convenience we consider a regular design in $\mathcal{S}$, i.e. $s_{ijr} = s_r$ for all i and j , but the ideas are not restricted to this assumption. The implementation could be easily modified to account for irregular dense grids $\{s_{ijr}: r=1, \dots, r_{ij}\}$.

Our objective is to develop a general modelling framework to regress time-varying responses $Y_{ij} = Y_i(t_{ij})$ onto time-varying functional covariates observed with noise that allows reconstruction of the response trajectory $Y_i(\cdot)$ and that accommodates non-normal responses. Denote by $\text{EF}\{\mu_i(t_{ij}), \eta\}$ the exponential family with mean $\mu_{ij} = \mu_i(t_{ij})$ for some smooth unknown function $\mu(\cdot)$ and dispersion parameter η . We propose LDFR models that account for a time-varying smooth effect:

$$Y_{ij} | t_{ij}, X_i(\cdot, t_{ij}) \sim \text{EF}\{\mu_i(t_{ij}), \eta\},$$

$$g\{\mu_i(t_{ij})\} = \alpha(t_{ij}) + \int_{\mathcal{S}} X_i(s, t_{ij}) \gamma(s, t_{ij}) ds + Z_{b,ij} b_i, \quad (1)$$

where $g(\cdot)$ is a known, monotone link function, $\alpha(\cdot)$ is an unknown intercept function and $\gamma(\cdot, t_{ij})$ is an unknown regression coefficient function that quantifies the time-varying effect of $X_i(\cdot, t_{ij})$ on the conditional mean response of Y_{ij} at time t_{ij} . Also b_i is a subject-specific q -dimensional vector and $Z_{b,ij}$ is its associated $(1 \times q)$ -dimensional random-design matrix. It is assumed that b_i are IID as $N_q(\mathbf{0}_q, D)$ where $\mathbf{0}_q$ is the q -dimensional vector of 0s and D is a $q \times q$ covariance matrix. Both the intercept function $\alpha(\cdot)$ and the regression coefficient function $\gamma(\cdot, \cdot)$ are assumed smooth. Just like in the functional linear model, the regression coefficient $\gamma(\cdot, t)$ in model (1) is only identifiable up to the space that is spanned by the $X_i(\cdot, t)$ s. For example, if the space that is spanned by $X_i(\cdot, t)$ is finite and $h(\cdot, t)$ is some bivariate

function lying in the orthogonal complement of this space for every t , then both $\gamma(\cdot, t)$ and $\gamma(\cdot, t) + h(\cdot, t)$ yield the same association with the response. Nevertheless, the integral term is identifiable. In contrast, if $X(\cdot, t)$ spans the whole space for a dense set of t s then $\gamma(\cdot, \cdot)$ is identifiable. Model (1) was introduced in Kundu et al. (2016) for Gaussian responses; here we extend the framework to accommodate also non-Gaussian responses that are in the exponential family. For Gaussian responses, our model framework is different from that of Kundu et al. (2016) in the assumptions that are made on the time-varying effect, $\gamma(\cdot, \cdot)$. Kundu et al. (2016) assumed that the unknown function $\gamma(s, t)$ has a simple dependence of t but a very complex dependence of s . The unknown function $\gamma(s, \cdot)$ is assumed to be accurately approximated by a linear combination of *known temporal functions*, which are arbitrarily specified, e.g. $\gamma(s, t) = \gamma_0(s) + t\gamma_1(s) + t^2\gamma_2(s)$ with unknown functions $\gamma_0(\cdot)$, $\gamma_1(\cdot)$ and $\gamma_2(\cdot)$. In contrast, we impose no such restrictions and allow the unknown bivariate function $\gamma(\cdot, \cdot)$ to vary smoothly. We propose a parsimonious modelling framework and estimation methodology that enable the reconstruction of the response trajectory, has competitive prediction performance and is computationally efficient.

2.1. Statistical modelling

Our modelling approach consists of two parts. First, let $\{\phi_k(\cdot): k \geq 1\}$ be a time invariant orthonormal basis in $L^2(\mathcal{S})$, i.e. $\int_{\mathcal{S}} \phi_k(s)\phi_{k'}(s)ds$ equals 1 if $k = k'$ and 0 otherwise. We consider this basis to represent both the unknown time-varying coefficient $\gamma(s, t)$ and the subject-specific latent signal, $X(s, t)$. We write $\gamma(s, t) = \sum_{k=1}^{\infty} \beta_k(t)\phi_k(s)$, where $\beta_k(\cdot)$ is an unknown smooth function of time and defined uniquely by $\beta_k(t) = \int_{\mathcal{S}} \gamma(s, t)\phi_k(s)ds$. As well we write $X(s, t) = \sum_{k \geq 1} \xi_{ik}(t)\phi_k(s)$, where $\xi_{ik}(t) = \int_{\mathcal{S}} X_i(s, t)\phi_k(s)ds$ are zero-mean smooth processes over t . It follows that $\int_{\mathcal{S}} X_i(s, t)\gamma(s, t)ds \approx \sum_{k=1}^K \xi_{ik}(t)\beta_k(t)$; for practical purposes the infinite summation is truncated at some finite level, say K . The choice of K is affected by the type of orthogonal basis that is used and is discussed in Section 2.2. Let $\xi_{ijk} = \xi_{ik}(t_{ij})$ by an abuse of notation; the linear predictor in model (1) is approximated by $g(\mu_{ij}^K)$, where

$$g(\mu_{ij}^K) = \alpha(t_{ij}) + \sum_{k=1}^K \xi_{ijk}\beta_k(t_{ij}) + Z_{b,ij}b_i. \quad (2)$$

In this framework, we propose to describe the smoothness of $\gamma(\cdot, \cdot)$ by combining two popular approaches in non-parametric regression. The truncation K controls the degree of smoothness of the estimate of $\gamma(s, t)$ in the s -direction: a small value of K results in an overly smooth function in this direction, whereas a large value gives a wiggly behaviour. To characterize the smoothness of the estimate of γ in the t -direction, we use a different approach, based on the total curvature. Assuming that $\gamma(\cdot, \cdot)$ is a second-order differentiable function, we describe the smoothness in the t -direction by the total curvature along this direction, $\|\gamma(\cdot, \cdot)/t^2\|^2 = \int \int \{\gamma(s, t)/t^2\}^2 ds dt$, and use this quantity to regularize a goodness-of-fit criterion. We propose to estimate $\gamma(\cdot, \cdot)$ by minimizing the penalized criterion

$$-2 \sum_{i=1}^I \sum_{j=1}^{n_i} \log \{f(Y_{ij} | \xi_{ij1}, \dots, \xi_{ijK}, b_i)\} - 2 \sum_{i=1}^I \log \{f_b(b_i)\} + \lambda_0 \|\alpha''(\cdot)\|^2 + \lambda \|\partial^2 \gamma(\cdot, \cdot)\|^2 / \partial t^2 \|^2$$

where $f(Y_{ij}|b_i)$ is the density that is specified by the exponential family model (1) and approximated via truncation K by equation (2), and $f_b(b_i)$ is the density that is assumed for the random terms b_i . In other words, the penalized criterion is approximated by

$$\begin{aligned} & -2 \sum_{i=1}^I \sum_{j=1}^{n_i} \log \{f(Y_{ij} | \xi_{ij1}, \dots, \xi_{ijK}, b_i)\} - 2 \sum_{i=1}^I \log \{f_b(b_i)\} + \lambda_0 \|\alpha''(\cdot)\|^2 \\ & + \lambda \sum_{k=1}^K \|\beta_k''(\cdot)\|^2, \end{aligned} \quad (3)$$

as $\|\partial^2 \gamma(\cdot, \cdot) / \partial t^2\|^2 \approx \sum_{k=1}^K \|\beta_k''(\cdot)\|^2$ because of the orthogonality of the basis $\phi_k(\cdot)$ s.

Typically $f_b(b) = \exp(-b^T D^{-1} b/2)$, corresponding to a multivariate normal distribution with mean vector $\mathbf{0}$ and variance-covariance matrix D , and ignoring the multiplicative term $\det(2\pi D)^{-1/2}$ where $\det(D)$ is the determinant of D . Here $\lambda_0 > 0$ is a smoothing parameter that controls the amount of smoothing of the unknown intercept function $\alpha(\cdot)$ and $\lambda > 0$ is an unknown parameter that controls the smoothness of $\gamma(s, t)$ in the direction t , relative to the goodness of fit, quantified by the likelihood term. The penalized criterion (3) essentially controls the smoothness of the unknown bivariate function $\gamma(\cdot, \cdot)$ through two parameters— K and λ —but in a distinct way from common practice (Ruppert et al., 2003; Wood, 2006a, b). Combining different approaches to control the smoothness of a multivariate function is inspired from Kim et al. (2017) in the context of function-on-function regression.

If the orthogonal basis $\{\phi_k(\cdot)\}_k$ and $\xi_{ij1}, \dots, \xi_{ijK}$ were known, minimizing expression (3) has been studied in the literature under the topic of time-varying coefficient regression (Hastie and Tibshirani, 1993; Hoover et al., 1998). In this regard, we model $\alpha(\cdot)$ and $\beta_k(\cdot)$ as linear combinations of known basis functions, say the truncated power basis, B -splines, etc. Let $\alpha(t) = B_0(t)^T \beta_0$ and $\beta_k(t) = B_k(t)^T \beta_k$, where the $\{B_0(t)\}_s$ and $\{B_k(t)\}_s$ are finite dimensional bases functions and β_0 and the β_k s are the vector of basis coefficients for $\alpha(t)$ and $\beta_k(t)$ respectively. The norm of the second derivatives, $\|\alpha''(\cdot)\|^2$ and $\|\beta_k''(\cdot)\|^2$, can be written as quadratic forms in the basis coefficients, $\beta_0^T D_0 \beta_0$ and $\beta_k^T D_k \beta_k$ respectively, where D_0 and D_k are known squared positive semidefinite matrices. Having the same tuning parameter λ for all the penalties $\|\beta_k''(\cdot)\|^2$ essentially implies that the functions $\beta_k(\cdot)$ have the same smoothness property. Given λ_0 and λ there is a unique minimizer of expression (3), and it is usually computed by a penalized version of iteratively reweighted least squares (see Wood (2006a), chapter 3.2). The tuning parameters can be selected by using generalized cross-validation or a variant of restricted maximum likelihood (REML).

A more appealing approach to study this approximate model is to use a Bayesian framework. As in Wahba (1983) and Silverman (1985) it is convenient to reformulate the penalties as prior beliefs about the vector parameters β_0 and β_k for $k = 1, \dots, K$. Specifically,

assume the generally improper priors $\beta_0 \sim N\{0, (\eta / \lambda_0) D_0^-\}$ and $\beta_k \sim N\{0, (\eta / \lambda_0) D_k^-\}$, where D_0^- denotes the generalized inverse of D_0 and $N\{0, (\eta / \lambda_0) D_0^-\}$ denotes improper Gaussian distributions with zero mean and positive semidefinite covariance $(\eta / \lambda_0) D_0^-$. The same convention is used for D_k^- and $N\{0, (\eta / \lambda_0) D_k^-\}$. The parameters λ_0 and λ control the dispersion of the priors; they are related to the smoothness of the functions $\alpha(\cdot)$ and $\beta_k(\cdot)$ respectively. Then the minimizer of expression (3) can be obtained as the posterior mode of the distribution $(\beta, b) | Y$ corresponding to the model

$$\begin{aligned} Y_{ij} | \xi_{ij1}, \dots, \xi_{ijK} &\sim \text{EF}(\mu_{ij}^K, \eta), \\ g(\mu_{ij}^K) &= B_0(t_{ij})^T \beta_0 + \sum_{k=1}^K \xi_{ijk} B_k(t_{ij})^T \beta_k + Z_{b,ij} b_i, \quad j = 1, \dots, n_i, \\ \beta_0 &\sim N\{0, (\eta / \lambda_0) D_0^-\}, \\ \beta_k &\sim N\{0, (\eta / \lambda) D_k^-\}, \quad k = 1, \dots, K, \\ b_i &\sim N(0, D), \end{aligned} \quad (4)$$

$\beta_0, \beta_1, \dots, \beta_K$ and b_i 's are mutually independent. The improperness of the distributions is due to the penalty matrices D_0 and D_k 's, which are rank deficient. This issue is well known in the penalized smoothing literature and can be remedied by using reparameterizations that separate the basis coefficients for each function into terms that are not penalized, treated as fixed effects, and terms that are penalized, treated as random effects with proper distribution; Wand (2003) and Ruppert et al. (2003), chapter 4.9. Such an approach enables us to construct Bayesian credible intervals for the unknown coefficients and any quantity that is derived from them. The tuning parameters can be estimated by Laplace-approximated REML as detailed in Wood (2011) and Wood et al. (2016). Incorporating additional covariates via the modelling framework (1) carries on in a straightforward manner to expressions (2) and (3), irrespective of whether the covariate's effects are modelled linearly or by a smooth dependence.

In our situation, however, neither the orthogonal basis $\{\phi_k(\cdot)\}_k$ nor the basis coefficients $\xi_{ij1}, \dots, \xi_{ijK}$ are known. In what follows we discuss the selection of these components, along with the choice of the truncation K , from the observed noisy functional covariates W_{ijr} 's.

2.2. Selection of the orthogonal basis

There are several options to select $\phi_k(\cdot)$'s. One approach is to use a prespecified basis similarly to Zhou et al. (2008). Another option is to use the eigenbasis of some appropriately chosen covariance function; see Park and Staicu (2015). We consider the latter approach and select $\phi_k(\cdot)$'s as the eigenbasis of the marginal covariance function induced by the functional covariates.

In our model, the latent subject-specific signal for the i th subject and time t , $X_i(s, t)$, is observed through $W_{ij}(s_r) = \tau(s_r, t_{ij}) + X_i(s_r, t_{ij}) + \epsilon_{ij}^W(s_r)$ at time t_{ij} and point s_r . Without loss of generality, we assume that $\tau(\cdot, \cdot) \equiv 0$. The terms $\epsilon_{ij}^W(s)$ are zero mean and IID, and encompass various sources of measurement error. If the true signal is observed only with white noise, then $\epsilon_{ij}^W(s) = \epsilon_{ij}^n(s)$, where $E[\epsilon_{ij}^n(s) \epsilon_{ij}^n(s')] = \mathbf{1}(s = s')\sigma_n^2$. If the true signal is observed only with smooth noise, then $\epsilon_{ij}^W(s) = \epsilon_{ij}^{\text{sm}}(s)$, where $E[\epsilon_{ij}^{\text{sm}}(s) \epsilon_{ij}^{\text{sm}}(s')] = \Gamma(s, s')$ for some smooth covariance function $\Gamma(\cdot, \cdot)$. In this paper we assume that $\epsilon_{ij}^W(s) = \epsilon_{ij}^{\text{sm}}(s) + \epsilon_{ij}^n(s)$, where the subscript, ‘sm’ is for ‘smooth’ and ‘n’ is for ‘white noise’.

Let $\Sigma(s, s') = E[X_i(s, T)X_i(s', T)]$ denote the marginal covariance of $X_i(s, T)$, by accounting for the randomness of the visit time, T ; if we denote by $h(t)$ the density that is assumed for T , the iterative expectation formula yields $\Sigma(s, s') = \int_{\mathcal{T}} E[X_i(s, t)X_i(s', t)]h(t)dt$. Define $\Xi(s, s') = \Sigma(s, s') + \Gamma(s, s')$ as the sum between the marginal covariance of $X_i(s, T)$ and the covariance of the smooth error component, $\epsilon_{ij}^{\text{sm}}(s)$. Note that $\Xi(s, s')$ is a proper covariance function (i.e. positive semidefinite and symmetric); see Park and Staicu (2015) for details. We refer to $\Xi(s, s')$ as the ‘marginal covariance function’ induced by the smooth part of the noisy observed functional covariates. In the absence of the smooth measurement error, $\Xi(s, s')$ reduces to $\Sigma(s, s')$ but, in general, this is not so. One advantage of the marginal covariance $\Xi(s, s')$ is that, in spite of its definition, this quantity is estimable from observed data W_{ij} ; see Section 3. This does not hold for the marginal covariance of the latent signal, $\Sigma(s, s')$,

We propose to select $\{\phi_k(\cdot)\}_{k \geq 1}$ as the eigenbasis of the marginal covariance $\Xi(s, s')$. Let $\{\phi_k(\cdot), \lambda_k\}_{k \geq 1}$ be the eigencomponents of $\Xi(s, s')$ with $\lambda_1 > \lambda_2 > \dots \geq 0$ and $\{\phi_k(\cdot)\}_{k \geq 1}$ an orthogonal basis in $L^2(\mathcal{S})$, and let K be a finite truncation such that $\Xi(s, s') \approx \sum_{k=1}^K \phi_k(s)\phi_k(s')\lambda_k$. This choice of basis may not be entirely informed by the latent signal due to the presence of a possible smooth noise. We can view this basis as a compromise between a preset choice of basis, such as B -spline or Fourier, and a data-informed basis such as the eigenbasis of the marginal covariance that is induced by the latent signal. As shown in Park and Staicu (2015), this basis is optimal for representing the smooth part of W_{ij} s as $W_{ij}^K(s) = \sum_{k=1}^K \phi_k(s)\xi_{W,ijk}$, where by an abuse of notation $\xi_{W,ijk} = \int_{\mathcal{S}} W_{ij}(s)\phi_k(s)ds$ for $k = 1, \dots, K$, in terms of minimizing a weighted mean-square-error criterion. In the last integral we referred to $W_{ij}^K(s)$ as the smooth part of W_{ij} s.

Let $X_i^K(s, t) = \sum_{k=1}^K \phi_k(s)\xi_{ik}(t)$, where $\xi_{ik}(t) = \int_{\mathcal{S}} X_i(s, t)\phi_k(s)ds$. One can show that, if K is chosen such that the approximation $\sum_{k \geq 1} \lambda_k \approx \sum_{k=1}^K \lambda_k$ is accurate, then $E\|X_i(\cdot, T) - X_i^K(\cdot, T)\|^2 = \int_{\mathcal{T}} E\|X_i(\cdot, t) - X_i^K(\cdot, t)\|^2 h(t)dt$ is negligible, where $\|f(\cdot)\|^2 = \int_{\mathcal{S}} f(s)^2 ds$ is the usual L^2 -norm on $\mathcal{S}$. Furthermore, this implies that

$$E \left[\int_{\mathcal{S}} \{X_i(s, T) - X_i^K(s, T)\} \beta(s, T) ds \right]^2 \quad (5)$$

is negligible, also, implying the accuracy of the finite approximation based on this basis,

$$\int_{\mathcal{S}} X_i(s, t) \beta(s, t) ds \approx \int_{\mathcal{S}} X_i^K(s, t) \beta(s, t) ds = \sum_{k=1}^K \xi_{ik}(t) \beta_k(t).$$

Here $|\mathcal{S}|$ denotes the length of the interval $\mathcal{S}$. Result (5) relies on the assumption that the regression function $\beta(s, t)$ is bounded and continuous over $\mathcal{S} \times \mathcal{T}$. The technical details are in the on-line supplementary material, section A1.

We end this section with a brief discussion about the size of K . If the eigenbasis $\{\phi_k(\cdot)\}_{k \geq 1}$ is also the eigenbasis for the marginal covariance $\Sigma(s, s')$, then the selection of K is optimal; see Park and Staicu (2015). This is so when both the smooth error covariance, $\Gamma(s, s')$ and $\Sigma(s, s')$ share the same basis. When the magnitude of Γ is sufficiently small in Hilbert–Schmidt norm relative to the smallest pairwise difference of the first K eigenvalues of Σ , $\Xi(\cdot, \cdot)$ can be viewed as a perturbation of $\Sigma(\cdot, \cdot)$. Using arguments from perturbation theory for linear operators (Kato, 2013), one can show that the leading eigencomponents of Σ can be recovered with negligible error from the leading eigencomponents of Ξ . Nevertheless, such a situation, although interesting theoretically, is difficult to verify in practical applications. In practice K is expected to be larger than a corresponding truncation based on the eigenbasis of the latent signal.

2.3. Basis coefficients for the latent covariate signal

For simplicity of exposition we assume that there is no white noise in the observed functional data W_{ij} , i.e. the latent signal is only contaminated with smooth measurement error and, by an abuse of notation, denote by $W_{ij}(\cdot)$ the smooth functional covariates.

Using the definition of the time-varying coefficients of the observed functional covariates, $\xi_{W,ijk} = \int_{\mathcal{S}} W_{ij}(s) \phi_k(s) ds$, and the representation $W_{ij}(s) = X_i(s, t) + \epsilon_{ij}^{\text{sm}}(s)$ it follows that $\xi_{W,ijk}$ can be written as $\xi_{W,ijk} = \xi_{ik}(t_{ij}) + e_{ijk}$. Here $\xi_{ik}(t_{ij}) = \int_{\mathcal{S}} X_i(s, t_{ij}) \phi_k(s) ds$ has zero mean and varies smoothly over t and is IID over i , whereas $e_{ijk} = \int_{\mathcal{S}} \epsilon_{ij}^{\text{sm}}(s) \phi_k(s) ds$ has zero mean and finite variance and is IID over both i and j . The properties of both terms, $\xi_{ik}(\cdot)$ and e_{ijk} , follow from the assumptions that are made on $X_i(s, \cdot)$ and $\epsilon_{ij}^{\text{sm}}(s)$, respectively. Thus, once $\{\xi_{W,ijk}; i, j\}$ have been obtained, we can use common techniques in functional data literature to recover the full trajectories $\xi_{ik}(\cdot)$.

Fix k arbitrarily, $k = 1, \dots, K$. One approach is to model the pseudodata

$\{(\xi_{W,ijk}, t_{ij}); j = 1, \dots, n_i\}_i^I$ as specified earlier and to use methods for sparse functional data (Yao et al., 2005). Specifically, denote by $G_k(t, t') = E[\xi_{ik}(t) \xi_{ik}(t')]$ the covariance of the latent smooth process $\xi_{ik}(\cdot)$ and let $\{\eta_{kb}, \psi_{kb}(\cdot)\}$ be the pair of eigenvalues and

eigenfunctions of $G_k(\cdot, \cdot)$, where $\eta_{k1} \geq \eta_{k2} \geq \dots > 0$, and $\{\psi_k(\cdot)\}_{k \geq 1}$ are mutually orthogonal, and have unit norm in $L^2(\mathcal{T})$. Sometimes a direction $\phi_k(\cdot)$ would correspond to a variance of e_{ijk} , say σ_k^2 , that is much larger than the variance of the signal, equal to $\int_{\mathcal{T}} G_k(t, t)dt$, thus implying that there is only noise along it; we discard such a direction from the original basis. When this is not so, the projected signal $\xi_{ik}(t)$ is modelled via the truncated Karhunen–Loève expansion $\xi_{ik}(t) = \sum_{l=1}^L \zeta_{ikl} \psi_{kl}(t)$, where $\zeta_{ikl} = \int_{\mathcal{T}} \xi_{ik}(t) \psi_{kl}(t) dt$ have zero mean and variance equal to η_{kl} and are uncorrelated across l . It follows that $\xi_{ik}(t)$ can be recovered by the finite dimensional representation $\tilde{\xi}_{ik}(t) = \sum_{l=1}^{L_k} \tilde{\zeta}_{ikl} \psi_{kl}(t)$, for some finite truncation L_k where $\tilde{\zeta}_{ikl}$ are estimated by the best linear unbiased predictors, $E[\zeta_{ikl} | \xi_{W,1k}, \dots, \xi_{W,ink}]$; see section A2 of the on-line supplementary material for technical details. Alternatively, $\{(\xi_{W,ijk}, t_{ij}) : j = 1, \dots, n_i\}_i^I$ can be modelled by assuming a parametric covariance model for $(\xi_{W,ijk}, t_{ij})$, such as exponential, or Matérn, or random-effects-based models; standard methods in longitudinal data analysis can be used to recover the full trajectory $\xi_{ik}(t)$.

3. Estimation

The above ideas are based on the true marginal covariance $\Xi(s, s')$. In practice, this covariance is unknown, but it can be estimated directly from W_{ijr} s by using the method of moments. The approach has been developed in Park and Staicu (2015) and we briefly describe it here for completeness. In practice, the true mean function $\tau(s, t)$ may not be 0; we first estimate it and then subtract it off. We model the mean of W_{ijr} at time t_{ij} and point s_r as a bivariate smooth function of both s and t , and estimate it by fitting a bivariate smoother under a working independence assumption (Marx and Eilers, 2005; Wood, 2006a; Xiao et al., 2013). Let $\tilde{W}_{ijr}$ be the de-measured data and let

$\tilde{\Xi}(s, s') = \sum_{i=1}^I \sum_{j=1}^{n_i} \tilde{W}_{ijr} \tilde{W}_{ijr'} / \sum_{i=1}^I n_i$ be the pooled sample covariance; $\tilde{\Xi}(s, s')$ has expected value equal to $\Xi(s, s')$ for $s \neq s'$ and equal to $\Xi(s, s') + \sigma_n^2$ for $s = s'$. The diagonal elements are inflated by the variance term σ_n^2 , when the observed functional covariates include white noise. For this reason, we refer to the pooled covariance as a ‘raw estimator’ of the marginal covariance, $\Xi(s, s')$. As in Staniswalis and Lee (1998) (Yao et al., 2005), the pooled covariance estimator is smoothed out by removing the diagonal elements to estimate the smooth marginal covariance $\Xi(s, s')$; we use the bivariate smoothing that is described in Xiao et al. (2013) in our numerical investigation. Let $\hat{\Xi}(s, s')$ denote the estimator of the marginal covariance obtained in this way. Park and Staicu (2015) showed that the pooled covariance $\tilde{\Xi}(s, s')$ is a consistent estimator for the true marginal covariance, when the observed functional covariates W_{ijr} include possibly smooth measurement error, but no white noise. In the presence of white noise, similar arguments to Yao et al. (2005) could be made to conclude that the smooth estimator $\hat{\Xi}(s, s')$ obtained is a consistent estimator of $\Xi(s, s')$; however, no formal proof of this result exists, to the best of our knowledge.

We estimate the eigenfunctions by first discretizing the estimated smooth covariance $\widehat{\Xi}(s, s')$ and then performing its spectral analysis. Denote the resulting K eigencomponents by $\{\hat{\phi}_k(\cdot), \hat{\lambda}_k\}_{k=1}^K$; we have $\int_S \hat{\phi}_k(s) \hat{\phi}_{k'}(s) ds$ equals 1 if $k = k'$ and 0 otherwise, and $\hat{\lambda}_1 > \hat{\lambda}_2 \dots \geq 0$. The truncation K is determined on the basis of a prespecified level for the percentage of variance explained. Specifically, K is chosen as the smallest integer such that $\sum_{k=1}^K \hat{\lambda}_k / \sum_{k \geq 1} \hat{\lambda}_k$ is larger than the prespecified percentage of variance explained value. As the choice of K will directly affect the components of the true regression coefficient function that could be recovered, we recommend the use of a larger percentage of variance explained threshold. This *ad hoc* criterion is a very popular technique (Di et al., 2009; Staicu et al., 2010) to obtain a low rank approximation of the covariance $\widehat{\Xi}(s, s')$ and we use it as well. Next, we use the methodology that was described in Section 2.3, by substituting the eigenbasis $\{\phi_k(s)\}_k$ by the estimated $\{\hat{\phi}_k(s)\}_k$ to recover the basis coefficients trajectories of the latent signal, say $\hat{\xi}_{ik}(t)$. Finally, using the estimates $\hat{\xi}_{ik}(t_{ij})$ for ξ_{ijk} , we can fit model (1) by employing the methodology that was described in Section 2.1. Because of dropping off some directions, the resulting basis size may be smaller than K ; for simplicity of exposition, however, we assume that the basis size is still K .

Using the parameter estimates from equations (3) or (4) we obtain estimates of the intercept function $\hat{\alpha}(t) = B_0(t)^T \hat{\beta}_0$ and of the regression bivariate function

$$\hat{\gamma}(s, t) = \sum_{k=1}^K \hat{\phi}_k(s) \hat{\beta}_k(t), \quad (6)$$

where $\hat{\beta}_k(t) = B_k(t)^T \hat{\beta}_k$. Conditionally on the estimates from the preprocessing of the functional covariates, uncertainty quantification for $\hat{\alpha}(\cdot)$ and $\hat{\gamma}(\cdot, \cdot)$ is readily available. However, valid inference for $\gamma(s, t)$ is complicated by the fact that our estimation procedure relies on low rank approximation of the latent signal covariance, and thus estimator is not identifiable. Instead we focus on response prediction and quantifying its uncertainty.

4. Prediction and inference

4.1. Prediction of response trajectories

One main goal is to predict response trajectories in two settings: for an existing data subject i , who has been observed at few sparse time points t_{ij} , and for a new subject i^* for whom only the associated noisy functional covariates are available. In general, the response for an existing data subject at an observed time t_{ij} , Y_{ij} , can be predicted by directly substituting the estimates of the parameters and the predicted random subject effects $\hat{b}_i$ into equation (2); the specifics depend on the form of the link function. Irrespectively of whether prior response data have been observed for the subject, prediction of a subject's response trajectory $Y_i(t)$ for all t requires estimation of the subject's time-varying basis coefficients trajectories $\xi_{ik}(t)$, where $\xi_{ijk} = \xi_{ik}(t_{ij})$.

The prediction of the i th response trajectory $Y_i(t)$ implicitly assumes the model $Y_i(t) \sim \text{EF}\{\mu_i(t), \eta\}$ for all t , where $g(\cdot)$ is as in model (1) and is used to relate the linear predictor to the mean response for every t , $g\{\mu_i(t)\} = \alpha(t) + \int \gamma(s, t)X_i(s, t)ds + Z_{b,ii}b_i$. Here $Z_{b,ii}$ is the random subject-specific design vector corresponding to a generic time t . For example, in the case of a random-slope model, we have $Z_{b,ii} = (1, t)$, $b_i = (b_{i0}, b_{i1})^T$ and thus $Z_{b,ii}b_i = b_{i0} + b_{i1}t$. Consider an existing data subject; its subject mean response trajectory $\mu_i(t)$ is predicted by

$$\hat{\mu}_i^K(t) = g^{-1}\left\{\hat{\alpha}(t) + \sum_{k=1}^K \hat{\xi}_{ik}(t)\hat{\beta}_k(t) + Z_{b,ii}\hat{b}_i\right\}, \quad (7)$$

where g^{-1} is the inverse function of g and the response trajectory $Y_i(t)$ is then predicted by using the assumed exponential family model $Y_i(t) \approx \text{EF}\{\hat{\mu}_i^K(t), \hat{\eta}\}$. When the responses are Gaussian we have $g(\mu) = \mu$ and expression (7) can be used to predict the subject trajectories $\hat{Y}_i(t) = \hat{\mu}_i^K(t)$. We conjecture that the predicted trajectory $\hat{Y}_i(t)$ consistently estimates $\tilde{Y}_i(t) = E[Y_i(t) | \mathbf{W}_i]$, which is $\tilde{Y}_i(t) = \alpha(t) + \sum_{k \geq 1} \xi_{ik}(t)\beta_k(t) + Z_{b,ii}b_i$, where $\tilde{\xi}_{ik}(t) = \sum_{l \geq 1} \psi_{kl}(t)\tilde{\xi}_{ikl}$, $\tilde{\xi}_{ikl} = E[\xi_{ikl} | \mathbf{W}_i]$ and $\mathbf{W}_i$ is the Rm_i -dimensional vector of $W_{ij}s$. When the responses are non-Gaussian, prediction of the subject-specific trajectories is not always clearly defined. For Bernoulli responses a common approach is to predict $\hat{Y}_i(t) = 1$ if $\hat{\mu}_i(t) \geq 0.5$ and $\hat{Y}_i(t) = 0$ if $\hat{\mu}_i(t) < 0.5$.

Next, consider a new subject i^* , for whom the available information consists of $\{[(W_{i^*j}^*, s_{ijp}): r], t_{i^*j}^*: j]\}$. The mean response trajectory is predicted as $\hat{\mu}_{i^*}^K(t) = g^{-1}\{\hat{\alpha}(t) + \sum_{k=1}^K \hat{\xi}_{i^*k}(t)\hat{\beta}_k(t)\}$, where the time-varying trajectories $\hat{\xi}_{i^*k}(t)$ are obtained as in Section 2.3 by using the noisy pseudodata $\xi_{W, i^*jk} = \int_{\mathcal{S}} \tilde{\mathbf{W}}_{i^*j}^*(s)\hat{\phi}_k(s)ds$, where $\tilde{\mathbf{W}}_{i^*j}^*(s)$ is detrended by using the estimated mean function, and using best linear unbiased prediction in the working model $\xi_{W, i^*jk} = \sum_{l=1}^{L_k} \psi_{kl}(t_{i^*j}^*)\zeta_{i^*kl} + \epsilon_{i^*jk}$, where $\zeta_{i^*kl} \sim (0, \hat{\eta}_{kl})$ and $\epsilon_{i^*jk} \sim (0, \hat{\sigma}_{\epsilon_k}^2)$. Circumflexes refer to the model estimates based on the training data. The response trajectory does not involve the subject-specific effects b_{i^*} , as their estimation requires availability of the response data for the i^* subject.

4.2. Asymptotic prediction bands

Once prediction of the response trajectory has been obtained, a natural follow-up goal is to quantify the uncertainty of this prediction. In this section we attempt to address this objective, when the response is Gaussian. The uncertainty in prediction is measured by the prediction error (Ruppert et al., 2003) $\hat{Y}_i(t) - Y_i(t)$ and thus will be different according to whether the subject is an existing data subject or a new subject. Regardless of the situation, the predicted trajectory $\hat{Y}_i(t)$ involves

- a. the term $\hat{\alpha}(t) + \sum_k \hat{\xi}_{ik}(t) \hat{\beta}_k(t)$, whose variance depends on the variance of the regression estimators $\hat{\alpha}(t)$ and $\hat{\beta}_k(t)$, given $\hat{\xi}_{ik}(t)$, but also,
- b. intrinsically through $\hat{\xi}_{ik}(t)$, the latter term has variance depending on the variability of the bases $\hat{\phi}_k(s)$ and $\hat{\psi}_{kl}(t)$, of the eigenvalues $\hat{\eta}_{kl}$, and of the finite truncations K and L_k .

As in standard functional data analysis, the uncertainties that are associated with terms (b) are difficult to quantify analytically. We discuss prediction bands by treating the terms as known, and ignoring the variability in estimating them. Let $\mathbf{W}_i$ be defined as before and $\xi_{W,ik}$ be the vector of the pseudodata $\xi_{W,ijk}$. In the case of an existing subject i , for time point $t \notin \{t_{i1}, \dots, t_{in_i}\}$ we have

$$\begin{aligned} \text{var}\{\hat{Y}_i(t) - Y_i(t) \mid \mathbf{W}_i\} &= \text{var}\{\hat{Y}_i(t) \mid \xi_{W,i1}, \xi_{W,i2}, \dots, \xi_{W,iK}\} + \text{var}(\epsilon_{it}) \\ &+ \text{var}\left\{\sum_{k=1}^K \beta_k(t) \xi_{ik}(t) \mid \mathbf{W}_i\right\}. \end{aligned} \quad (8)$$

In practice we construct the prediction band, by replacing the unknown quantities by the estimated counterparts and using truncation K for the infinite sum in equation (8). The variance of the white noise ϵ_{it} and the other variance parameters are estimated by using REML; see Section 2.1. The conditional variance of $\hat{Y}_i(t)$ is estimated with standard approaches in longitudinal data analysis (Ruppert et al., 2003; Wood, 2006a, b); the estimation is implemented in various computer packages and we discuss it in the on-line supplementary material, section B1. The last term is non-trivial to quantify; we do it by modelling W_{ijr} , using the longitudinal functional data analysis framework of Park and Staicu (2015) that was briefly detailed in Section 2.3 and additionally using a working independence assumption of $\xi_{ik}(t)$ across k . The analytical expression is derived in the supplementary material. Finally, in the case of a new data subject, we have

$$\text{var}\{\hat{Y}_i(t) - Y_i(t) \mid \mathbf{W}_i\} = \text{var}\{\hat{Y}_i(t) \mid \mathbf{W}_i\} + Z_{b,it} \text{var}(b_i) Z_{b,it}^T + \text{var}(\epsilon_{it}) + \text{var}\left\{\sum_{k=1}^K \beta_k(t) \xi_{ik}(t) \mid \mathbf{W}_i\right\}.$$

A $100(1 - \alpha)\%$ pointwise prediction interval for $Y_i(t)$ is $\hat{Y}_i(t) \pm z_{\alpha/2} \widehat{\text{SE}}\{\hat{Y}_i(t) - Y_i(t)\}$, where $z_{\alpha/2}$ is the $\alpha/2$ upper quantile of the standard normal distribution. Here $\widehat{\text{SE}}\{\hat{Y}_i(t) - Y_i(t)\}$ is the estimated conditional standard error of $\hat{Y}_i(t) - Y_i(t)$ and is calculated as the square root of the estimated conditional variance of $\hat{Y}_i(t) - Y_i(t)$. In Section 6.2 we assess the performance of the predicted response trajectory, as well as that of the proposed pointwise prediction bands, for both existing data subjects and new subjects.

5. Data application

Our motivating application is a lactating sow study where the primary objective is to investigate the effect of thermal environment (i.e. temperature T) on the feed intake of the lactating sows. This study is very important for several reasons.

- a. Ambient temperatures above the evaporative critical temperature decreases the amount of food intake which, as a result, deteriorates the reproductive performance and hinders the growth rate of piglets of lactating sows (Black et al., 1993).
- b. Also, poor feed intake of the lactating sows leads to increased body weight loss during lactation and reduced milk yield, and is further associated with compromised weight gain of their litter (Johnston et al., 1999; Renaudeau and Noblet, 2001).
- c. Thirdly, heat stress results in a reduction in the farrowing rate (the percentage of sows that become pregnant and farrow a litter of piglets) and the total number of pigs born alive (Bloemhof et al., 2013) which in turn has a negative effect on the total production of pork meat per year.
- d. Fourth, pigs from sows that are raised in an unfavourable thermal environment will be fatter than those reared in favourable cooler environments and this makes pork meat fattier (Baumgard, 2015).
- e. Fifth, because of heat stress associated with a hot climactic thermal environment, the swine industry in the USA incurs a total estimated loss worth of \$300 million per year on average (St-Pierre et al., 2003).

Thus insight into how the feeding behaviour changes over time with the ambient temperature has the potential to assist better in calibrating the room thermal conditions of the lactating sows, and thus positively impacting the swine industry.

The experimental study was carried out during July–October in 2013 in a 2600-sow commercial research unit in Oklahoma (Rosero et al., 2016) and involved 480 PIC Camborough sows. The sows were kept in the farrowing facility where they gave birth to piglets. Depending on the number of previous pregnancies (parity levels), sows were classified into younger (parity equal to 0 or 1) or older (parity equal to 2 or higher). The sows are brought to the farrowing crates when they are approximately 5 days before they are due to give birth. They arrive in groups; the study involves 21 groups of about 21–23 sows. Sows are observed during their 20–21-day lactation period and their respective food consumption is monitored. Each sow is provided food individually with a computerized feeding system (Howema, Big Dutchman, Germany). The amount of food offered, FO, was recorded at 2.00 p.m. on each day and feed refusal, FR, was measured the following day before any subsequent food addition; feed intake, FI, was calculated as $FI = FO - FR$ in kilograms. Minute-by-minute information about the ambient air temperature (degrees Celsius) and humidity (per cent) of the farrowing facility were recorded by data loggers (LogTag, MicroDAQ Ltd, Contoocook, New Hampshire). The experimenters removed information of five sows because of unreliable measurements and thus we had available information for 475 sows. The facility ambient temperature was controlled by a ventilation system; the barns have cool cells that pull fresh air through wet corrugated material to provide further cooling of air. There are some missingness observations for temperature profiles due to machine failure which qualifies the pattern of missingness as missingness

completely at random. Our objective is to study the effect of temperature on the feed intake of the sows during their lactation period.

Let i index the sows, j index the repeated instances for the same sow and t_{ij} denote the lactation day of the i th sow, corresponding to the j th instance at which the sow is observed; for many sows we have $t_{ij} = j$ for $j = 1, \dots, 21$, but this is not always so. The ‘time’ is defined as the 24-h time window from 2.00 p.m. to 1.59 p.m. Furthermore let g_i index the group of the i th sow, $g_i = 1, \dots, 21$; typically, the sows within the same group give birth closer to one another. Denote by $nTemp_{ij}(\cdot) = nTemp(\cdot, t_{ij})$ the daily temperature profile observed at the t_{ij} lactation day of the i th sow; the measurements typically include noise: hence the prefix ‘n’. Later we use notation $Temp_{ij}(\cdot)$ for the true temperature profile corresponding to $nTemp_{ij}(\cdot)$. Fig. 1(a) shows the daily temperature corresponding to the first 21 days for a random sow. Let FI_{ij} be FI for the i th sow at its j th repeated occasion, lactation day t_{ij} .

We assume that the relationship between the feed intake and the temperature is described by the LDFR model

$$FI_{ij} = \beta_{p_i}(t_{ij}) + \int_S Temp_{ij}(s) \gamma(s, t_{ij}) ds + b_{g_i} + b_{0i(g_i)} + \varepsilon_{ij}, \quad (9)$$

where $\beta_{p_i}(\cdot)$ is the mean feed intake for group p_i where $p_i = 0$ (young sows) and $p_i = 1$ (old sows) and $\gamma(s, \cdot)$ quantifies the time-varying effect of the temperature on FI; the integral reflects the aggregated effect during the course of the 24 h. The term $b_{g_i} + b_{0i(g_i)}$ models the dependence of the responses within the same sow as well as the dependence of the responses of the sows who are in the same group. The random terms b_{g_i} are a group-specific effect and $b_{0i(g_i)}$ is the sow within-group-specific intercept. We assume that $b_{g_i} \sim N(0, \sigma_g^2)$ and

$b_{0i(g_i)} \sim N(0, \sigma_0^2)$ are mutually independent. Finally, it is assumed that the measurement errors ε_{ij} are independent and distributed as $N(0, \sigma_e^2)$. Model (9), which models the current day feed intake, does not account for the previous day feed intake; we argue this choice in Section 5.1.

To fit model (9), we begin with de-meaning the functional covariates. We estimate the mean by passing a bivariate smoother through the data (Wood, 2001) using the tensor product of cubic splines (−20 in the s -direction and 10 in the t -direction); this is implemented in `gam()` of the `mgcv` package in R and uses generalized cross-validation as default method to estimate the tuning parameters. To select the orthonormal basis, we set the percentage of variance explained to 95%, which resulted in $K=6$ basis functions. The model fitting continues as detailed in Section 3 and also employed in Section 6. In the final part, the parameter functions $\beta_k(\cdot)$ are modelled by using 10 cubic splines and the smoothing parameters are selected by using REML; this also is implemented in the `gam()` function.

5.1. Estimation of the model components: prediction of response trajectories

Plots of the residuals are included in the on-line supplementary material, section C.3. Also there we discuss the lack of evidence for including the feed intake of the previous day as a

covariate in the model. Fitting model (9) indicates much lower variability in the feed intake at the group level ($\hat{\sigma}_g^2 = 0.03$) compared with that at the sow level ($\hat{\sigma}_0^2 = 0.63$). Fig. 2 shows the estimates of the model parameter functions. Specifically Fig. 2(a) depicts the estimated mean feed intake for old sows (the dotted curve), $\hat{\beta}_1(t)$, and for young sows (the chain curve), $\hat{\beta}_0(t)$, along with 95% pointwise confidence bands. It appears that the feed intake is about the same in the first couple of days, for both young and old sows, but shortly afterwards the older sows eat more than their younger relatives. By the fourth lactation day the older sows have an advantage of feed intake of up to 1–1.5 kg per day and they maintain this advantage for the remaining duration of the lactation period. Fig. 2(b) shows the estimated regression coefficient $\hat{\gamma}(s, t)$, which quantifies the minute-by-minute (s) effect of the temperature on the feed intake during the first 21 lactation days (t). The estimated effect shows that the association between temperature and feed intake changes during the lactation duration; interpreting this effect more is more challenging because of the lack of identifiability in estimation. Using a version of R^2 for this setting, the model proposed accounts for about 55.2% of the variation in the data. Sections C.1–C.3 of the on-line supplementary material include additional results for the data analysis, whereas section C.4 considers an additional simulation experiment inspired by the data application and compares the prediction performance of our methods with competitors.

Fig. 3 shows the predicted full trajectories of the feed intake for two young sows (Figs 3(a) and 3(b)) and one old sow (Fig. 3(c)) selected at random from different groups along with their pointwise prediction bands constructed as detailed in Section 4. The predicted trajectories are obtained from $\widehat{\text{FI}}_i(t) = \hat{\beta}_{p_i}(t) + \int \widehat{\text{Temp}}_i(s, t) \hat{\gamma}(s, t) ds + \hat{b}_{g_i} + \hat{b}_{0i(g_i)}$ for every day $t = 1, \dots, 21$, where $\widehat{\text{Temp}}_i(\cdot, t)$ are the smooth and de-measured temperature profiles observed in relation to sow i , and $\hat{b}_{g_i}$ and $\hat{b}_{0i(g_i)}$ are the predicted random effects. These results indicate also greater feed intake for older sows relative to their younger counterparts; the prediction intervals are wider to account for estimated measurement error.

5.2. Fit assessment

We first assess the prediction performance of the method proposed and compare it with the available competitors, LPFR and LPEER. In this regard we split the data into a training set, which is used to build the model, and a test set on which the prediction performance is evaluated; we replicate the test–train split for 25 times. We consider two ways of forming the test data:

- a. randomly select 350 sows of the total of 475 sows and include only the measurements corresponding to their last 50% lactation days observed;
- b. take all the 475 sows and include only the measurements corresponding to about 20% of their lactation days that are selected at random.

The remaining data form the training set. Approach (a) involves data on fewer sows than does approach (b). At the same time, approach (a) assesses the performance of prediction at

‘future’ lactation days, whereas approach (b) evaluates the prediction performance at random lactation days within the 1–21 days at which the sows are observed.

For completeness we describe the implementation of the competitive approaches. LPFR assumes that temperature has a constant effect across the lactation days of a sow and models this time invariant effect by using 30 truncated linear splines basis functions and the tuning parameters are estimated by REML. The covariance structure is specified as in model (9) and the model is fitted by using the `lpfr()` function that is available in the `refund` package (Huang et al., 2015). LPEER uses polynomial functions of time of degree $d = 0, 1, \dots, 4$ to model $\gamma(s, t) = \gamma_1(s) + t\gamma_1(s) + \dots + t^d\gamma_d(s)$ and selects the optimal d by Akaike’s information criterion; the model is fitted with `lpeer()` in `refund`. The covariance structure is specified by using a subject-specific random intercept and random group effect; it is not clear how to modify the existing code to accommodate a subject-specific slope effect.

Table 1 reports the results. The findings show that LDFR and LPEER perform relatively similarly in terms of in-sample and out-of-sample prediction in the two situations that were considered. Nevertheless our methodology yields computations that are an order of magnitude faster. The prediction results with LPFR are inferior, and they indicate a lack of appropriateness of a time invariant effect model for our lactating sow application, based on the observations from our simulation investigation; see Section 6.

6. Simulation

6.1. Description of the settings

We use Monte Carlo simulations to assess the numerical performance of the method proposed (LDFR). Whenever applicable we compare it with two other competing approaches: LPFR (Goldsmith et al., 2012) and LPEER (Kundu et al., 2016). The data $[t_{ij}, Y_{ij}, \{(W_{ijr}, s_r) : r = 1, \dots, R\} : j = 1, \dots, n_i]_{i=1}^I$ are generated according to the following scenarios. The latent signal is described by $X_\lambda(s, t) = 2\cos(2\pi s)\xi_{\lambda 1}(t) + 2\sin(2\pi s)\xi_{\lambda 2}(t)$, where $\xi_{\lambda 1}(t) = 2\cos(2\pi t)\zeta_{\lambda 11} + 2\sin(2\pi t)\zeta_{\lambda 12}$ and $\xi_{\lambda 2}(t) = 2\cos(4\pi t)\zeta_{\lambda 21} + 2\sin(4\pi t)\zeta_{\lambda 22}$; here $\zeta_{\lambda 11}, \zeta_{\lambda 12}, \zeta_{\lambda 21}$ and $\zeta_{\lambda 22}$ are assumed to be mutually IID from $\mathcal{N}(0, 1.75)$, $\mathcal{N}(0, 1)$, $\mathcal{N}(0, 1.5)$ and $\mathcal{N}(0, 0.75)$ respectively. Let $W_{ijr} = W_i(s_r, t_{ij}) = \tau(s_r, t_{ij}) + X_i(s_r, t_{ij}) + \epsilon_{ij}^{\text{sm}}(s_r) + \epsilon_{ijr}^n$, for $\tau(s, t) = 1 + 2s + 3t + 4st$, $\epsilon^{\text{sm}}(s) = 2\cos(2\pi s)\epsilon_{1,jj} + 2\sin(2\pi s)\epsilon_{2,jj} + 2\cos(4\pi s)\epsilon_{3,jj}$ for $\epsilon_{1,jj} \stackrel{\text{IID}}{\sim} \mathcal{N}(0, 1)$, $\epsilon_{2,jj} \stackrel{\text{IID}}{\sim} \mathcal{N}(0, 0.3)$, $\epsilon_{3,jj} \stackrel{\text{IID}}{\sim} \mathcal{N}(0, \sigma^2)$ and $\epsilon_{ijr} \stackrel{\text{IID}}{\sim} \mathcal{N}(0, \sigma_n^2)$. The smooth error $\epsilon_{ij}^W(\cdot)$ has the same modes of variation as the latent signal $X_\lambda(\cdot, t)$ for $\sigma^2 = 0$ solely. This implies that, for $\sigma^2 = 0$, the covariance function, $\Xi(s, s')$, defined in Section 2.2, has the same eigenbasis as the marginal covariance that is induced by the latent signal, $\Sigma(s, s')$; as σ^2 grows larger, $\Xi(s, s')$ has increasingly different eigenbasis. Consider the scenarios $\sigma^2 = 0$ (scenario A1) and $\sigma^2 \in \{0.5, 4, 9, 50\}$ (scenario A2). The variance of the white noise is set to $\sigma_n^2 = 1$ (scenario B1) and $\sigma_n^2 = 9$ (scenario B2). The grid for s is a set of 101 equidistant points in $[0, 1]$; we use two sampling designs for t_{ij} s. In each case, $\{t_{i1}, \dots, t_{in_i}\}$ are randomly chosen from a set of 41 equidistant points in $[0, 1]$.

- a. *Sparse design* (design C1): the number of repeated responses per subject is $n_i \in \{11, \dots, 15\}$.
- b. *Moderately sparse design* (design C2): $n_i \in \{21, \dots, 25\}$.

We generate Y_{ij} in the exponential family as follows.

- a. Gaussian responses with mean $\mu_{ij} = \mu_i(t_{ij})$ defined by $\mu_i(t) = \alpha(t) + \int X_i(s, t) \gamma^\delta(s, t) ds$, where $\alpha(t) = 7 \sin(3\pi t)$ and random deviation ϵ_{ij} (scenario D1): consider three dependence structures:
 - i. independent,
 - ii. compound symmetric and
 - iii. a random-effects model.

The detailed description is given in the on-line supplementary material, section B1.

- b. Binary-valued responses (scenario D2): we consider $P(Y_{ij} = 1) = \exp(\omega_{ij}) / \{1 + \exp(\omega_{ij})\}$, where $\omega_{ij} = \alpha(t_{ij}) + \int_{\mathcal{S}} X_i(s, t_{ij}) \gamma^\delta(s, t_{ij}) ds + b_{i0} + b_{i1} t_{ij}$, and the subject-specific random effects b_{i0} and b_{i1} are used to model the dependence across the repeated measurements and are generated as in dependence structure (iii) above. The choice for $\alpha(\cdot)$ described in scenario D1 varies with time. Although a time-varying intercept presents no issues for our proposed method, such a choice does not seem to be accommodated by LPFR. Thus, for binary-valued responses, to compare the results of our method with LPFR, we consider $\alpha(\cdot) = 2$.

The functional coefficient is defined as $\gamma^\delta(s, t) = 2\exp(-\delta t)\cos(2\pi s) + 2\delta t \sin(\delta t) \sin(2\pi s)$ (function E1) and a polynomial function of t (function E2): $\gamma^\delta(s, t) = 2(1 + \delta t) \cos(2\pi s) + 2(1 - \delta t + \delta t^2) \sin(2\pi s)$, where δ controls the departure from a time invariant effect. For example in function E1 when $\delta = 0$ we have that $\gamma^\delta(s, t) = 2\cos(2\pi s)$, whereas, when $\delta \neq 0$, $\gamma^\delta(s, t)$ varies with time t . We investigate the cases $\delta = \{0, 1, 5, 10\}$. Details of the implementation in R for both our method and the competitive methods are included in the on-line supplementary material, section B1.

To assess the prediction performance of the method proposed, we divided each simulated data set into a training and a test set, and calculated the root mean prediction error for the two sets, as well as the root prediction error of the full response trajectory. Additionally, we evaluate the pointwise prediction bands' accuracy by the estimated coverage probability and expected length. For binary-valued responses, we assess the accuracy in estimating the linear predictor trajectory. For brevity we include the details on calculating these measures, as well as the other evaluation criteria in the supplementary material section B.1.

The results are based on 1000 independent samples for each combination of the simulation settings. In our numerical investigation, we used an Intel(R) Core(TM) i7-4770, 3.40-GHz processor with 8.0 Gbytes random-access memory in 64-bit operating system.

6.2. Prediction performance assessment and comparison

6.2.1. Prediction accuracy—Prediction accuracy is assessed for both training and test sets. The two sets are obtained first at the subject level, by splitting the data of each subject into a training and a test set, and then collecting the training sets and test sets respectively, across all available subjects. The residual mean-square error for the training set is presented as in-sample prediction error, IN_{PE} , whereas for the test set it is referred to as out-of-sample prediction error, OUT_{PE} . We first consider Gaussian responses and compare the proposed method with existing alternatives, when data are generated from a compound symmetric covariance structure D1, part (ii), and the sparse longitudinal design C1. Table 2 illustrates the median prediction error and interquartile range IQR for LDFR with LPEER; because of the heavy computational burden of the latter approach, we limit our investigation to 100 Monte Carlo samples per setting. The comparison with LPFR is included in Table 1 of the on-line supplementary material. Overall, the results suggest similar prediction accuracy for both in and out of sample, across all three methods, when the effect is time invariant ($\delta = 0$). However, when the association is time varying ($\delta > 0$), prediction quality suffers from not accounting for the time dependence of the regression coefficient; both LDFR and LPEER have better prediction compared with LPFR. Table 2 shows generally similar accuracy between LDFR and LPEER for various values of $\delta > 0$. Computationally, however, the two methods are different: LDFR is over an order of magnitude faster than LPEER; as seen in the sixth and seventh and 13th and 14th columns of Table 2. Furthermore, the numerical study shows that the accuracy of our method increases with the number of repeated measurements per subject; this is expected as, in this case, the estimation of the within-subject covariance improves. Also the prediction performance is not affected by the magnitude of the white noise in the functional covariates.

For binary responses with logit link (scenario D2) we compare LDFR with LPFR—the only available competitor for prediction at times points t_{ij} at which the functional covariates are observed. The data are generated assuming both a random intercept and a random slope, and we fit the model with subject-specific random intercept (model misspecification). The results are similar to those before and for brevity we include them in Table 8 of the on-line supplementary material, section B.4. The method proposed allows prediction of the full linear predictor trajectory $g\{\mu^K(t)\}$ and the prediction performance shows similar properties to those for Gaussian responses, as shown in Table 3.

6.2.2. Prediction accuracy of response trajectory—One of the main contributions of this work is prediction of the full response trajectory. The prediction of the entire trajectory is assessed by using the root mean prediction error of $\hat{Y}_i(\cdot)$, defined as

$$RMPE_{trj} = \left(\frac{1}{I} \sum_{i=1}^I \left[\frac{1}{n} \sum_{j=1}^n \{Y_i(t_j) - \hat{Y}_i(t_j)\}^2 \right] \right)^{1/2},$$

where $\{t_1, \dots, t_n\}$ is an equally spaced grid of 41 points in $[0, 1]$ and $Y_i(t_j)$ is obtained by using the generating model. The fifth and 12th columns of Table 2 show the prediction accuracy of the full response trajectory for various sample sizes, when the responses are

Gaussian with between 11 and 15 response measures per subject. First we remark that, as the magnitude of δ increases, the difficulty of the problem increases and the prediction accuracy for all cases suffers. The accuracy improves both as the number of repeated measurements per subject increases and when the sample size increases, with the former factor having higher impact; these comments are based on the results summarized in the on-line supplementary material, Figs 2 and 3.

6.2.3. Accuracy of the prediction bands—We examine the performance of the response prediction bands in the case where the responses are Gaussian and when the subjects have been previously observed (Y_j), as well as when the subjects have not been observed (Y_{j*}). Table 3 reports the actual coverage and expected length for two nominal levels 90% and 95%. Whereas the actual coverage probability is conservative for the subjects previously observed, for new subjects the level is quite liberal. This should not be surprising, as we are constructing the prediction intervals by ignoring few sources of variability, as detailed in Section 4.2. Similar behaviour is observed in standard functional data analysis in prediction bands for trajectories, when the variability in eigenbasis components is not accounted for (Goldsmith et al., 2013). As expected, the width of the prediction bands is larger when predicting for the new subjects compared with existing subjects and it increases with the complexity of the regression coefficient.

7. Discussion

In this paper we consider longitudinal dynamic functional regression for scalar responses and functional covariates observed in a longitudinal design. We propose a flexible way to model the time-varying bivariate regression coefficient function by combining ideas from functional data analysis and longitudinal data analysis. As one reviewer asserted, this clever combination enables us to tackle a challenging problem that has not previously been solved in this generality. The methodology relies on the assumption that the basis that is used in reducing the dimensionality of the functional covariates, although parsimonious, is sufficiently large to include all the directions along which there is non-negligible association between the response and the functional predictors. In this regard, we recommend the use of the number of leading eigenfunctions based on a generous percentage of explained variance.

The approach is applicable to Gaussian or non-Gaussian responses and can directly accommodate additional vector covariates and non-linear effects of vector covariates, as well as multiple functional covariates observed on diverse sampling designs and with noise. The methods enable us to recover the full response trajectories. The methodology can be easily implemented by using existing freely available software. Numerical results show that the prediction performance of our approach is competitive or superior to existing alternative approaches in situations when alternatives are available. In spite of the increased flexibility, this method is computationally efficient; in fact it is orders of magnitude faster than its closest competitor. The methodology was motivated by and applied to an animal science study where we can predict the feed intake of lactating sows during their lactation period on the basis of daily temperature profiles.

This research leads to a few avenues for future research. First, the estimation of the regression coefficient is unidentifiable, which, although it does not affect the response prediction, limits any inference about the regression coefficient function. One possibility is to pursue this study along the ideas of Scheipl and Greven (2016). Second, our methodology relies on the implicit assumption that the current response is related to the current functional predictor only, i.e. $E[Y_{ij}|X_{i1}(\cdot), \dots, X_{in}(\cdot)] = E[Y_{ij}|X_{ij}(\cdot)]$. This assumption makes sense for our application, but it may not be reasonable for other situations. One possible approach to account for the past functional covariates is to consider a regression model like the historical functional linear model (see Malfait and Ramsay (2003), Pomann et al. (2016), Scheipl et al. (2015) and Kim et al. (2011)).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

Staicu's research was funded by National Science Foundation grant DMS 1454942 and National Institutes of Health grants R01 NS085211 and P01 CA142538-09. The data that were used originated from work supported in part by the North Carolina Agricultural Foundation, Raleigh, North Carolina. The authors thank the Associate Editor and two reviewers whose comments and suggestions greatly improved the paper.

References

- Baumgard L (2015) Assessing the impact of seasonal loss of productivity. National Pork Board Animal Science Webinar.
- Black J, Mullan B, Lorsch M and Giles L (1993) Lactation in the sow during heat stress. *Livestock Prod Sci.*, 35, 153–170.
- Bloemhof S, Mathur PK, Knol EF and Van der Waaij EH (2013) Effect of daily environmental temperature on farrowing rate and total born in dam line sows. *J. Anim. Sci.*, 91, 2667–2679. [PubMed: 23482580]
- Cai TT and Hall P (2006) Prediction in functional linear regression. *Ann. Statist.*, 34, 2159–2179.
- Cardot H, Ferraty F and Sarda P (1999) Functional linear model. *Statist. Probab. Lett.*, 45, 11–22.
- Cardot H, Ferraty F and Sarda P (2003) Spline estimators for the functional linear model. *Statist. Sin.*, 13, 571–591.
- Cardot H and Sarda P (2008) Varying-coefficient functional linear regression models. *Commun. Statist. Theory Meth.*, 37, 3186–3203.
- Di C-Z, Crainiceanu CM, Caffo BS and Punjabi NM (2009) Multilevel functional principal component analysis. *Ann. Appl. Statist.*, 3, 458–488.
- Fan J and Zhang J-T (2000) Two-step estimation of functional linear models with applications to longitudinal data. *J. R. Statist. Soc. B.*, 62, 303–322.
- Gertheiss J, Goldsmith J, Crainiceanu C and Greven S (2013) Longitudinal scalar-on-functions regression with application to tractography data. *Biostatistics*, 14, 447–461. [PubMed: 23292804]
- Goldsmith J, Crainiceanu CM, Caffo B and Reich DS (2012) Longitudinal penalized functional regression for cognitive outcomes on neuronal tract measurements. *Appl. Statist.*, 61, 453–469.
- Goldsmith J, Crainiceanu CM, Caffo BS and Reich DS (2011) Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage*, 57, 431–439. [PubMed: 21554962]
- Goldsmith J, Greven S and Crainiceanu CM (2013) Corrected confidence bands for functional data using principal components. *Biometrics*, 69, 41–51. [PubMed: 23003003]
- Hastie T and Tibshirani R (1993) Varying-coefficient models. *J. R. Statist. Soc. B.*, 55, 757–796.

- Hoover DR, Rice JA, Wu CO and Yang L-P (1998) Nonparametric smoothing estimates of time-varying coefficient models with longitudinal data. *Biometrika*, 85, 809–822.
- Huang L, Scheipl F, Goldsmith J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C and Reiss P (2015) refund: regression with functional data. R Package Version 0.1-13.
- Ivanescu AE, Staicu A-M, Scheipl F and Greven S (2015) Penalized function-on-function regression. *Computnl Stat.*, 30, 539–568.
- Johnston L, Ellis M, Libal G, Mayrose V and Weldon W (1999) Effect of room temperature and dietary amino acid concentration on performance of lactating sows: NCR-89 Committee on Swine Management. *J. Anim. Sci*, 77, 1638–1644. [PubMed: 10438007]
- Kato T (2013) *Perturbation Theory for Linear Operators*. New York: Springer Science and Business Media.
- Kim K, entürk D and Li R (2011) Recent history functional linear models for sparse longitudinal data. *J. Statist. Planng Inf*, 141, 1554–1566.
- Kim JS, Staicu A-M, Maity A, Carroll RJ and Ruppert D (2017) Additive function-on-function regression. *J. Computnl Graph. Statist*, 27, 234–244.
- Kundu MG, Harezlak J and Randolph TW (2016) Longitudinal functional models with structured penalties. *Statist. Modllng*, 16, 114–139.
- Malfait N and Ramsay JO (2003) The historical functional linear model. *Can. J. Statist*, 31, 115–128.
- Marx BD and Eilers PH (2005) Multidimensional penalized signal regression. *Technometrics*, 47, 13–22.
- McLean MW, Hooker G, Staicu A-M, Scheipl F and Ruppert D (2014) Functional generalized additive models. *J. Computnl Graph. Statist*, 23, 249–269.
- Morris JS and Carroll RJ (2006) Wavelet-based functional mixed models. *J. R. Statist. Soc. B*, 68, 179–199.
- Müller H-G (2005) Functional modelling and classification of longitudinal data. *Scand. J. Statist*, 32, 223–240.
- Müller H-G and Yao F (2012) Functional additive models. *J. Am. Statist. Ass*, 103, 1534–1544.
- Park SY and Staicu A-M (2015) Longitudinal functional data analysis. *Stat*, 4, 212–226. [PubMed: 26594358]
- Pomann G, Staicu A-M, Lobaton EJ, Mejia AF, Dewey BE, Reich DS, Sweeney E and Shinohara RS (2016) A lag functional linear model for prediction of magnetization transfer ratio in multiple sclerosis lesions. *Ann. Appl. Statist*, 10, 2324–2348.
- Ramsay J and Silverman BW (1997) *Functional Data Analysis*, 1st edn. Berlin: Springer.
- Ramsay JO and Silverman BW (2005) *Functional Data Analysis*, 2nd edn. Berlin: Springer.
- Reiss PT and Ogden RT (2007) Functional principal component regression and functional partial least squares. *J. Am. Statist. Ass*, 102, 984–996.
- Renaudeau D and Noblet J (2001) Effects of exposure to high ambient temperature and dietary protein level on sow milk production and performance of piglets. *J. Anim. Sci*, 79, 1540–1548. [PubMed: 11424692]
- Rosero DS, Boyd RD, McCulley M, Odle J and van Heugten E (2016) Essential fatty acid supplementation during lactation is required to maximize the subsequent reproductive performance of the modern sow. *Anim. Reprodn Sci*, 168, 151–163.
- Ruppert D, Wand MP and Carroll RJ (2003) *Semiparametric Regression*. New York: Cambridge University Press.
- Scheipl F and Greven S (2016) Identifiability in penalized function-on-function regression models. *Electron. J. Statist*, 10, 495–526.
- Scheipl F, Staicu A-M and Greven S (2015) Functional additive mixed models. *J. Computnl Graph. Statist*, 24, 477–501.
- Silverman BW (1985) Some aspects of the spline smoothing approach to non-parametric regression curve fitting (with discussion). *J. R. Statist. Soc. B*, 47, 1–52.
- Staicu A-M, Crainiceanu CM and Carroll RJ (2010) Fast methods for spatially correlated multilevel functional data. *Biostatistics*, 11, 177–194. [PubMed: 20089508]

- Staniswalis JG and Lee JJ (1998) Nonparametric regression analysis of longitudinal data. *J. Am. Statist. Ass.*, 93, 1403–1418.
- St-Pierre N, Cobanov B and Schnitkey G (2003) Economic losses from heat stress by US livestock industries. *J Dairy Sci.*, 86, E52–E77.
- Wahba G (1983) Bayesian “confidence intervals” for the cross-validated smoothing spline. *J. R. Statist. Soc. B*, 45, 133–150.
- Wand MP (2003) Smoothing and mixed models. *Computnl Statist.*, 18, 223–249.
- Wood S (2001) mgcv: GAMS and generalized ridge regression for R. *R News*, 1, 20–25.
- Wood S (2006a) *Generalized Additive Models: an Introduction with R*. Boca Raton: CRC Press.
- Wood S (2006b) On confidence intervals for generalized additive models based on penalized regression splines. *Aust. New Zeal. J. Statist*, 48, 445–464.
- Wood SN (2011) Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *J R. Statist. Soc. B*, 73, 3–36.
- Wood SN, Pya N and Säfken B (2016) Smoothing parameter and model selection for general smooth models. *J. Am. Statist. Ass.*, 111, 1548–1563.
- Xiao L, Li Y and Ruppert D (2013) Fast bivariate P -splines: the sandwich smoother. *J. R. Statist. Soc. B*, 75, 577–599.
- Yao F, Müller H-G and Wang J-L (2005) Functional data analysis for sparse longitudinal data. *J. Am. Statist. Ass.*, 100, 577–590.
- Zhou L, Huang JZ and Carroll RJ (2008) Joint modelling of paired sparse functional data using principal components. *Biometrika*, 95, 601–619. [PubMed: 19396364]

Fig. 1. Data for a randomly chosen sow: (a) total feed intake and (b) minute-by-minute temperature profiles (the colour code corresponds to two arbitrarily chosen days)

Fig. 2.

Parameter estimates in the lactation sow application: (a) estimated intercept function for the old (.....) and young (---) sows with 95% pointwise confidence intervals; (b) estimated regression coefficient $\hat{\gamma}(\cdot, t)$ for all lactation days $t = 1, \dots, 21$

Fig. 3. Predicted full feed intake trajectories for (a), (b) two young and (c) one old sows: — — — —, 95% pointwise prediction bands based on LDFR

Table 1.Median response prediction accuracy and computing time for LDFR and LPFR and LPEER[†]

<i>IN_{PE}</i>			<i>OUT_{PE}</i>			<i>Computing time (s)</i>		
<i>LDFR</i>	<i>LPEER</i>	<i>LPFR</i>	<i>LDFR</i>	<i>LPEER</i>	<i>LPFR</i>	<i>LDFR</i>	<i>LPEER</i>	<i>LPFR</i>
<i>(a)</i>								
1.26 (0.01)	1.32 (0.01)	1.40 (0.01)	1.60 (0.04)	1.61 (0.02)	1.69 (0.03)	77.23 (18.92)	4243.31 (1084.48)	534.01 (132.08)
<i>(b)</i>								
1.33 (0.01)	1.39 (0.01)	1.45 (0.01)	1.45 (0.05)	1.47 (0.02)	1.54 (0.02)	66.56 (8.08)	4548.67 (872.37)	724.33 (284.67)

[†]Results refer to prediction for (a) their last 50% lactation days observed or for (b) 20% random days that were not included in the model fitting. Interquartile ranges are given in parentheses.

Table 2.

Gaussian responses D1, part (ii), C1, B1 and E1 [†]

Results for $\delta = 1$						Results for $\delta = 5$					
IN_{PE}			OUT_{PE}			$RMPE, LDFR$			$Run\ time\ (s)$		
LDFR	LPEER	LDFR	LDFR	LPEER	LDFR	LDFR	LPEER	LDFR	LDFR	LPEER	LDFR
$I = 100$											
0.62 (0.04)	0.82 (0.03)	0.78 (0.09)	0.94 (0.05)	0.75 (0.08)	14.74	380.75	1.35 (0.16)	1.78 (0.15)	1.81 (0.35)	2.05 (0.15)	1.74 (0.30)
$I = 300$											
0.61 (0.03)	0.81 (0.02)	0.73 (0.07)	0.92 (0.03)	0.71 (0.07)	97.00	1270.51	1.29 (0.13)	1.78 (0.07)	1.64 (0.29)	2.02 (0.07)	1.58 (0.27)
											63.75
											4542.27

[†]The model is fitted assuming a compound symmetric type of dependence (correct) structure. Median prediction errors and interquartile ranges in parentheses are reported.

Table 3.

Gaussian responses D1, part (ii), C2, B1 and E1 [†]

		Results for 1 - α = 0.95			Results for 1 - α = 0.90		
		$\delta = 1$	$\delta = 5$	$\delta = 10$	$\delta = 1$	$\delta = 5$	$\delta = 10$
<i>I</i> = 100							
Y_i	0.97 (0.00) [1.33]	0.99 (0.00) [2.72]	0.99 (0.00) [4.02]	0.93 (0.01) [1.12]	0.98 (0.01) [2.29]	0.98 (0.01) [3.37]	
Y_i^*	0.87 (0.03) [2.35]	0.89 (0.06) [3.33]	0.90 (0.07) [4.55]	0.79 (0.03) [1.97]	0.83 (0.07) [2.80]	0.85 (0.08) [3.82]	
<i>I</i> = 300							
Y_i	0.97 (0.00) [1.31]	0.99 (0.00) [2.61]	0.99 (0.00) [3.84]	0.93 (0.01) [1.10]	0.98 (0.00) [2.19]	0.98 (0.00) [3.22]	
Y_i^*	0.87 (0.02) [2.35]	0.89 (0.05) [3.23]	0.90 (0.06) [4.39]	0.79 (0.03) [1.97]	0.83 (0.05) [2.71]	0.84 (0.07) [3.68]	

[†]The model is fitted assuming a CS type of dependence (correct) structure. Reported are the average coverage probabilities of 95% and 90% pointwise prediction bands, standard errors (in parentheses) and average lengths (in square brackets) for the observed (Y_i) and unobserved subjects (Y_i^*). Results are based on 1000 Monte Carlo simulations.