Response-Adaptive Regression for Longitudinal Data

Shuang Wu & Hans-Georg Müller

Department of Statistics, University of California, Davis, CA 95616 USA

Version July 2010

ABSTRACT

We propose a response-adaptive model for functional linear regression, which is adapted to sparsely sampled longitudinal responses. Our method aims at predicting response trajectories and models the regression relationship by directly conditioning the sparse and irregular observations of the response on the predictor, which can be of scalar, vector or functional type. This obliterates the need to model the response trajectories, a task that is challenging for sparse longitudinal data and was previously required for functional regression implementations for longitudinal data. The proposed approach turns out to be superior compared to previous functional regression approaches in terms of prediction error. It encompasses a variety of regression settings that are relevant for the functional modeling of longitudinal data in the life sciences. The improved prediction of response trajectories with the proposed response-adaptive approach is illustrated for a longitudinal study of Kiwi weight growth and by an analysis of the dynamic relationship between viral load and CD4 cell counts observed in AIDS clinical trials.

Key Words: AIDS clinical trials, Functional data analysis, Functional linear regression, Growth curves, Kiwi growth, Prediction, Repeated measurements, Sparse data, Viral load.

Correspondence: Shuang Wu, *Email*: swu@wald.ucdavis.edu, *Address*: Department of Statistics, One Shields Ave., University of California, Davis, CA 95616, USA.

1. Introduction

Functional linear regression models have been widely developed over recent years (see Ramsay and Silverman, 2005). Such models come in many flavors, as either predictors, responses or both may be functional. We interpret here "functional" in the wide sense, i.e., it suffices that the data observed for one subject can be thought of as having been generated by underlying smooth trajectories. For longitudinal studies in the life sciences, completely observed smooth trajectories are rather the exception; a much more common situation is that one has noisy repeated measurements that are obtained at irregular time points. We focus in the following on the case where the responses in a regression situation are of this longitudinal type.

Consider a clinical study regarding viral load and CD4 cell counts, which are important markers for evaluating antiviral therapies in treating AIDS. We are interested in the dependence of trajectories of viral load, which often is hard to measure, on trajectories of CD4 cell counts. Typically, viral load and CD4 cell counts are scheduled to be measured simultaneously on pre-specified days in such studies, but due to missing and modified visiting times, the actual data are often unbalanced, with varying timing of the measurements among different patients. In the clinical trial we study, the number of measurements available for each subject ranges from 4 to 8 with a median of 7, representing a sparse data scheme. Individual trajectories of CD4 cell counts and viral load (on the $\log_{10}$ scale) are shown in Figure 1. Generally, the cross-sectional trend for CD4 cell counts is upwards and that for viral load is downwards, reflecting the effect of the treatment. However, these graphs do not reveal the nature of the functional relationship between pairs of CD4 and viral load trajectories that pertain to the same subject. This relationship is of interest for predicting the time course of viral load from that of CD4 counts on at the individual level, as well as for a better understanding of the underlying time-dynamic relationship between CD4 counts and viral load. These considerations motivate a functional regression analysis and careful implementation of such an analysis is of interest for the case of sparse and irregular data

1

as presented in this example (see Section 4).

Among functional regression models, linear models with functional predictors and scalar responses have been the most studied model type to date (Cardot et al., 1999, 2003, 2007; James and Silverman, 2005; Cai and Hall, 2006; Hall and Horowitz, 2007; Crambes et al., 2009). The case of functional responses was introduced in Ramsay and Dalzell (1991), see also Faraway (1997). Investigations of functional linear regression models with functional responses typically employ a basis representation of the response trajectories (see, e.g. Rice and Wu, 2001; Sy et al., 1997). Specifically aiming at regression analysis for data from longitudinal studies, where measurements often are sparsely sampled and noisy, Yao et al. (2005b) proposed a version of the functional linear regression model, in which both predictors and responses are functions.

Using functional principal component analysis, the functional linear model can be decomposed into a series of simple linear regressions between the functional principal components (FPCs) of the response trajectories and those of the predictor trajectories, both of which may be estimated through conditional expectations (Yao et al., 2005a; Müller et al., 2008). While this model is quite flexible and allows for widely varying patterns regarding the timing of the measurements that are obtained for predictor and response trajectories, using the conditional response FPCs in the estimating process is suboptimal, because one can only estimate the functional principal component FPCs conditionally on the observed response measurements, as the true FPCs are not available and cannot be directly targeted in the case of sparse response measurements. This initial conditioning step has nothing to do with the targeted regression relation and the conditional FPCs contain less information about the regression than the original data, so that this conditioning step is not advisable if the aim is prediction in a functional regression framework.

Motivated by these considerations, we propose response-adaptive functional regression (RARE) as an alternative. The guiding principle of RARE is to model the regression relationship by conditioning the sparsely observed response measurements directly on the predictors, thereby omitting

the preliminary and actually counterproductive intermediate step of modeling the response trajectories. In this way, one bypasses the usually employed basis or other representation of the response trajectories, making it possible to target the regression relationship of the response trajectories with the predictors directly. We demonstrate in simulations and applications to growth curves and AIDS longitudinal studies that the proposed RARE approach leads to improved prediction of response trajectories for predictors that are scalar, multivariate or functional.

The paper is organized as follows. In Section 2 we provide a brief review of established functional linear regression methods. The proposed response-adaptive regression model and estimators are introduced in Section 3. Simulation studies that demonstrate the improvements in prediction for the RARE model can be found in Section 4. This is followed in Section 5 by a report on applications of the proposed model to Kiwi weight growth data and an analysis of the dynamic relationship between viral load and CD4 cell counts in AIDS clinical trials.

2. Review of Functional Linear Regression Model for Sparse and Irregular Data

In the functional linear regression models we consider here, predictors X and responses Y are assumed to be non-stationary smooth random functions in $L^2(\mathcal{S})$ and $L^2(\mathcal{T})$, respectively, with smooth mean functions $\mu_X(s) = \mathrm{E}X(s)$, $\mu_Y(t) = \mathrm{E}Y(t)$ and smooth auto-covariance functions $G_X(s_1, s_2) = \mathrm{cov}(X(s_1), X(s_2))$, $G_Y(t_1, t_2) = \mathrm{cov}(Y(t_1), Y(t_2))$. It is assumed that processes X and Y possess Karhunen-Loève expansions (Ash and Gardner, 1975) with representations

$$X(s) = \mu_X(s) + \sum_{j=1}^{\infty} \xi_j \phi_j(s), \quad Y(t) = \mu_Y(t) + \sum_{k=1}^{\infty} \zeta_k \psi_k(t), \tag{1}$$

where ϕ_j and ψ_k are sequences of orthonormal eigenfunctions of associated auto-covariance operators that form a basis of the function space and are associated with sequences of non-increasing eigenvalues λ_j and τ_k, satisfying $\sum_j \lambda_j < \infty$, $\sum_k \tau_k < \infty$, $G_X(s_1, s_2) = \sum_j \lambda_j \phi_j(s_1) \phi_j(s_2)$, $s_1, s_2 \in \mathcal{S}$ and $G_Y(t_1, t_2) = \sum_k \tau_k \psi_k(t_1) \psi_k(t_2)$, $t_1, t_2 \in \mathcal{T}$. The coefficients ξ_j and ζ_k are referred to as functional principal components (FPCs). They are sequences of uncorrelated random vari-

ables, respectively, with means $E(\xi_j) = 0$, $E(\zeta_k) = 0$ and variances $var(\xi_j) = \lambda_j$, $var(\zeta_k) = \tau_k$.

The functional linear regression model with response function Y and functional predictor X is

$$E(Y(t)|X) = \beta_0(t) + \int_{\mathcal{S}} \beta(s,t)X(s)ds, \tag{2}$$

where the bivariate regression coefficient function $\beta(s,t)$ is smooth and square integrable. Under regularity conditions (He et al., 2000), the regression coefficient surface β has the representation

$$\beta(s,t) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \frac{E(\xi_j\zeta_k)}{E(\xi_j^2)}\phi_j(s)\psi_k(t) = \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \beta_{jk}\phi_j(s)\psi_k(t). \tag{3}$$

In view of (1) and (3), model (2) can be further written as

$$E(Y(t)|X) = \alpha(t) + \sum_{k=1}^{\infty}\sum_{j=1}^{\infty} \beta_{jk}\xi_j\psi_k(t), \tag{4}$$

where $\alpha(t) = \beta_0(t) + \int_{\mathcal{S}} \beta(s,t)\mu_X(s)ds$ serves as the regression intercept.

Since the information contained in predictors X and responses Y is completely reflected by the sequences of FPCs $\{\xi_1, \xi_2, ...\}$ and $\{\zeta_1, \zeta_2, ..\}$ and the FPCs for X, respectively Y, are uncorrelated random variables, model (4) can be decomposed into a series of simple linear regressions between the FPCs of response processes and those of predictor processes, i.e., $E(\zeta_k|\xi_j) = E\big[E(\zeta_k|X)|\xi_j\big] = E\big[E(\zeta_k|\xi_j, \xi_2, ...)|\xi_j\big] = \beta_{jk}\xi_j$. Therefore, the coefficients β_{jk} in the representation of $\beta(s,t)$ are the slopes of simple linear regressions through the origin between FPCs ζ_k and ξ_j.

Estimates of the mean functions and covariance functions for processes X and Y may be obtained by applying one- and two-dimensional local linear smoothers to the aggregated data from all subjects, respectively. The eigenfunctions and eigenvalues are then estimated through spectral decomposition of the discretized covariance surface estimates (Rice and Silverman, 1991). From (1), the FPCs can be represented as $\xi_j = \int(X(s)-\mu_X(s))\phi_j(s)ds$ and $\zeta_k = \int(Y(t)-\mu_Y(t))\psi_k(t)dt$, respectively. These FPCs have traditionally been estimated by numerical integration. This method works well if the trajectories are fully or at least densely observed, but fails to provide reasonable approximations to the true FPCs when only sparse and irregular data are available.

In the latter case, we may obtain FPCs ξ_j and ζ_k by functional principal component analysis through conditional expectation (PACE), as described in Yao et al. (2005a).

With estimates $\hat{\phi}_j$, $\hat{\psi}_k$, $\hat{\xi}_{ij}$, $\hat{\zeta}_{ik}$ in hand, we infer the regression coefficient function $\beta(s,t)$ from

$$\hat{\beta}(s,t) = \sum_{k=1}^{K} \sum_{j=1}^{J} \hat{\beta}_{jk} \hat{\phi}_j(s) \hat{\psi}_k(t), \tag{5}$$

where $\hat{\beta}_{jk}$ is the estimated slope of regressing $\hat{\zeta}_k$ on $\hat{\xi}_j$ without intercept, for $j = 1, ..., J$ and $k = 1, ..., K$ (Müller et al., 2008); positive integers J and K are the number of principal components included in the model for predictors X and responses Y, respectively. Methods to choose J and K include one-curve-leave-out cross-validation, AIC and BIC type of criteria (Yao et al., 2005a) and FVE (fraction of variation explained; see Liu and Müller, 2008). For a newly observed subject with sparse and irregular measurements, the predicted response trajectory becomes

$$\hat{Y}_{J,K}^*(t) = \hat{\alpha}(t) + \sum_{k=1}^{K} \sum_{j=1}^{J} \hat{\beta}_{jk} \hat{\xi}_j^* \hat{\psi}_k(t), \tag{6}$$

where $\hat{\alpha}(t) = \hat{\mu}_Y(t)$ and $\hat{\xi}_j^*$, $j = 1, ..., J$, are the FPCs of the newly observed predictor trajectory, which can be estimated through conditional expectation. If the predictors are scalars or vectors $\boldsymbol{z} = (Z_1, ..., Z_P)$, $P \geq 1$, with means $\boldsymbol{\mu_z} = (\mu_{Z_1}, ..., \mu_{Z_P})$, the regression model can be written as

$$\mathrm{E}(Y(t)|\boldsymbol{z}) = \gamma(t) + \sum_{p=1}^{P} \delta_p(t) Z_p, \tag{7}$$

where $\delta_p(t)$ is the regression coefficient function associated with the pth predictor, $p = 1, ..., P$.

As in the case of functional predictors, one may employ basis representations $\delta_p(t) = \sum_{k=1}^{\infty} \delta_{pk} \psi_k(t)$, $p = 1, ..., P$, and then decompose (7) into a series of multiple linear regressions of the response FPCs against the vector predictors

$$\mathrm{E}(\zeta_k|\boldsymbol{z}) = \sum_{p=1}^{P} \delta_{pk}(Z_p - \mu_{Z_p}), \quad k = 1, 2, \dots . \tag{8}$$

Since the predictors may be correlated with each other, this model cannot be further decomposed into simple linear regressions between each ζ_k and Z_p, as was possible by using functional

principal components of predictor functions due to their uncorrelatedness. Otherwise, estimation and prediction procedures are similar to those for the case of functional predictors.

The strength of this approach to linear functional regression lies in its flexibility to handle varying patterns of randomness in the timing of the measurements of predictor and response functions and its capability to handle sparse measurements that may be contaminated with additional measurement errors. However, the use of response FPCs $\hat{\zeta}_k$ in estimating $\hat{\beta}_{jk}$ (resp. $\hat{\delta}_{pk}$ in the case of scalar or vector predictors) is suboptimal, because the $\hat{\zeta}_k$ are estimates for the conditional expectations of the true FPCs ζ_k, given the observed sparse and noisy response measurements, while the true response FPCs ζ_k are not estimable in the case of sparse data. The conditioning step to obtain $\hat{\zeta}_k$ leads to information loss and to inefficient estimation and prediction in the framework of functional regression. These considerations motivate the response-adaptive regression approach for the case of sparsely sampled functional responses.

3. Response-Adaptive Regression for Sparsely Sampled Functional Responses

3.1. Response-Adaptive Regression Model

The central idea of the response-adaptive RARE approach is to model the regression relationship through conditioning the observed response measurements directly on the predictors, rather than conditioning the response FPCs on the predictor. Our objective is to predict an unknown response trajectory based on measurements of a new subject. We assume that the information available for the functional responses consists of noisy, sparse and irregular repeated measurements of the response trajectories. We first consider the case of functional predictors. The repeated observations that are available for these functional predictors can be either dense regular or sparse irregular; we provide a more detailed description for the latter case.

Let U_{il} be the repeated measurements of the ith subject for the predictor and V_{im} for the response, observed at times $S_{i1}, ..., S_{iL_i}$ and $T_{i1}, ..., T_{iM_i}$, respectively. The distributions of these

measurement times need not be uniform, but are assumed to have a density that is bounded away from zero, and the observed measurement times are assumed to be independent from the shapes of the trajectories. The number of measurements available for the i-th subject are denoted by L_i (for predictors) and M_i (for responses) and are assumed to be i.i.d. random variables, distributed as count variables L and M, which may be correlated but are independent of all other random observations. It is assumed that $U_{il} = X(S_{il}) + \varepsilon_{il}$, $S_{il} \in \mathcal{S}$, $1 \le l \le L_i$ and $V_{im} = Y(T_{im}) + \epsilon_{im}$, $T_{im} \in \mathcal{T}$, $1 \le m \le M_i$, $1 \le i \le n$, where the measurement errors ε_{il} and ϵ_{im} are i.i.d. with zero means and constant variances, σ_{ε}^2 and σ_{ϵ}^2, respectively.

Conditioning on the observed number of measurements L_i and M_i, $1 \le i \le n$, model (4), adapted to the vector $\boldsymbol{v}_i = (V_{i1}, ..., V_{iM_i})^T$ of sparsely observed responses for the ith subject, is

$$\mathrm{E}(\boldsymbol{v}_i|X) = \mathrm{E}(\boldsymbol{v}_i|\, \xi_1, \xi_2, ...) = \left(\alpha(T_{im}) + \sum_{k}^{\infty} \sum_{j}^{\infty} \beta_{jk} \xi_j \psi_k(T_{im}) \right)^T_{1 \le m \le M_i}, \quad i = 1, ..., n. \quad (9)$$

If the numbers of principal components chosen to represent the predictor and response processes are J and K, respectively, model (9) can be approximated by

$$\mathrm{E}(\boldsymbol{v}_i|\, \xi_1, ..., \xi_J) = \left(\alpha(T_{im}) + \sum_{k}^{K} \sum_{j}^{J} \beta_{jk} \xi_j \psi_k(T_{im}) \right)^T_{1 \le m \le M_i} = \widetilde{\boldsymbol{\alpha}}_i + \boldsymbol{D}_i \widetilde{\boldsymbol{\beta}}, \quad i = 1, ...n, \quad (10)$$

with $\widetilde{\boldsymbol{\alpha}}_i = (\alpha(T_{i1}), ..., \alpha(T_{iM_i}))^T$, $\widetilde{\boldsymbol{\beta}} = (\beta_{jk})_{k=1,...,K;\, j=1,...,J} = (\beta_{11}, ..., \beta_{1K}, ..., \beta_{J1}, ...\beta_{JK})^T$ and design matrix $\boldsymbol{D}_i = \widetilde{\boldsymbol{\xi}} \otimes \boldsymbol{C}_i$, where $\otimes$ is the Kronecker product, $\widetilde{\boldsymbol{\xi}} = (\xi_1, ..., \xi_J)$ and $\boldsymbol{C}_i$ is a $M_i \times K$ matrix with elements $\left(\psi_k(T_{im}) \right)_{1 \le m \le M_i;\, 1 \le k \le K}$.

Considering the observed measurements from all subjects, let $\boldsymbol{v} = (\boldsymbol{v}_1^T, ..., \boldsymbol{v}_n^T)^T$, $\widetilde{\boldsymbol{\alpha}} = (\widetilde{\boldsymbol{\alpha}}_1^T, ..., \widetilde{\boldsymbol{\alpha}}_n^T)^T$ and $\boldsymbol{D} = (\boldsymbol{D}_1^T, ..., \boldsymbol{D}_n^T)^T$ and stack (10) for all n subjects, to obtain

$$\mathrm{E}(\boldsymbol{v}|\, \xi_1, ..., \xi_J) = \widetilde{\boldsymbol{\alpha}} + \boldsymbol{D} \widetilde{\boldsymbol{\beta}}. \quad (11)$$

This is a multiple linear regression model with response $\boldsymbol{v}$ and design matrix $\boldsymbol{D}$. We then need to estimate the coefficient $\widetilde{\boldsymbol{\beta}}$ and the intercept $\widetilde{\boldsymbol{\alpha}}$ to obtain the regression coefficient surface estimate $\hat{\beta}(s, t)$ and obtain predicted responses for newly observed predictors.

3.2. Estimation of the Model Components

Suppose the unknown intercept function $\alpha(t)$ is initially approximated by an arbitrary function $\eta(t)$. Since $\{\psi_k(t),\ k = 1, 2, ...\}$ is an orthogonal basis of $L^2(\mathcal{T})$, we may expand the difference in the form $\alpha(t) - \eta(t) = \sum_{k=1}^{\infty} \theta_k \psi_k(t)$. We further assume that for a sufficiently good approximation function η, $\alpha(t) - \eta(t)$ can be well approximated by the first $\widetilde{K}$ terms of this expansion. Choosing $\eta(t) = \hat{\mu}_Y(t)$, model (10) becomes

$$
\mathrm{E}(\boldsymbol{v}_i | \xi_1, ..., \xi_J) = \left(\hat{\mu}_Y(T_{im}) + \sum_{k}^{\widetilde{K}} \theta_k \psi_k(T_{im}) + \sum_{k}^{K} \sum_{j}^{J} \beta_{jk} \xi_j \psi_k(T_{im}) \right)_{1 \le m \le M_i}^{T}. \tag{12}
$$

Selecting $\widetilde{K} = K$, we may simplify the above model to obtain

$$
\mathrm{E}(\boldsymbol{v}_i | \xi_1, ..., \xi_J) = \widetilde{\boldsymbol{\mu}}_{Yi} + \boldsymbol{R}_i (\widetilde{\boldsymbol{\theta}}^T, \widetilde{\boldsymbol{\beta}}^T)^T, \tag{13}
$$

where $\widetilde{\boldsymbol{\mu}}_{Yi} = (\hat{\mu}_Y(T_{i1}), ..., \hat{\mu}_Y(T_{iM_i}))^T$, $\widetilde{\boldsymbol{\theta}} = (\theta_1, ..., \theta_K)^T$ and $\boldsymbol{R}_i = (\boldsymbol{C}_i, \boldsymbol{D}_i)$. With $\widetilde{\boldsymbol{\mu}}_Y = (\widetilde{\boldsymbol{\mu}}_{Y1}^T, ..., \widetilde{\boldsymbol{\mu}}_{Yn}^T)^T$ and $\boldsymbol{R} = (\boldsymbol{R}_1^T, ..., \boldsymbol{R}_n^T)^T$, stacking models (13) leads to

$$
\mathrm{E}(\boldsymbol{v} | \xi_1, ..., \xi_J) = \widetilde{\boldsymbol{\mu}}_Y + \boldsymbol{R} (\widetilde{\boldsymbol{\theta}}^T, \widetilde{\boldsymbol{\beta}}^T)^T. \tag{14}
$$

While the design matrix $\boldsymbol{R} = (\boldsymbol{R}_1^T, ..., \boldsymbol{R}_n^T)^T$ contains unknown elements ξ_j and ψ_k through matrices $\boldsymbol{C}_i$ and $\boldsymbol{D}_i$, $j = 1, ..., J$, $k = 1, ..., K$, $i = 1, ..., n$, these matrices can be easily estimated by plugging in estimates $\hat{\psi}_k$ of response eigenfunctions and $\hat{\xi}_j$ of predictor FPCs, whence the estimated design matrix becomes $\widehat{\boldsymbol{R}} = (\widehat{\boldsymbol{R}}_1^T, ..., \widehat{\boldsymbol{R}}_n^T)^T$ with $\widehat{\boldsymbol{R}}_i = (\widehat{\boldsymbol{C}}_i, \widehat{\boldsymbol{D}}_i)$, $i = 1, ..., n$. The required estimates $\hat{\xi}_j$ and $\hat{\psi}_k$, and also an estimate $\hat{\mu}_Y$ of the mean function μ_Y, can easily be obtained by applying functional principal component analysis to predictors and responses, respectively, as described in Section 2. The included number of principal components J for predictors and K for responses can be chosen by any of the selection methods reviewed in Section 2, in addition to one-subject-leave out cross-validation based on predicting responses. For sparsely observed predictors, the FPCs $\hat{\xi}_j$ are estimates of the predictor FPCs conditional on

the observed data. Since the conditional FPCs may deviate from the true FPCs in very sparse situations, this may be lead to smaller improvement in the prediction of the responses by RARE as compared to situations with densely observed predictors.

The fact that the components of the response vectors $\boldsymbol{v}$ collected for a subject are correlated, suggests to use weighted least squares to implement model (14). The optimal least squares weight matrix is $\boldsymbol{\Sigma} = \mathrm{cov}(\boldsymbol{v}, \boldsymbol{v}) = \mathrm{diag}\{\boldsymbol{\Sigma}_1, ..., \boldsymbol{\Sigma}_n\}$, a block diagonal matrix, where the i-th block $\boldsymbol{\Sigma}_i = \mathrm{cov}(\boldsymbol{v}_i, \boldsymbol{v}_i)$ corresponds to the covariance matrix for the ith subject. These matrices can be easily estimated by plugging in the estimated covariance surface of the response trajectory Y, which can be obtained as described in Section 2, in conjunction with a modified estimator for the variances along the diagonal as described in Appendix A.1 of Liu and Müller (2009). The coefficients of model (14) are then computed through weighted least squares, yielding

$$(\widetilde{\boldsymbol{\theta}}^T, \widetilde{\boldsymbol{\beta}}^T)^T = (\widehat{\boldsymbol{R}}^T \widehat{\boldsymbol{\Sigma}}^{-1} \widehat{\boldsymbol{R}})^{-1} \widehat{\boldsymbol{R}}^T \widehat{\boldsymbol{\Sigma}}^{-1} (\boldsymbol{v} - \widetilde{\boldsymbol{\mu}}_Y). \tag{15}$$

The estimate of the intercept function $\alpha(t)$ in model (4) is $\hat{\alpha}(t) = \hat{\mu}_Y(t) + \sum_{k=1}^{K} \hat{\theta}_k \hat{\psi}_k(t)$ and the estimate of the coefficient function $\beta(s,t)$ is given by (5), with $\hat{\beta}_{jk}$ obtained as in (15). Analogous to the coefficient of determination R^2 in standard linear regression (Draper and Smith, 1998), the strength of a functional linear regression may be quantified by

$$R^2 = 1 - \frac{\sum_{i=1}^{n} \int_{\mathcal{T}} [Y_i(t) - \mathrm{E}(Y_i(t)|X_i)]^2 dt}{\sum_{i=1}^{n} \int_{\mathcal{T}} [Y_i(t) - \mu_Y(t)]^2 dt}, \tag{16}$$

see Müller and Yao (2008). For sparsely observed responses, (16) suggests a sample based "quasi" R^2 measure, with the following empirical version,

$$Q = 1 - \frac{\sum_{i=1}^{n} \sum_{m=1}^{M_i} [V_{im} - \widehat{\mathrm{E}}(Y_i(T_{im})|X_i)]^2 / M_i}{\sum_{i=1}^{n} \sum_{m=1}^{M_i} [V_{im} - \mu_Y(T_{im})]^2 / M_i}, \tag{17}$$

where $\widehat{\mathrm{E}}(Y_i(t)|X_i) = \hat{\alpha}(t) + \sum_{k=1}^{K} \sum_{j=1}^{J} \hat{\beta}_{jk} \hat{\xi}_{ij} \hat{\psi}_k(t)$ is the fitted response trajectory. For a new subject where one has a vector of repeated measurements for the predictor trajectory X^*, the predicted response trajectory Y^* is then given by (6), using estimates $\hat{\alpha}(t)$ and $\hat{\beta}_{jk}$.

3.3. Model for Scalar or Vector Predictors

We now consider the RARE model for the case of scalar or vector predictors. From (7), one has

$$\mathrm{E}(Y(t)|\boldsymbol{z}) = \gamma(t) + \sum_{k=1}^{\infty} \sum_{p=1}^{P} \delta_{pk} Z_p \psi_k(t). \tag{18}$$

Analogous to (10), model (18) adapted to sparsely sampled responses for the ith subject is

$$\mathrm{E}(\boldsymbol{v}_i|\boldsymbol{z}) = \widetilde{\boldsymbol{\gamma}}_i + \boldsymbol{W}_i \widetilde{\boldsymbol{\delta}}, \quad i = 1, ...n, \tag{19}$$

where $\widetilde{\boldsymbol{\gamma}}_i = (\gamma(T_{i1}), ..., \gamma(T_{iM_i}))^T$, $\widetilde{\boldsymbol{\delta}} = (\delta_{pk})_{k=1,...,K;\, p=1,...,P} = (\delta_{11}, ..., \delta_{1K}, ..., \delta_{P1}, ...\delta_{PK})^T$, and where K is the number of principal components that is chosen to represent the response Y and the design matrix $\boldsymbol{W}_i = \boldsymbol{z} \otimes \boldsymbol{C}_i$.

Stacking equations (19) across all subjects leads to the RARE model for the case of scalar or vector predictors. The estimation of model components and predictions for newly observed predictor vectors $\boldsymbol{z}^*$ is similar to the procedures described in Section 3.2; details are omitted.

4. Simulation Studies

Simulation studies were carried out to compare the proposed response-adaptive regression (RARE) model and the functional linear regression (FLIN) model reviewed in Section 2 in terms of prediction error. In each simulation, we constructed 200 pairs of responses and predictors, where the response trajectories were sparsely observed, with the number of measurements for each response uniformly selected from {2,3,4,5} to reflect a high level of sparseness. Once the number of observations was determined, the locations of the measurement times were generated uniformly on $[0, 5]$, separately for each response. We also considered responses with denser observations, with predictive performance for these cases reported in the supplementary material.

Simulations for the vector predictor case. We generated the vector predictors $\boldsymbol{z} = (Z_1, Z_2)$ as multivariate normal random variables with means $\mu_{Z_1} = 3$ and $\mu_{Z_2} = 2$, variances $\sigma^2_{Z_1} = 4$ and $\sigma^2_{Z_2} = 1$ and correlation $\rho = 0.35$. Setting $\delta_{11} = 1$, $\delta_{12} = 0.8$, $\delta_{21} = -1$ and $\delta_{22} = 0.5$,

the regression coefficient functions were constructed as $\delta_p(t) = \sum_{k=1}^{2} \delta_{pk}\psi_k(t)$, $p = 1, 2$, with $\psi_1(t) = -\sqrt{2/5}\cos(2\pi t/5)$ and $\psi_2(t) = \sqrt{2/5}\sin(2\pi t/5)$, $0 \leq t \leq 5$. We chose the response trajectories $Y(t)$ to be the same as $\mathrm{E}(Y(t)|\boldsymbol{z})$, which were computed following (7) with predictor dimension $p = 2$ and $\gamma(t) = t + \sin(t)$, $t \in [0, 5]$. The sparse observations for each response were then obtained as $V_{im} = Y(T_{im}) + \epsilon_{im}$, $1 \leq m \leq M_i$, $i = 1, ..., n$, with additive measurement errors $\epsilon_{im} \sim \mathcal{N}(0, 0.25)$.

Simulations for the functional predictor case. Here we considered two scenarios: (a) Pairing sparsely sampled response trajectories with densely and regularly sampled predictor functions; (b) Pairing sparsely sampled responses with sparsely sampled predictors. For scenario (a), measurement locations for the predictor trajectories were generated equidistantly on $[0, 10]$, with 20 measurements for each predictor. For scenario (b), the number of observations made for each predictor trajectory was uniformly selected from $\{2,3,4,5\}$ and measurement locations were then selected on $[0, 10]$ following a uniform distribution.

In both functional predictor scenarios (a) and (b), we choose the mean function as $\mu_X(s) = s + 2\sin(s)$, $0 \leq s \leq 10$, and a covariance function constructed from three eigenfunctions $\phi_1(s) = -\cos(\pi s/5)/\sqrt{5}$, $\phi_2(s) = \sin(\pi s/5)/\sqrt{5}$ and $\phi_3(s) = -\cos(2\pi s/5)/\sqrt{5}$, with eigenvalues $\lambda_1 = 9$, $\lambda_2 = 4$ and $\lambda_3 = 1$, respectively. Each predictor trajectory was generated with FPCs $\xi_{ij} \sim \mathcal{N}(0, \lambda_j)$, $j = 1, 2, 3$, and the observations $U_{il} = X(S_{il}) + \varepsilon_{il}$, $1 \leq l \leq L_i$, $i = 1, ..., n$, were additively contaminated with measurement errors $\varepsilon_{il} \sim \mathcal{N}(0, 0.25)$. With $(\beta_{jk})_{k=1,2,3; j=1,2,3} = (1, 0.8, 0.8, -1, 0.5, -0.5, 0.8, -0.5, 0.2)$, we simulated the response trajectories through $Y_i(t) = \alpha(t) + \sum_{k=1}^{3}\sum_{j=1}^{3} \beta_{jk}\xi_{ij}\psi_k(t)$, where $\alpha(t) = t + \sin(t)$, ψ_1 and ψ_2 are as defined before and $\psi_3(t) = -\sqrt{2/5}\cos(4\pi t/5)$, $t \in [0, 5]$. Similar to the simulations for vector predictors, the response measurements were again generated to be contaminated with errors $\epsilon_{im} \sim \mathcal{N}(0, 0.25)$.

For each simulation, the number of principal components J and K for predictors and responses were the same for the RARE and FLIN models, both chosen by the BIC criterion that was

mentioned in Section 2. To compare the predictive performance of the RARE and FLIN models, we generated 100 new predictors $X_i^*(s)$ with repeated measurements for each predictor simulated as U_{ij}^* (resp. $\boldsymbol{z}_i^*$ in the case of vector predictors) and the corresponding response trajectories $Y_i^*(t)$. The quality of predicted response curves was measured by the relative mean squared prediction error, defined as

$$\text{RMSPE} = \frac{1}{n} \sum_{i=1}^{n} \frac{\int_0^5 (\hat{Y}_{i,J,K}^*(t) - Y_i^*(t))^2 dt}{\int_0^5 (Y_i^*(t))^2 dt}. \tag{20}$$

The means and medians of the relative prediction errors for 500 Monte Carlo simulations are displayed in Table 1. The RARE model consistently outperforms the FLIN model in all cases, especially for vector predictors, where the decrease in prediction error exceeds 50%. The gains in the dense predictor case are also quite large, approaching a 40% decrease in prediction error. For the sparse predictor case, the improvement is somewhat smaller, but still corresponds to a 10% decrease in prediction error. As discussed in Section 3.2, the reason for smaller gains in the case of sparse predictors is that the predictor FPCs $\hat{\xi}_j$ used in the design matrix $\widehat{\boldsymbol{R}}$ are targeting the FPCs conditional on the available data, while the true FPCs are not estimable when only sparse data are available. The results show that the RARE model clearly improves upon the FLIN model and leads to better prediction of response trajectories.

5. Applications

5.1. Kiwi Weight Growth Data

Brown Kiwi is a species of birds endemic to New Zealand. Due to its status as a threatened species, monitoring growth and development of these birds is of great interest. In an ecological study carried out in New Zealand, a sample of $n = 165$ birds was monitored, of which 70 were wild birds and 60 were involved in the Operation Nest Egg (ONE) program. In this ecological program, Brown Kiwi eggs and chicks are removed from the wild and raised in captivity or a predator-free environment, on a so-called "crèche island", which in this case was the island Motuora near

Auckland (see Colbourne et al., 2005), and then are released back to the wild. An additional 35 birds were kept in captivity and monitored throughout their lives (Jones et al., 2009). For each bird, its weight (in grams) was recorded at multiple ages (days). Due to the uncertainty of recapture, the number and timing of measurements varied substantially between different birds. In addition to weight and age, sex and status (captivity/crèche/wild) were also recorded for each bird. The complete data set is available at http://www-ist.massey.ac.nz/GJones.

We are interested in modeling and comparing the weight growth curves for birds in the wild and in captive status. Sex is a known covariate for weight growth so must be considered as well. For birds released into the wild during the ONE program, only the weight measurements recorded while they were kept in captivity were included. Five birds started their lives at crèche island without being raised in captivity and were excluded from the analysis. We further truncated the weight trajectories at age 1000 days, because there were no data recorded for birds in captivity between days 1000 and 2500; one wild bird with no weight records before day 1000 was also excluded, so that the data for $n = 159$ birds were entered into the analysis. Of these, 69 are wild, 90 captive, 81 male and 78 female. The number of available weight measurements for each bird ranges from 1 to 69, with a median of 9 measurements, so the recordings of the weight trajectories are quite sparse and also very irregularly spaced in terms of age.

We consider the regression model (7) with log-transformed weight trajectories as responses and vector predictors $\boldsymbol{z} = (Z_1, Z_2)$, where Z_1 and Z_2 are indicator variables for status and gender, respectively, with $Z_1 = Z_2 = 0$ indicating male wild birds. The smoothed mean function $\hat{\mu}_Y(t)$ for the weight trajectories is displayed in the left panel of Figure 2. We find that after a rapid increase during the first 250 days of age, the weight increase slows down. The BIC criterion chose $K = 2$ principal components for response trajectories, explaining 97.8% of total variation. The estimates of the first two eigenfunctions are shown in the right panel of Figure 2. While the first eigenfunction (explaining 84.0% of total variation) is relatively flat, and

thus reflects a bird's overall weight level, the second eigenfunction (explaining 13.8% of total variation) has a prominent peak around day 500. Large positive values of the second functional principal component are associated with a lower starting level followed by a sharper increase in weight during the first 500 days, as compared to the mean function, followed by a subsequent larger degree of slowing of further weight increases compared to the average; negative values are associated with the opposite changes, relative to the mean trajectory.

The fitted log-weight trajectories obtained with the RARE approach, in comparison with those from FLIN, for the four groups are displayed in Figure 3. We find that fits obtained with RARE generally appear to track the trajectories more closely, as compared to those obtained by fitting from the FLIN model, especially for the female captive birds. The improved fit is also reflected in the value of the quasi-R^2 measure Q (defined in (17)), which increases from 0.107 for FLIN to 0.231 for RARE.

To further investigate predictive performance, we computed the leave-one-subject-out relative prediction errors

$$\text{RPE}^{(-i)} = \frac{\sum_{m=1}^{M_i}[\hat{Y}_i^{(-i)}(T_{iM_i})) - Y_i(T_{iM_i})]^2}{\sum_{m=1}^{M_i} Y_i^2(T_{iM_i})}, \tag{21}$$

where $\hat{Y}_i^{(-i)}$ is the predicted response for the ith bird, based on the model fitted with the data for the ith bird removed. The median of $\text{RPE}^{(-i)}$ was found to be 0.00137 for FLIN and 0.00056 for RARE. This substantial decrease in prediction error clearly demonstrates the improvement in prediction by that one may gain by applying the RARE approach.

5.2. Relationship between Viral Load and CD4 Cell Counts in AIDS Clinical Trials

As HIV-1 RNA copies (viral load) and CD4 cell counts are important virologic and immunologic markers for HIV-1 infection, their dynamic relationship during antiviral treatments is of interest. As the viral load measurement is more difficult to obtain and therefore often one has only longitudinal CD4 data available, in which case it is of interest to predict the associ-

14

ated time course of viral load. The data for the following analysis were collected in accordance with AIDS Clinical Trials Group (ACTG) protocol 315 (Lederman et al., 1998). These data are available at http://www.urmc.rochester.edu/biostat/people/faculty/WuSite/index.htm and have also been studied in Liang et al. (2003) and Wu and Liang (2004), who based their analysis on varying-coefficient models. The data consist of $n = 46$ patients with moderately advanced HIV-1 infection, for whom measurements of viral load and CD4 cell counts were available for the first 24 weeks of treatment.

Due to extreme sparsity of available measurements between the last two recorded treatment days, days 84 and 168, we considered the data in the first 90 days only. The mean function estimates for CD4 cell counts and viral load are displayed in Figure 1. One finds opposite trends for the two estimated mean functions, indicating a roughly inverse relationship between mean viral load and mean CD4 cell counts. The mean function of viral load appears to be oversmoothed, which is caused by the fact that the data after day 30 are very sparse. Using this mean function estimate $\hat{\mu}_Y(t)$ as the intercept in the FLIN model may lead to biased prediction of the response trajectories. This potential bias problem resulting from the oversmoothing is seen to be nicely resolved when using the estimated intercept function $\hat{\alpha}(t) = \hat{\mu}_Y(t) + \sum_{k=1}^{K} \hat{\theta}_k \hat{\psi}_k(t)$ in the RARE model. The additional terms in $\hat{\alpha}(t)$ adequately capture the rapid decrease of viral load during the first 30 days, as indicated by the dashed line in the right panel of Figure 1.

The BIC criterion led to the choice of two eigenfunctions for CD4 cell counts and three for viral load (accounting for 99.7% and 99.8% of total variation, respectively). As seen in the left panel of Figure 4, the first eigenfunction of CD4 cell counts (explaining 97.4% of total variation) is roughly constant, so the first functional principal component is associated with the overall level of CD4 cell counts. The second eigenfunction of CD4 cell counts (explaining 2.4% of total variation) increases from negative to positive during the first 30 days and is approximately constant subsequently. Large positive values of the second functional principal component then

correspond to smaller starting values for CD4 cell counts at the beginning of the study, followed by faster increases up to day 30, as compared to the mean function.

The first eigenfunction of viral load in the right panel of Figure 4, explaining 91% of total variation, also is approximately constant, corresponding to an overall level of viral load. The second eigenfunction (explaining 5.9% of total variation) is associated with opposite changes in viral load before and after day 40, which means if viral load decreases fast at the beginning, it may decrease more slowly subsequently or vice versa. The third eigenfunction (explaining 3% of total variation) adds to the mean function with smaller than average viral load during the first 20 days and an attenuated decrease after day 50, with slightly increased levels in between, if the third principal component is positive, or effects the opposite change if it is negative.

The fitted viral load trajectories, obtained from both RARE and FLIN approaches and over-laid with the observed measurements for 9 randomly selected patients, are displayed in Figure 5. Overall, the RARE fits are clearly outperforming FLIN; this is in line with an observed substantial increase of quasi-R^2 from 0.053 for the FLIN model to 0.202 for the RARE model. In Figure 5, the fitted curves generally are seen to track the observed measurements quite well in the beginning but not as well after about 20 or 30 days. This is consistent with our findings about the weakening relationship after day 20, which we deduce from the shape of the estimated regression coefficient surface $\hat{\beta}(s,t)$ (see supplementary materials). Lastly, the median of the leave-one-subject-out prediction errors is 0.0264 for FLIN and 0.0196 for RARE, demonstrating the improvement in prediction that can be achieved by using the RARE approach.

6. Discussion

The proposed response-adaptive method for modeling functional linear regression (2) (or (7)) is a flexible approach that is applicable to the cases of scalar, multivariate or functional predictors. Although the method is presented in the framework of sparsely sampled longitudinal responses,

its strength remains if the responses have denser observations (see supplementary materials). The case of sparse responses is particularly important for biomedical applications and longitudinal studies in many fields. Few flexible methodologies are available for this challenging case.

The proposed RARE model treats the entire response curve $Y(t)$, $t \in \mathcal{T}$ as an object and aims at predicting Y from newly observed predictors X. In biomedical applications, this is useful when a variable of interest Y is hard to measure while another variable X which is closely associated with Y is readily available. The RARE model can also be extended to a version where predictor time intervals and response time intervals do not overlap, in analogy to the "historical" functional linear model (Malfait and Ramsay, 2003); the RARE approach is applied for an increasing sequence $t \in \mathcal{T}$, with predictors $X(u)$ constrained to $u \in [0, t]$ and responses $Y(v)$ to $v \in [t, T]$, where T is the right endpoint of $\mathcal{T}$. Then the estimated regression coefficients depend on t, and one obtains updated predictions of the not yet observed parts of the response trajectories as time t progresses, possibly invoking an additional smoothing step across t.

Bootstrap inference is useful to empirically assess the significance of the functional regression (see, e.g., Müller et al., 2008). Formal test procedures for the regression coefficients may be developed under Gaussian assumptions, extending the approach described for confidence regions in Yao et al. (2005a). Another extension of interest is the inclusion of multiple functional predictors, simply by including the principal components of multiple predictors in the design matrix $\boldsymbol{R}$ in (14).

References

Ash, R. B. and Gardner, M. F. (1975). *Topics in stochastic processes.* Academic Press [Harcourt Brace Jovanovich Publishers], New York. Probability and Mathematical Statistics, Vol. 27.

Cai, T. and Hall, P. (2006). Prediction in functional linear regression. *The Annals of Statistics*

34, 2159–2179.

Cardot, H., Crambes, C., Kneip, A., and Sarda, P. (2007). Smoothing splines estimators in functional linear regression with errors-in-variables. *Computational Statistics & Data Analysis* **51,** 4832–4848.

Cardot, H., Ferraty, F., and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45,** 11–22.

Cardot, H., Ferraty, F., and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13,** 571–591.

Colbourne, R., Bassett, S., Billing, T., McCormick, H., McLennan, J., Nelson, A., and Robertson, H. (2005). The development of operation nest egg as a tool in the conservation management of kiwi. *Science for Conservation* **259,**.

Crambes, C., Kneip, A., and Sarda, P. (2009). Smoothing splines estimators for functional linear regression. *The Annals of Statistics* **37,** 35–72.

Draper, N. R. and Smith, H. (1998). *Applied Regression Analysis.* Wiley, New York.

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

Hall, P. and Horowitz, J. L. (2007). Methodology and convergence rates for functional linear regression. *The Annals of Statistics* **35,** 70–91.

He, G., Müller, H.-G., and Wang, J.-L. (2000). Extending correlation and regression from multivariate to functional data. In Puri, M. L., editor, *Asymptotics in statistics and probability*, pages 301–315. VSP International Science Publishers.

James, G. M. and Silverman, B. W. (2005). Functional adaptive model estimation. *Journal of the American Statistical Association* **100,** 565–576.

Jones, G., Leung, J., and Robertson, H. (2009). A mixed model for investigating a population of asymptotic growth curves using restricted b-splines. *Journal of Agricultural, Biological, and Environmental Statistics* **14,** 66–78.

Lederman, M. M., Connick, E., Landay, A., Kuritzkes, D. R., Spritzler, J., Clair, M. S., Kotzin, B. L., Fox, L., Chiozzi, M. H., Leonard, J. M., Rousseau, F., Wade, M., D'Arc Roe, J., Martinez, A., and Kessler, H. (1998). Immunologic responses associated with 12 weeks of combination antiretroviral therapy consisting of zidovudine, lamivudine, and ritonavir: Results of aids clinical trials group protocol 315. *Journal of Infectious Diseases* **178,** 70–79.

Liang, H., Wu, H., and Carroll, R. J. (2003). The relationship between virologic and immunologic responses in AIDS clinical research using mixed-effects varying-coefficient models with measurement error. *Biostatistics* **4,** 297–312.

Liu, B. and Müller, H.-G. (2008). Functional data analysis for sparse auction data. In Jank, W. and Shmueli, G., editors, *Statistical Methods in eCommerce Research*, pages 269–290. Wiley, New York.

Liu, B. and Müller, H.-G. (2009). Estimating derivatives for samples of sparsely observed functions, with application to on-line auction dynamics. *Journal of the American Statistical Association* **104,** 704–714.

Malfait, N. and Ramsay, J. O. (2003). The historical functional linear model. *Canadian Journal of Statistics. La Revue Canadienne de Statistique* **31,** 115–128.

Müller, H.-G., Chiou, J.-M., and Leng, X. (2008). Inferring gene expression dynamics via functional regression analysis. *BMC Bioinformatics* **9,** 60.

Müller, H.-G. and Yao, F. (2008). Functional additive models. *Journal of the American Statistical Association* **103,** 1534–1544.

Ramsay, J. O. and Dalzell, C. J. (1991). Some tools for functional data analysis. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53,** 539–572.

Ramsay, J. O. and Silverman, B. W. (2005). *Functional data analysis.* Springer Series in Statistics. Springer, New York, second edition.

Rice, J. A. and Silverman, B. W. (1991). Estimating the mean and covariance structure non-parametrically when the data are curves. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **53,** 233–243.

Rice, J. A. and Wu, C. O. (2001). Nonparametric mixed effects models for unequally sampled noisy curves. *Biometrics* **57,** 253–259.

Sy, J. P., Taylor, J. M. G., and Cumberland, W. G. (1997). A stochastic model for the analysis of bivariate longitudinal AIDS data. *Biometrics* **53,** 542–555.

Wu, H. and Liang, H. (2004). Backfitting random varying-coefficient models with time-dependent smoothing covariates. *Scandinavian Journal of Statistics* **31,** 3–19.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005a). Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association* **100,** 577–590.

Yao, F., Müller, H.-G., and Wang, J.-L. (2005b). Functional linear regression analysis for longitudinal data. *The Annals of Statistics* **33,** 2873–2903.

Table 1: Simulation results for relative mean squared prediction error (RMSPE, eq. (20)), for vector predictors and functional predictors with either densely sampled noisy or sparsely sampled noisy observations, comparing the proposed response-adaptive regression (RARE) with the established functional linear regression (FLIN) approach.

Regression	Vector Predictors		Dense Functional Predictors		Sparse Functional Predictors	
	Mean	Median	Mean	Median	Mean	Median
RARE	0.0067	0.0056	0.0388	0.0368	0.124	0.122
FLIN	0.0157	0.0148	0.0615	0.0599	0.138	0.135
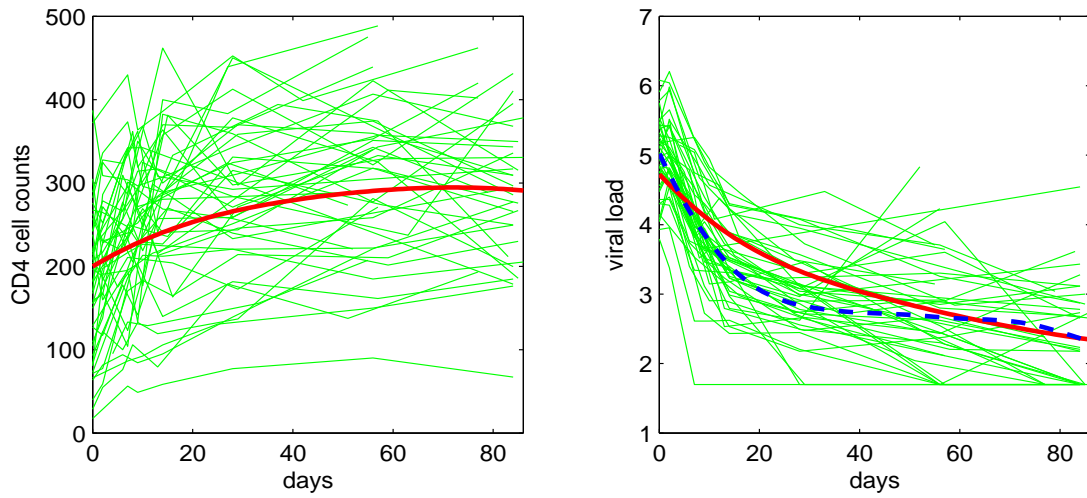
Figure 1: Estimates of the mean functions for CD4 cell counts (left panel) and viral load (solid, right panel) with corresponding observed individual trajectories. The dashed line in the right panel is the estimated intercept function $\hat{\alpha}(t)$ in the RARE model.
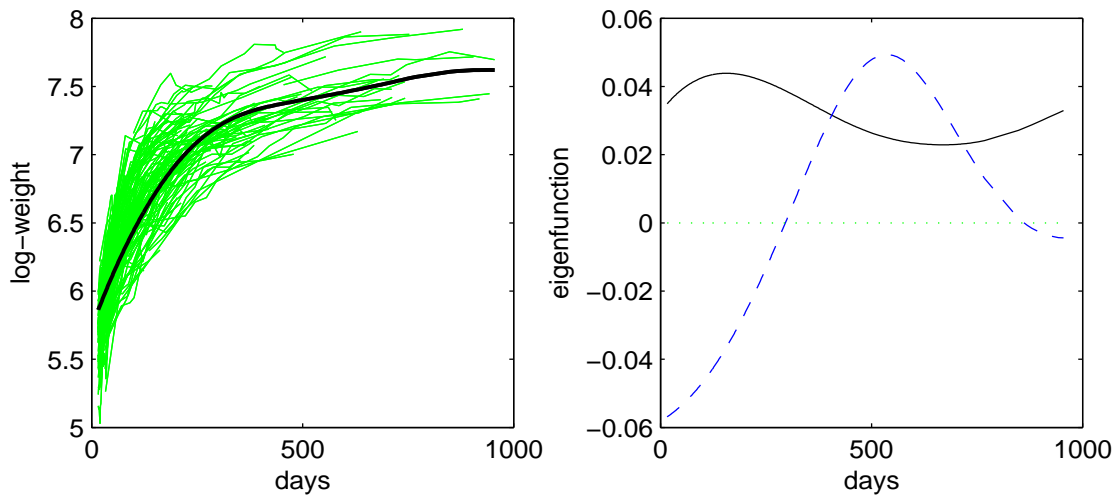

Figure 2: Log-transformed weight trajectories overlaid with the smoothed mean function $\hat{\mu}_Y(t)$ (left panel) and estimates of the first (solid) and second (dashed) eigenfunctions (right panel) for Brown Kiwi weight measurements.
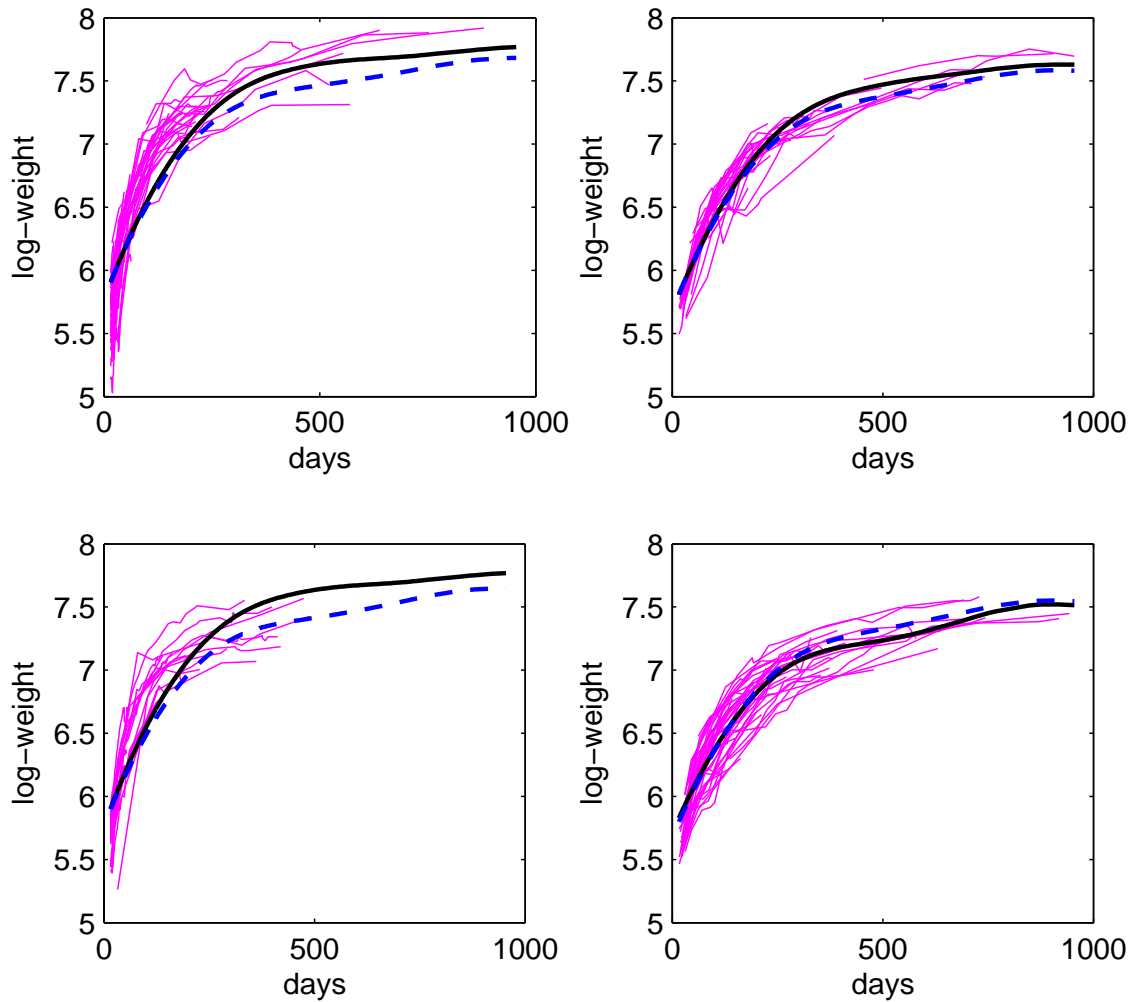
Figure 3: Fitted log-weight curves obtained from the proposed response-adaptive regression (solid) and from functional linear regression (dashed) for four groups: female captive (top left), female wild (top right), male captive (bottom left) and male wild (bottom right).
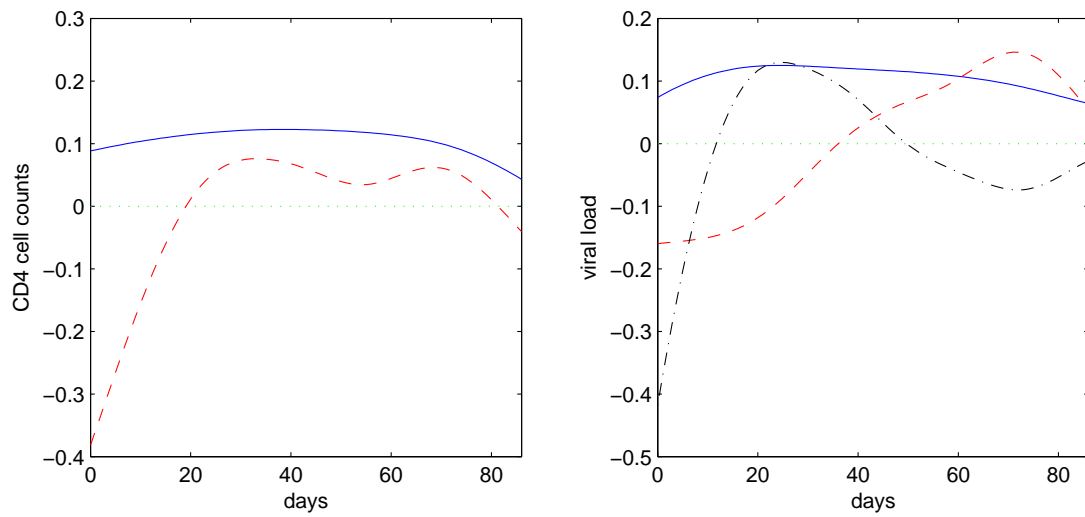
Figure 4: Estimated eigenfunctions for CD4 cell counts (left panel) and viral load (right panel), showing first (solid), second (dashed) and third (dash-dot, for viral load only) eigenfunctions.
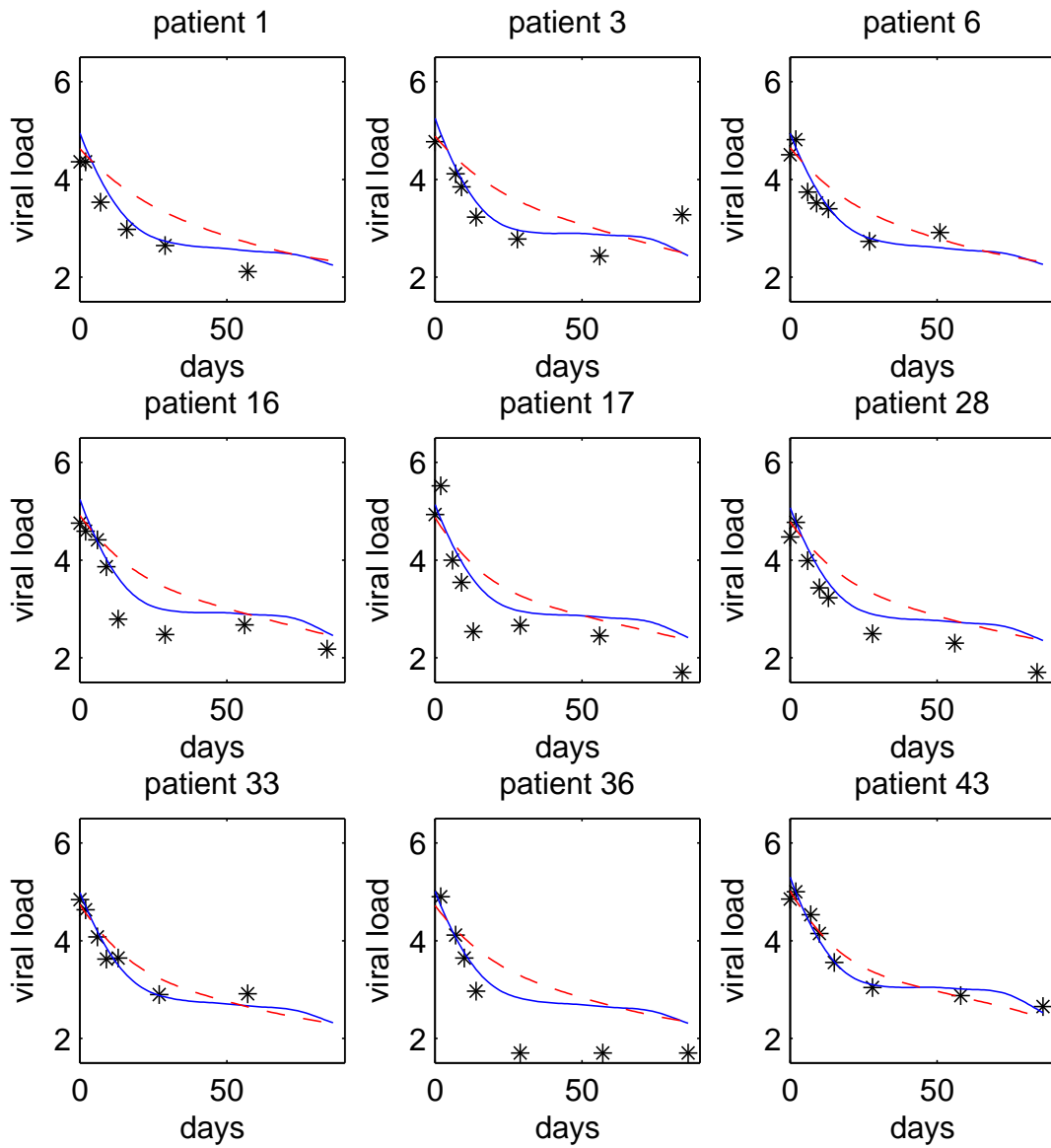
Figure 5: Fitted trajectories of viral load, comparing the fits obtained from the proposed response-adaptive regression (RARE) approach (solid) with those for the established functional linear regression model (dashed) for nine randomly selected patients.