RESEARCH ARTICLE

Dynamic prediction in functional concurrent regression with an application to child growth

Andrew Leroux[1] | Luo Xiao[2] | Ciprian Crainiceanu[1] | William Checkley[3]

[1]Department of Biostatistics, Johns Hopkins University, Baltimore, MD 21205, USA

[2]Department of Statistics, North Carolina State University, Raleigh, NC 27606, USA

[3]School of Medicine, Johns Hopkins University, Baltimore, MD 21205, USA

Correspondence
Luo Xiao, Department of Statistics, North Carolina State University, Raleigh, NC 27606, USA.
Email: lxiao5@ncsu.edu

In many studies, it is of interest to predict the future trajectory of subjects based on their historical data, referred to as dynamic prediction. Mixed effects models have traditionally been used for dynamic prediction. However, the commonly used random intercept and slope model is often not sufficiently flexible for modeling subject-specific trajectories. In addition, there may be useful exposures/predictors of interest that are measured concurrently with the outcome, complicating dynamic prediction. To address these problems, we propose a dynamic functional concurrent regression model to handle the case where both the functional response and the functional predictors are irregularly measured. Currently, such a model cannot be fit by existing software. We apply the model to dynamically predict children's length conditional on prior length, weight, and baseline covariates. Inference on model parameters and subject-specific trajectories is conducted using the mixed effects representation of the proposed model. An extensive simulation study shows that the dynamic functional regression model provides more accurate estimation and inference than existing methods. Methods are supported by fast, flexible, open source software that uses heavily tested smoothing techniques.

KEYWORDS
covariance function, face, fPCA, longitudinal data, mixed effects, penalized splines, sparse functional data

1 | INTRODUCTION

In many biological and epidemiological studies, sampling is conducted at multiple time points resulting in longitudinal data that exhibit within-subject correlation. Traditionally, longitudinal data have been analyzed using either marginal models[1] or conditional mixed effect models.[2-4] Both approaches are parametric and are not designed to account for subtle or strong departures from the assumed parametric trends. This problem can manifest in a number of ways in longitudinal data, including autocorrelation in the residuals of random intercept/slope models.[5] To address such challenges, one may consider using methods for functional data, which allow more flexible modeling of subject-specific random curves. A random functional intercept model can be understood as a special case of a broader class of models referred to as functional mixed effects models. However, functional mixed models are complex, and the computational burden for fitting them is nontrivial.

A computationally feasible approach to estimating functional mixed effects models is to combine semiparametric regression techniques[6] with functional data analysis.[7] Arguably, the functional mixed effects framework was first

introduced by Guo[8] and was further developed in subsequent years.[9-19] Recently, Scheipl et al[20] provided a general framework for fitting functional mixed effects models based on the idea of incorporating nonparametric smoothing approaches into the standard mixed effects modeling framework.[6]

Although the aforementioned advances in methodology are substantial, the release of accompanying software has been limited. As a result, use of functional mixed effects models in scientific studies has lagged behind traditional methods in spite of the wide range of data types amenable to this modeling framework. To the best of our knowledge, only the R package *refund*[21] has provided functions that fit standard functional mixed effects models. Despite the varied models that can be fit using the *refund* package, the estimation of a functional concurrent regression (FCR) model with both irregularly measured response and predictors is not currently possible. Moreover, neither *refund* nor the standard *R* packages for fitting mixed effects models, *lme4*[22] and *nlme*,[23] are able to readily handle dynamic prediction. Specifically, the current software packages do not allow for predictions on subjects not included in the model fitting procedure. With reproducibility and dissemination of our methods as a primary goal, we propose a dynamic FCR model applied to growth curve data in children and provide an accompanying *R* package, *fcr*, capable of fitting this class of models.

Dynamic prediction in functional data analysis is relatively new. Recently, several authors have proposed methods, which allow for prediction of partially observed functional data via functional principal component analysis.[24-28] However, these works involve only functional responses and do not consider subject-specific predictors of any form. The proposed work fills this gap using a functional concurrent model. Indeed, in our application, subject predictions are based on time-invariant covariates (sex), time-varying covariates (weight), and past length measurements. The effects of both time-varying and time-invariant covariates are modeled using time-varying coefficients.[29] A child's past length measurements and their length at a future age are modeled by a child-specific functional random intercept. While a random intercept/random slope model could be used to characterize child-specific growth trajectories of height, such an approach is not well suited for the nonlinear growth trajectories of children (see Figure 2).

The rest of the paper is organized as follows. In Section 2, we describe the motivating data and provide an exploratory analysis. In Sections 3 and 4, we introduce the dynamic FCR model and propose an estimation method. In Section 5, we apply the method to the CONTENT growth data and discuss our findings. In Section 6, we conduct a simulation study to evaluate the performance of the proposed functional mixed effects model. We conclude the paper with a discussion in Section 7.

2 | THE CONTENT CHILD GROWTH DATA

The CONTENT study was conducted in 2 peri-urban shanty towns approximately 25 km south of central Lima in Peru. The goal of the study was to assess the impact of a particular bacterial infection, *Helicobacter pylori*, on child growth.[30] To that end, anthropometric measurements on 215 children (106 boys and 109 girls) were collected periodically starting from birth. We focus on data collected at 547 unique time points between 0 and 24 months. Children have an average of 39 anthropometric measurements each, with measurements observed more frequently during the first few months of life.

We examine 3 World Health Organization (WHO)-defined anthropometric measurements: length-for-age z-score (LAZ), weight-for-length z-score, and weight-for-age z-score (WAZ).[31] The z-scores are calculated using the age- and sex-specific WHO standard references. This z-scoring procedure is standard in growth curve modeling and provides data on a scale relative to a standard of growth, with a z-score of 0 representing the WHO-defined average for that age and sex. Z-scores greater (smaller) than 0 indicate above (below) average measurement relative to the WHO standard in that category. As a result, the z-scores can be compared within- and between-children over time. For example, a z-score of 0 at 3 months compared with a z-score of 1 at 6 months for the same child indicates that the child is at the WHO average length at 3 months, but above the WHO average by 6 months.

Figure 1A-C displays the marginal distributions of anthropometric measurements binned into monthly intervals. Several features of the data can be observed. First, these children are, on average, below the WHO-standard average length and above the WHO-standard average weight. These growth characteristics have been reported in similar populations.[32,33] In addition, the trend in z-scores for boys and girls is similar across all ages. Furthermore, there appears to be a disproportionate weight gain relative to length gain for the 0- to 3-month period (see Figure 1C).

Figure 1D displays the estimated correlation between subjects' LAZ(t) and WAZ(s) for all months t, s. Correlations are generally high between measurements taken close in time and are higher closer to birth (notice the shades of red along the main diagonal and the darker red shades in the lower left corner). Correlations remain high even at larger time lags (months 20+) with values above 0.4 for many time pairs.

Unlike nonstandardized length (height) and weight measurements, which increase continuously for most children, the trend in *z*-scores is quite different. Figure 2 displays the LAZ and WAZ scores for 4 children, indicating strong nonlinear characteristics of the curves. These 4 subjects are representative of subjects' diverse growth patterns. Subject 1 has a fairly steady increasing trajectory up to 12 months followed by a period of relatively stable growth from 12 to 24 months. In contrast, subjects 2 and 3 experience steep declines (increases) followed by steep increases (declines) in early life. Following a period of relatively consistent growth, near 12 months, subject 3 undergoes a period of rapid decline in standardized length, suggesting some degree of stunting. Accounting for the diverse and markedly nonlinear growth patterns observed requires flexible models. Regarding the relationship between LAZ and WAZ, Figure 2 suggests these 2 *z*-scores exhibit positive correlations, reinforcing the observed population level correlations displayed in Figure 1.

Since it is known that delayed growth development is associated with a number of negative health outcomes, it is of interest to identify children at risk of delayed growth and/or growth stunting as early as possible. Such identification via growth trajectories can only be achieved using a dynamic approach. Indeed, a post hoc analysis can only identify that certain children were stunted and cannot inform the development of potential interventions. Correspondingly, the primary goal of our application is to build a model that can dynamically predict length-for-age *z*-scores in the first 2 years of life. As there is a clear association between weight and length in children, we aim to build a model that addresses this relationship that is compatible with dynamic prediction. The procedure requires the use of historical information available up to a particular time point to predict the future trajectory. In Figure 3, we provide an illustration of the dynamic prediction of LAZ based on observed LAZ and WAZ. Finally, we propose an inferential framework that provides confidence intervals both for curve predictions and model parameters.
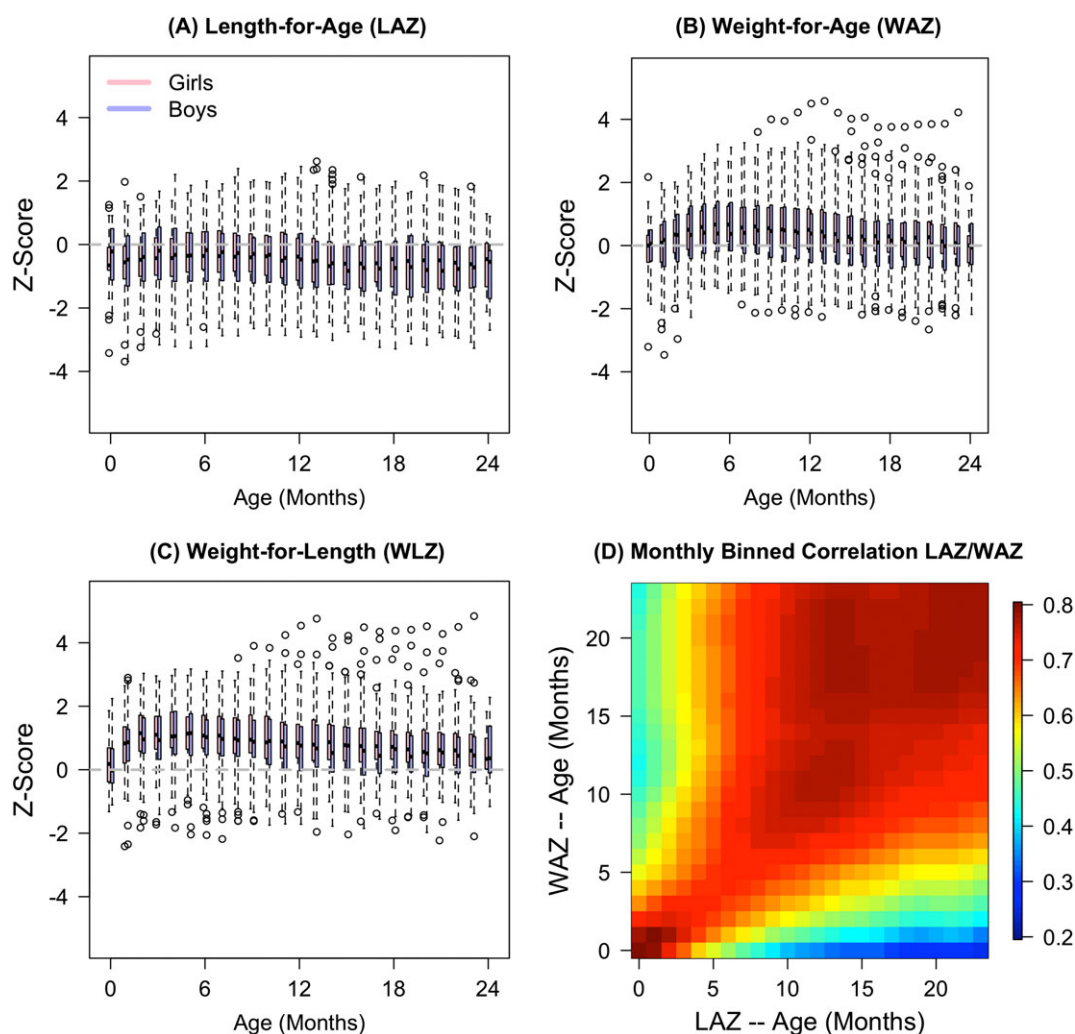
FIGURE 1 Distribution of, A, length-for-age *z*-scores (LAZ), B, weight-for-age *z*-scores (WAZ), C, weight-for-length *z*-scores (WLZ) binned into monthly categories, and, D, estimated empirical correlation between LAZ and WAZ at each month. Correlations were calculated using Z-scores projected onto an evenly spaced grid of ages {0.5, 1.5, … , 23.5} for all subjects using *face::face.sparse()* applied separately to LAZ and WAZ
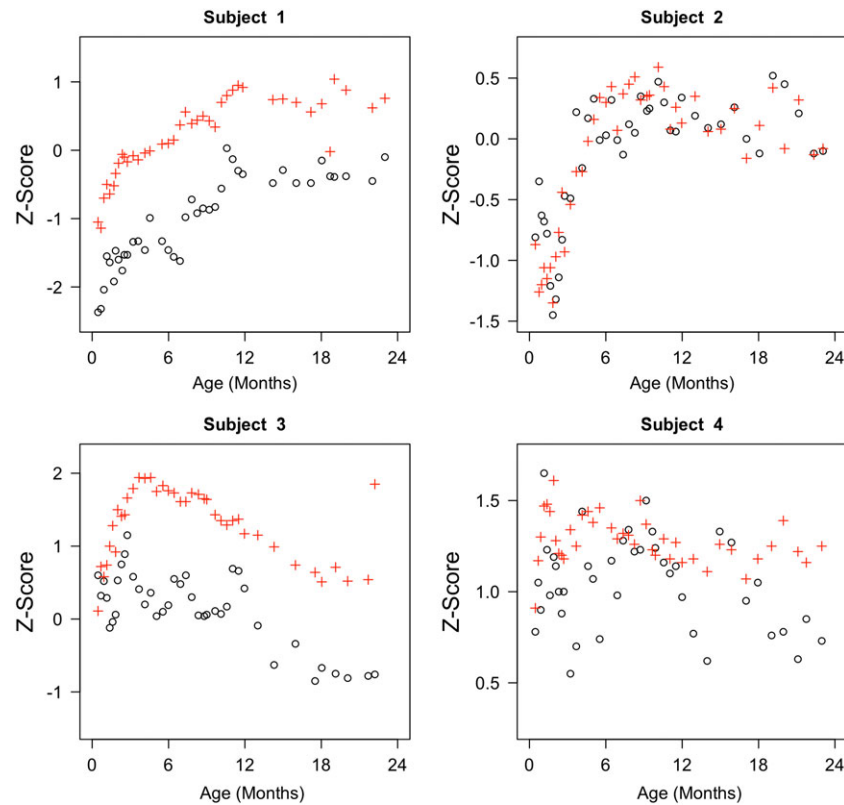
FIGURE 2 Length-for-age z-score and weight-for-age z-score curves for 4 children. Length-for-age z-score is presented as ∘ and weight-for-age z-score is presented as +
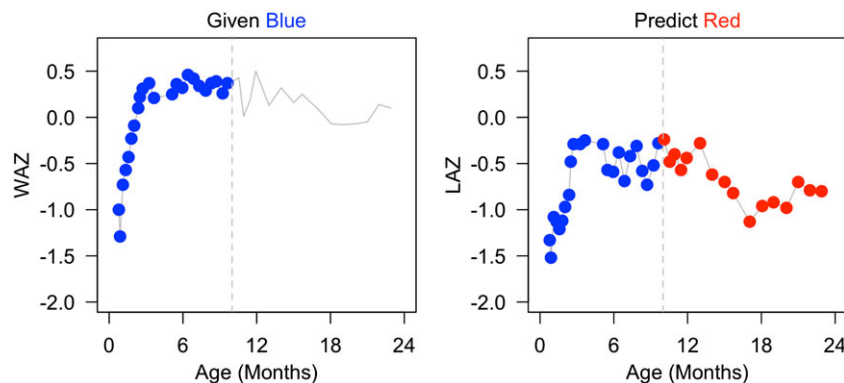
FIGURE 3 Illustration of dynamic prediction using a subject from the CONTENT study. Conditional on the weight-for-age z-score (WAZ) and length-for-age z-score (LAZ) data in blue, the interest is to predict the length-for-age z-score (LAZ) in red

3 | MODEL

3.1 | Notation

Let t denote age in months. For subject i, let $Y_i(t)$ be the LAZ at age t, $Z_i(t)$ be the WAZ at age t, and X_i be the sex indicator (1 for boys and 0 for girls). The observed data are $[\{t_{ij}, Y_{ij} = Y_i(t_{ij}), Z_{ij} = Z_i(t_{ij}), X_i\}, 1 \leq j \leq m_i, 1 \leq i \leq N]$, where t_{ij} denotes the jth visit time for subject i, m_i is the number of observations for subject i, and N is the total number of subjects.

3.2 | Model

We consider the following model:

$$Y_i(t) = f_0(t) + X_i f_1(t) + Z_i(t) f_2(t) + b_i(t) + \epsilon_i(t), \tag{1}$$

where $f_0(\cdot)$ is the population intercept function, $f_1(\cdot)$ models the sex difference, $f_2(\cdot)$ models the concurrent association of WAZ with LAZ, $b_i(\cdot)$ models the subject-specific random functional deviation of subject i from $f_0(\cdot)$, and $\epsilon_i(t)$ are independent random errors. More specifically, we model $b_i(t)$ as a zero-mean Gaussian process with a covariance function $\text{cov}\{b_i(s), b_i(t)\} = C(s, t)$ and assume that $\epsilon_i(t)$ are independent and identically distributed as $N(0, \sigma_\epsilon^2)$. We assume that $b_i(\cdot)$ and $\epsilon_i(t_{ij})$ are mutually independent across subjects.

The WAZ process is also measured with error, and we assume that for subject i, the observed WAZ is $Z_{i,\text{obs}}(t_{ij}) = Z_i(t_{ij}) + \epsilon_i^z(t_{ij})$. Here, $\epsilon_i^z(t_{ij})$ is the term for measurement errors and is assumed to be independent and identically distributed as $N(0, \sigma_{\epsilon z}^2)$. We let $Z_i(t_{ij}) = \mu_z(t_{ij}) + b_{iz}(t_{ij})$, where $\mu_z(\cdot)$ is the population mean of WAZ and $b_{iz}(\cdot)$ is the subject-specific random functional deviation from the population mean. We model $b_{iz}(\cdot)$ as a stochastic process with zero mean and a covariance function $\text{cov}\{b_{iz}(s), b_{iz}(t)\} = C_z(s, t)$ and assume that $b_{iz}(\cdot)$ are independent across subjects. It is possible to use additional predictors in estimating $Z_i(t_{ij})$, although this option is not explored here.

Model (1) is a concurrent effect model for the time-varying covariate $Z_i(t_{ij})$. We show how this concurrent modeling framework can be used for dynamic prediction in Section 4. The historical information is incorporated implicitly in model (1) via the functional random effect $b_i(t)$. We have found this implicit approach to provide an excellent alternative to the explicit modeling of historical functional data. Thus, our proposal will add to the literature and will provide alternative inferential approaches that are especially useful in the case of sparse longitudinal data.

The explicit modeling of historical functional data was introduced by Malfait and Ramsay,[34] while Şentürk and Müller[35] allowed the response to depend on smooth functions of arbitrarily many lagged covariate values. Building on the idea of historical models, Şentürk and Müller[36] used the concept of a "*history index*" to allow for a response to depend only on the recent historical values of a covariate. Given the existence of these models, it is tempting to consider a full historical effect term $\int_0^t \beta(s, t) Z_i(s) \mathrm{d}s$, where $\beta(s, t)$ is an unknown bivariate function. We do not adopt such a specification because it did not improve fit to the CONTENT data.

3.3 | Model estimation

One of our main contributions is to provide explicit methods and formulas to fit model (1). This is not trivial, as such models have not been implemented before, and they require some heavy notation and attention to detail. First, we conduct a functional principal component analysis on the $Z_{i,\text{obs}}(t_{ij})$s using the R package *face*. This results in estimates $\hat{\mu}_z(t)$, $\hat{C}_z(s, t)$, and $\hat{\sigma}_{\epsilon z}^2$. Then, conditional on the observed data $\mathbf{Z}_{i,\text{obs}} = \{Z_{i,\text{obs}}(t_{i1}), \dots, Z_{i,\text{obs}}(t_{im_i})\}'$, we predict $Z_i(t)$ by its best linear unbiased predictor (BLUP), $\tilde{Z}_i(t) = \hat{\mu}_z(t) + \hat{C}_z(t, \mathbf{t}_i)'\{\hat{C}_z(\mathbf{t}_i, \mathbf{t}_i) + \hat{\sigma}_{\epsilon z}^2 \mathbf{I}_{m_i}\}^{-1}\{\mathbf{Z}_{i,\text{obs}} - \hat{\mu}_z(\mathbf{t}_i)\}$,[37] where $\mathbf{t}_i = (t_{i1}, \dots, t_{im_i})' \in \mathbb{R}^{m_i}$, $\hat{C}_z(t, \mathbf{t}_i) = \{C_z(t, t_{i1}), \dots, C_z(t, t_{im_i})\}' \in \mathbb{R}^{m_i}$, $\hat{\mu}_z(\mathbf{t}_i) = \{\mu_z(t_{i1}), \dots, \mu_z(t_{im_i})\}' \in \mathbb{R}^{m_i}$, and $\hat{C}_z(\mathbf{t}_i, \mathbf{t}_i) = \{\hat{C}_z(t_{ij}, t_{ik})\}_{1 \le j, k \le m_i} \in \mathbb{R}^{m_i \times m_i}$. We then substitute $\tilde{Z}_i(t_{ij})$ for $Z_i(t_{ij})$ in model (1).

For notational convenience, we rewrite model (1) as

$$Y_{ij} = f_0(t_{ij}) + \sum_{p=1}^{P} z_{p,ij} f_p(t_{ij}) + b_i(t_{ij}) + \epsilon_{ij}, \tag{2}$$

where $P = 2$, $z_{1,ij} = X_i$, $z_{2,ij} = Z_{ij}$. Models for functional responses have been studied in Ivanescu et al[38] and Scheipl et al[20] and are implemented via the *pffr* function in the R package *refund*. However, *pffr* currently accommodates only time-varying predictors, which are measured on a regular grid. As a result, *pffr* cannot be used to implement the method described above using the CONTENT data. Below we detail our estimation method using P-splines,[39] and we provide an R package that extends the *pffr* function for concurrent models with irregularly measured time-varying covariates.

We approximate $f_p(\cdot)$ by a linear combination of B-spline basis functions. More specifically, let $\mathbf{B}(t) = \{B_1(t), \dots, B_c(t)\}'$ be a sequence of B-spline basis functions evaluated at t, where c is the number of interior knots plus the order of the B-spline basis functions. We use cubic B-splines and place interior knots at the quantiles of the observed time points.[40] For simplicity, we use the same basis functions for all f_p, $p = 0, 1, \dots, P$. Then we let $f_p(t) = \mathbf{B}(t)' \theta_p$, where $\theta_p = (\theta_{p1}, \dots, \theta_{pc})'$ is the vector of coefficients.

We also approximate the random subject-specific curves $b_i(t)$ using B-spline basis functions with $b_i(t) = \sum_{k=1}^{c} u_{ik} B_k(t)$, where $\mathbf{u}_i = (u_{i1}, \dots, u_{ic})' \in \mathbb{R}^c$. Unlike Chen and Wang[12] who used a small number of basis functions, we use a relatively large number of basis functions to allow for increased modeling flexibility of subject-specific functions. We assume that

$\mathbf{u}_i$ follows a multivariate normal distribution with zero mean and covariance $\mathbf{\Gamma} = (\gamma_{k\ell})_{1 \le k, \ell \le c} \in \mathbb{R}^{c \times c}$. This implies that $C(s, t) = \sum_{1 \le k, \ell \le c} \gamma_{k\ell} B_k(s) B_\ell(t)$. The covariance matrix $\mathbf{\Gamma}$ is estimated from the data. Details are provided in Section 3.4.

For each i, we let $\mathbf{y}_i = (Y_{i1}, \ldots, Y_{im_i})' \in \mathbb{R}^{m_i}$, $\mathbf{z}_{p,i} = (z_{p,i1}, \ldots, z_{p,im_i})' \in \mathbb{R}^{m_i}$, $\mathbf{B}_i = [\mathbf{B}(t_{i1}), \ldots, \mathbf{B}(t_{im_i})]' \in \mathbb{R}^{m_i \times c}$, and $\mathbf{X}_i = [\mathbf{B}_i, \mathrm{diag}(\mathbf{z}_{1,i})\mathbf{B}_i, \ldots, \mathrm{diag}(\mathbf{z}_{P,i})\mathbf{B}_i] \in \mathbb{R}^{m_i \times (P+1)c}$. With this notation, we have $\mathbb{E}(\mathbf{y}_i) = \mathbf{X}_i \theta$, where $\theta = (\theta_0, \ldots, \theta_P)' \in \mathbb{R}^{(P+1)c}$, and $\mathbf{V}_i = \mathrm{Cov}(\mathbf{y}_i) = \mathbf{B}_i \mathbf{\Gamma} \mathbf{B}_i' + \sigma_\epsilon^2 \mathbf{I}_{m_i}$. The generalized least squares estimate for θ is then

$$\hat{\theta} = \arg\min_\theta N^{-1} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \theta)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \theta).$$

Because regression splines tend to over-fit, we enforce smoothness in the estimated coefficient functions by imposing a penalty on θ as suggested by Eilers and Marx.[39] Specifically, we let $\mathbf{P} = \mathrm{blockdiag}(\mathbf{P}_0, \mathbf{P}_1, \ldots, \mathbf{P}_P)$ and for $p \in \{0, 1, \ldots, P\}$, $\mathbf{P}_p = \lambda_p \mathbf{D}' \mathbf{D}$, where $\mathbf{D}$ is the $(c - o) \times c$ differencing matrix and o is the order of the difference penalty.[41] We use $o = 2$ throughout the paper. Thus, we estimate θ by penalized generalized least squares

$$\hat{\theta} = \arg\min_\theta N^{-1} \sum_{i=1}^N (\mathbf{y}_i - \mathbf{X}_i \theta)' \mathbf{V}_i^{-1} (\mathbf{y}_i - \mathbf{X}_i \theta) + \theta' \mathbf{P} \theta$$

and we obtain

$$\hat{\theta} = \left(N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i + \mathbf{P} \right)^{-1} \left(N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{y}_i \right). \tag{3}$$

Let $\hat{\theta} = (\hat{\theta}_0, \ldots, \hat{\theta}_P)'$ such that $\hat{\theta}_p$ is the corresponding estimate of θ_p. Then each coefficient function can be estimated as $\hat{f}_p(t) = \mathbf{B}(t)' \hat{\theta}_p$.

Using simple algebra, it follows that

$$\mathrm{Cov}(\hat{\theta}) = \left(N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i + \mathbf{P} \right)^{-1} \left(N^{-2} \sum_{i=1}^N \mathbf{X}_i \mathbf{V}_i^{-1} \mathbf{X}_i \right) \left(N^{-1} \sum_{i=1}^N \mathbf{X}_i' \mathbf{V}_i^{-1} \mathbf{X}_i + \mathbf{P} \right)^{-1}.$$

Then we can easily construct pointwise standard error for $\hat{f}_p(t)$, the details of which are omitted here.

3.4 | Estimation of variance components

We adopt a 4-step procedure for estimating $C(s, t)$, i.e., $\mathbf{\Gamma}$, and σ_ϵ^2. In step 1, we get an initial estimate $\hat{\theta}^0$ of θ by letting $\mathbf{V}_i = \mathbf{I}_{m_i}$. In step 2, we obtain residuals $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\theta}^0$, $1 \le i \le N$, which are then used to estimate $C(s, t)$ and σ_ϵ^2. In step 3, given the estimates $\hat{\mathbf{\Gamma}}$ and $\hat{\sigma}_\epsilon^2$ of $\mathbf{\Gamma}$ and σ_ϵ^2, respectively, we update $\hat{\theta}$ using (3). Then, in step 4, we again obtain residuals $\mathbf{e}_i = \mathbf{y}_i - \mathbf{X}_i \hat{\theta}$, $1 \le i \le N$ which are used to produce final estimates of $C(s, t)$ and σ_ϵ^2. While this procedure can be iterated to obtain updated estimates of variance components and fixed effects, in our application, multiple iterations did not result in meaningful changes to estimated quantities.

We use the R package *face*[42] for estimating $C(s, t)$ and σ_ϵ^2 using the residuals $\mathbf{e}_i$, $1 \le i \le N$. The *face* package is based on Xiao et al,[37] where $C(s, t)$ is decomposed as $\sum_{1 \le k, \ell \le c} B_k(s) B_k(t) \gamma_{k\ell}$.

3.5 | Selection of smoothing parameters

Either restricted maximum likelihood or generalized cross validation can be used to select the smoothing parameters $\lambda_k (0 \le k \le p)$ contained in the penalty matrix $\mathbf{P}$.[43] Due to the size of the CONTENT data and the large number of smoothing parameters requiring estimation, we estimate our model using the *bam* function in the *mgcv* package, which uses estimation procedures designed to be much faster for very large datasets.[44]

4 | DYNAMIC PREDICTION OF LAZ

Consider a new subject with sex covariate X and observed growth data $\{t_j, Y_j, \tilde{Z}_j, 1 \le j \le m\}$. For simplicity, we drop the subject index and assume $t_1 < t_2 < \cdots < t_m$. We would like to predict $Y(t)$ for $t_m < t < T_{\max}$ where

$T_{\max} = \max(t_{1,m_1}, t_{2,m_2}, \ldots, t_{i,m_i})$. That is, $T_{\max}$ denotes the time of the latest observed outcome among all subjects on which the model is fit. Extrapolating beyond all observed data is clearly infeasible in the absence of additional information or assumptions on the features of trajectories beyond the observed domain.

From model (1), the model for the new subject is

$$Y_j = f_0(t_j) + X f_1(t_j) + Z_j f_2(t_{ij}) + b(t_j) + \epsilon_j, \quad 1 \le j \le m.$$

Let $\mathbf{y} = (Y_1, \ldots, Y_m)'$, $\mathbf{z}_1 = X\mathbf{1}_m$, $\mathbf{z}_2 = (Z_1, \ldots, Z_m)'$, $\mathbf{B}_* = [\mathbf{B}(t_1), \ldots, \mathbf{B}(t_m)]'$, and $\mathbf{X} = [\mathbf{B}_*, \mathrm{diag}(\mathbf{z}_1)\mathbf{B}_*, \mathrm{diag}(\mathbf{z}_2)\mathbf{B}_*]$. Let $b(t) = \mathbf{B}(t)'\mathbf{u}$, where $\mathbf{u} \in \mathbb{R}^c$. It can then be derived that the BLUP of $\mathbf{u}$ is $\tilde{\mathbf{u}} = \mathbb{E}(\mathbf{u}|\mathbf{y}) = \boldsymbol{\Gamma}\mathbf{B}_*'(\mathbf{B}_*\boldsymbol{\Gamma}\mathbf{B}_*' + \sigma_\epsilon^2 \mathbf{I}_m)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$.[6] Thus, the BLUP of $b(t)$ is $\mathbf{B}(t)'\tilde{\mathbf{u}}$, and the BLUP of $Y(t)$ is $\tilde{Y}(t) = \mathbf{X}\boldsymbol{\theta} + \mathbf{B}(t)'\boldsymbol{\Gamma}\mathbf{B}_*'(\mathbf{B}_*\boldsymbol{\Gamma}\mathbf{B}_*' + \sigma_\epsilon^2\mathbf{I}_m)^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\theta})$. The variance for $\tilde{Y}(t) - Y(t)$ is $\mathbf{B}(t)'\left\{\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{B}_*'(\mathbf{B}_*\boldsymbol{\Gamma}\mathbf{B}_*' + \sigma_\epsilon^2\mathbf{I}_m)^{-1}\mathbf{B}_*\boldsymbol{\Gamma}\right\}\mathbf{B}(t) + \sigma_\epsilon^2$, which can be used to construct the approximate 95% pointwise prediction interval for $Y(t)$ as

$$\tilde{Y}(t) \pm 2\sqrt{\mathbf{B}(t)'\left\{\boldsymbol{\Gamma} - \boldsymbol{\Gamma}\mathbf{B}_*'(\mathbf{B}_*\boldsymbol{\Gamma}\mathbf{B}_*' + \sigma_\epsilon^2\mathbf{I}_m)^{-1}\mathbf{B}_*\boldsymbol{\Gamma}\right\}\mathbf{B}(t) + \sigma_\epsilon^2}.$$

In practice, we use the estimated BLUP, where we plug in the estimates of parameters. In particular, we replace $Z_i = Z(t_i)$ by their respective BLUP estimates $\tilde{Z}(t_i)$.

Note that the width of the prediction intervals will depend on both $\boldsymbol{\Gamma}$ and σ_ϵ^2. Intuitively, the higher the correlation between $Y(t)$ and observed data, the narrower the prediction intervals will be regardless of the temporal distance t and the observed time points. However, in many biological applications, correlation decreases consistently as temporal distance increases. In such scenarios, we expect prediction intervals to generally increase as one attempts to predict further in time from observed time points.

5 | DATA APPLICATION

We apply the proposed model (1) to the CONTENT child growth data introduced in Section 2. The estimated coefficient functions are displayed in Figure 4. The baseline/intercept function is negative and generally decreasing. The sex effect is estimated to be near zero at all ages, which is not surprising because the z-scores already account for sex differences. WAZ is positively associated with LAZ with the strongest association near birth and around ages 18 to 24 months.

Figure 5 displays key features of the estimated covariance function: the correlation function, the variance function $\hat{C}(t, t), t \in (0, 24)$, and the top 5 estimated eigenfunctions. The within-curve correlation of the LAZ is relatively high (>0.5 at most pairs of time points), and the variation of LAZ peaks around 6 months. Interestingly, this period corresponds with the time where WAZ has the smallest estimated association with LAZ, suggesting some potential unobserved factor influencing LAZ trajectories during early life. The total variance in the estimated covariance function, which is equal to the sum of the eigenvalues of $\hat{C}(s, t)$, is approximately 0.43, and the estimated error variance, $\hat{\sigma}_\epsilon^2$, is 0.033. Accordingly, about 93% of variation in LAZ unexplained by WAZ and sex can be modeled by the random subject-specific functions.

To illustrate the idea of dynamic prediction, in Figure 6, we dynamically predict LAZ curves conditioned on up to 6, 12, and 18 months of observed LAZ and WAZ for 4 subjects. This allows for a direct evaluation of future LAZ predictions using partial subject data as a proxy for true out-of-sample subject prediction accuracy. Here, each column corresponds to a
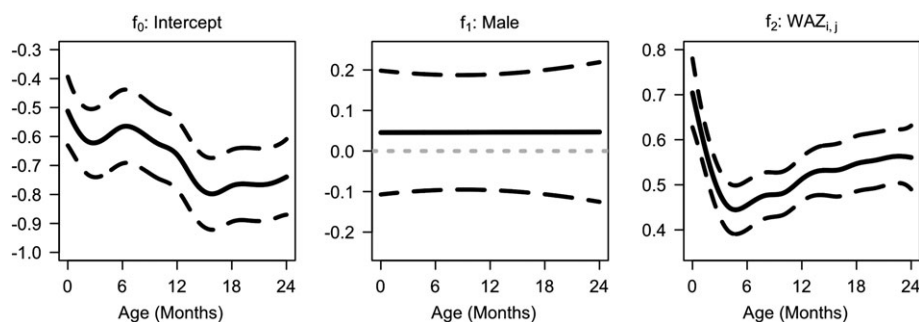
FIGURE 4 Estimated coefficient functions (solid lines) and associated 95% pointwise confidence intervals (dashed lines). WAZ, weight-for-age z-score
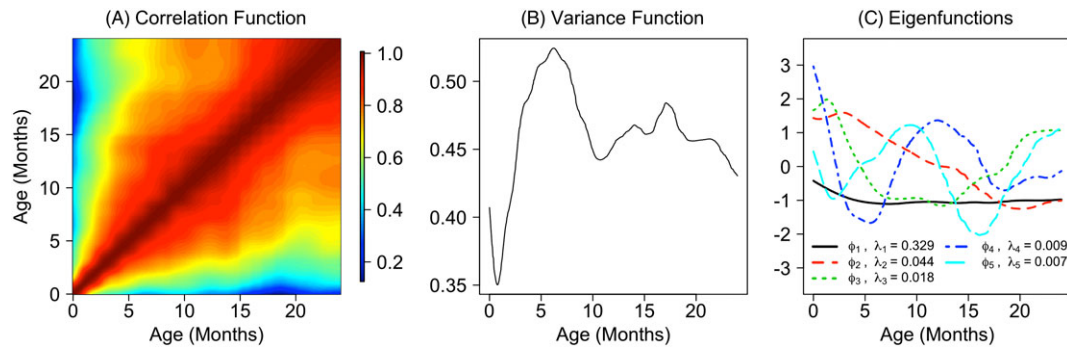
FIGURE 5 Left to right: A, estimated correlation function; B, estimated variance function; C, first 5 estimated eigenfunctions (ϕ) and corresponding eigenvalues (λ)
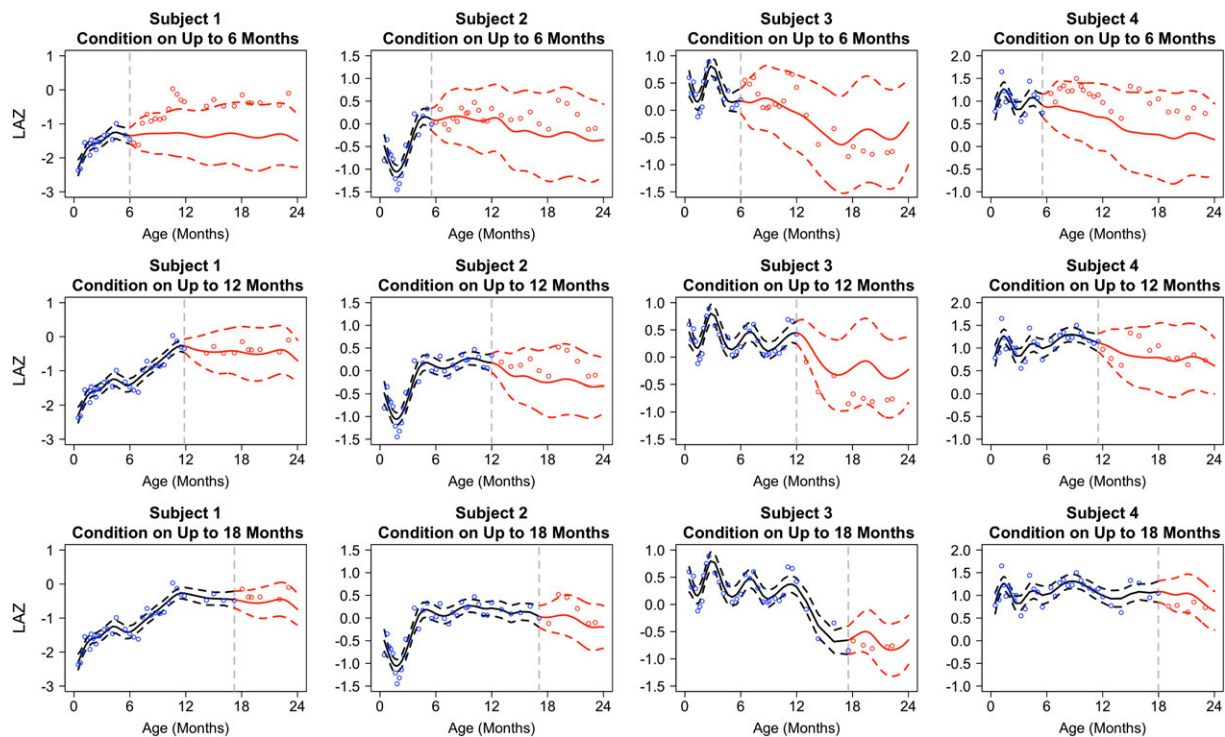
FIGURE 6 Example of dynamic prediction for 4 subjects. Points represent subjects' observed length-for-age z-score (LAZs), solid black/red lines represent the predicted curves, dashed black/red lines indicate 95% pointwise confidence intervals for the trajectories. Only observed data on and to the left of the vertical grey dashed line (blue points) are used for prediction

subject. Only the points on the left of the vertical dashed grey line are used to predict the entire curve. The predicted curves with associated confidence intervals at future time points are shown in red. Unsurprisingly, we see that predictions closer in time to the observed data are generally more accurate. Predictions based on fewer observed data points and farther into the future are generally less accurate and associated with larger uncertainty (wider confidence bands). Moreover, the dynamic prediction intervals generally contain the future actual observed data, indicating the capability of the proposed model to predict future LAZ.

6 | A SIMULATION STUDY

We conduct a simulation study to investigate the performance of the functional concurrent model in estimating the fixed effects and predicting the individual response curves. For simplicity, hereafter, we denote our proposed model as FCR. We compare our model with an additive model with the same mean structure, but no subject random effects (AM), an additive

TABLE 1 Various models

AM:	$\text{LAZ}_{ij} = f_0(t_{ij}) + f_1(t_{ij})\text{Male}_i + f_2(t_{ij})Z_{ij} + \epsilon_{ij}$
AMM:	$\text{LAZ}_{ij} = f_0(t_{ij}) + f_1(t_{ij})\text{Male}_i + f_2(t_{ij})Z_{ij} + b_{0i} + b_{1i}t_{ij} + \epsilon_{ij}$
FCR:	$\text{LAZ}_{ij} = f_0(t_{ij}) + f_1(t_{ij})\text{Male}_i + f_2(t_{ij})Z_{ij} + b_i(t_{ij}) + \epsilon_{ij}$
FRI:	$\text{LAZ}_{ij} = f_0(t_{ij}) + b_i(t_{ij}) + \epsilon_{ij}$

Abbreviations: AM, additive model with the same mean structure; AMM, additive model with scalar random effects; FCR, functional concurrent regression; FRI, functional random intercept; LAZ, length-for-age z-score.

model with scalar random effects (AMM), and a model with functional random intercept but no covariates (FRI). We find that FCR provides more accurate and efficient estimators of coefficient functions and prediction of subject-specific curves in terms of both in-sample and out-of-sample dynamic prediction.

6.1 | Settings

Table 1 presents the list of models we consider. For the AMM, b_{0i} and b_{1i} are subject-specific random intercepts and slopes, respectively. More specifically, (b_{0i}, b_{1i}) are assumed to follow a bivariate normal distribution. For the FCR, $b_i(\cdot)$ are subject-specific random functions. In all models, ϵ_{ij} are independent and identically distributed random errors with a normal distribution $N(0, \sigma_\epsilon^2)$. We let the generating covariance and coefficient functions be the estimates from applying FCR to the real data. Note that the estimates used in the simulation are slightly different than those presented in Section 5 as we reduced the number of basis functions used to estimate $C(s, t)$, resulting in fewer estimated eigenfunctions and faster computation times for the simulations. This reduced number of basis functions was also used for simulating WAZ as described below.

We compare the FCR to the AM, which ignores within-subject correlation entirely, the AMM, which allows for subject-specific random intercepts and slopes for age, and FRI, for which only time is used. We use *mgcv*::gam() for fitting the AM, *mgcv*::gamm() for fitting the AMM, and *face*::face.sparse() for fitting the FRI.

The number of subjects is either $N = 100$ or 200, the number of observations m_i is sampled from either a Uniform(15,25) or a Uniform(25,35), and $\sigma_\epsilon = 0.18$ or 0.37. In the data, σ_ϵ is estimated to be 0.18 for the FCR. Given m_i, the observation times t_{ij} are sampled uniformly without replacement from an evenly spaced grid of 500 points in the unit interval. The male indicator variable is sampled from a Bernoulli distribution with $p = 0.5$.

Both weight-for-age z-scores WAZ_{ij} and subject-specific random functional intercepts $b_i(t_{ij})$ are simulated based on the results of applying FRI to the real data. For example, WAZs are generated by $\text{WAZ}_{ij} = \mu_w(t_{ij}) + \sum_{k=1}^K \xi_{ik}\phi_k(t_{ij}) + \epsilon_{ij}^z$, where $\mu_w(t)$ is the mean function estimate of WAZ, ξ_{ik} is generated from $N(0, \lambda_k)$, and λ_k, ϕ_k are the estimated eigenvalues and eigenfunctions, and $\epsilon_{ij}^z \sim N(0, \sigma_z^2)$ where $\sigma_z = 0.16$ is the square root of the estimated error variance of WAZ in the CONTENT data. Finally, 500 datasets are generated for each setting. Additional simulation scenarios considering alternative covariance structures and varied signal-to-noise ratios yield similar results and hence are included in the Supporting Information.

6.2 | Results

6.2.1 | Estimation of coefficient functions

To evaluate the performance of the FCR for estimating coefficient functions, we use integrated squared error (ISE) defined as $\int_0^1 \left\{ f_k(t) - \hat{f}_k(t) \right\}^2 dt$ where $k = 0, 1, 2$. We compare the FCR with the AM and the AMM fits. For each coefficient function, we calculate the medians and interquartile ranges of ISEs of the estimates from each model.

Table 2 indicates that FCR consistently provides more accurate estimates of all coefficient functions than AMM and AM. Table 3 provides the average pointwise coverage probability of estimated coefficient functions across 500 evenly spaced points between 0 and 1. The FCR and AMM perform reasonably well, while the AM performs poorly. The FCR and AMM provide comparable coverage probabilities for estimating the coefficient functions of the time-invariant covariates ($f_0(\cdot)$ and $f_1(\cdot)$). However, FCR provides substantially higher coverage probabilities for estimating the coefficient functions of the time-varying covariate ($f_2(\cdot)$). All methods compared here seem to give lower than 95% coverage probabilities for f_2. Additional simulations (not shown) indicate that this poor performance can be attributed to the magnitude of the f_2 coefficient and increasing the magnitude of the association between WAZ and the outcome results in close to 95% confidence interval coverage probabilities for FCR.

TABLE 2 100×Median (interquartile range) of integrated squared error for predicting coefficient functions using FCR, AMM, and AM across 500 simulations for each combination of (N, σ_ϵ)

		N = 100			N = 200		
		FCR	**AMM**	**AM**	**FCR**	**AMM**	**AM**
$m_i \sim$ Unif[15, 25]							
$\sigma_\epsilon = 0.18$	f_0	0.48 (0.71)	0.57 (0.76)	0.63 (0.82)	0.25 (0.40)	0.27 (0.40)	0.34 (0.44)
	f_1	0.74 (1.73)	1.03 (1.83)	1.24 (2.01)	0.38 (0.69)	0.43 (0.82)	0.55 (0.96)
	f_2	0.26 (0.30)	0.30 (0.34)	0.42 (0.53)	0.14 (0.16)	0.15 (0.18)	0.25 (0.24)
$\sigma_\epsilon = 0.37$	f_0	0.54 (0.84)	0.57 (0.90)	0.65 (0.95)	0.28 (0.39)	0.28 (0.40)	0.36 (0.43)
	f_1	0.92 (1.68)	1.12 (1.92)	1.29 (1.87)	0.43 (0.84)	0.51 (0.86)	0.66 (1.00)
	f_2	0.30 (0.30)	0.33 (0.38)	0.49 (0.47)	0.16 (0.17)	0.17 (0.17)	0.26 (0.30)
$m_i \sim$ Unif[25, 35]							
$\sigma_\epsilon = 0.18$	f_0	0.50 (0.74)	0.60 (0.80)	0.71 (0.91)	0.22 (0.38)	0.26 (0.42)	0.32 (0.44)
	f_1	0.87 (1.46)	1.12 (1.83)	1.31 (1.94)	0.36 (0.73)	0.49 (0.73)	0.57 (0.87)
	f_2	0.21 (0.25)	0.29 (0.37)	0.38 (0.50)	0.11 (0.13)	0.14 (0.18)	0.21 (0.27)
$\sigma_\epsilon = 0.37$	f_0	0.54 (0.73)	0.55 (0.84)	0.61 (0.82)	0.27 (0.40)	0.27 (0.42)	0.32 (0.46)
	f_1	0.81 (1.60)	0.92 (1.68)	1.11 (1.73)	0.39 (0.79)	0.50 (0.88)	0.59 (0.93)
	f_2	0.26 (0.26)	0.30 (0.35)	0.44 (0.50)	0.12 (0.13)	0.13 (0.15)	0.23 (0.24)

Abbreviations: AM, additive model with the same mean structure; AMM, additive model with scalar random effects; FCR, functional concurrent regression.

TABLE 3 Average coverage probabilities for 95% confidence bands using FCR, AMM, and AM under various simulation scenarios

		N = 100			N = 200		
		FCR	**AMM**	**AM**	**FCR**	**AMM**	**AM**
$m_i \sim$ Unif[15, 25]							
$\sigma_\epsilon = 0.18$	f_0	0.93	0.94	0.56	0.93	0.93	0.60
	f_1	0.91	0.93	0.53	0.93	0.94	0.58
	f_2	0.84	0.71	0.63	0.84	0.72	0.64
$\sigma_\epsilon = 0.37$	f_0	0.93	0.93	0.60	0.93	0.93	0.63
	f_1	0.93	0.94	0.57	0.94	0.95	0.57
	f_2	0.86	0.79	0.65	0.88	0.80	0.67
$m_i \sim$ Unif[25, 35]							
$\sigma_\epsilon = 0.18$	f_0	0.91	0.92	0.50	0.93	0.94	0.55
	f_1	0.92	0.92	0.47	0.94	0.94	0.49
	f_2	0.84	0.64	0.58	0.85	0.65	0.61
$\sigma_\epsilon = 0.37$	f_0	0.92	0.94	0.55	0.93	0.94	0.58
	f_1	0.93	0.95	0.53	0.93	0.94	0.51
	f_2	0.86	0.75	0.59	0.90	0.78	0.63

Abbreviations: AM, additive model with the same mean structure; AMM, additive model with scalar random effects; FCR, functional concurrent regression. Coverage is assessed at 500 equally spaced points between 0 and 24 and then averaged across the 500 points for each simulation scenario.

6.2.2 | Dynamic curve prediction

Regarding individual curve predictions, we compare the FCR with the functional principal component model (FRI) and the AMM described in Table 1. We do not include a comparison with the AM because it focuses on the estimation of fixed effects and is therefore not suitable for individual prediction. For each simulated dataset, we calculate the mean ISE (MISE) for predicting all individual curves. The evaluation criteria are the median and interquartile range of the MISEs over all simulated datasets. Mean ISE is calculated using 50 subjects not included during model fitting. Dynamic prediction is performed on 3 nonoverlapping intervals ([8/24, 12/24], [14/24, 18/24], [20/24, 24/24]) using data conditional on up to 3 time points (6/24, 12/24, 18/24). This results in 6 observed dynamic prediction periods, with MISE calculated for each period separately (see Table 4). For ease of comparison with the CONTENT application, we rescale these time periods to months on the interval 0 to 2 years.

Table 4 reports the results for dynamic prediction error for one simulation scenario ($m_i \sim$ Unif(25, 35), $N = 100, \sigma_\epsilon =$ 0.18). Dynamic prediction accuracy increases with the amount and proximity of subject data used in prediction. For

TABLE 4 10×Median (interquartile range) of mean integrated squared error for dynamically predicting subject-specific curves using the FCR, fPCA, and AMM across 500 simulations for $N = 100$, $m_i \sim \text{Unif}(25, 35)$, $\sigma_\epsilon = 0.18$

		Prediction Time Window			
		8-12 mo	14-18 mo	20-24 mo	
Observed Data	0-6 mo	2.06 (0.56)	3.41 (0.97)	3.74 (1.20)	FCR
		2.25 (0.64)	3.66 (1.07)	3.88 (1.17)	FRI
		2.21 (0.60)	3.68 (1.07)	4.71 (1.55)	AMM
	0-12 mo		1.80 (0.49)	2.42 (0.63)	FCR
			1.91 (0.56)	2.54 (0.70)	FRI
			2.17 (0.63)	4.77 (1.47)	AMM
	0-18 mo			1.18 (0.36)	FCR
				1.27 (0.36)	FRI
				2.50 (0.76)	AMM

Abbreviations: AMM, additive model with scalar random effects; FCR, functional concurrent regression; FRI, functional random intercept; fPCA, functional principal component analysis. All dynamic prediction errors are assessed using 50 subjects that were not included in the model fitting procedure are used to evaluate the dynamic prediction error presented here. Note that although simulations were performed on the unit interval, the time periods have been rescaled to reflect the time domain of the CONTENT data.

example, prediction accuracy for 20 to 24 months using data up to 18 months is better than both: (1) prediction accuracy for 14 to 18 months using data up to 12 months and (2) prediction accuracy for 20 to 24 months using data up to 12 months. Comparing the FCR to the AMM, the improvement in dynamic prediction using up to 12 months of data is substantial, with reductions in MISE between 17% and 53%. However, note that while FCR outperforms FRI, the difference is consistently less than 10%.

Additional simulations (not reported) show that increasing the magnitude of the association between covariates and the outcome increases the relative performance of both FCR and AMM relative to FRI. This makes sense as FRI does not include any covariate information in the model. We therefore recommend using simulation to assess the relative performance of AMM, FRI, and FCR in a particular application under a range of plausible true data generating mechanisms.

7 | DISCUSSION

We have proposed a dynamic FCR model and provided accompanying code that is easy to use. We applied our method directly to irregularly spaced growth curve data to study the association between length and weight of children and proposed a procedure for performing dynamic prediction. Using the motivating data as a basis for a comprehensive simulation study, our method showed superior performance compared to traditional methods for analyzing growth curve data in terms of estimation and inference on fixed effects as well as both in-sample and dynamic subject predictions.

As we focus heavily on individual predictions, we would like to underline that simply predicting a new subject for standard mixed effects models is not currently available in the most widely used *R* packages for fitting mixed models (*lme4* and *nlme*), or *refund* for functional mixed effects models. As a result, not only is our method more accurate in terms of subject predictions, it is also easier to make out-of-sample predictions from the perspective of an end user. This is particularly relevant for users who want to perform dynamic prediction as the BLUP estimates for the random effects must be recalculated at every point of interest.

In addition to providing a user-friendly interface, the underlying software has been extensively tested by research groups working on child growth data. While the estimation procedure described in this paper is similar to that proposed in Cederbaum et al,[19] the procedure has not been implemented in any peer-reviewed publication. Moreover, our covariance estimation procedure[37] is fast, accurate, and was designed specifically for sparse functional covariance estimation.

Although our primary focus is on dynamic prediction, the methods are not limited to the specific functional concurrent model considered in the paper. Indeed, it can be used for performing general functional regression with a random functional intercept. This class of models is applicable to a wide variety of study designs and data types. Ultimately, the use of statistical methods depends strongly on the degree to which these methods are accessible. Here, we provide both the method and an estimation procedure that is relatively computationally inexpensive. The accessibility of our code may

help facilitate adoption in practice. An *R* package *fcr* has been compiled to fit the proposed dynamic FCR models and will be distributed to CRAN soon.

ACKNOWLEDGEMENTS

ORCID

Luo Xiao 🔟 http://orcid.org/0000-0001-8707-0914

REFERENCES

1. Liang K, Zeger S. Longtitudinal data analysis using generalized linear models. *Biometrika*. 1986;73(1):13-22.
2. Laird N, Ware J. Random-effects models for longitudinal data. *Biometrics*. 1982;38(4):963-974.
3. Lindstrom M, Bates D. Nonlinear mixed effects models for repeated measures data. *Biometrics*. 1990;46(3):673-687.
4. Davidian M, Giltinan D. *Nonlinear Models for Repeated Measurement Data*. New York: Chapman and Hall; 1995.
5. Grajeda LM, Ivanescu A, Saito M, et al. Modelling subject-specific childhood growth using linear mixed-effect models with cubic regression splines. *Emerg Themes Epidemiol*. 2016;13(1).
6. Ruppert D, Wand MP, Carroll RJ. *Semiparametric Regression*. Cambridge: Cambridge university press; 2003.
7. Ramsay J, Silverman B. *Functional Data Analysis*. 2nd ed. New York: Springer; 2005.
8. Guo W. Functional mixed effects models. *Biometrics*. 2002;58(1):121-128.
9. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *J Am Stat Assoc*. 2005;100(470):577-590.
10. Morris JS, Carrol RJ. Wavelet-based functional mixed models. *J R Stat Soc Ser B (Stat Method)*. 2006;68(2):179-199.
11. Zhang D, Lin X, Sowers M. Two-stage functional mixed models for evaluating the effect of longitudinal covariate profiles on a scalar outcome. *Biometrics*. 2007;63(2):351-362.
12. Chen H, Wang Y. A penalized spline approach to functional mixed effects model analysis. *Biometrics*. 2011;67(3):861-870.
13. Di C, Crainiceanu CM, Caffo BS, Punjabi NM. Multilevel functional principal component analysis. *Ann Appl Stat*. 2009;3(1):458-488.
14. Zhou L, Huang JZ, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ. Reduced rank mixed effects models for spatially correlated hierarchical functional data. *J Am Stat Assoc*. 2010;105(489):390-400.
15. Zhu H, Brown PJ, Morris JS. Robust, adaptive functional regression in functional mixed model framework. *J Am Stat Assoc*. 2011;106(495):1167-1179.
16. Yuan Y, Gilmore JH, Geng X, et al. Fmem: Functional mixed effects modeling for the analysis of longitudinal white matter tract data. *NeuroImage*. 2014;84:753-764.
17. Hadjipantelis PZ, Aston JAD, Müller HG, Evans JP. Unifying amplitude and phase analysis: a compositional data approach to functional multivariate mixed-effects modeling of Mandarin Chinese. *J Am Stat Assoc*. 2015;110(510):545-559.
18. Reimherr M, Nicolae D. Estimating variance components in functional linear models with applications to genetic heritability. *J Am Stat Assoc*. 2016;111(513):407-422.
19. Cederbaum J, Pouplier M, Hoole P, Greven S. Functional linear mixed models for irregularly or sparsely sampled data. *Stat Model*. 2016;16(1):67-88.
20. Scheipl F, Staicu AM, Greven S. Functional additive mixed models. *J Comput Graphical Stat*. 2015;24(2):477-501.
21. Huang L, Scheipl F, Goldsmith J, et al. refund: Regression with functional data. R package version 0.1-16. 2016. https://CRAN.R-project.org/package=refund. Accessed August 4, 2016.
22. Bates D, Martin M, Bolker B, Walker S. Fitting linear mixed-effects models using lme4. *J Stat Software*. 2015;67(1):1-48.
23. Pinheiro J, Bates D, DebRoy S, Sarkar D, R Core Team. nlme: linear and nonlinear mixed effects models. R package version 3.1-124. 2016. Available at https://CRAN.R-project.org/package=nlme. Accessed January 21, 2016.
24. Chiou JM. Dynamical functional prediction and classification, with application to traffic flow prediction. *The Ann Appl Stat*. 2012;6(4):1588-1614.
25. Goldberg Y, Ritov Y, Mandelbaum A. Predicting the continuation of a function with applications to call center data. *J Stat Plann Inference*. 2014;147:53-65.
26. Kraus D. Components and completion of partially observed functional data. *J R Stat Soc Ser B*. 2015;77(4):777-801.
27. Delaigle A, Hall P. Approximating fragmented functional data by segments of Markov chains. *Biometrika*. 2016;103(4):779-799.
28. Shang HL. Functional time series forecasting with dynamic updating: an application to intraday particulate matter concentration. *Econom and Stat 1*. 2017;1:184-200.

29. Hastie T, Tibshirani R. Varying-coefficient models. *J R Stat Soc Ser B (Stat Methodol)*. 1993;55(4):757-796.

30. Jaganath D, Saito M, Gilman RH, et al. First detected Helicobacter pylori infection in infancy modifies the association between diarrheal disease and childhood growth in Peru. *Helicobacter*. 2014;19(4):272-279.

31. Borghi E, de Onis M, Garza C, et al. Construction of the world health organization child growth standards: selection of methods for attained growth curves. *Stat Med*. 2006;25(2):247-265.

32. Trowbridge F, Marks JS, Lopez del Romana G, Madrid S, Boutton TW, Klein PD. Body composition of Peruvian children with short stature and high weight-for-height. II. Implications for the interpretation for weight-for-height as an indicator of nutritional status. *Am J Clin Nutr*. 1987;46(3):411-418.

33. Sterling R, Miranda J, Gilman R, et al. Early anthropomorphic indices predict short stature and overweight status in a cohort of peruvians in early adolescence. *Am J Phys Anthropol*. 2012;148(3):451-461.

34. Malfait N, Ramsay JO. The historical functional linear model. *Can J Stat*. 2003;31(2):115-128.

35. Şentürk D, Müller HG. Generalized varying coefficient models for longitudinal data. *Biometrika*. 2008;95(3):653-666.

36. Şentürk D, Müller HG. Functional varying coefficient models for longitudinal data. *J Am Stat Assoc*. 2010;105(491):1256-1264.

37. Xiao L, Li C, Checkley W, Crainiceanu C. Fast covariance estimation for sparse functional data. *Stat Comput*. 2017. https://doi.org/10.1007/s11222-017-9744-8

38. Ivanescu AE, Staicu AM, Scheipl F, Greven S. Penalized function-on-function regression. *Comput Stat*. 2015;30(2):539-568.

39. Eilers PH, Marx BD. Flexible smoothing with b-splines and penalties. *Stat Sci*. 1996;11(2):89-102.

40. Ruppert D. Selecting the number of knots for penalized splines. *J Comput Graphical Stat*. 2002;11(4):735-757.

41. Whittaker ET. On a new method of graduation. *Proc Edinburgh Math Soc*. 1923;41:63-75.

42. Xiao L, Li C, Checkley W, Crainiceanu C. face: Fast covariance estimation for sparse functional data. R package version 0.1-2. 2016. Available at http://CRAN.R-project.org/package=face. Accessed January 3, 2017.

43. Wood SN. *Generalized Additive Models: An Introduction with R*. 2nd ed. Boca Raton: CRC Press; 2017.

44. Wood SN, Li Z, Shaddick G, Augustin NH. Generalized additive models for gigadata: Modeling the U.K. black smoke network daily data. *J Am Stat Assoc*. 2017;112(519):1199-1210.

SUPPORTING INFORMATION

Additional Supporting Information may be found online in the supporting information tab for this article.