Dynamic Prediction of Non-Guassian Outcome with fast Generalized Functional Principal Analysis

Ying Jin   Andrew Leroux

April 04, 2023

Introduction

Biomedical investigators are often interested in predicting future observations of subjects based on their historical data, typically referred to as dynamic prediction. Traditionally, this type of data has been modeled either marginally, such as generalized estimating equations, or constitutionally on specific subjects, such as mixed effect models (Laird and Ware 1982; LIANG and ZEGER 1986; Lindstrom and Bates 1990; "Nonlinear models for repeated measurement data" 2003). For subject-specific models, predictions are made based on the correlation between repeated measures from the same subject and covariates that can be either fixed or varying. However, most of these methods are based on linear model structure, which is very limited in terms of flexibility of correlation. When sample size is large or the density of repeated measures is high, they also tend to cause severe computational burden (Rizopoulos 2022). In addition, with generalized, non-Gaussian outcomes, these methods lack the ability to handle out-of-sample prediction because there is usually no close-form solution for subject-specific random effects

To address these problems, one may turn to functional mixed effect models instead when measures are dense across the domain of measures. Such methods accommodate more flexible correlation structure by modeling subject-specific random effects as a function. Non-parametric smoothing (Scheipl et al. 2014) are often incorporated to speed up the computation, where random effects are approximated by spline basis functions or eigenfunctions from functional principal compo-

nent analysis (FPCA). The introduction of basis functions also makes out-of-sample prediction more straightforward. Instead of estimating subject-specific random effects of new observations, we can simply estimate coefficients/loadings on the basis function used for smoothing. Existing research on dynamic prediction with functional data analysis methods has been focusing on continuous/Gaussian outcomes, usually modelling subject-specific random effects with FPCA (Chiou 2012; Goldberg et al. 2014; Shang 2017). Kraus (2015) has used this approach to predict missing observations in partially observed function tracks, and Delaigle and Hall (2016) achieved similar goals using Markov Chains. While methods mentioned above used only partial observations for prediction with an intercept-only model, Leroux et al. (2018) proposed Functional Concurrent Regression (FCR) framework which can incorporate the effect of subject-specific predictors.

Unfortunately, fewer papers have focused on its extension to generalized, non-Gaussian outcomes, such as series of binary or count outcomes. Existing methods also tend to be very computationally intensive. For example, Chen et al. (2013) proposed approaches to fit marginal functional models that is compatible to multi-level, generalized outcomes. Goldsmith et al. (2015) established a model framework that takes into account the fixed effect of time-invariant covariates, with parameters estimated with Bayesian method in *Stan*. Gertheiss et al. (2016) identified bias introduced by directly applying FPCA methods to generalized functions, and proposed to address this problem using a two-stage joint estimation strategy. Linde (2009) used an adapted Bayesian variational algorithm for FPCA of binary and count data. In terms of implementation, Wrobel et al. (2019) proposed a fast, efficient way to fit GFPCA on binary data using EM algorithm, accompanied by the an open source R package *registr*.

In this paper, we aim to develop a fast, scalable method for dynamic prediction of generalized functional outcome based on functional mixed effect model with FPCA smoothing. Specifically, "Generalized" indicates that the outcome follows non-Gaussian exponential family distribution, and the term "functional" indicates that repeated measure are collected densely across the study domain (e.g., time). Section 2 explains the implementation procedure of the proposed method. In Section 3, we illustrate the performance and efficiency of our proposed method in a simulation study. In Section 4, we apply this method to a real world dataset from a weight loss study. Section 5 presents a discussion of advantages and limitation of the proposed method.

Method

The observed data for a single subject i is $(t, Y_i(t))$, where t consists of dense, discrete points along the domain of data collection. In practice, t is usually a time index along the duration of study. We hereafter refer to the range of t as the "functional domain". For simplicity, we further assume that outcomes are collected on a regular grid over the functional domain, meaning that t are evenly distributed and does not change across subjects.

$Y_i(t)$ is the non-Gaussian outcome observed at t. We assume that the outcome $Y_i(t)$ can be characterized by a latent continuous function $\eta_i(t)$. That is, at a specific t, $Y_i(t)$ follows a exponential family distribution such that:

$$g[E(Y_i(t))] = \eta_i(t) = \beta_0(t) + b_i(t)$$

where g is an appropriate link function, $\beta_0(t)$ is the population mean of latent function, and $b_i(t)$ is a subjects-level random effect function following a zero-mean Gaussian process.

According to the Karhunen-Loève theorem, the unobserved random effect function $b_i(t)$ can be represented as the linear combination of a infinite set of orthogonal functions: $b_i(t) = \sum_{k=1}^{\infty} \xi_{ik}\phi_k(t)$. We approximate this function by a finite set of $\phi_k(t)$, such that $b_i(t) = \sum_{k=1}^{K} \xi_{ik}\phi_k(t)$.

In practice, this approximation could be implemented through an FPCA procedure on $\eta_i(t)$. In this case $\phi_k(t)$, $k \in \{1, ..., K\}$ are the first K eigenfunctions that explain the most variation in $\eta_i(t)$. These eigenfunctions, also referred to as Principal Component (PC) functions, are orthogonal are can be orthonormal with proper scaling. ξ_{ik} are subject-specific scores (or loadings) on each PC function. According to properties of FPCA method, these scores are mutually independent over both subject (i) and eigenfunctions (k). That is, each ξ_{ik} follows normal distribution $N(0, \lambda_k)$, where λ_k is the kth eigenvalues.

Fast Generalized FPCA Based on the problem set up above, we propose the following algorithm for fast, generalizable implementation of FPCA (fGFPCA) on the unobserved latent process $\eta_i(t)$:

1. Bin the observed outcomes in to small, non-overlapping, equal length intervals. We hereafter index the bins by their midpoints s. The number of observations from the same subject in bin s is n_s, and t_{sj} refers to the jth observation point in bin s.

2. Fit a local Generalized Mixed Model at every bin. Specifically, at bin s, we fit $g[E(Y_i(t_{sj}))] = \beta_0(s) + b_i(s), j \in \{1, ..., n_s\}$. From this series of models we can estimate the individual latent functions at every bin: $\hat{\eta}_i(s) = \hat{\beta}_0(s) + \hat{b}_i(s)$.

3. Fit FPCA on the estimated latent functions: $\hat{\eta}_i(s) = f_0(s) + \sum_{k=1}^{K} \xi_{ik}\phi_k(s) + \epsilon_i(s)$, where $\epsilon_i(t)$ indicates random noise introduced by the binning process in previous steps. This random noise is assumed to follow a zero-mean Gaussian process, such that at each specific point t, $\epsilon_i(t) \sim N(0, \sigma^2)$. The FPCA output includes estimates of basis functions $\mathbf{\Phi} = \{\phi_1(s), ..., \phi_k(s)\}$, eigenvalues $\hat{\lambda}_1...\hat{\lambda}_k$, mean function $\hat{f}_0(s)$ and residual variance $\hat{\sigma}^2$.

Dynamic prediction Now assume we have a new observations with partially observed outcome $Y_{new}(t)$, $t \leq t_m$. The maximum observed point is t_m, which is a point along the functional domain and belongs to bin s_m on the binned grid from step 1 above. With components extracted from the FPCA model in step 3, we will be able to make predictions on the future, unobserved bins. Specifically, with the partially observed binary function $Y_{new}(t)$, we will be able to calculate the maximum likelihood estimate (MLE) of the subject-specific PC scores $\hat{\xi}_{ik}$. Then the value of latent functions at unobserved points can be estimated as $\hat{\eta}_i(s) = \hat{f}_0(s) + \sum_{k=1}^{K} \hat{\xi}_{ik}\phi_k(s)$

Following the algorithm above, predictions of individual latent functions are made on the binned grid based on partially observed tracks on the un-binned grid. Since the bins are set up to be small in length, the binned grid would still be dense. However, in situations where predictions on the original, un-binned grid is needed, linear interpolation of estimated latent function tracks turns out to be a fast, convenient way with good performance for prediction at points between bins.

Simulation

In this section, we illustrate the predictive performance and computational efficiency of the proposed method through a simulation study. We simulated 50 datasets, each with 500 subjects.

For every subject, we generate 1000 binary outcomes $Y_i(t) \in \{0,1\}$ across functional domain $t \in (0,1000]$, where the distribution of outcome is characterized by a continuous latent function. The data generation mechanism can be expressed as follows:

$$Y_i(t) \sim Binomial(\frac{exp(\eta_i(t))}{1+exp(\eta_i(t))})$$

$$\eta_i(t) = f_0(t) + \xi_{i1}\sqrt{2}sin(2\pi t) + \xi_{i2}\sqrt{2}cos(2\pi t) + \xi_{i3}\sqrt{2}sin(4\pi t) + \xi_{i4}\sqrt{2}cos(4\pi t)$$

In this simulation, we set $f_0(t) = 0$. ξ_{ik} are mutually independent normal random variables $\xi_{ik} \sim N(0, \lambda_k)$. Here we set the values of λ_k to be 0.5^{k-1}, $k \in (1, ..., 4)$. In addition, for simplicity of presentation, we generate data on a regular grid, which means observations points are equally distributed across $(0,1000]$ and are the same for all subjects.

We use two metrics to evaluate the out-of-sample predictive performance: integrated squared error (ISE) and Area-Under-the-Receiver-Operator-Curve (AUC). ISE assesses the prediction accuracy of latent continuous functions. It is evaluated on the binned grid at midpoints of each unobserved bin. With observations up to bin s_m, ISE is defined as $\frac{1}{N}\sum_{i=1}^{N}\sum_{s>s_m}(\hat{\eta}_i(s) - \eta_i(s))^2$. The second metric, AUC, focuses on evaluation of prediction of the binary outcome. Since the binary outcomes are generated on the original, un-binned grid, we evaluated AUC on this grid by estimating values of latent functions between bins with linear interpolation.

As a reference method, we compare our method to Generalized Linear Mixed Models using Adaptive Gaussian Quadrature (GLMMadaptive). This is one of the fastest existing method developed for dynamic prediction of repeated generalized outcomes. Just like many mixed models, this method is very limited in terms of flexibility. For example, the model used for prediction of our simulated datset would simply be an linear model with one covariate indicating observation time: $g(E(Y_i)) = \beta_0 + \beta_1 t + b_{i0} + b_{i1}t$. While the flexibility of this mixed model can be increased using spline functions, the dimension of spline basis is also restricted by computational ability, and is unfeasible to implement under the scale of our simulated data or the complexity of our proposed method.

The average ISE and AUC across all simulation is presented in Table 1, and Figure 1 presents the predicted latent function of four randomly selected subjects from the first simulated dataset.

The prediction is made conditioning on different length of observed track (specifically, with observations up to t = 200, 400, 600, 800 respectively), and evaluation is made on equal-length time windows on the unobserved tracks following the maximum observation time. As Table 1 reveals, the proposed method (fGFPCA) outperforms GLMMadaptive under every scenario, which was expected since the linear model structure of GLMM is too simple for the cyclic data generation mechanism. In addition, fGFPCA also took less computation time. For one simulated dataset, fGFPCA spent 4.94 minutes on model fitting and out-of-sample prediction overall, while GLMMadaptive took 5.11 minutes.

Because of the flexibility of our proposed model framework, the accuracy of prediction at specific time points would improve as more observed data is collected. This is revealed in Table 1 as ISE decreases and AUC increases with maximum observed time (left-to-right), also in Figure 1 as predicted curves get closer to the true latent function with longer observed track. However the same tendency is not observed with GLMMadaptive models, as a result of restricted model flexibility. A linear model that fits well to a specific part can fit very badly to other parts along the underlying latent function tracks, especially when the shapes of functions are wiggly or cyclic. Therefore, more observations do not necessarily make the model more predictive.
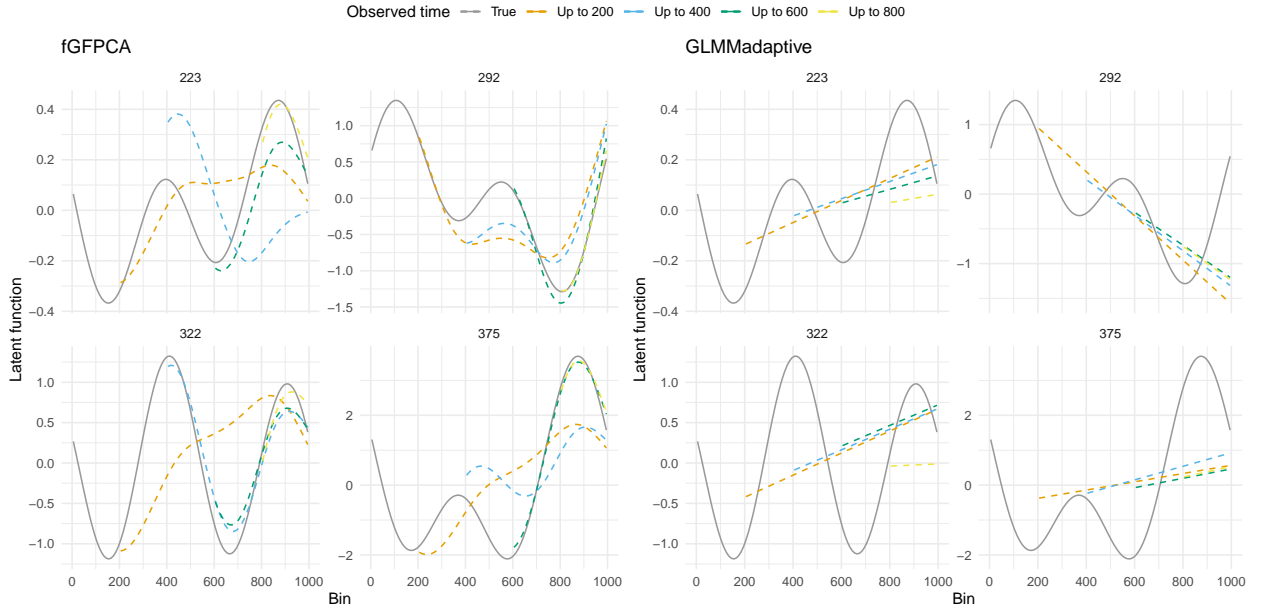

Figure 1: Predicted track of four randomly selected subjects from one simulated dataset. The grey solid line indicates the true latent continuous function. The dashed lines indicated predicted latent function tracks, and color indicates different observation time.

Table 1: Predictive performance of fGFPCA and GLMM adaptive on the simulated datasets. ISE and AUC are average values across all 50 simulations.

	Maximum observed time							
	fGFPCA				GLMMadaptive			
	200	400	600	800	200	400	600	800
Prediction time window								
ISE								
(200, 400]	15.29				38.20			
(400, 600]	18.68	8.53			28.59	26.46		
(600, 800]	22.53	5.51	1.65		31.32	27.97	27.40	
(800, 1000]	11.34	8.72	1.90	1.30	55.19	46.89	58.57	59.25
AUC								
(200, 400]	0.74				0.59			
(400, 600]	0.66	0.73			0.52	0.59		
(600, 800]	0.71	0.79	0.80		0.67	0.70	0.69	
(800, 1000]	0.74	0.75	0.78	0.78	0.52	0.56	0.53	0.57

Data application

In this section, we illustrate the performance of our proposed method using data from a randomized trial on weight loss at University of Colorado Anschutz Medical Campus (Bothwell et al. 2022; Ostendorf et al. 2022). The study originally aimed to study the effect of dietary strategies on weight loss, and participants were given a smart scale for daily at-home weighing. This record was then scaled into a binary indicator of participant adherence to the daily-weigh-in procedure, where 1 indicates participants followed the requirement and 0 otherwise. In this section, we will use this series of daily indicators as the binary functional outcome, and use the proposed method to dynamically predict participant adherence on future days given historical track.

The dataset includes 55 participants aged 22-56 at baseline. Adherence to daily weigh-in is available along 400 consecutive days for all participants. In other words, for each individual, 400 binary measurements were taken on a regular grid along the functional domain (number of days into the study). We will fit both fGFPCA and GLMMadaptive models and compared their performance on this data. For fGFPCA, every 10 observations are binned together, resulting in 40 bins for final model fitting. Linear interpolation was used for extension from the binned grid back to the un-binned grid. Since the latent continuous function tracks are not known in this case, prediction

performance on evaluated with AUC alone.

Table 2 compares the predictive performance of both models. Similar to the simulatio study, prediction is made conditioning on observations up to the 100th, 200th and 300th day respectively, and evaluation is made every 100 days following the maximum observation time. As is revealed, when the observed time is short (up to 100th day), the two method performed similarly when predicting recent days, however fGFPCA did better in predicting further days along the trial (prediction time window of $(300, 400]$). As more observations coming in, fGFPCA sees increasingly higher AUC compared to GLMMadaptive model. This could be a result of the difference in model flexibility of the two methods. While the linear relationship in underlying latent functions are not likely to change immediately following the observed track, the whole track could be much more complex. Therefore, while fGFPCA performs similarly to GLMMadaptive right after the end of observation, it does a much better job capturing the underlying pattern along the whole track. Figure 2 further illustrate such difference using predicted latent function tracks of four randomly drawn subjects.
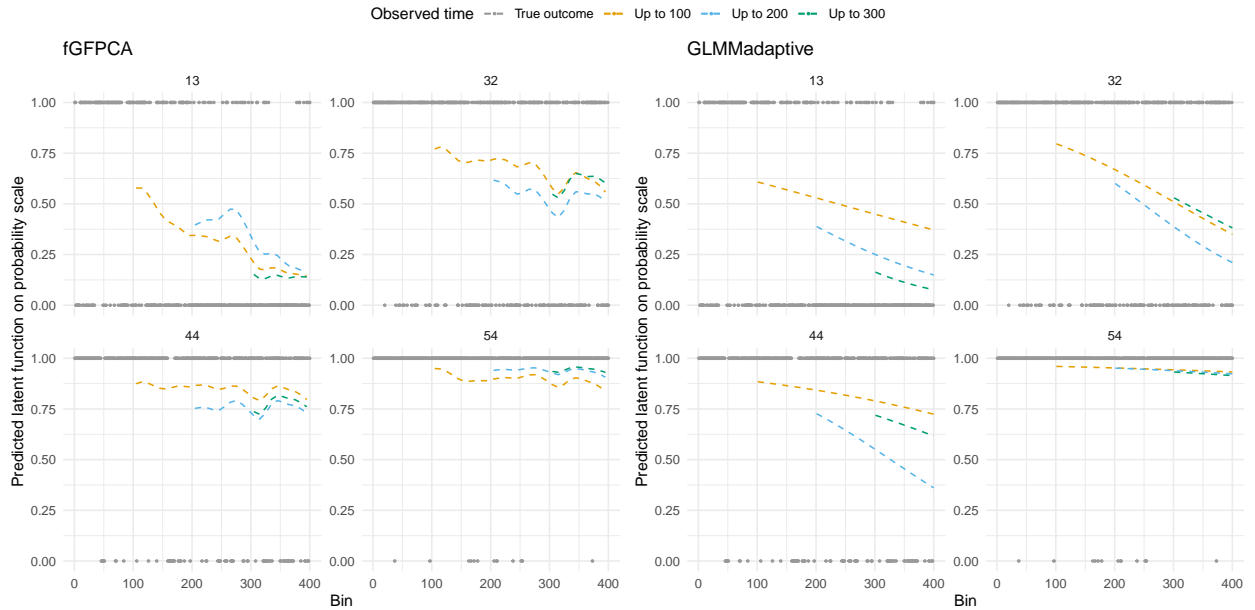

Figure 2: Predicted track of four randomly selected subjects from the daily weigh-in dataset. The grey solid line indicates the true latent continuous function. The dashed lines indicated predicted latent function tracks, and color indicates different observation time.

Table 2: AUC of fGFPCA and GLMM adaptive on the daily weigh-in dataset

	Maximum observed time					
	fGFPCA			GLMMadaptive		
	100	200	300	100	200	300
Prediction time window						
(100, 200]	0.75			0.76		
(200, 300]	0.77	0.82		0.76	0.80	
(300, 400]	0.76	0.81	0.84	0.74	0.78	0.82

Discussion

The simulation study and case study of daily weigh-in data above have revealed the proposed method can achieve better predictive performance with much less computational burden. Compared to existing methods based on linear mixed model, fGFPCA can accommodate flexible correlation structure between repeated measure, which is particularly useful when outcome is generated from latent functions with wiggly or cyclic pattern. As in the simulation setting in Section 3, the binary outcomes are generated from a cyclic process. fGFPCA has much lower ISE and higher AUC than GLMM in this case, and its prediction accuracy also improves as more outcomes are collected, which is a desirable property for dynamic prediction where the datasets would be periodically updated with newly collected data. Moreover, fGFPCA also significantly reduce time spent on model fitting and predict, which is beneficial when repeated measures are dense or sample size is large. In fact, as far as we are aware of, the proposed method is the only feasible method that can handle the scale of the simulated datasets with compatible flexibility.

While the proposed method achieved good accuracy of point prediction, it is more complicated to develop a straightforward, interpretable metric to evaluate the precision of prediction. The challenge arises from the fact that fitting local GLMMs (Step 2) and FPCA (Step 3) can both introduce uncertainty to the final prediction. Since they are implemented consecutively, it is not clear how the uncertainty from both procedures should be integrated into one single interval estimator. In practice, one could consider developing interval predictions conditional on the local GLMMs, quantifying uncertainty of individual score estimates from FPCA model alone. Since the score $\hat{\zeta}_{ik}$ is essentially a maximum likelihood estimate, it is natural to estimate its variance based on

likelihood theory with observed information $I(\hat{\xi}_{ik})$. Therefore, the variation of prediction interval is:

$$Var(\hat{\eta}_i(s) - \eta_i(s)) = \boldsymbol{\Phi}(s)I(\hat{\boldsymbol{\xi}}_i)\boldsymbol{\Phi}^T(s) + \hat{\sigma}_\epsilon^2$$

Where $\boldsymbol{\Phi}(s) = (\phi_1(s)...\phi_K(s))$ and $\hat{\boldsymbol{\xi}}_i = (\hat{\xi}_{i1}, ..., \hat{\xi}_{iK})$.

The hyperparameters introduced by the binning procedure in Step 1 can also affect the final predictive performance, thus further investigations are needed to quantify such effects. Predictive accuracy and computational speed can change with bin width, number of observations in each bin, whether the bins overlap with each other, also whether the bins are equally spaced along the functional grid. Moreover, if the bins are very narrow with only a few observations, unidentifiability issues can happen when fitting local mixed models. While this may not necessarily affect point predictions, it can cause numeric failure to estimate standard error and prediction intervals. In practice, we expect this to happen when the measurements are sparse or sample size is small. Under such scenarios, one may choose traditional mixed model framework instead to get valid estimates at the expense of computational time.

It should also be noted that the binning procedure in Step 1 changes the density of repeated measures. We essentially assume that latent functions are smooth in small intervals, and treat adjacent observations as if they are generated from the same time point. The final prediction of latent continuous function tracks are made on the binned grid, integrating information from several adjacent observations in the same bin. In Section 3 and 4, this sparser grid was transformed to the original, un-binned grid using linear interpolation of the predicted latent function. However, this is by no means the only method to be considered. Such transformation can also be made by projection to a higher-dimensional space, similar as general additive model with a large number of basis functions. We suspect that this method would work better than linear interpolation when the original grid is sparse or irregular, but take longer time for computation.

References

Bothwell, S., Kaizer, A., Catenacci, V., and Wrobel, J. (2022), "Pattern-based clustering of daily weigh-in trajectories using dynamic time warping," *Biometrics*. https://doi.org/10.1111/biom.13773.

Chen, H., Wang, Y., Paik, M. cho, and Choi, H. A. (2013), "A marginal approach to reduced-rank penalized spline smoothing with application to multilevel functional data," *J Am Stat Assoc.*, 108, 1216–1229. https://doi.org/10.1080/01621459.2013.826134.

Chiou, J.-M. (2012), "Dynamical functional prediction and classification, with application to traffic flow prediction," *The Annals of Applied Statistics*, Institute of Mathematical Statistics, 6, 1588–1614. https://doi.org/10.1214/12-AOAS595.

Delaigle, A., and Hall, P. (2016), "Approximating fragmented functional data by segments of markov chains," *Biometrika*, 103, 779–799. https://doi.org/10.1093/biomet/asw040.

Gertheiss, J., Goldsmith, J., and Staicu, A. (2016), "A note on modeling sparse exponential-family functional response curves," *Comput Stat Data Anal*, 105, 46–52. https://doi.org/10.1016/j.csda.2016.07.010.

Goldberg, Y., Ritov, Y., and Mandelbaum, A. (2014), "Predicting the continuation of a function with applications to call center data," *Journal of Statistical Planning and Inference*, 147, 53–65. https://doi.org/https://doi.org/10.1016/j.jspi.2013.11.006.

Goldsmith, J., Zipunnikov, V., and Schrack, J. (2015), "Generalized multilevel function-on-scalar regression and principal component analysis," *Biometrics*, 71, 344–53. https://doi.org/10.1111/biom.12278.

Hall, P., Müller, H.-G., and Yao, F. (2008), "Modelling sparse generalized longitudinal observations with latent gaussian processes," *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70, 703–723. https://doi.org/https://doi.org/10.1111/j.1467-9868.2008.00656.x.

Kraus, D. (2015), "Components and completion of partially observed functional data," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, [Royal Statistical Society, Wiley], 77, 777–801.

Laird, N. M., and Ware, J. H. (1982), "Random-effects models for longitudinal data," *Biometrics*, [Wiley, International Biometric Society], 38, 963–974.

Leroux, A., Crainiceanu, C. M., and Wrobel, J. (n.d.). "Fast generalized functional principal component analysis."

Leroux, A., Xiao, L., Crainiceanu, C., and Checkley, W. (2018), "Dynamic prediction in functional concurrent regression with an application to child growth," *Statistics in medicine*, 37, 1376–1388.

LIANG, K.-Y., and ZEGER, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13–22. https://doi.org/10.1093/biomet/73.1.13.

Linde, van der (2009), "A bayesian latent variable approach to functional principal components analysi with binary and count data," *A StA Adv Stat Anal*, 307–333. https://doi.org/10.1007/s10182-009-0113-6.

Lindstrom, M. J., and Bates, D. M. (1990), "Nonlinear mixed effects models for repeated measures data," *Biometrics*, [Wiley, International Biometric Society], 46, 673–687.

"Nonlinear models for repeated measurement data: An overview and update" (2003), [International Biometric Society, Springer], 8, 387–419.

Ostendorf, D. M., Caldwell, A. E., Zaman, A., Pan, Z., Bing, K., Wayland, L. T., Creasy, S. A., Bessesen, D. H., MacLean, P., Melanson, E. L., and Catenacci, V. A. (2022), "Comparison of weight loss induced by daily caloric restriction versus intermittent fasting (DRIFT) in individuals with obesity: Study protocol for a 52-week randomized clinical trial," *Trials*, 23, 718. https://doi.org/10.1186/s13063-022-06523-2.

Rizopoulos, D. (2022), *GLMMadaptive: Generalized linear mixed models using adaptive gaussian quadrature*.

Scheipl, F., Staicu, A.-M., and Greven, S. (2014), "Functional additive mixed models," *J Comput Graph Stat*, 24, 447–501. https://doi.org/10.1080/10618600.2014.901914.

Shang, H. L. (2017), "Functional time series forecasting with dynamic updating: An application to intraday particulate matter concentration," *Econometrics and Statistics*, 1, 184–200. https://doi.org/https://doi.org/10.1016/j.ecosta.2016.08.004.

Suresh, K., Taylor, J. M. G., Spratt, D. E., Daignault, S., and Tsodikov, A. (2017), "Comparison of joint modeling and landmarking for dynamic prediction under an illness-death model," *Biom J*, 59, 1277–1300. https://doi.org/10.1002/bimj.201600235.

Wrobel, J., Zipunnikov, V., Schrack, J., and Goldsmith, J. (2019), "Registration for exponential family functional data," *Biometrics*, 75, 48–57. https://doi.org/https://doi.org/10.1111/biom.

12963.