Bayesian Sparse Functional Principal Components Analysis Models Dynamic Temporal Changes in Longitudinal Microbiome Studies

Lingjing Jiang [1], Yoshiki Vázquez-Baeza [2], Antonio Gonzalez [3]
Loki Natarajan[1,4], Rob Knight *[3,5,6], Wesley K. Thompson*[1]

[1]Division of Biostatistics, UC San Diego, La Jolla, California, USA
[2]Jacobs School of Engineering, UC San Diego, La Jolla, California, USA
[3]Department of Pediatrics, UC San Diego, La Jolla, California, USA
[4]Moores UC San Diego Cancer Center, UC San Diego, La Jolla, California, USA
[5]Department of Computer Science, UC San Diego, La Jolla, California, USA
[6]Director of the Center for Microbiome Innovation, UC San Diego, La Jolla, California, USA

Abstract
Longitudinal microbiome studies provide valuable information about dynamic interactions between the microbiome and host, by capturing both between-individual differences and within-subject dynamics. As the costs of DNA sequencing have decreased, microbiome researchers have a greater opportunity to perform longitudinal studies to better understand microbial changes in response to an intervention in real time. However, microbial communities can change abruptly in response to small perturbations. Current approaches for longitudinal microbiome analysis are not sufficient to capture this dynamic temporal variation, especially with the additional challenges of irregular sampling intervals, limited sample size, missing values and dropouts. We developed a Bayesian Sparse Functional Principal Components Analysis (SFPCA) methodology to meet the growing need to model dynamic temporal change and to detect its dependence on biological covariates in longitudinal microbiome analysis. We show in simulations and in real data applications that Bayesian SFPCA is able to overcome the above challenges in longitudinal microbiome analysis, and is more sensitive for capturing temporal variations and detecting differences due to biological covariates than existing methods. We therefore expect Bayesian SFPCA to be a valuable tool to the microbial community for longitudinal analysis.

Key Words: Microbiome, longitudinal, Bayesian, Functional PCA, dimension reduction

1. Introduction

The microbiome is inherently dynamic, driven by interactions among microbes, with the host, and with the environment. These complex dynamics begin at birth, as the infant is colonized with microbes (1, 2), and continue in the healthy adult as microbial populations vary with hormonal and seasonal cycles (3–5) and a myriad of other host and environmental factors such as geography, diet and aging (6–8). At any point in life, human microbiome can be dramatically altered, either transiently or long term, by

diseases such as infections (1), medical interventions such as antibiotics (9) or even daily routines such as handwashing (10). Since the human microbiome is highly dynamic and personalized, longitudinal microbiome studies that sample human-associated microbial communities repeatedly over time provide valuable information for researchers to observe both inter- and intra-individual variability, or to measure changes in response to an intervention in real time (5, 11–16).

As the costs of DNA sequencing have decreased, microbiome researchers have a greater opportunity to perform longitudinal studies. While longitudinal data with multiple timepoints provide more information than cross-sectional (single-timepoint) data, the computational tools to analyze longitudinal microbiome studies still lag behind, despite recent developments such as q2-longitudinal package (17). To perform proper longitudinal microbiome analysis, researchers need to overcome the following challenges: temporal changes are highly dynamic and may follow arbitrary curves, time points are usually sampled irregularly to better capture biological variation at certain time intervals, sample size is limited and can be highly unbalanced between compared groups, and missing values and dropouts are common due to sequencing error or sample collection. In some studies, researchers have collapsed samples across time points to average individual's signals or they have summarized first with multivariate approaches that condense the initial observations (e.g., (18, 19)). Although several studies were able to make important discoveries using these approaches, statistical power can be increased by using additional information content and directionality of the temporal axis. Frequently, microbiome researchers apply classic longitudinal methods, such as the Linear Mixed Effects Model, to analyze their data (17, 20), neglecting the highly non-linear temporal variation. To address this gap in data analysis, a new longitudinal method called splinectomeR has been recently introduced (21), which used a loess spline to model the data that may not follow any classical model or shape, and then perform hypothesis testing of categorical variables in longitudinal studies with permutation test. However, we find that splinectomeR tends to over-linearize the dynamic temporal variation, has limited capacity to detect group differences over time, and its results can be impacted by outliers, particularly in sparse datasets.

To meet the growing need of longitudinal analysis in microbiome studies and overcome the challenges mentioned above, we developed Bayesian Sparse Functional Principal Components Analysis (SFPCA), a particularly useful method to model the highly flexible temporal change, and capture the dominant patterns of variation, even when data are sparse with measurements for individuals occurring at possibly differing time points. Bayesian SFPCA is a nonparametric method that requires very limited assumptions to be made about the data and is therefore widely applicable to various longitudinal microbiome studies. We model individual trajectories over time under the framework of functional data analysis (FDA), which is to express discrete observations arising from time series in the form of a function that represents the entire measured curve as a single observation (22). Sparse Functional Principal Components Analysis (SFPCA) is to use dimension reduction via principal components to model curves that are observed longitudinally over irregular and sparse set of time points (23). We implemented the SFPCA model in a Bayesian framework by specifying priors on the parameters and using Hamiltonian Markov Chain Monte Carlo (MCMC) sampling algorithm. This Bayesian model allows researchers to perform model selection using leave-one-out cross-validation with Pareto-smoothed important sampling (PSIS-LOO) (24) and assess model fit with posterior predictive checking (25). We demonstrate in both simulations and experimental microbiome data that Bayesian SFPCA is better at modeling dynamic temporal variation

and thus more sensitive in detecting dependence of longitudinal observations on biological covariates of interest than splinectomeR.

2. Bayesian SFPCA model for longitudinal microbiome data

The basic idea behind Functional Data Analysis (FDA) is to express discrete observations arising from time series in the form of a smooth function that represents the entire time course as a single observation, to represent this function as a linear combination of a low-dimensional but flexible set of basis functions, and then to applying statistical concepts from multivariate data analysis to this representation (22). This approach is highly flexible in the sense that temporal patterns can be described by continuous smooth dynamics, and the timing intervals for data observations can vary across individuals. Moreover, the FDA approach requires limited assumptions to be made about the data: it makes no parametric assumptions about mean time effects or subject-level deviations from the mean trend. Therefore, the FDA approach is widely applicable to a variety of scenarios. Goals for FDA on longitudinal data may include describing the major modes of variation in the data, exploring the individual variation of curves from overall mean trajectories, and characterizing the dependence of curves on covariates of interest (26).

The classical assumption of FDA is that each curve or function has been measured at all time points or, more realistically, over a densely sampled grid. However, in practice, especially in microbiome studies, curves are often measured at an irregular and sparse set of time points which can differ widely across individuals. James et al (23) proposed Sparse Functional Principal Components Analysis (SFPCA), a principal component model for handling this more difficult scenario of sparse functional data using a reduced rank mixed effects framework. Let $Y_i(t)$ be the measurement at time t for the ith individual or curve. Let $\mu(t)$ be the overall mean function, let f_j be the jth principal component function and let $f = (f_1, f_2, \dots, f_k)^T$, where k is the number of principal components. Then,

$$Y_i(t) = \mu(t) + \sum_{j=1}^{k} f_j(t)\,\alpha_{ij} + \varepsilon_i(t), \qquad i = 1, \dots, N$$
$$= \mu(t) + f(t)^T \alpha_i + \varepsilon_i(t)\,, \; i = 1, \dots, N$$

subject to the orthogonality constraint $\int f_j\, f_l = \delta_{jl}$, the Kronecker δ. The random vector α_i gives the relative weights on the principal component functions for the ith individual and $\varepsilon_i(t)$ is random measurement error. Since the data are measured at only a finite number of time points, μ and f are represented using a basis of spline functions in order to place some restrictions on the form of the mean and the principal component curves. Let $b(t)$ be a spline basis with dimension q. Let Θ and θ_μ be, respectively, a q by k matrix and a q- dimensional vector of spline coefficients. The resulting restricted model has the form

$$Y_i(t) = b(t)^T \theta_\mu + b(t)^T \Theta \alpha_i + \varepsilon_i(t), \qquad i = 1, \dots, N$$

For each individual i, let $t_{i1}, t_{i2}, \dots, t_{in_i}$ be the possibly different time points at which measurements are available. Then

$$Y_i = (Y_i(t_{i1}), \dots, Y_i(t_{in_i}))^T,$$
$$B_i = (b(t_{i1}), \dots, b(t_{in_i}))^T \text{ subject to the orthogonality constraint } \mathrm{B}^T\mathrm{B} = \mathrm{I}$$

The final model can be simplified as

$$Y_i(t) = B_i\theta_\mu + B_i\Theta\alpha_i + \varepsilon_i, \qquad i = 1, \ldots, N,$$
$$\Theta^T\Theta = \mathrm{I}, \qquad \varepsilon_i \sim N(0, \sigma^2 I), \qquad \alpha_i \sim N(0, D)$$

We implemented this SFPCA model in a Bayesian framework to allow direct model selection and model assessment. A difficulty in implementing the Bayesian SFPCA model is that there is no unique set of FPC loadings Θ. For any k by k orthogonal rotation matrix P, if $\Theta^* = \Theta P^T$, then ${\Theta^*}^T\Theta^* = \mathrm{P}\,\Theta^T\,\Theta P^T = I$, and hence the loadings are unidentifiable. A technique to identify loadings is to sample from the communality and derive loadings from it (27). The communality defined as $\mathrm{W} = \Theta^T\,\Theta$ is identified, but the loadings Θ are not. Because the communality is not full rank, standard conjugate or other widely used priors from random positive definite symmetric matrices cannot be used. We therefore place exchangeable prior distributions on the elements of the loading matrix Θ, ensuring the induced prior for W invariant to permutations of the data dimensions (27). Then we apply *post-hoc* rotation to recover loadings matrix Θ and scores α_i, with Θ be the matrix of the first k eigenvectors of the communality W. We implemented this Bayesian model using Hamilton Markov Chain Monte Carlo (MCMC) sampling algorithm in Stan (28). This enables researchers to use leave-one-out cross-validation with Pareto-smoothed important sampling (PSIS-LOO) (24) to perform model selection on the number of principal components k, the number of basis functions q, and the placement of knots in the spline bases. Moreover, we can simulate posterior predictive replicated data to assess model fit with posterior predictive checking (25). Given that SFPCA model can capture the highly non-linear temporal variation in sparse longitudinal microbiome data, and our Bayesian implementation provides additional model selection and assessment procedures, this Bayesian SFPCA method is particularly useful for longitudinal microbiome analysis. The R code of this method is available on GitHub (https://github.com/knightlab-analyses/bayesianSFPCA).

3. Simulations

The ideal longitudinal methods for microbiome studies should be able to capture the highly non-linear temporal patterns in the data, and be sensitive to detecting the relationship between temporal variations and biological covariates of interest, under the challenges of irregularly- and sparsely-sampled time points, missing values and dropouts, and potentially limited sample size. To demonstrate how usefully Bayesian SFPCA can model the dynamic microbial temporal patterns with the practical complications listed above, we first adopted the simulated data set from splinectomeR to assess its performance in modeling time series data with different perturbations, and then simulated data with various numbers of samples and proportions of missing data to evaluate its robustness in obtaining proper inference regardless of small sample size and missing data. In both scenarios, Bayesian SFPCA was better at capturing the internal temporal dynamics, and more sensitive in discovering its association with covariates of interest than splinectomeR.

3.1 Simulated data with perturbations

In the simulations of splinectomeR, the response variable was perturbed at one or two regions of the time series, and at three magnitudes of change (Figures 1a, c). When comparing the baseline with the perturbed data, SplinectomeR failed to detect the

difference between shift_1x and baseline in the single region perturbed data (p-value = 0.3; Figure 1b), and the difference between shift_1x or shift_2x with baseline (p = 0.79, p = 0.36 for 1x, 2x shifts respectively; Figure 1d) in the double region perturbed data. The estimated mean curves for each version were trimmed from the beginning and the end because the spline that splinectomeR fits at the two ends are susceptible to outliers (Figure 1b, d). Moreover, the mean curves for baseline in the singled region perturbed data, and for baseline, 1x and 2x shifts in the double region perturbed data were over-linearized, thus losing the underlying dynamic patterns (Figure 1b, d).

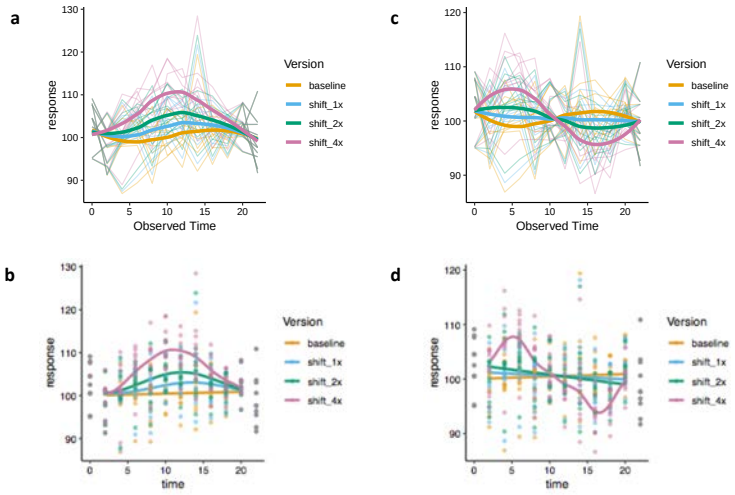

Figure 1: Results of splinectomeR on simulated data with perturbations **(a)** A spaghetti plot showing the simulated trajectories of single region perturbed data with loess mean curves for each version. **(b)** Estimated mean curves of splinectomeR on simulated single region perturbed data points. **(c)** A spaghetti plot showing the simulated trajectories of double region perturbed data with loess mean curves for each version. **(d)** Estimated mean curves of splinectomeR on simulated double region perturbed data points.

The Bayesian SFPCA approach utilizes PC (principal component) functions to describe the major modes of temporal variation in the data, and PC scores to estimate weights of each individual on major modes of variation, and then to examine associations between temporal patterns and biological covariates of interest. Therefore, Bayesian SFPCA is able to capture temporal patterns beyond the typical mean trajectories that researchers often look at and has more power in detecting temporal associations. For example, there was more variation than the observed mean curve ("bell" shape) in single region perturbed data, and the "S" shape in double region perturbed data, as indicated by the observed trajectories of different shapes shown in Figure 1a, c. Specifically, patient 3 and 5 exhibited more variation between time point 5 and 15 than the "bell" shaped mean curve (Figure 2a, b); patient 4 had dramatic decrease and increase at the two ends (Figure 3a), and patient 8 had more ups and downs between time point 5 and 10, which were both not shown in the "S" shaped mean curve (Figure 3b). The PC curves estimated by Bayesian SFPCA were able to capture all these various patterns and the statistical inference based on the PC scores were sensitive enough to detect associations between temporal patterns and relevant covariates. For the simulated data with single region

perturbations, the PC 1 curve represented a similar pattern as the mean bell curve but with different magnitudes and slight variation at the end, which explained 68.64% of the variation in the data (Figure 2c). The PC 2 curve explained 31.36% of the variation and emphasized variations between time point 5 and 15 (Figure 2d), capturing the patterns as seen in patient 3 and 5 (Figure 2a, b). The boxplot of PC scores demonstrated a statistically significant increasing magnitude of perturbation in PC1 with p-values of 3.6e-06 for 1x shift and of less than 2.2e-16 for 2x and 4x shifts (versus p = 0.3, 0.005, 0.001 for 1x, 2x, and 4x shifts from splinectomeR) (Figure 2e). The PC 2 scores showed a decreasing trend, which was statistically different between baseline and 2x or 4x shifts (Figure 2f). For the double region perturbed data, the mean curves for all versions were almost flat with little curvature, due to the different magnitudes of perturbations in each version. The PC 1 curve characterized the observed "S" shape of the mean curve and explained 72.81% of the variation (Figure 3c), and the PC 2 curve explained 27.19% of the variation and captured the patterns deviating from the "S" shape (Figure 3d), as observed in patient 4 and 8 (Figure 3a, b). The PC scores perfectly captured a statistically significant increasing magnitude of perturbation in PC1 with p-values of 0.0013 for 1x, 3e-10 for 2x and less than 2.22e-16 for 4x shifts (versus p = 0.79, 0.36, 0.002 for 1x, 2x, and 4x shifts from splinectomeR) (Figure 3e). The PC2 scores indicated that the PC 2 pattern was shared among different versions (Figure 3f). These results demonstrate that Bayesian SFPCA can have much greater power in unveiling temporal patterns beyond the typical mean shape, and in detecting the association between temporal variation and covariates of interest, compared to other methods like splinectomeR.

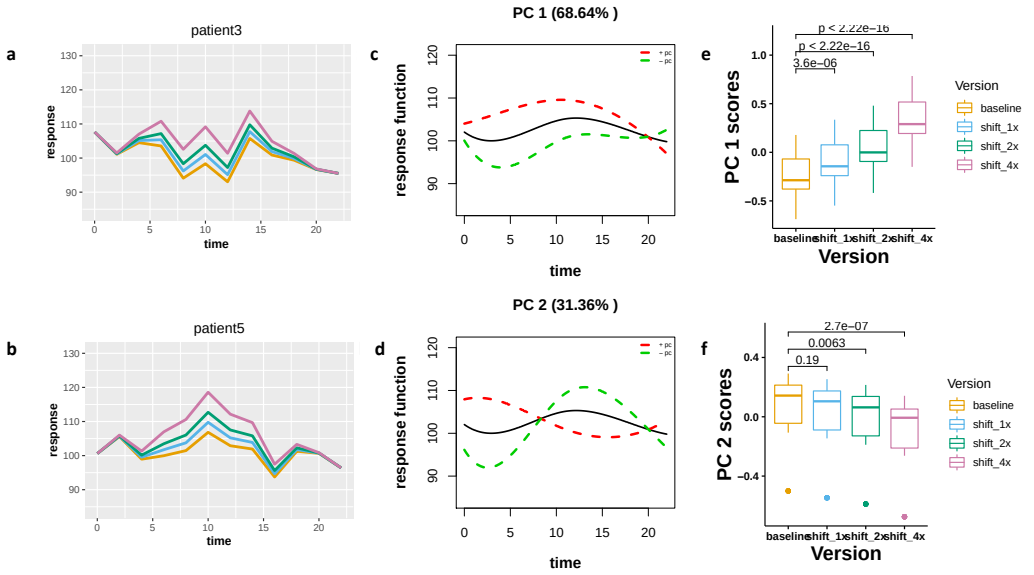

Figure 2: Results of Bayesian SFPCA on simulated single region perturbed data **(a)** Observed trajectories for patient 3 with different magnitude of perturbations. **(b)** Observed trajectories for patient 5 with different magnitude of perturbations. **(c)** Mean response curve and the effects of adding (red) and subtracting (green) PC 1 function. **(d)** Mean response curve and the effects of adding (red) and subtracting (green) PC 2 function. **(e)** Boxplot of estimated PC 1 scores with p-values from two-sample t tests. **(f)** Boxplot of estimated PC 2 scores with p-values from two-sample t tests.

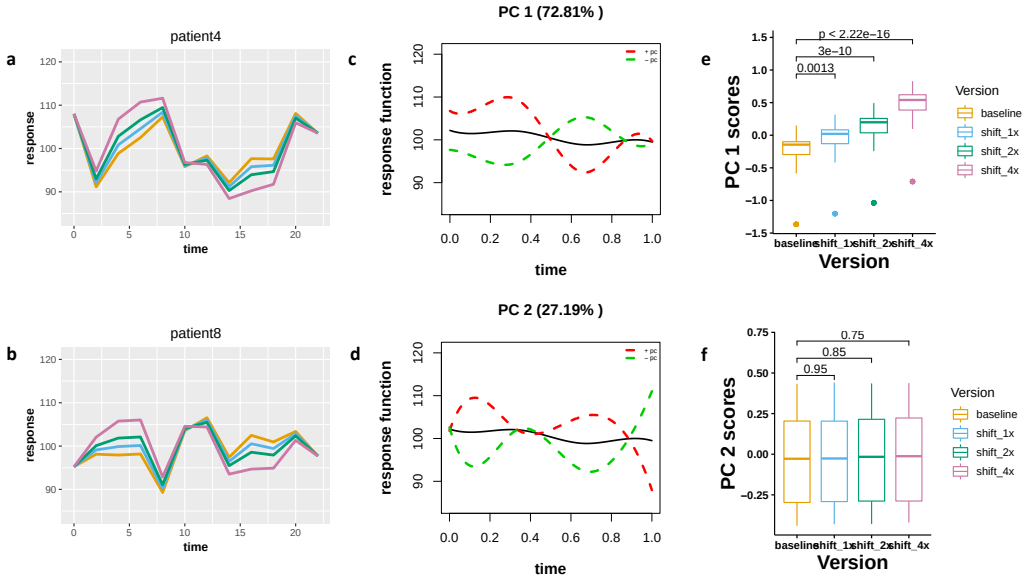

Figure 3: Results of Bayesian SFPCA on simulated double region perturbed data **(a)** Observed trajectories for patient 3 with different magnitude of perturbations. **(b)** Observed trajectories for patient 5 with different magnitude of perturbations. **(c)** Mean response curve and the effects of adding (red) and subtracting (green) PC 1 function. **(d)** Mean response curve and the effects of adding (red) and subtracting (green) PC 2 function. **(e)** Boxplot of estimated PC1 scores with p-values from two-sample t tests. **(f)** Boxplot of estimated PC 2 scores with p-values from two-sample t tests.

3.2 Simulated data with missing values and varied sample size

Due to potential sequencing error and sample collection procedures, missing data and dropouts are the norm rather than the exception in longitudinal microbiome studies. Moreover, despite large-scale cross-sectional microbiome studies such as the Human Microbiome Project (29, 30) and American Gut Project (5), studies characterizing human associated microbial communities over time often have relatively small sample size (9, 11, 31). Hence, it is important to assess the performance of Bayesian SFPCA in simulations with various sample size and different proportions of missing data.

Using the PC functions and population mean curve from a real microbiome dataset (32), we simulated longitudinal trajectories for two groups of subjects by arbitrarily assigning group one and two to have positive and negative PC scores on PC 1 respectively and no difference on PC 2 (supplemental Figure S1). We varied the total number of subjects at 100, 50, 25, 10, and the proportion of missing values at 0%, 20%, 50% and 80%. The estimated mean curves for each group from splinectomeR are highly susceptible to the decreasing sample size and increasing loss of information (Supplemental Figure S2).

To fully evaluate the performance of Bayesian SFPCA, we wanted to examine how well it could recover the underlying temporal patterns (including both the population mean trajectory and two PC functions) and distinguish the two groups based on PC 1 yet not on PC 2 scores. With 100 total samples, even though the proportion of missing data increases from 0% to 80%, the estimated overall mean curves and PC curves almost

perfectly recovered the ground truth in both scenarios (Figure 4a, b, d, e), and the two groups were accurately separated based on PC 1 scores yet not on PC 2 (Figure 4c, f). Concerning the scenarios with 50 or 25 total samples with 80% missing data, the estimated mean curves were close to the true ones, except for the slight deviation toward the end due to the large portion of missing data there (Figure 5a, d). The PC curves were estimated well for the case of 50 samples on both PCs (Figure 5b), but good on only PC1 for the case of 25 samples (Figure 5d). An artificial curvature occurred on the estimated PC 2 curve for the case of 25 samples (Figure 5e), because few data existed at the beginning (Figure 5d). Despite these small defects, the ability to distinguish between the two groups based on PC scores was still maintained (Figure 5c, f). As for the challenging scenarios of 5 samples per group with 50 or 80% missing data, the estimated mean curves in both scenarios and the PC curves for the scenario with 50% missingness were still robust (Figure 6 a, d, b). However, for the case with 80% missing data, the estimated PC 1 curve was a vertical shift from the true PC 2 (not PC 1) curve, and the estimated PC 2 curve exhibited a different shape from the ground truth (Figure 6e). A close look at the simulated trajectories (Figure 6d) indicated that few trajectories exhibited the increasing trend in PC 1 due to the loss of data, hence the deviated estimation were caused by the limitation of the underlying data; but these did not affect the separation of groups based on estimated PC scores in both scenarios (Figure 6c, f). In summary, the performance of Bayesian SFPCA is extremely robust to limited sample size and a high proportion of missing data.

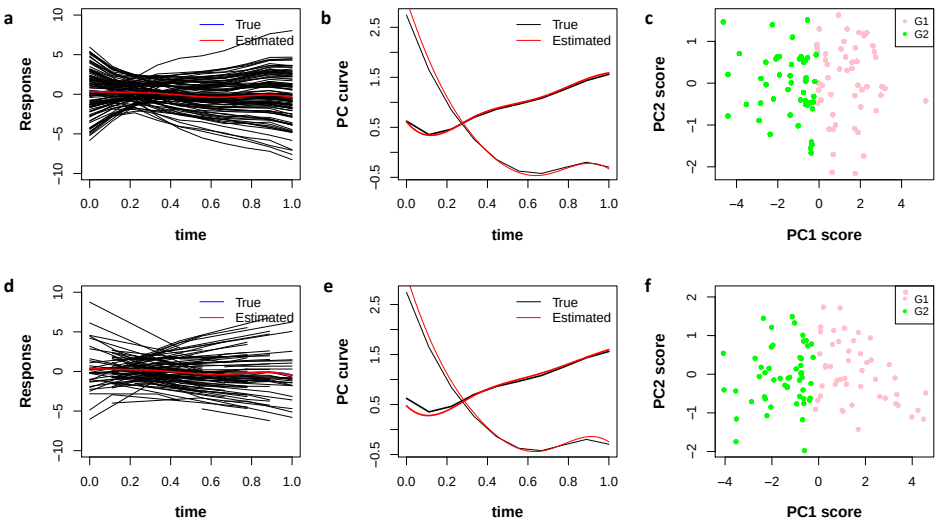

Figure 4: Results of Bayesian SFPCA on simulated data with 100 total samples of 0% vs. 80% missing data **(a)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 0% missing data. **(b)** Estimated (red) vs. true (black) PC curves on simulated data with 0% missing data. **(c)** Estimated PC scores for two groups on simulated data with 0% missing data. **(d)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 80% missing data. **(e)** Estimated (red) vs. true (black) PC curves on simulated data with 80% missing data. **(f)** Estimated PC scores for two groups on simulated data with 80% missing data.

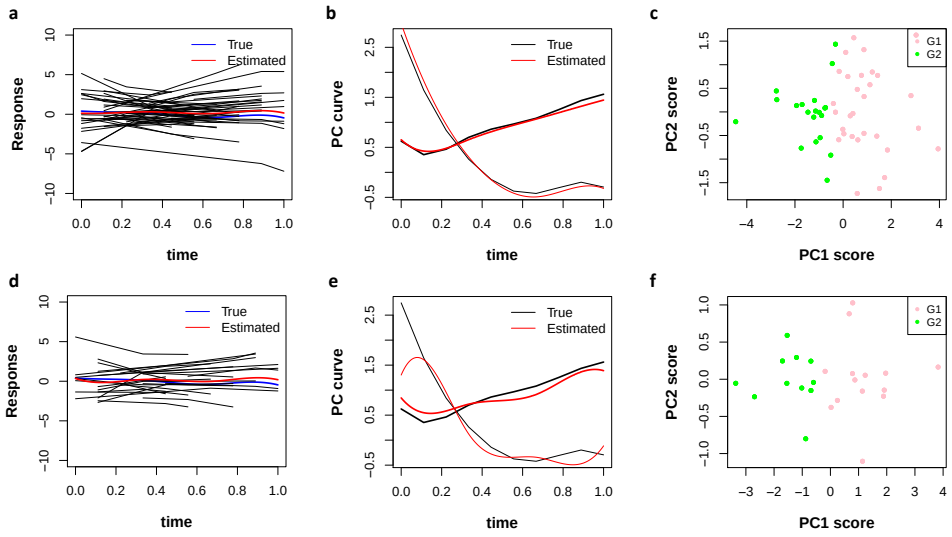

Figure 5: Results of Bayesian SFPCA on simulated data with 50 vs. 25 total samples of 80% missing data **(a)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 50 samples. **(b)** Estimated (red) vs. true (black) PC curves on simulated data with 50 samples. **(c)** Estimated PC scores for two groups on simulated data with 50 samples. **(d)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 25 samples. **(e)** Estimated (red) vs. true (black) PC curves on simulated data with 25 samples. **(f)** Estimated PC scores for two groups on simulated data with 25 samples.

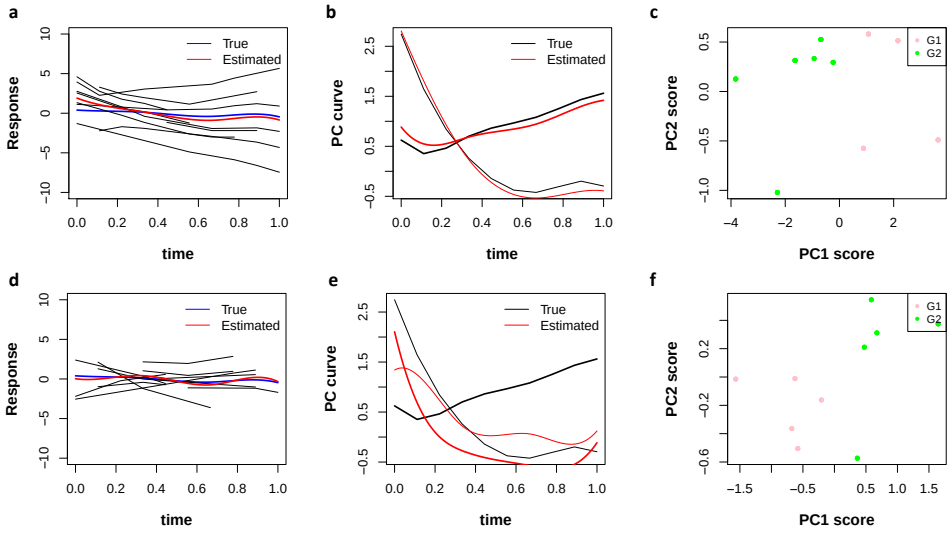

Figure 6: Results of Bayesian SFPCA on simulated data with 10 total samples of 50% vs. 80% missing data **(a)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 50% missing data. **(b)** Estimated (red) vs. true (black) PC curves on simulated data with 50% missing data. **(c)** Estimated PC scores for two groups on

simulated data with 50% missing data. **(d)** Estimated (red) vs. true (blue) overall mean curve on simulated trajectories (black) with 80% missing data. **(e)** Estimated (red) vs. true (black) PC curves on simulated data with 80% missing data. **(f)** Estimated PC scores for two groups on simulated data with 80% missing data.

4. Real data applications

To demonstrate the value of Bayesian SFPCA in analyzing real longitudinal microbiome studies, we present a re-analysis of data from the early childhood and the microbiome (ECAM) study (20). This study tracked the microbiota compositions, as measured by 16S rRNA gene amplicon profiling, of 43 infants in the United States sampled at irregular intervals from birth to two years of age and associations between antibiotic exposure, delivery mode, and predominant diet on microbiota composition and development. Here, we focus on modeling the maturation of the gut microbiome based on Shannon alpha diversity, as well as its associations with antibiotics, delivery modes and diets.

4.1 Maturation of the gut microbiome

Shannon diversity exhibits a steadily increasing trend during infants' first two years of life, despite drastic interpersonal differences at the first 3 months (Figure 7a). The estimated mean curve from Bayesian SFPCA nicely captured a temporal pattern of microbial maturation, as well as three different stages of increase: slowest growth rate from birth to 5 month, most rapid increase between 5 and 10 month (probably due to the introduction of solid foods), and intermediate growth after 10 month (Figure 8a). Moreover, the PC curves featured three major patterns of temporal variation in the data. PC 1 curve explained 62.37% of the variation and captured variation from the mean Shannon curve before 5 month and after 15 month (Figure 8b). An infant with a positive score on this component experienced an initial drop of bacterial diversity in the first 5 months, increasing diversity afterwards and decelerating after 15 months. In contrast, an infant with a negative score had increasing diversity throughout the entire 2 years. PC 2 curve explained 28.5% of the variation and captured variation from birth until 15 months and (Figure 8c). PC 3 captured a strong up and down pattern during the first 10 months and explained 9.13% of the variation (Figure 8d).

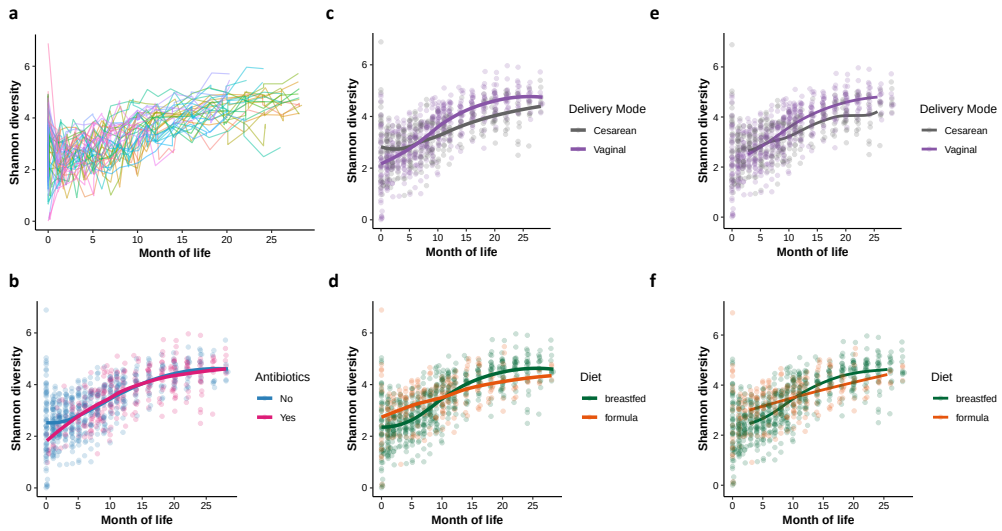

Figure 7: Observed and splinectomeR estimated group mean curves on ECAM data **(a)** Observed Shannon diversity trajectories **(b)** Observed Shannon diversity with loess mean curves on antibiotic exposure. **(c)** Observed Shannon diversity with loess mean curves on delivery mode. **(d)** Observed Shannon diversity with loess mean curves on diet. **(e)** Observed Shannon diversity with splinectomeR estimated mean curves on delivery mode. **(f)** Observed Shannon diversity with splinectomeR estimated mean curves on diet.

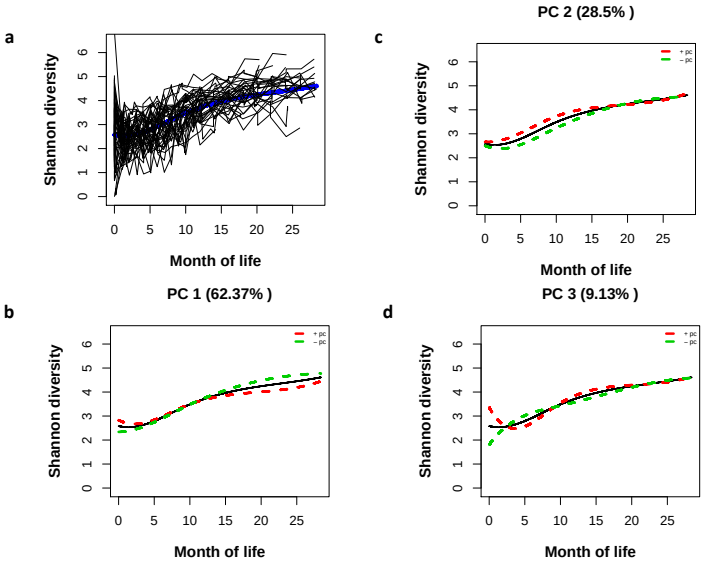

Figure 8: Estimated overall mean curve and PC curves from Bayesian SFPCA on ECAM data **(a)** Observed Shannon diversity (black trajectories) with estimated overall mean curve (blue curve). **(b)** Mean Shannon diversity curve (black) and the effects of adding (red) and subtracting (green) PC 1 function. **(d)** Mean Shannon diversity curve (black) and the effects of adding (red) and subtracting (green) PC 2 function. **(e)** Mean Shannon

diversity curve (black) and the effects of adding (red) and subtracting (green) PC 3 function.

4.2 Antibiotic exposure

The observed infants' trajectories of Shannon diversity showed that antibiotic use diminished alpha diversity immediately following birth but accelerated their rates of increase during the first 5 months of life (Figure 7b). However, infants with antibiotic exposure exhibited similar growth trends as those without after 5 months (Figure 7b). Since antibiotic exposure was a time-varying covariate, splinectomeR could not perform inference in this scenario. Bayesian SFPCA found a trending difference (two sample t-test p = 0.069) in Shannon diversity between infants with and without antibiotic exposure on PC 3 scores (Figure 9a), but not with the other two PCs (p = 0.29 on PC 1 and p = 0.12 on PC 2). This suggested that the two groups may differ slightly in the PC 3 pattern, where the greatest variation occurred before 5 months (Figure 8d), agreeing with the difference shown in the loess mean curves of observed data (Figure 7b).

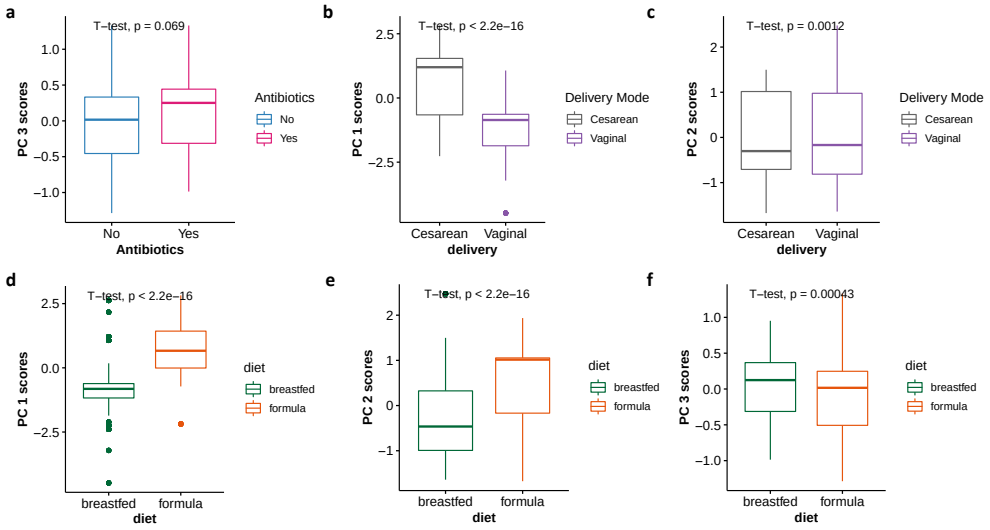

Figure 9: Significant associates inferred from PC scores of Bayesian SFPCA on ECAM data **(a)** Boxplot of estimated PC 3 scores on antibiotic exposure. **(b)** Boxplot of estimated PC 1 scores on delivery mode. **(c)** Boxplot of estimated PC 2 scores on delivery mode. **(d)** Boxplot of estimated PC 1 scores on diet. **(e)** Boxplot of estimated PC 2 scores on diet. **(f)** Boxplot of estimated PC 3 scores on diet.

4.3 Delivery Mode

Cesarean-delivered infants bypass the microbial exposure in the birth canal; this significantly altered the bacterial diversity compared to vaginally-born infants (2, 20), and we observed a slower growth rate of Shannon diversity throughout the first two years of life (Figure 7c). Although splinectomeR detected this significant difference (p = 0.02), estimated mean curves from splinectomeR only captured the difference after 8 months and missed the differential trajectories in the earlier period. Bayesian SFPCA found that

cesarean-delivered infants were significantly different from vaginally born infants in Shannon diversity on both PC 1 and PC 2 modes of variation ($p < 2.2e-16$ on PC 1, $p = 0.0012$ on PC 2; Figure 9b, c). The separation between two groups was clearest on PC 1 with cesarean-delivered infants having generally positive component scores while vaginally born infants negative scores (Figure 9b). This indicated that cesarean-delivered infants were most different from vaginal born infants before month 5 (decrease in cesarean vs. increase in vaginal) and after month 15 (higher growth rate in vaginal) (Figure 8b), which concurred with the observed difference (Figure 7c).

4.4 Infant Diet

Two major dietary groups were compared in this cohort: infants who were predominantly (> 50% of feedings) breast-fed or predominantly formula-fed for the first three months of life. The growth rate of Shannon diversity was significantly decreased in formula-fed children during 12-24 months of life compared to breast-fed ones ((20); Figure 7d). However, splinectomeR found no significant difference between the two groups ($p = 0.11$; Figure 7f). In contrast, Bayesian SFPCA detected significant difference in all three PCs with p-values all less than 0.001 (Figure 9d-f), suggesting that the difference between formula- and breastfed- infants can belong to any of the temporal pattern shown in three PCs (Figure 8b-d) but with PC 3 the least distinguished.

In summary, Bayesian SFPCA was able to find significant effect of antibiotics exposure, delivery mode and infant diets on Shannon diversity in the ECAM study as in the original paper (20). Moreover, it provided more insight into the temporal dynamics that drove these differences. Similar analyses could be done on beta diversities (e.g. UniFrac distance from following time points to birth or Jaccard distance between infants and mothers) (20, 33) or specific bacteria taxa using additive or centered log-ratio transformations of relative abundance (34).

5. Conclusion

In both simulations and real datasets, we have shown that Bayesian SFPCA is a powerful tool for capturing dynamic temporal patterns, as well as detecting their associations with biological covariates of interest in longitudinal microbiome studies. Furthermore, Bayesian SFPCA performance is robust to irregularly-sampled intervals, limited sample size, and missing data or dropouts. Future work to be done includes extending Bayesian SFPCA to model multiple longitudinal measurements simultaneously so that it can be applied to multi-omics longitudinal studies.

Acknowledgments

We thank Robin R. Shields-Cutler for helping with the application of splinectomeR. We thank Se Jin Song, Tomasz Kosciolek, Anupriya Tripathi and Shi Huang for discussions on the microbiome applications of Bayesian SFPCA. This work was supported by the Emerald Foundation (grant 265), and Janssen Pharmaceuticals, Inc (grant 20175015). YVB was funded by the Janssen Human Microbiome Institute through UC San Diego's Center for Microbiome Innovation.

References

1. Koenig JE, Spor A, Scalfone N, Fricker AD, Stombaugh J, Knight R, Angenent LT, Ley RE. 2011. Succession of microbial consortia in the developing infant gut microbiome. Proc Natl Acad Sci 108:4578–4585.
2. Dominguez-Bello MG, Costello EK, Contreras M, Magris M, Hidalgo G, Fierer N, Knight R. 2010. Delivery mode shapes the acquisition and structure of the initial microbiota across multiple body habitats in newborns. PNAS 107:11971–11975.
3. Gajer P, Brotman RM, Bai G, Sakamoto J, Schutte UME, Zhong X, Koenig SSK, Fu L, Ma Z, Zhou X, Abdo Z, Forney LJ, Ravel J. 2012. Temporal Dynamics of the Human Vaginal Microbiota. Sci Transl Med 4:132ra52-132ra52.
4. Smits SA, Leach J, Sonnenburg ED, Gonzalez CG, Lichtman JS, Reid G, Knight R, Manjurano A, Changalucha J, Elias JE, Dominguez-Bello MG, Sonnenburg JL. 2017. Seasonal cycling in the gut microbiome of the Hadza hunter-gatherers of Tanzania. Science (80-) 357:802–806.
5. McDonald D, Hyde E, Debelius JW, Morton JT, Gonzalez A, Ackermann G, Aksenov AA, Behsaz B, Brennan C, Chen Y, DeRight Goldasich L, Dorrestein PC, Dunn RR, Fahimipour AK, Gaffney J, Gilbert JA, Gogul G, Green JL, Hugenholtz P, Humphrey G, Huttenhower C, Jackson MA, Janssen S, Jeste D V., Jiang L, Kelley ST, Knights D, Kosciolek T, Ladau J, Leach J, Marotz C, Meleshko D, Melnik A V., Metcalf JL, Mohimani H, Montassier E, Navas-Molina J, Nguyen TT, Peddada S, Pevzner P, Pollard KS, Rahnavard G, Robbins-Pianka A, Sangwan N, Shorenstein J, Smarr L, Song SJ, Spector T, Swafford AD, Thackray VG, Thompson LR, Tripathi A, Vázquez-Baeza Y, Vrbanac A, Wischmeyer P, Wolfe E, Zhu Q, Knight R, Mann AE, Amir A, Frazier A, Martino C, Lebrilla C, Lozupone C, Lewis CM, Raison C, Zhang C, Lauber CL, Warinner C, Lowry CA, Callewaert C, Bloss C, Willner D, Galzerani DD, Gonzalez DJ, Mills DA, Chopra D, Gevers D, Berg-Lyons D, Sears DD, Wendel D, Lovelace E, Pierce E, TerAvest E, Bolyen E, Bushman FD, Wu GD, Church GM, Saxe G, Holscher HD, Ugrina I, German JB, Caporaso JG, Wozniak JM, Kerr J, Ravel J, Lewis JD, Suchodolski JS, Jansson JK, Hampton-Marcell JT, Bobe J, Raes J, Chase JH, Eisen JA, Monk J, Clemente JC, Petrosino J, Goodrich J, Gauglitz J, Jacobs J, Zengler K, Swanson KS, Lewis K, Mayer K, Bittinger K, Dillon L, Zaramela LS, Schriml LM, Dominguez-Bello MG, Jankowska MM, Blaser M, Pirrung M, Minson M, Kurisu M, Ajami N, Gottel NR, Chia N, Fierer N, White O, Cani PD, Gajer P, Strandwitz P, Kashyap P, Dutton R, Park RS, Xavier RJ, Mills RH, Krajmalnik-Brown R, Ley R, Owens SM, Klemmer S, Matamoros S, Mirarab S, Moorman S, Holmes S, Schwartz T, Eshoo-Anton TW, Vigers T, Pandey V, Treuren W Van, Fang X, Zech Xu Z, Jarmusch A, Geier J, Reeve N, Silva R, Kopylova E, Nguyen D, Sanders K, Salido Benitez RA, Heale AC, Abramson M, Waldispühl J, Butyaev A, Drogaris C, Nazarova E, Ball M, Gunderson B. 2018. American Gut: an Open Platform for Citizen Science Microbiome Research. mSystems 3.
6. Yatsunenko T, Rey FE, Manary MJ, Trehan I, Dominguez-Bello MG, Contreras M, Magris M, Hidalgo G, Baldassano RN, Anokhin AP, Heath AC, Warner B, Reeder J, Kuczynski J, Caporaso JG, Lozupone CA, Lauber C, Clemente JC, Knights D, Knight R, Gordon JI. 2012. Human gut microbiome viewed across age and geography. Nature 486:222–227.
7. Turnbaugh PJ, Ridaura VK, Faith JJ, Rey FE, Knight R, Gordon JI. 2009. The effect of diet on the human gut microbiome: a metagenomic analysis in

humanized gnotobiotic mice. Sci Transl Med 1:6ra14.
8. Claesson MJ, Cusack S, O'Sullivan O, Greene-Diniz R, de Weerd H, Flannery E, Marchesi JR, Falush D, Dinan T, Fitzgerald G, Stanton C, van Sinderen D, O'Connor M, Harnedy N, O'Connor K, Henry C, O'Mahony D, Fitzgerald AP, Shanahan F, Twomey C, Hill C, Ross RP, O'Toole PW. 2011. Composition, variability, and temporal stability of the intestinal microbiota of the elderly. Proc Natl Acad Sci 108:4586–4591.
9. Dethlefsen L, Relman DA. 2011. Incomplete recovery and individualized responses of the human distal gut microbiota to repeated antibiotic perturbation. Proc Natl Acad Sci U S A 108 Suppl 1:4554–61.
10. Fierer N, Hamady M, Lauber CL, Knight R. 2008. The influence of sex, handedness, and washing on the diversity of hand surface bacteria. Proc Natl Acad Sci U S A 105:17994–9.
11. Flores GE, Caporaso JG, Henley JB, Rideout JR, Domogala D, Chase J, Leff JW, Vázquez-Baeza Y, Gonzalez A, Knight R, Dunn RR, Fierer N. 2014. Temporal variability is a personalized feature of the human microbiome. Genome Biol 15:531.
12. DiGiulio DB, Callahan BJ, McMurdie PJ, Costello EK, Lyell DJ, Robaczewska A, Sun CL, Goltsman DSA, Wong RJ, Shaw G, Stevenson DK, Holmes SP, Relman DA. 2015. Temporal and spatial variation of the human microbiota during pregnancy. Proc Natl Acad Sci U S A 112:11060–5.
13. Kostic AD, Gevers D, Siljander H, Vatanen T, Hyötyläinen T, Hämäläinen A-M, Peet A, Tillmann V, Pöhö P, Mattila I, Lähdesmäki H, Franzosa EA, Vaarala O, de Goffau M, Harmsen H, Ilonen J, Virtanen SM, Clish CB, Orešič M, Huttenhower C, Knip M, Xavier RJ, Xavier RJ. 2015. The Dynamics of the Human Infant Gut Microbiome in Development and in Progression toward Type 1 Diabetes. Cell Host Microbe 17:260–273.
14. Halfvarson J, Brislawn CJ, Lamendella R, Vázquez-Baeza Y, Walters WA, Bramer LM, D'Amato M, Bonfiglio F, McDonald D, Gonzalez A, McClure EE, Dunklebarger MF, Knight R, Jansson JK. 2017. Dynamics of the human gut microbiome in inflammatory bowel disease. Nat Microbiol 2:17004.
15. Smarr L, Hyde ER, McDonald D, Sandborn WJ, Knight R. 2017. Tracking Human Gut Microbiome Changes Resulting from a Colonoscopy. Methods Inf Med 56:442–447.
16. Weingarden A, González A, Vázquez-Baeza Y, Weiss S, Humphry G, Berg-Lyons D, Knights D, Unno T, Bobr A, Kang J, Khoruts A, Knight R, Sadowsky MJ. 2015. Dynamic changes in short- and long-term bacterial composition following fecal microbiota transplantation for recurrent Clostridium difficile infection. Microbiome 3:10.
17. Bokulich NA, Dillon MR, Zhang Y, Rideout JR, Bolyen E, Li H, Albert PS, Caporaso JG. 2018. q2-longitudinal: Longitudinal and Paired-Sample Analyses of Microbiome Data. mSystems 3.
18. David LA, Maurice CF, Carmody RN, Gootenberg DB, Button JE, Wolfe BE, Ling A V, Devlin AS, Varma Y, Fischbach MA, Biddinger SB, Dutton RJ, Turnbaugh PJ. 2014. Diet rapidly and reproducibly alters the human gut microbiome. Nature 505:559–63.
19. Zhou Y, Shan G, Sodergren E, Weinstock G, Walker WA, Gregory KE. 2015. Longitudinal Analysis of the Premature Infant Intestinal Microbiome Prior to Necrotizing Enterocolitis: A Case-Control Study. PLoS One 10:e0118632.
20. Bokulich NA, Chung J, Battaglia T, Henderson N, Jay M, Li H, D Lieber A, Wu F, Perez-Perez GI, Chen Y, Schweizer W, Zheng X, Contreras M, Dominguez-

Bello MG, Blaser MJ. 2016. Antibiotics, birth mode, and diet shape microbiome maturation during early life. Sci Transl Med 8:343ra82.
21. Shields-Cutler RR, Al-Ghalith GA, Yassour M, Knights D. 2018. SplinectomeR Enables Group Comparisons in Longitudinal Microbiome Studies. Front Microbiol 9:785.
22. Ramsay JO, Silverman BW. 1997. Functional Data Analysis. Springer New York, New York, NY.
23. James GM, Hastie TJ, Sugar CA. Principal Component Models for Sparse Functional Data. Biometrika. Oxford University PressBiometrika Trust.
24. Vehtari A, Gelman A, Gabry J. 2015. Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC.
25. Gelman A, Carlin J, Stern H, Dunson D, Vehtari A, Rubin D. 2013. Bayesian data analysis, 3rd ed. Chapman and Hall/CRC.
26. Thompson WK, Rosen O. 2008. A Bayesian Model for Sparse Functional Data. Biometrics 64:54–63.
27. Lockwood JR, Savitsky TD, McCaffrey DF. 2015. Inferring constructs of effective teaching from classroom observations: An application of Bayesian exploratory factor analysis without restrictions. Ann Appl Stat 9:1484–1509.
28. Carpenter B, Gelman A, Hoffman MD, Lee D, Goodrich B, Betancourt M, Brubaker M, Guo J, Li P, Riddell A. 2017. Stan: A Probabilistic Programming Language. J Stat Softw 76:1–32.
29. Turnbaugh PJ, Ley RE, Hamady M, Fraser-Liggett CM, Knight R, Gordon JI. 2007. The human microbiome project. Nature 449:804–10.
30. Consortium THMP, Methé BA, Nelson KE, Pop M, Creasy HH, Giglio MG, Huttenhower C, Gevers D, Petrosino JF, Abubucker S, Badger JH, Chinwalla AT, Earl AM, FitzGerald MG, Fulton RS, Hallsworth-Pepin K, Lobos EA, Madupu R, Magrini V, Martin JC, Mitreva M, Muzny DM, Sodergren EJ, Versalovic J, Wollam AM, Worley KC, Wortman JR, Young SK, Zeng Q, Aagaard KM, Abolude OO, Allen-Vercoe E, Alm EJ, Alvarado L, Andersen GL, Anderson S, Appelbaum E, Arachchi HM, Armitage G, Arze CA, Ayvaz T, Baker CC, Begg L, Belachew T, Bhonagiri V, Bihan M, Blaser MJ, Bloom T, Bonazzi VR, Brooks P, Buck GA, Buhay CJ, Busam DA, Campbell JL, Canon SR, Cantarel BL, Chain PS, Chen I-MA, Chen L, Chhibba S, Chu K, Ciulla DM, Clemente JC, Clifton SW, Conlan S, Crabtree J, Cutting MA, Davidovics NJ, Davis CC, DeSantis TZ, Deal C, Delehaunty KD, Dewhirst FE, Deych E, Ding Y, Dooling DJ, Dugan SP, Jr WMD, Durkin AS, Edgar RC, Erlich RL, Farmer CN, Farrell RM, Faust K, Feldgarden M, Felix VM, Fisher S, Fodor AA, Forney L, Foster L, Francesco V Di, Friedman J, Friedrich DC, Fronick CC, Fulton LL, Gao H, Garcia N, Giannoukos G, Giblin C, Giovanni MY, Goldberg JM, Goll J, Gonzalez A, Griggs A, Gujja S, Haas BJ, Hamilton HA, Harris EL, Hepburn TA, Herter B, Hoffmann DE, Holder ME, Howarth C, Huang KH, Huse SM, Izard J, Jansson JK, Jiang H, Jordan C, Joshi V, Katancik JA, Keitel WA, Kelley ST, Kells C, Kinder-Haake S, King NB, Knight R, Knights D, Kong HH, Koren O, Koren S, Kota KC, Kovar CL, Kyrpides NC, Rosa PS La, Lee SL, Lemon KP, Lennon N, Lewis CM, Lewis L, Ley RE, Li K, Liolios K, Liu B, Liu Y, Lo C-C, Lozupone CA, Lunsford RD, Madden T, Mahurkar AA, Mannon PJ, Mardis ER, Markowitz VM, Mavrommatis K, McCorrison JM, McDonald D, McEwen J, McGuire AL, McInnes P, Mehta T, Mihindukulasuriya KA, Miller JR, Minx PJ, Newsham I, Nusbaum C, O'Laughlin M, Orvis J, Pagani I, Palaniappan K, Patel SM, Pearson M, Peterson J, Podar M, Pohl C, Pollard KS, Priest ME, Proctor LM, Qin X, Raes J, Ravel J, Reid JG, Rho M, Rhodes R, Riehle KP, Rivera MC, Rodriguez-Mueller B, Rogers Y-H, Ross

MC, Russ C, Sanka RK, Sankar P, Sathirapongsasuti JF, Schloss JA, Schloss PD, Schmidt TM, Scholz M, Schriml L, Schubert AM, Segata N, Segre JA, Shannon WD, Sharp RR, Sharpton TJ, Shenoy N, Sheth NU, Simone GA, Singh I, Smillie CS, Sobel JD, Sommer DD, Spicer P, Sutton GG, Sykes SM, Tabbaa DG, Thiagarajan M, Tomlinson CM, Torralba M, Treangen TJ, Truty RM, Vishnivetskaya TA, Walker J, Wang L, Wang Z, Ward D V., Warren W, Watson MA, Wellington C, Wetterstrand KA, White JR, Wilczek-Boney K, Wu YQ, Wylie KM, Wylie T, Yandava C, Ye L, Ye Y, Yooseph S, Youmans BP, Zhang L, Zhou Y, Zhu Y, Zoloth L, Zucker JD, Birren BW, Gibbs RA, Highlander SK, Weinstock GM, Wilson RK, White O. 2012. A framework for human microbiome research. Nature 486:215–221.

31. Caporaso JG, Lauber CL, Costello EK, Berg-Lyons D, Gonzalez A, Stombaugh J, Knights D, Gajer P, Ravel J, Fierer N, Gordon JI, Knight R. 2011. Moving pictures of the human microbiome. Genome Biol 12:R50.
32. Dominguez-Bello MG, De Jesus-Laboy KM, Shen N, Cox LM, Amir A, Gonzalez A, Bokulich NA, Song SJ, Hoashi M, Rivera-Vinas JI, Mendez K, Knight R, Clemente JC. 2016. Partial restoration of the microbiota of cesarean-born infants via vaginal microbial transfer. Nat Med 22:250–3.
33. Lozupone C, Knight R. 2005. UniFrac: a new phylogenetic method for comparing microbial communities. Appl Environ Microbiol 71:8228–35.
34. Aitchison J (John). 1986. The statistical analysis of compositional data. Chapman and Hall.

Supplemental figures

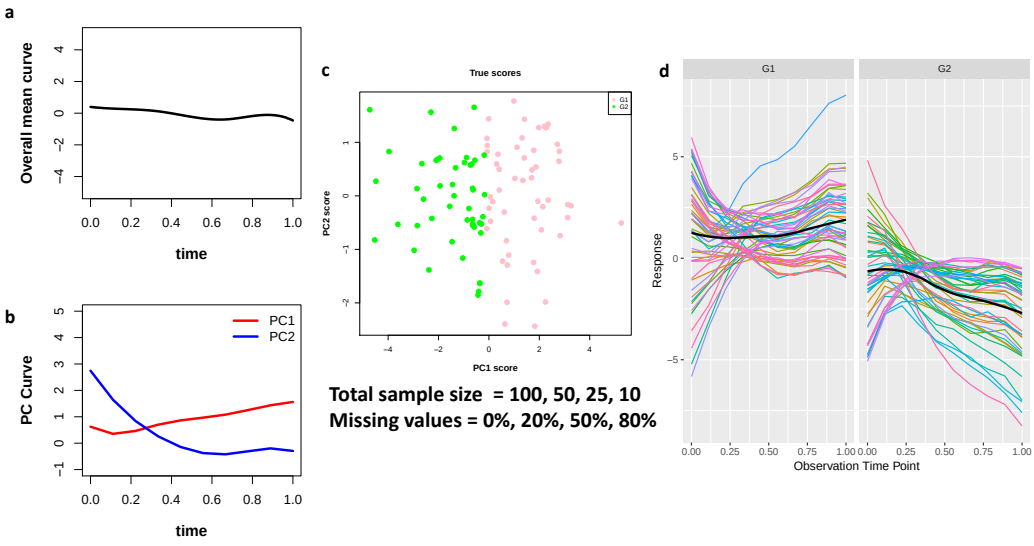

Figure S1: Simulation settings with missing data and varied sample size **(a)** underlying overall mean curve from a real microbiome dataset. **(b)** true PC curves from a real microbiome dataset. **(c)** assignment of two groups of subjects based on separation of PC 1 scores but not PC 2 scores. **(d)** simulated trajectories for subjects in two groups with black line indicating the loess mean curves for each group.

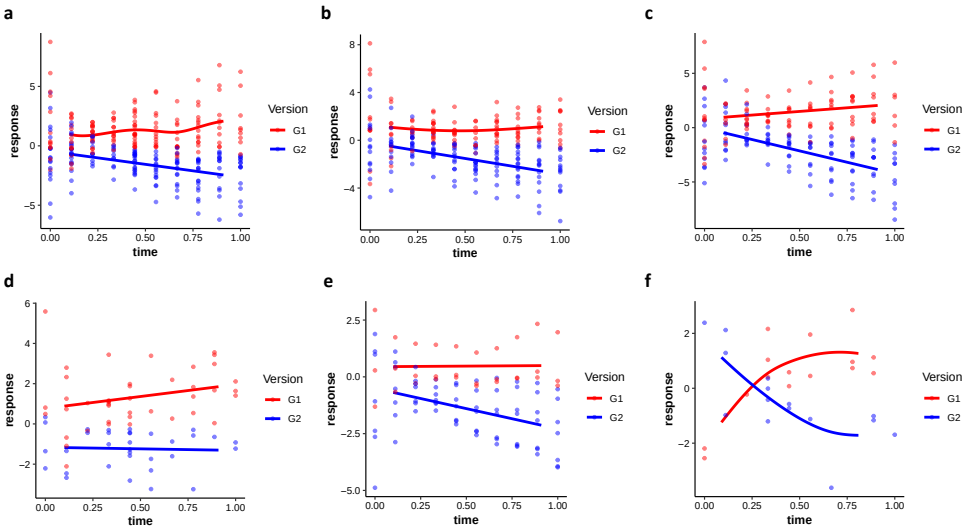

Figure S2: Performance of splinectomeR on simulated data with varied sample size and missingness **(a)** Estimated group mean curves on simulated observations with 100 total samples and 80% missing data. **(b)** Estimated group mean curves on simulated observations with 50 total samples and 50% missing data. **(c)** Estimated group mean curves on simulated observations with 25 total samples and 20% missing data. **(d)** Estimated group mean curves on simulated observations with 25 total samples and 80% missing data. **(e)** Estimated group mean curves on simulated observations with 10 total samples and 0% missing data. **(f)** Estimated group mean curves on simulated observations with 10 total samples and 80% missing data.