



Published in final edited form as:

Comput Stat Data Anal. 2017 January ; 105: 46–52. doi:10.1016/j.csda.2016.07.010.

A note on modeling sparse exponential-family functional response curves

Jan Gertheiss^{a,*}, Jeff Goldsmith^b, and Ana-Maria Staicu^c

Jan Gertheiss: jan.gertheiss@tu-clausthal.de; Jeff Goldsmith: jeff.goldsmith@columbia.edu; Ana-Maria Staicu: astaicu@ncsu.edu

^aInstitute of Applied Stochastics and Operations Research, Clausthal University of Technology, Clausthal-Zellerfeld, Germany

^bDepartment of Biostatistics, Columbia University, New York, U.S.A

^cDepartment of Statistics, North Carolina State University, Raleigh, U.S.A

Abstract

Non-Gaussian functional data are considered and modeling through functional principal components analysis (FPCA) is discussed. The direct extension of popular FPCA techniques to the generalized case incorrectly uses a marginal mean estimate for a model that has an inherently conditional interpretation, and thus leads to biased estimates of population and subject-level effects. The methods proposed address this shortcoming by using either a two-stage or joint estimation strategy. The performance of all methods is compared numerically in simulations. An application to ambulatory heart rate monitoring is used to further illustrate the distinctions between approaches.

Keywords

Binomial Data; Functional Principal Components; Longitudinal Data; Mixed Models; Smoothing; Sparse Sampling Design

1. Introduction

Standard methods to analyze repeatedly observed exponential family data can be separated into two main categories: *marginal models*, or population-average models, and *conditional models*, also known as mixed effects or subject-specific models. The former models focus on inference for population-level effects, and are traditionally estimated using a generalized estimating equations (GEE) framework [1, 2, 3, 4]. In this note, we focus on the latter framework, which focuses on the within-subject associations, because this perspective is most commonly employed in functional data analysis (FDA). The conditional model

*Corresponding author. Institute of Applied Stochastics and Operations Research, Clausthal University of Technology, Erzstr. 1, 38678 Clausthal-Zellerfeld, Germany. Phone: +49 5323 72 2954. Fax: +49 5323 72 2304.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

framework is particularly attractive in FDA because it models the dependence inherent in the functional observation for a subject as the realization of a latent random process, and this realization is often a quantity of interest. Unlike the common approach used in a longitudinal data analysis, which assumes parametric subject-specific effects in a mixed model, the dependence of the latent process in a functional data analysis is assumed unknown.

Functional data are commonly defined as observations on subjects that one can imagine as arising from the evaluation of a subject-specific real-valued curve at a finite grid of points. In many cases, however, the grid is irregular across subjects or sparse at the subject level. There has been tremendous interest in the analysis of functional data during the past few decades; see, e.g., Ramsay and Silverman [5] or Ferraty and Vieu [6] for book-length overviews, and Sørensen et al. [7] for a brief introduction. One popular technique, used to describe the variability in a sample of curves, is functional principal component analysis (FPCA). The conceptual FPCA model for real-valued response curves $Y_i(t)$ for $t \in \mathcal{T}$ with smooth mean function $\mu(t)$ and covariance function $\Sigma(t, t')$ is

$$Y_i(t) = \mu(t) + \sum_{k \geq 1} \xi_{ik} \psi_k(t) + \varepsilon_i(t), \quad (1)$$

where $\xi_{ik} \sim (0, \lambda_k)$ are subject-specific zero-mean loadings that are uncorrelated and have variance λ_k , $\{\lambda_k, \psi_k(t)\}$ is the pair of eigenvalue/eigenfunction of the covariance $\Sigma(\cdot, \cdot)$ with $\lambda_1 \geq \lambda_2 \geq \dots \geq 0$, and $\varepsilon_i(\cdot)$ is some zero-mean white noise process. This is commonly known as the Karhunen-Loève (KL) expansion and is a very common approach to model functional data [see, e.g., 8, 9, 10]. In practice, the sum in (1) is truncated such that

$$E[Y_i(t) | \xi_i] = \mu(t) + \sum_{k=1}^M \xi_{ik} \psi_k(t), \text{ where } M \text{ is a finite truncation.}$$

Model (1) decomposes curves $Y_i(t)$ using shared basis functions $\psi_k(\cdot)$ and subject-specific scores ξ_{ik} , and is very similar to a mixed model: indeed, setting $M=1$ and taking $\psi_k(t) = 1$, $\forall t$ yields the well-known random intercept model. However, FPCA describes the directions of variation that appear in the sample and gives a parsimonious decomposition of the complex variance structure $\Sigma(\cdot, \cdot)$. For sparse data in particular, FPCA borrows strength across subjects to estimate major patterns of variation and improve estimation of subject-specific curves [11]. Estimation techniques for model (1) have been proposed and studied for dense grids, irregular observations, and sparsely observed functional data [see 5, 11, and many others]. A popular algorithm for FPCA, for both dense and sparse data, consists of the following main steps: (i) estimate the marginal mean by assuming independence across subjects and across grid points within a subject; (ii) estimate the marginal covariance of subject-specific deviations, again by assuming independence, and take its spectral decomposition to obtain estimates of the eigenfunctions $\psi_k(\cdot)$ and eigenvalues λ_k ; and (iii) conditioning on the estimated mean function and eigenfunctions, estimate the scores ξ_{ik} in a mixed model.

Analyzing repeated exponential-family outcomes using functional data approaches is currently an area of intensive research [12, 13, 14, 15, 16, 17, 18]. Hall et al. [12] extended

model (1) to handle non-Gaussian functional data. Here, only the *latent* process is assumed to be Gaussian. Their proposed generalized FPCA (GFPCA) model is

$$E[Y_i(t)|\xi_i] = h\left\{\mu(t) + \sum_{k=1}^M \xi_{ik} \psi_k(t)\right\}, \quad (2)$$

with a known response function $h(\cdot)$; more recently, this model framework has been extended to account for rare events [14] and repeated functional observations on each subject [17]. As in model (1), it is assumed that, conditional on the subject-specific scores ξ_{ik} , the responses $Y_i(t)$'s are independent over t . The estimation method proposed by Hall et al. for model (2) directly extends the ideas from Gaussian response FPCA. However, for non-Gaussian curves, using a marginal approach to estimate the mean and covariance in what is inherently a conditional model, and then estimating subject-specific effects based on this mean and covariance, results in poor performance: the marginal mean is biased for the mean of the specified model and, as a consequence, the estimation of basis functions and subject effects is affected negatively. This is not an issue for model (1), of course, because the marginal and conditional mean are the same.

To gain more intuition, consider binary response data arising from model (2) using a logistic link function. In this case, the marginal mean $\alpha(t) = E[Y_i(t)]$ does not equal $h(\mu(t))$ if $\mu(t) \neq 0$. More specifically (see also the Appendix), for some $t_0 \in \mathcal{T}$,

$$\alpha(t_0) < h(\mu(t_0)), \text{ if } \mu(t_0) > 0, \text{ and } \alpha(t_0) > h(\mu(t_0)), \text{ if } \mu(t_0) < 0. \quad (3)$$

The amount of bias introduced by using a marginal estimate of the conditional mean depends on the amount of variability in the latent subject-specific process.

Example

To illustrate this point, Figure 1 shows two simulated samples of curves generated from model (2) with $M=1$, $\mu(t) = 2 \sin(2\pi t)$ and $\psi_1(t) = \sin(2\pi t)$; in one sample of curves $\lambda_1 = 1$, and in the other $\lambda_1 = 4$. The left panel shows the curves on the logit scale, on which the data are generated and naturally interpreted; the right panel shows these transformed to the probability scale, along with the marginal means of both samples. Although the samples share a mean function $\mu(t)$ on the logit scale and share a conditional mean (given $\xi = 0$) on the probability scale, both marginal means are biased estimates of $h(\mu(t))$, and a larger bias is observed for the more variable sample. Indeed, when deriving their marginal approach to estimate parameters in model (2), Hall et al. [12] assumed the variation of the latent process around mean $\mu(t)$ to be relatively small. This assumption, however, is easily violated in practice, and Hall et al. do not provide suggestions for this case.

In this note, we argue that the GFPCA model (2) is correctly understood to have a conditional, rather than a marginal, interpretation; as a result, we propose new techniques for

estimation that are built on this distinction. We emphasize the use of these methods for sparse data due to the benefits of borrowing strength across subjects in this setting, but note that the same considerations exist in the dense case. Our approach is to make use of the generalized additive mixed model framework that inherently underlies model (2); we estimate quantities of interest either in a frequentist framework via a two-step procedure, or in Bayesian framework via a joint modeling algorithm. We show through numerical studies that accounting for the dependence in this way leads to improved mean estimation and to improved prediction of latent trajectories. The methods proposed are publicly available as an R package on <https://github.com/jeff-goldsmith/gfpca>, which is also compatible with [refund.shiny](https://www.refund.shiny) [19, 20].

The rest of the paper is organized as follows. In Section 2 we describe the new estimation strategies explicitly motivated by the correct conditional interpretation of FPCA. In Section 3 these are compared to the marginal approaches described above in simulation studies. In Section 4, a real world data set is analyzed. Section 5 concludes.

2. Alternative Proposals for Estimation and Prediction

Two-step GFPCA in a Generalized Additive Mixed Model

Assume for the moment that eigenfunctions $\psi_k(t)$ in model (2) are known. Notationally, let $Y_i(t_{ij})$, $j = 1, \dots, J_i$ be the observation for subject i at grid point t_{ij} . Then for observed data, the model

$$E[Y_i(t_{ij})] = h \left(\mu(t_{ij}) + \sum_{k=1}^M \xi_{ik} \psi_k(t_{ij}) \right) \quad (4)$$

is a generalized additive mixed model (GAMM) with unknown mean structure μ and subject-specific random effects ξ_{ik} . Excellent software is available for fitting such models, such as the R package `gamm4` [21]. Shortly, the mean function $\mu(t)$ is estimated using a rich spline basis with an explicit penalty to enforce smoothness. Taking advantage of the well-known connection between ridge-style penalties and mixed models, spline basis coefficients are treated as random effects [22, 23, 24]; the associated variance component, which controls the degree of penalization, as well as the variances of the random effects ξ_k are estimated via maximum likelihood. The terms ξ_k are considered “true” random effects in the sense that they are conceptualized as draws from a population [25].

The preceding estimation technique is consistent with the conditional interpretation of model (2). Of course, in practice the eigenfunctions ψ_k are unknown. Therefore we propose to estimate these using available tools, and then to plug estimates $\hat{\psi}_k(t)$ into model (4). Two approaches for obtaining estimates $\hat{\psi}_k(t)$ are considered. First, we directly apply FPCA for real-valued response curves to the observed data using the algorithms described in the Introduction, ignoring the distribution of the observed data. Second, we apply the method of Hall et al. [12] for generalized FPCA by computing a marginal mean and covariance using a Taylor approximation to the response function $h(\cdot)$. Given the estimates of $\hat{\psi}_k(t)$ produced

by these methods, it is possible to estimate the conditional mean $\mu(t)$ and scores ξ_k in the generalized additive mixed model (4).

One alternative approach, as suggested by an anonymous reviewer, is to use (i) rich splines at the subject level to estimate a conditional model, followed (ii) by rotation to obtain FPCs. By using the estimated covariance matrix of the basis coefficients, one could rotate the initial basis to obtain the FPC to be used in the two-step model. Recently, Scheipl et al. [18] considered (i) to model non-Gaussian functional data and found that doing so works well in terms of accurately predicting the subject-specific curves for densely observed data, but it may be problematic in the case of sparsely observed data. Specifically, using a relatively large number of basis functions, subject-specific basis coefficients are hard to estimate with a small number of irregularly sampled observations per subject.

Another alternative, along the same direction, is to use a reduced rank mixed effects model. Specifically, one could represent the FPC as rotations of user-specified spline functions and impose independence for the principal components' coefficients. Existing literature [26, 27] developed this approach for Gaussian functional data using a joint estimation of the mean and principal component functions by using a Monte Carlo EM algorithm. Recently, Huang et al. [15] extended these ideas to non-Gaussian functional data, additionally assuming data coming from different clusters with distinct mean and covariance structures. Here the number of principal components can be selected using a grid search employing a Bayesian information criterion. However, Huang et al. investigated the approach numerically for dense designs only, and performance with sparse data remains unclear. Furthermore, there is no code publicly available, and the authors communicated by email that they did not have usable code to share at the time this manuscript was prepared. Our joint Bayesian GFPCA described below, however, is closely related. Hence we will consider this in the simulation studies and real data application.

Bayesian GFPCA

Focusing exclusively on densely observed functional data, Goldsmith et al. [17] proposed to estimate parameters in model (2), including the mean $\mu(t)$, the scores ξ_k and the FPC basis functions $\psi_k(t)$, in a joint Bayesian framework [see also 28]. Briefly, after fixing M , the mean function and FPC basis functions are expressed using a rich spline basis. The vector of spline coefficients for each of these is given a multivariate normal prior distribution with a variance structure that enforces a smoothness penalty, analogous to the specification of the mean in the generalized additive model. Variance components controlling the degree of penalty are given inverse gamma priors and, in keeping with Bayesian PCA, scores ξ_k are assigned standard normal prior distributions. Given these priors, samples from the posterior distribution are drawn using Hamiltonian Monte Carlo using Stan [29]. An important disadvantage of this approach compared to the preceding two-step methods is the need to pre-specify the number of components M to be estimated, similarly to Huang et al. [15], rather than basing this on a percentage of outcome variance explained. However, as in Huang et al., it is possible to choose M using a grid search with an appropriately defined selection criterion. In this note, we extend this Bayesian approach to allow for sparsely observed data and explore the method's properties for such data.

3. Simulation Studies

We now conduct simulations to evaluate the techniques proposed in Section 2 for binary response data. We generate 250 datasets using model (2) in the following way. We set the mean function to be $\mu(t) = 8(t - 0.4)^2 - 3$, with $t \in \mathcal{T} = [0, 1]$, which gives $\mu(t)$ that is relatively far away from zero for most t . Principal component basis functions are $\psi_1(t) = \sqrt{2}\cos(2\pi t)$ and $\psi_2(t) = \sqrt{2}\sin(2\pi t)$, and scores are sampled using $\xi_{i1} \sim N(0, 1)$, $\xi_{i2} \sim N(0, 0.5)$. We consider two settings:

- Setting 1: we generate $n = 300$ subjects, and for each subject i between 7 and 10 time points $t_{ij} \in \mathcal{T}$, $j = 1, \dots, J_i$ are sampled, with J_i being uniformly distributed on $\{7, 8, 9, 10\}$.
- Setting 2: we generate $n = 600$ subjects and draw J_i uniformly from $\{14, 15, \dots, 20\}$, which is similar to the real data application in Section 4.

We use a logit link function and generate binary observations $Y_{ij}(t_{ij})$ by sampling from a binomial distribution with probability defined by the subject's latent process. For each generated dataset, we estimate the mean function $\mu(t)$ and the latent processes $b_i(t) = \sum_k \xi_{ik} \psi_k(t)$ for each subject using the marginal approach [12] and the methods described in Section 2. For GAMM-based methods using plug-in estimates of the $\psi_k(t)$, the number of FPCs is chosen such that at least 90% of the variation is explained. Because the Bayesian approach requires the number of basis functions to be specified prior to model fitting, we use 3 for all datasets to examine performance under reasonable misspecification. Finally, we include an “oracle” GAMM model by using the true eigenfunctions to provide a frame of reference.

The top row of Figure 2 shows the results of our simulation exercise for Setting 1. The estimated mean functions, averaged over the simulation runs, for each method appear in the left panel. For three datasets the refitting method using the HMY [12] approach to estimate eigenfunctions did not work due to the large number (8 or even 12) of eigenfunctions. Those datasets were excluded in Figure 2. As expected, the marginal approach gives a biased estimate for the true mean. Both of the two-step GAMMs reduce this bias considerably, as does the Bayesian approach. The “oracle” GAMM, using the true eigenfunctions, produces nearly unbiased estimates; any remaining bias is due to the use of penalization for the mean function estimate. The middle panel shows violin plots of the integrated mean squared error (IMSE = $\int [\mu(t) - \hat{\mu}(t)]^2 dt$) for the mean function $\mu(t)$ for each method, and the right panel shows similar plots for the average IMSE of the latent process $\mu(t) + \sum_k \xi_{ik} \psi_k(t)$ across subjects i . The Bayesian and the two-step GAMMs perform similarly on average; but for the latter, two extreme outliers are observed (with one not being shown in Figure 2). Analogous plots for Setting 2 are shown in the bottom row of Figure 2. Here the Bayesian approach appears to be superior to the two-step methods for the prediction of latent trajectories, perhaps because the plug-in estimates of the eigenfunctions are based on a marginal mean estimate and linear approximations. With respect to the mean function, the Bayesian approach gives virtually unbiased estimates. In the two-step approach, curves tend to be smoother (not visible from the averaged curves in Figure 2), which may lower the MSE (see middle panel), but induces some bias.

4. Real Data Application

In this Section we conduct an analysis of data collected to understand patterns in blood pressure and hypertension over the course of a day. Full details of the recruitment process and study population are given in Shimbo et al. [30] and described shortly here. A total of 813 adults with no history of cardiovascular disease, hypertension or other major diseases, and who were not taking anti-hypertensive drugs, were recruited from several large employers in New York City. The ambulatory blood pressure (ABP) of each subject was monitored using a cuff that recorded blood pressure at 28 minute intervals. For this analysis, we focus on hypertension at each recording as the binary outcome of interest, defined using established clinical thresholds as systolic BP ≥ 135 mm Hg or diastolic BP ≥ 90 mm Hg. Because blood pressure follows a circadian rhythm but clinical definitions of hypertension are not established, we consider recordings made during the hours 10:00am to 8:00pm. Data are aggregated into 5 minute intervals yielding a grid of length 121. Due to the sampling design, curves are sparsely observed within a subject but densely observed across subjects, and the median number of recordings per subject is 15.

We applied each of the methods examined in the simulations to this dataset. The left panel in Figure 3 shows the mean curve estimated by the marginal, two-step, and Bayesian methods. In this dataset hypertension is relatively rare, particularly in the evening when blood pressures tend to decline, and our results suggest that marginal approaches overestimate the probability of a hypertensive recording. The two-step and Bayesian methods are generally in agreement, although, as in our simulations, the Bayesian approach is more variable. The middle and right panels of Figure 3 show observed data (blue dots) and estimated subject-specific latent means based on the two-step and Bayesian approaches for 2 subjects. As for the mean, there is generally good agreement between methods although the Bayesian approach is more variable.

5. Concluding Remarks

The distinction between sparse functional data and traditional longitudinal data is a difficult one to make [see 11, 31, 12, 32, to name a few]. Generally, we feel that if the data are sufficient to support the added flexibility of functional approaches, like FPCA, then these tools should be used; however we hesitate to offer “rules of thumb” for the necessary number of observations per subject to qualify as “sparse functional data.” In any case, the lessons learned in longitudinal data analysis should not be forgotten when using a functional data context. One of these lessons is the distinction between conditional and marginal interpretations of the mean in generalized models.

Exponential family response curves are becoming more common, and standard tools in functional data analysis are being adapted to this setting. In this note we focused on the extension of FPCA, one of the most commonly used techniques for analyzing samples of curves, and emphasized the inherently conditional interpretation of the model. This has often been overlooked in previous estimation strategies, which instead use a mixture of marginal and conditional estimation strategies. We proposed both two-stage and joint estimation techniques for GFPCA and demonstrated the superiority of these methods over previous

approaches in simulations. Our application to real data further illustrated the bias that arises from marginal estimation in GFPCA. In particular the two-stage approach can be seen as a “quick-fix” for the marginal approach, which is still quite easy to implement, and works well in practice. We recall that some caution is needed when performing inference for FPCA and GFPCA, especially in the two-step approaches: uncertainty in the basis functions $\psi_k(t)$ should be incorporated into overall estimates of variability [33].

GFPCA is especially useful in the sparse data setting, where patterns of variability learned across subjects can inform the estimation of latent processes at the subject level even when relatively few points are observed. However, our proposed techniques are also applicable for densely observed data, and in that setting retain the benefits of parsimony and useful descriptions of the major patterns of variability in the population. That being said, dense data at the subject level allow the estimation of subject-specific deviations from the mean using, for example, a spline basis expansion [18] for each subject rather than an FPC expansion. Such an approach is expected to yield approximately unbiased estimates of the mean function for dense data, although it is not feasible for sparse data.

Acknowledgments

J. Goldsmith’s research was supported in part by Award R21EB018917 from the National Institute of Biomedical Imaging and Bioengineering and by Award R01HL123407 from the National Heart, Lung, and Blood Institute. A.-M. Staicu’s research was supported in part by US National Science Foundation grant number DMS 1454942 and US National Institute of Health grants R01NS085211 and R01MH086633. We want to thank the Editor, the Associate Editor and two anonymous referees for useful comments and suggestions that helped improve the paper.

References

1. Liang KY, Zeger SL. Longitudinal data analysis using generalized linear models. *Biometrika*. 1986; 73:13–22.
2. Zeger SL, Liang KY. Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*. 1986; 42:121–130. [PubMed: 3719049]
3. Liang KY, Zeger SL. Inference based on estimating functions in the presence of nuisance parameters. *Statistical Science*. 1995; 10:158–173.
4. Fitzmaurice GM, Laird NM, Rotnitzky AG. Regression models of discrete longitudinal responses (with discussion). *Statistical Science*. 1993; 8:284–309.
5. Ramsay, J.; Silverman, B. *Functional Data Analysis*. Springer; New York: 2005.
6. Ferraty, F.; Vieu, P. *Nonparametric Functional Data Analysis: Theory and Practice*. Springer; New York: 2006.
7. Sørensen H, Goldsmith J, Sangalli L. An introduction with medical applications to functional data analysis. *Statistics in Medicine*. 2013; 32:5222–5240. [PubMed: 24114808]
8. Di CZ, Crainiceanu C, Caffo BS, Punjabi NM. Multilevel functional principal components analysis. *The Annals of Applied Statistics*. 2009; 3:458–488. [PubMed: 20221415]
9. Greven S, Crainiceanu C, Caffo B, Reich D. Longitudinal functional principal components analysis. *Electronic Journal of Statistics*. 2010; 4:1022–1054. [PubMed: 21743825]
10. Jacques J, Preda C. Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*. 2014; 71:92–106.
11. Yao F, Müller HG, Wang JL. Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*. 2005; 100:577–590.
12. Hall P, Müller HG, Yao F. Modelling sparse generalized longitudinal observations with latent gaussian processes. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2008; 70:703–723.

13. Chen H, Wang Y, Paik MC, Choi HA. A marginal approach to reduced-rank penalized spline smoothing with application to multilevel functional data. *Journal of the American Statistical Association*. 2013; 108:1216–1229. [PubMed: 24497670]
14. Serban N, Staicu AM, Carroll RJ. Multilevel cross-dependent binary longitudinal data. *Biometrics*. 2013; 69:903–913. [PubMed: 24131242]
15. Huang H, Li Y, Guan Y. Joint modeling and clustering paired generalized longitudinal trajectories with application to cocaine abuse treatment data. *Journal of the American Statistical Association*. 2014; 109(508):1412–1424.
16. Gertheiss J, Maier V, Hessel EF, Staicu AM. Marginal functional regression models for analyzing the feeding behavior of pigs, *Journal of Agricultural, Biological, and Environmental Statistics*. 2015; 20:353–370.
17. Goldsmith J, Zippunikov V, Schrack J. Generalized multilevel functional-on-scalar regression and principal components analysis. *Biometrics*. 2015; 71:344–353. [PubMed: 25620473]
18. Scheipl F, Gertheiss J, Greven S. Generalized functional additive mixed models. *Electronic Journal of Statistics*. 2016; 10:1455–1492.
19. Wrobel J, Park SY, Staicu AM, Goldsmith J. Interactive graphics for functional data analyses. *Stat*. 2016; 5:108–118. [PubMed: 27182437]
20. Goldsmith J, Wrobel J. refund.shiny: Interactive plotting for functional data analyses. R package version 0.1. 2015
21. Wood, SN.; Scheipl, F. R package version 0.2–3. 2014. gamm4: Generalized additive mixed models using mgcv and lme4.
22. Ruppert, D.; Wand, MP.; Carroll, RJ. Semiparametric Regression. Cambridge University Press; Cambridge: 2003.
23. Wood SN. Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society Series B (Statistical Methodology)*. 2011; 73:3–36.
24. Scheipl F, Staicu AM, Greven S. Functional additive mixed models. *Journal of Computational and Graphical Statistics*. 2014; 24:477–502. [PubMed: 26347592]
25. Hodges, JS. Richly Parameterized Linear Models: Additive, Time Series, and Spatial Models Using Random Effects. CRC Press; Boca Raton: 2013.
26. James GM, Hastie TJ, Sugar CA. Principal component models for sparse functional data. *Biometrika*. 2000; 87:587–602.
27. Zhou L, Huang JZ, Carroll RJ. Joint modelling of paired sparse functional data using principal components. *Biometrika*. 2008; 95:601–619. [PubMed: 19396364]
28. Tidemann-Miller B, Reich B, Staicu A-M. Modeling multivariate mixed-response functional data. arXiv:1601.02461v1.
29. Stan Development Team. Stan Modeling Language User's Guide and Reference Manual, Version 2.10.0. 2016. URL <http://mc-stan.org/>
30. Shimbo D, Newman JD, Schwartz JE. Masked hypertension and prehypertension: diagnostic overlap and interrelationships with left ventricular mass: the Masked Hypertension Study. *American Journal of Hypertension*. 2012; 25:664–671. [PubMed: 22378035]
31. Hall P, Müller HG, Wang JL. Properties of principal component methods for functional and longitudinal data analysis. *The Annals of Statistics*. 2006; 34:1493–1517.
32. Li Y, Hsing T. Uniform convergence rates for nonparametric regression and principal component analysis in functional/longitudinal data. *The Annals of Statistics*. 2010; 38:3321–3351.
33. Goldsmith J, Greven S, Crainiceanu C. Corrected confidence bands for functional data using principal components. *Biometrics*. 2013; 69:41–51. [PubMed: 23003003]

Appendix

Proof of inequalities (3). Let's consider the latent process $X(t_0) = \mu(t_0) + Z(t_0)$ at some point t_0 . Since $\mu(t_0) = \mu_0$ is fixed and $Z(t_0) = Z_0$ is normal with mean zero and some variance $\sigma_0^2 > 0$

we have $X(t_0) = X_0 \sim N(\mu_0, \sigma_0^2)$. The marginal mean of the binary response Y at t_0 is $E\{E(Y(t_0)|X(t_0))\} = E(h(X_0))$. Let's first consider the case $\mu_0 > 0$, and let ϕ_0 denote the density of $N(\mu_0, \sigma_0^2)$, and Φ_0 the corresponding cdf. We have $E(h(X_0)) =$

$$= \int_{-\infty}^{+\infty} h(x) \phi_0(x) dx = \int_{-\infty}^0 h(x) \phi_0(x) dx + (\Phi_0(2\mu_0) - \Phi_0(0)) \int_0^{2\mu_0} \frac{h(x) \phi_0(x)}{\Phi_0(2\mu_0) - \Phi_0(0)} dx + \int_{2\mu_0}^{+\infty} h(x) \phi_0(x) dx.$$

Since $f(x) = \phi_0(x) / (\Phi_0(2\mu_0) - \Phi_0(0))$, if $x \in (0, 2\mu_0)$, and zero otherwise is the density of a random variable with mean μ_0 , and h is strictly concave on $(0, 2\mu_0)$ we know, due to Jensen's inequality that

$$E(h(X_0)) < \int_{-\infty}^0 h(x) \phi_0(x) dx + \int_{2\mu_0}^{+\infty} h(x) \phi_0(x) dx + (\Phi_0(2\mu_0) - \Phi_0(0)) h(\mu_0).$$

Since ϕ_0 is symmetric around μ_0 we can rewrite

$$\int_{2\mu_0}^{+\infty} h(x) \phi_0(x) dx = \int_{-\infty}^0 h(-x + 2\mu_0) \phi_0(x) dx, \text{ and thus have}$$

$$E(h(X_0)) < \int_{-\infty}^0 \delta(x) \phi_0(x) dx + (1 - 2\Phi_0(0)) h(\mu_0),$$

with $\delta(x) = h(x) + h(-x + 2\mu_0)$. About $\delta(x)$ we know (i) it is monotonically increasing on $(-\infty, 0]$ (first derivative $\delta'(x) \geq 0$ for $x \in (-\infty, 0]$), and (ii), again due to Jensen's inequality and h being strictly concave on $(0, 2\mu_0)$, $\frac{1}{2}\delta(0) = \frac{1}{2}(h(0) + h(2\mu_0)) < h((0+2\mu_0)/2) = h(\mu_0)$. As a consequence

$$E(h(X_0)) < \int_{-\infty}^0 2h(\mu_0) \phi_0(x) dx + (1 - 2\Phi_0(0)) h(\mu_0) = 2h(\mu_0) \Phi_0(0) + h(\mu_0) - 2h(\mu_0) \Phi_0(0) = h(\mu_0).$$

The proof of $E(h(X_0)) > h(\mu_0)$ for $\mu_0 < 0$ is analogous.

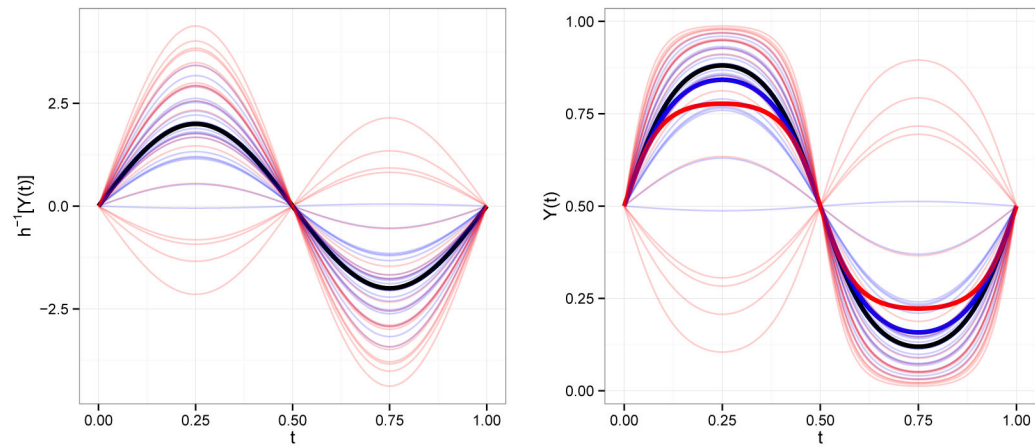
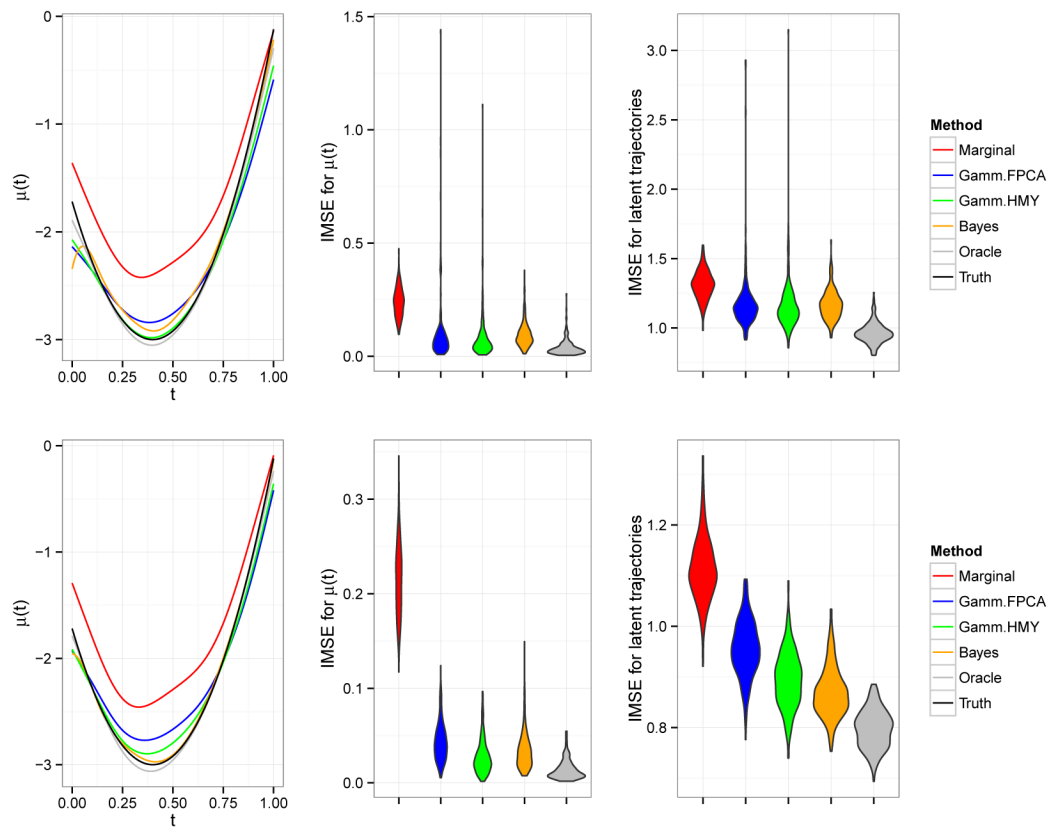


Figure 1.

Left panel shows two samples of latent curves on the logit scale with more and less variance, in red and blue respectively, around a common latent mean shown in black. The right panel shows these curves transformed to the probability scale; marginal means are shown in red and blue, while the transformation of the true mean is shown in black.

**Figure 2.**

Simulation results for Setting 1 (top row) and Setting 2 (bottom row). In each row, the left panel shows the average estimated mean function across all simulated datasets, the middle panel shows the IMSE for the mean function, and the right panel shows the average IMSE of latent processes across subjects. In case of Setting 1 for Gamm.FPCA and Gamm.HMY one extreme outlier is not shown.

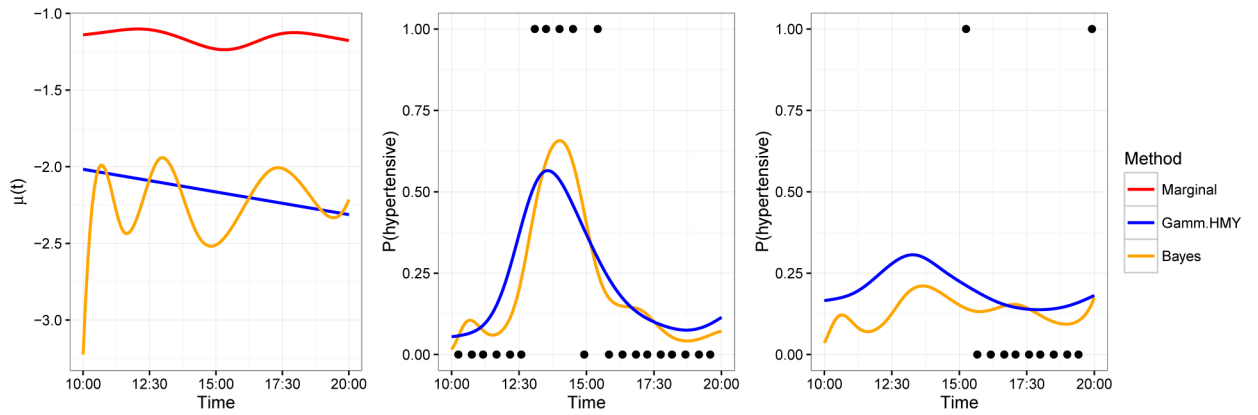


Figure 3.

Results of the analysis of the ambulatory blood pressure dataset. The left panel shows estimated mean functions from the marginal approach, the two-step GAMM-based approach, and the Bayesian approach. The middle and right panels show observed data and estimated latent means for two subjects.