SUPPLEMENT TO "MULTILEVEL FUNCTIONAL PRINCIPAL COMPONENT ANALYSIS"

By Chong-Zhi Di, Ciprian M. Crainiceanu,
Brian S. Caffo and Naresh M. Punjabi

Johns Hopkins University

1. Assessing the criterion for choosing the number of principal components via simulations. The criterion for choosing the number of components at level 1, N_1, is to choose the minimum number that satisfies the following two conditions: 1. the cumulative percentage of variance explained by the first N_1 components at level 1 is larger than a threshold P_1 (for instance 90%); and 2. the percentage of variance explained by any additional principal component is less than another threshold P_2 (for instance, $1/T$, where T is the number of grid points for each function). The number of principal components at level 2, N_2, was chosen similarly.

We evaluated the performance of this procedure via simulations in Case 2. We tried different thresholds, $\{0.6, 0.7, 0.8, 0.9\}$ for the first criterion, and $\{0.01, 0.03, 0.06, 0.08, 0.1\}$ for the second criterion. Figure 1 and 2 display results for the estimated number of dimensions at level 1 and 2, respectively. The true model had four principal components at each level. Results indicate that for these thresholds the criterion tends to be conservative and that the performance depends on the choice of the two thresholds. The stricter the criterion, the more often we overestimate the number of dimensions. For the simulation results presented in the paper, we used the threshold 90% for the cumulative proportion of variation, and 1% for the variance explained by each single additional component. Using this criterion we recovered the first four components in each of the 1000 simulations.

Figure 3 shows the estimated eigenvalues from 10 randomly selected simulations. For illustration purpose, we used the log scale for the Y axis. For level 1, the eigenvalues decay approximately exponentially (linearly on the log scale) for the first four to five components and even faster thereafter. Our procedure estimated the number of dimensions to be 4, 5, or 6, as shown. For level 2, similar patterns were observed. Our procedure estimated the number of dimensions to be 4 or 5 in most simulations.

In summary, our procedure performs well, but can be slightly conservative. Extensive simulations can solve this problem by fine tunning the thresholds.
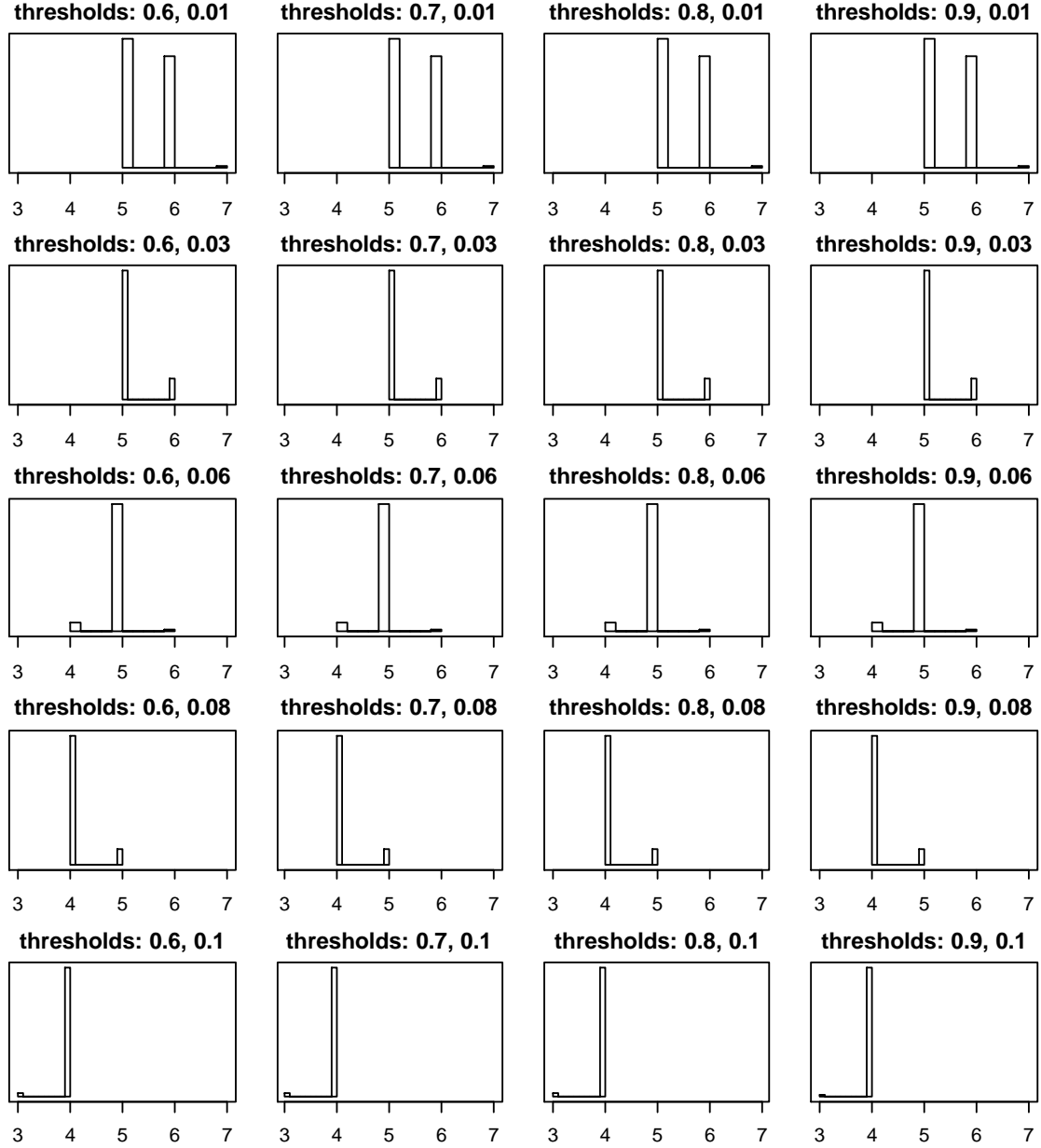
FIG 1. *Estimated number of principal components at level 1 from 1000 simulations. The first threshold is for the cumulative percentage of variance explained, while the second threshold is for the percentage of variance explained for each additional principal component.*
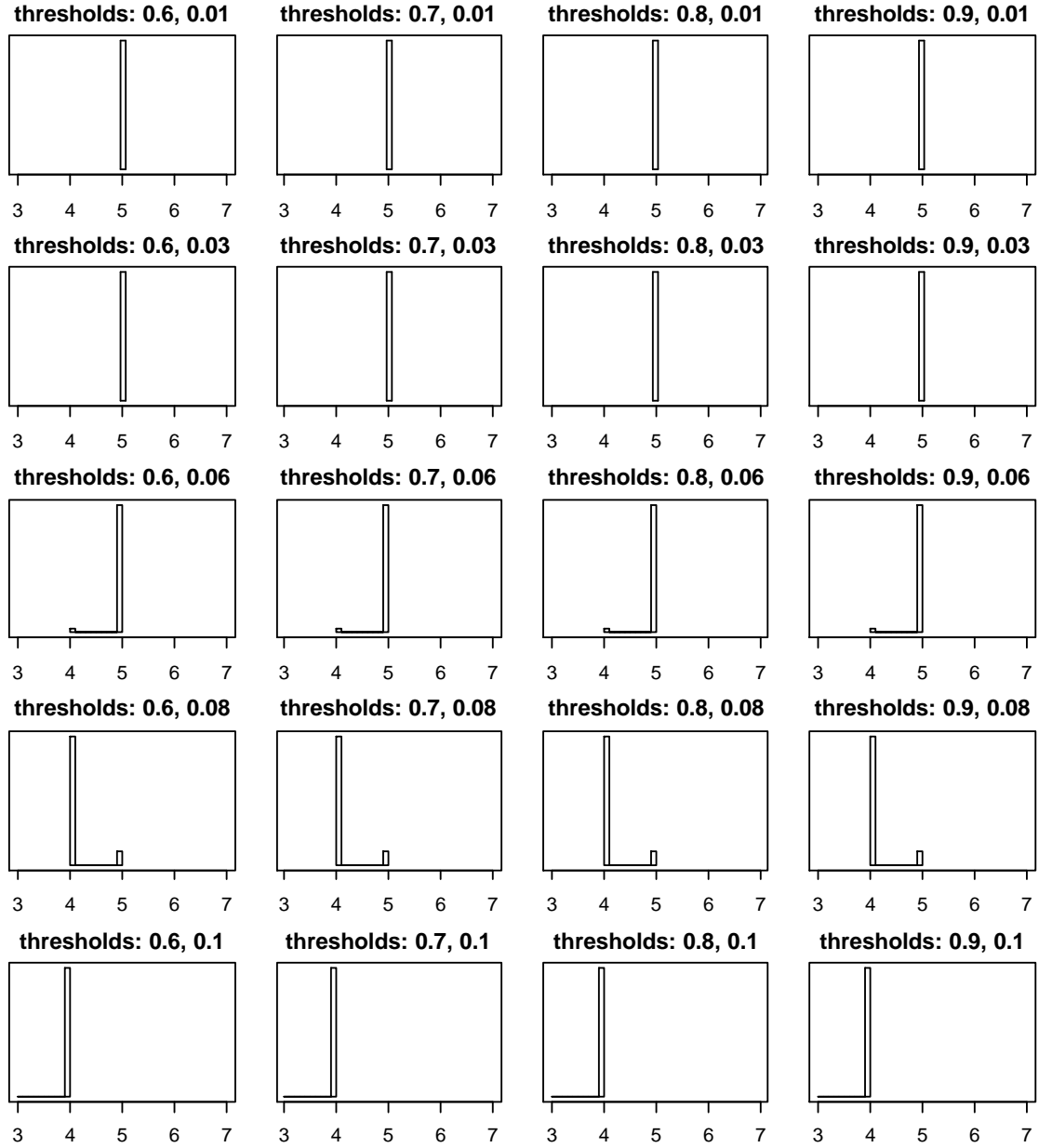
FIG 2. *Estimated number of principal components at level 2 from 1000 simulations. The first threshold is for the cumulative percentage of variance explained, while the second threshold is for the percentage of variance explained for each additional principal component.*
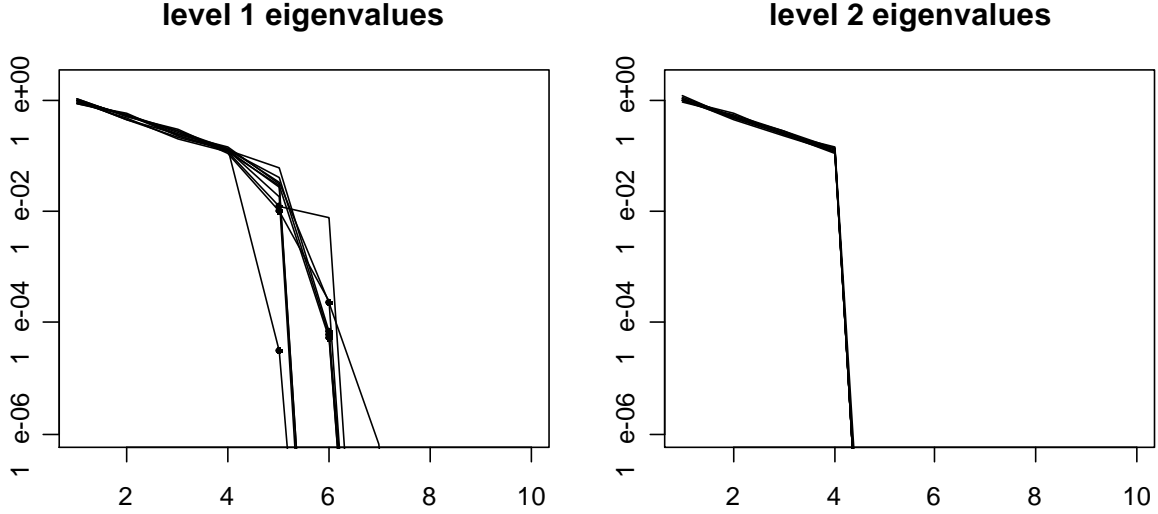
level 1 eigenvalues   ### level 2 eigenvalues

FIG 3. *Estimated level 1 and level 2 eigenvalues from 10 randomly selected simulations.*

2. Bayesian MCMC for principal component scores. As discussed in the paper, we used two linear mixed models, i.e, the full and the projection models, to estimate the principal component scores. In this section, we give details for the full model, including model specification, full conditionals, and simulation results.

The full model is

$$
(2.1) \quad
\begin{cases}
Y_{ij}(t_m) &= \mu(t_m) + \eta_j(t_m) + \sum_{k=1}^{N_1} \xi_{ik}\phi_k^{(1)}(t_m) + \sum_{l=1}^{N_2} \zeta_{ijl}\phi_l^{(2)}(t_m) + \varepsilon_{ij}(t_m) \\
\xi_{ik} &\sim N\{0, \lambda_k^{(1)}\}; \ \ \zeta_{ijl} \sim N\{0, \lambda_l^{(2)}\}; \ \ \varepsilon_{ij}(t_m) \sim N(0, \sigma^2) \\
\frac{1}{\sigma^2} &\sim Gamma(\alpha_1, \alpha_2)
\end{cases}
$$

When estimating the principal component scores, we plugged in estimates of eigenvalues, $\lambda_k^{(1)}$ and $\lambda_l^{(2)}$, and eigenfunctions, $\phi_k^{(1)}$ and $\phi_l^{(2)}$, and specify a Gamma prior for $1/\sigma^2$ with mean α_1/α_2 and variance α_1/α_2^2. We used $\alpha_1 = \alpha_2 = 0.0001$.

The full conditionals for this model are

1. $1/\sigma^2|others \sim Gamma(\alpha_1^{post}, \alpha_2^{post})$, where $\alpha_1^{post} = \alpha_1 + \frac{1}{2}IJT$, $\alpha_2^{post} = \alpha_2 + \frac{1}{2}\sum_{i,j,m} \epsilon_{ijm}^2$, and $\epsilon_{ijm} = Y_{ij}(t_m) - \mu(t_m) - \eta_j(t_m) - \sum_k \xi_{ik}\phi_k^{(1)}(t_m) - \sum_l \zeta_{ijl}\phi_l^{(2)}(t_m)$.

2. $\xi_{ik}|others \sim N(\mu_{\xi,ik}^{post}, v_{\xi,ik}^{post})$, where

$$\mu_{\xi,ik}^{post} = \frac{\lambda_k^{(1)} J \cdot \sum_m \phi_k^{(1)}(t_m)^2}{\lambda_k^{(1)} J \cdot \sum_m \phi_k^{(1)}(t_m)^2 + \sigma^2} \cdot \frac{\sum_{j,m} \phi_k^{(1)}(t_m) \left\{ \epsilon_{ijm} + \xi_{ik} \phi_k^{(1)}(t_m) \right\}}{J \cdot \sum_m \phi_k^{(1)}(t_m)^2}$$

and

$$v_{\xi,ik}^{post} = \frac{\lambda_k^{(1)} \sigma^2}{\lambda_k^{(1)} J \cdot \sum_m \phi_k^{(1)}(t_m)^2 + \sigma^2}.$$

Note that $\epsilon_{ijm} + \xi_{ik}\phi_k^{(1)}(t_m) = Y_{ij}(t_m) - \mu(t_m) - \eta_j(t_m) - \sum_{q \neq k} \xi_{iq}\phi_q^{(1)}(t_m) - \sum_l \zeta_{ijl}\phi_l^{(2)}(t_m)$
does not depend on ξ_{ik}.

3. $\zeta_{ijl}|others \sim N(\mu_{\zeta,ijl}^{post}, v_{\zeta,ijl}^{post})$, where

$$\mu_{\zeta,ijl}^{post} = \frac{\lambda_l^{(2)} \cdot \sum_m \phi_l^{(2)}(t_m)^2}{\lambda_l^{(2)} \cdot \sum_m \phi_l^{(2)}(t_m)^2 + \sigma^2} \cdot \frac{\sum_m \phi_k^{(1)}(t_m) \left\{ \epsilon_{ijm} + \zeta_{ijl} \phi_l^{(2)}(t_m) \right\}}{\sum_m \phi_l^{(2)}(t_m)^2}$$

and

$$v_{\zeta,ijl}^{post} = \frac{\lambda_l^{(2)} \sigma^2}{\lambda_l^{(2)} \cdot \sum_m \phi_l^{(2)}(t_m)^2 + \sigma^2}.$$

Note that $\epsilon_{ijm} + \zeta_{ijl}\phi_l^{(2)}(t_m) = Y_{ij}(t_m) - \mu(t_m) - \eta_j(t_m) - \sum_k \xi_{ik}\phi_k^{(1)}(t_m) - \sum_{q \neq l} \zeta_{ijq}\phi_q^{(2)}(t_m)$
does not depend on ζ_{ijl}.

If the grid points are dense, the following approximations $\sum_m \phi_k^{(1)}(t_m)^2 \approx T$ and $\sum_m \phi_l^{(2)}(t_m)^2 \approx T$ (because the eigenfunctions have unit norms, where T is the number of grid points) could help simplify formulas.

Consider the simulation framework for Case 2 described in Section 4.2. To study the convergence towards the target distribution we ran 5 Markov chains with different initial values. For each chain, we ran 2500 iterations, discarded the first 500 iterations as burn-in, and retained the last 2000 iterations for inference.

We show results for the full model with true $\sigma = 1$. Figure 4 displays the MCMC histories as well as the autocorrelation function of the chain for parameter σ. The chain mixes very well, with very small (less than 0.05) autocorrelations even at lag 1. Thus, 2000 iterations (with Monte Carlo error 0.0001, effective sample size 1900) are enough for this setting. We also tried different starting values for σ. In each case, after 500 iterations, the chain looked similar to the one in Figure 4. Figures 5 and 6 display the MCMC histories for level 1 and 2 scores of the first two
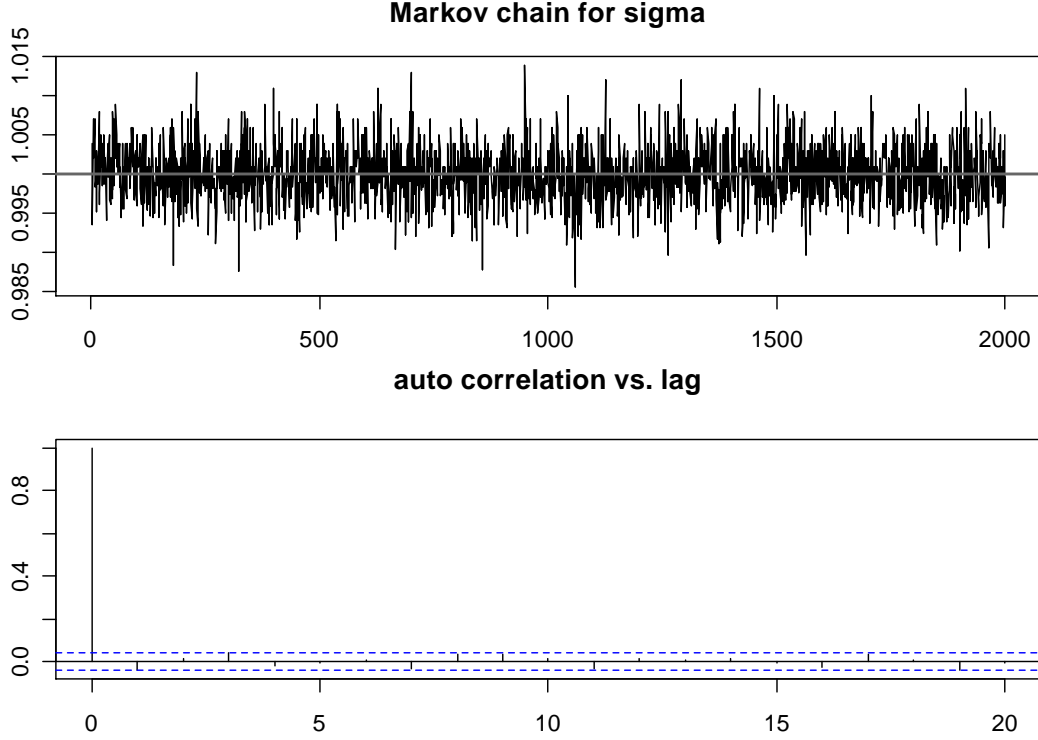
FIG 4. *MCMC histories for the magnitude of measurement error σ. The gray horizonal line represents the true $\sigma = 1$*

subjects. These chains also mix very well, with small auto correlations (with Monte Carlo error around 0.005, effective sample size 1800).

3. Simulation results.

3.1. *More results for Case 2.* Figure 7 displays estimated overall and visit-specific mean functions from 20 randomly selected simulations in the case when $\sigma = 0$. The estimated mean functions are close to the truth $\{\mu(t) = \eta_j(t) = 0\}$ in each simulation, relative to the range $(-6, 6)$ of the simulated functions. Results for the case with measurement error are similar and are not shown.

Figure 8 shows results for the second eigenfunction at both levels corresponding to all three levels of noise and for both estimation methods. Results for the other eigenfunctions show similar patterns.

3.2. *Simulation results for Case 1.* In the paper, we focused on Case 2. Here, we provide results for Case 1. Most of them show similar patterns to those for Case 2.
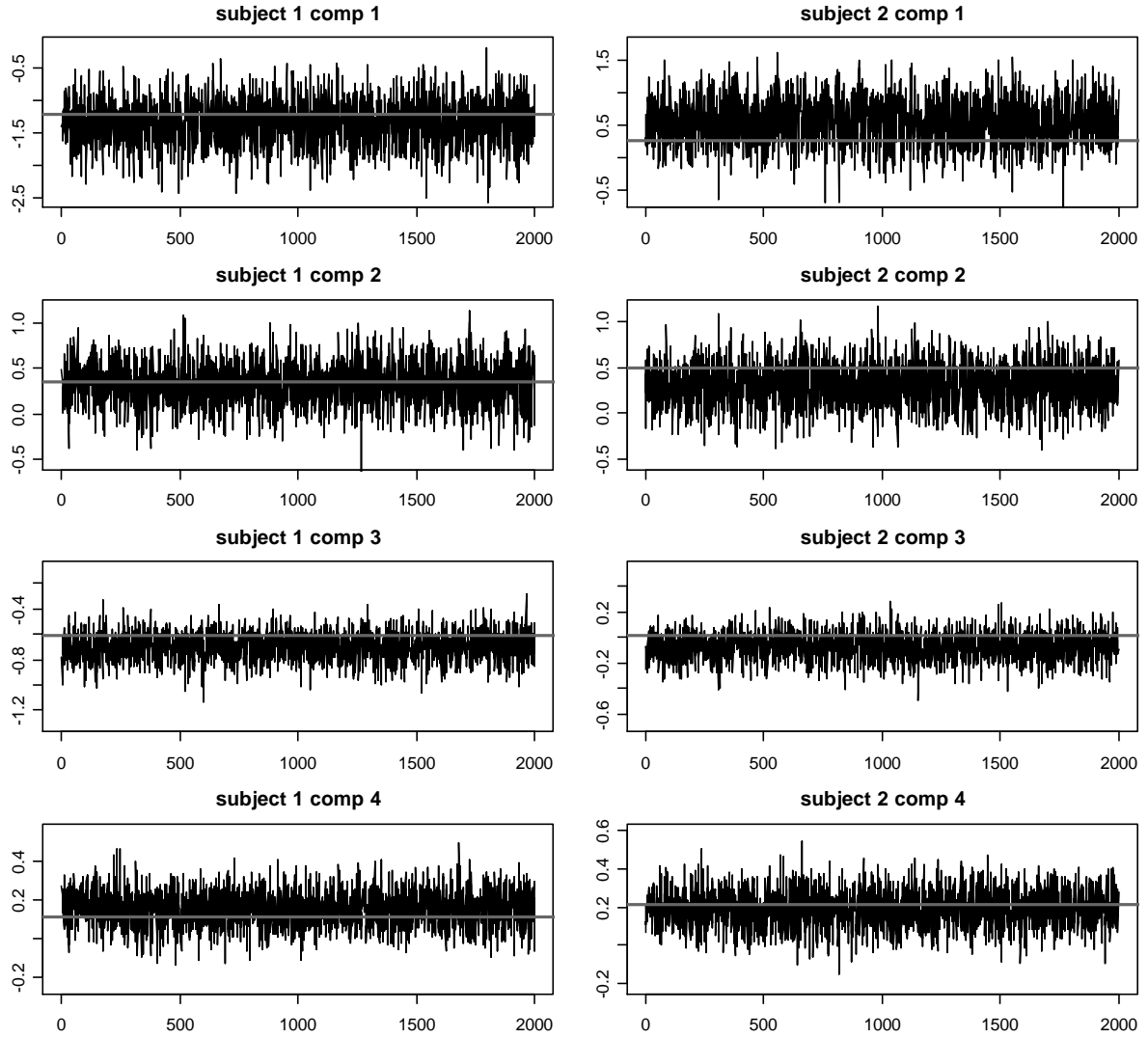
FIG 5. *MCMC histories for level 1 principal component scores (the true $\sigma = 1$). The gray horizonal line represents the true principal component scores.*
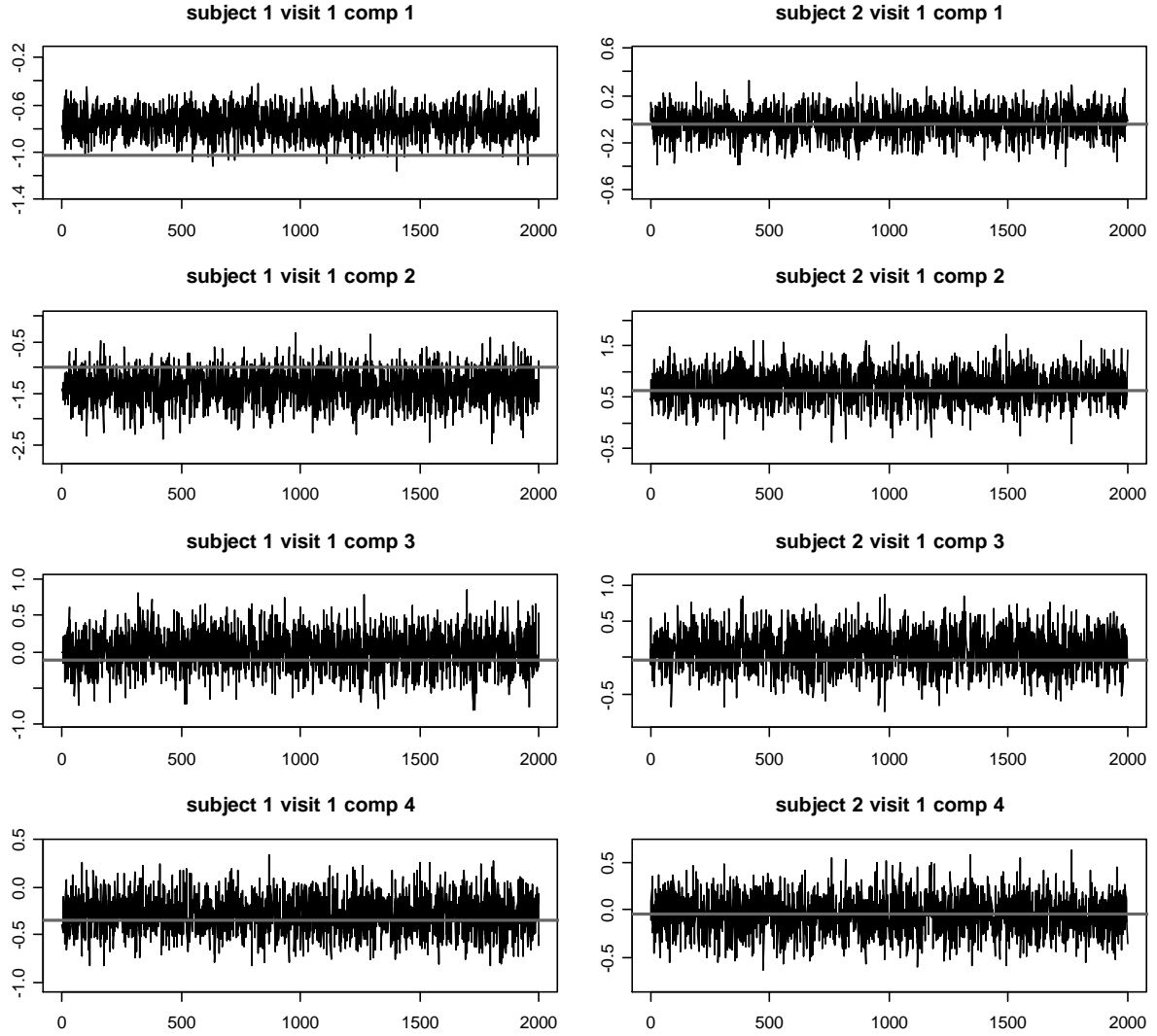
Fig 6. *MCMC histories for level 2 principal component scores (the true $\sigma = 1$). The gray horizonal line represents the true principal component scores.*
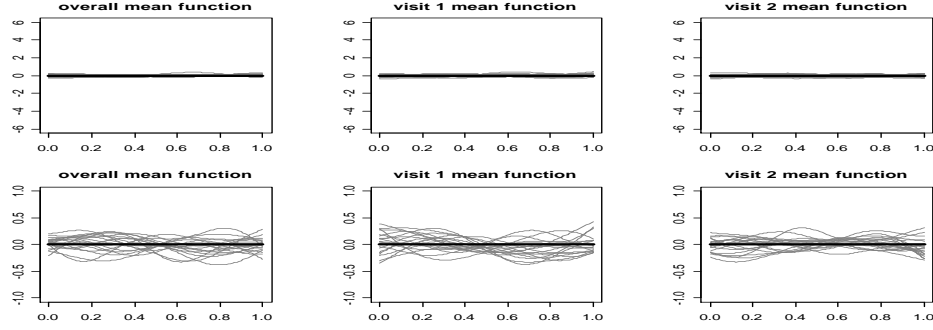
FIG 7. *Estimated overall and visit-specific mean functions from 20 randomly selected simulations when $\sigma = 0$. The thick black lines are the true mean functions. Figures in the upper panels are on a scale similar to Figure 2 in the paper, while those in the lower panels use a more refined scale.*
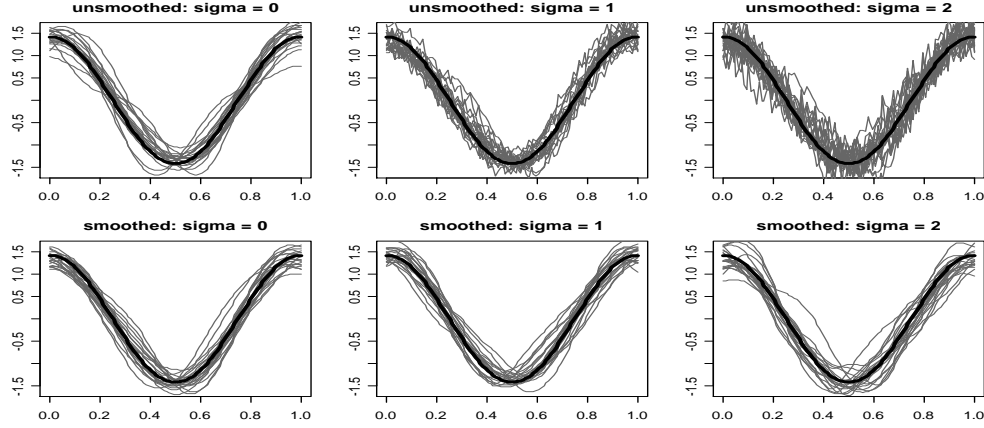

FIG 8. *Unsmoothed vs. smoothed MFPCA: estimated eigenfunctions (level 1 component 2) from 20 randomly selected simulations, when $\sigma = 0, 1, 2$. Thick black lines: true eigenfunctions.*

Figure 9 shows simulated functions for 4 subjects, with $\sigma = 0, 1, 2$ corresponding to the first, second and third row, respectively. As expected, when the amount of noise increases, the patterns at the subject-level functions become less obvious or hardly recognizable. Our proposed methodology can recover the true signals with 200 samples even in the extreme case when $\sigma = 2$.

Figure 10 shows boxplots of estimated level 1 and 2 eigenvalues for the three levels of noise using the unsmoothed MFPCA algorithm described in Section 2.2. The red solid line indicates the true eigenvalue. Figure 11 shows the same type of results as Figure 10 for the smooth MFPCA estimation algorithm. Those results are very similar to those of Case 2.

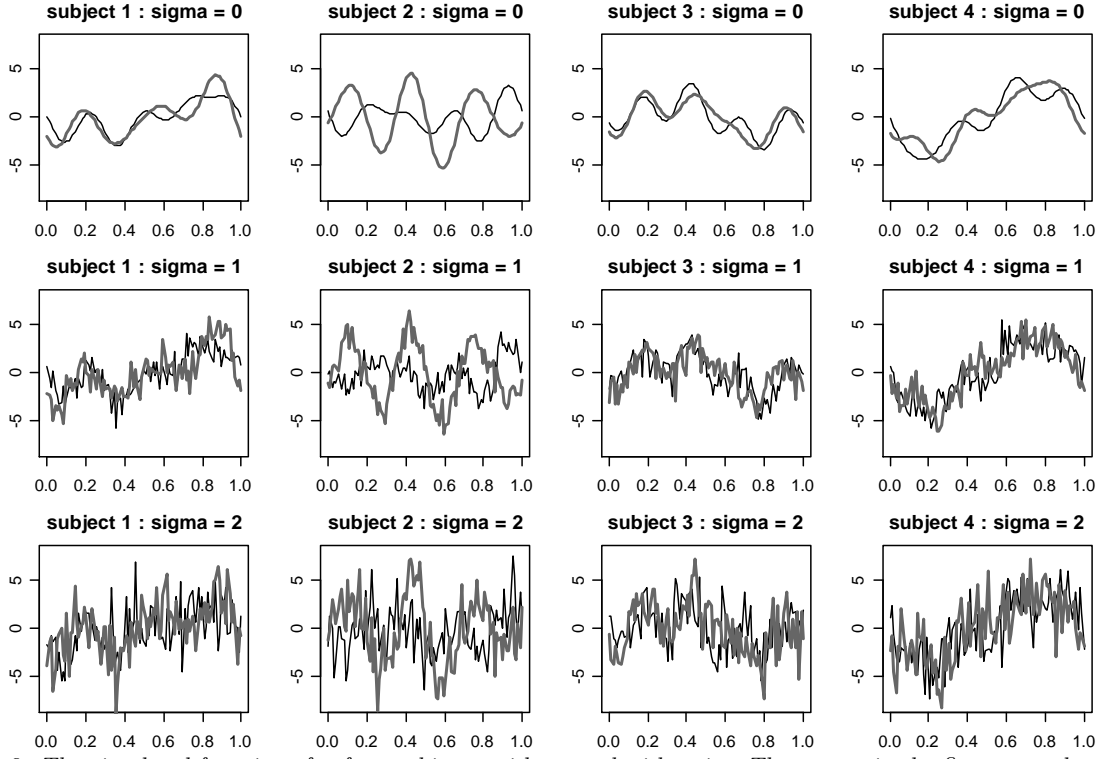Figure 12 displays estimated eigenfunctions from 20 randomly selected simulations. Simulated

FIG 9. *The simulated functions for four subjects: without and with noise. The curves in the first, second and third row are functions without noise, with noise $\sigma = 1$ and $\sigma = 2$, respectively. In each figure, thin gray lines represent curves at visit 1 and thick black lines represent curves at visit 2.*
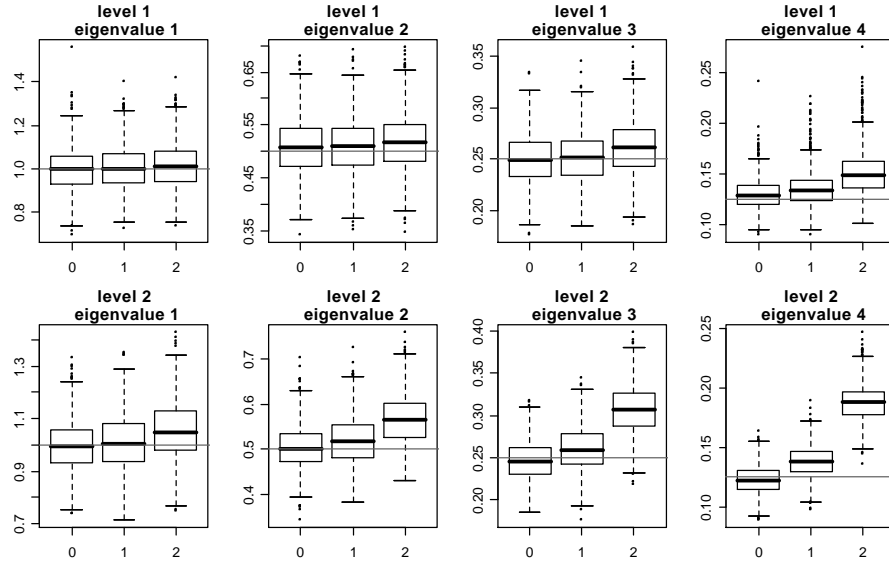

FIG 10. *Boxplots of estimated eigenvalues from MFPCA: the true functions are without and with noise. The gray lines are the true eigenvalues.*
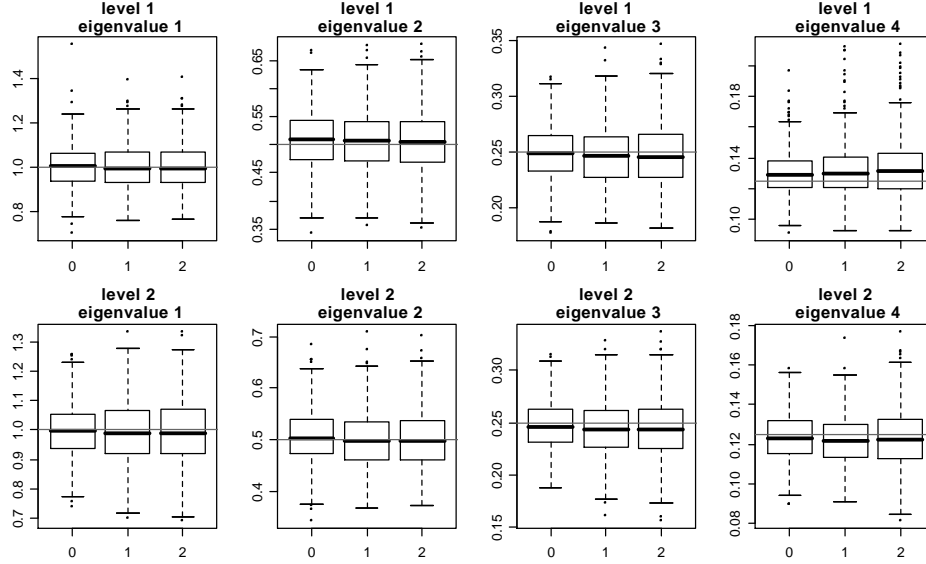
FIG 11. *Boxplots of estimated eigenvalues from the smooth MFPCA: the true functions are without and with noise. The gray lines are the true eigenvalues.*

data had no noise and we used the unsmooth version of our algorithm. Figure 13 shows results for the second eigenfunction at both levels corresponding to all three levels of noise and for both estimation methods. Results indicate that the estimation method successfully separates level 1 and 2 variation and correctly captures the shape of each individual eigenfunction. The smooth MFPCA algorithm provides smoother curves, with each individual curve approximating well the true shape of the function.

4. FPCA vs. MFPCA. In this section we explore the performance of standard FPCA when it is incorrectly used in the case when functions are clustered. Also, we investigate the performance of the MFPCA in the limit case when functions are not correlated.

To address the first problem we simulated functions with $\sigma = 0$. These clustered data were analyzed using standard FPCA, which ignored the within-cluster correlations. The top panel in Figure 14 displays the boxplots of the first four eigenvalue estimates in $1,000$ simulations. The bottom panel in Figure 14 displays 10 randomly selected eigenfunction estimates for each of the most important 4 eigenfunctions. The eigenvalues estimators are slightly biased for the first 4 eigenvalues and one may even argue that FPCA is reasonably good in this situation. However, a closer inspection of the bottom panels of Figure 14 reveals that the estimated eigenfunctions have large variability, probably due to the inability of the procedure to discriminate between level 1 and 2 eigenfunctions. For reference, these plots should be compared to the results of the
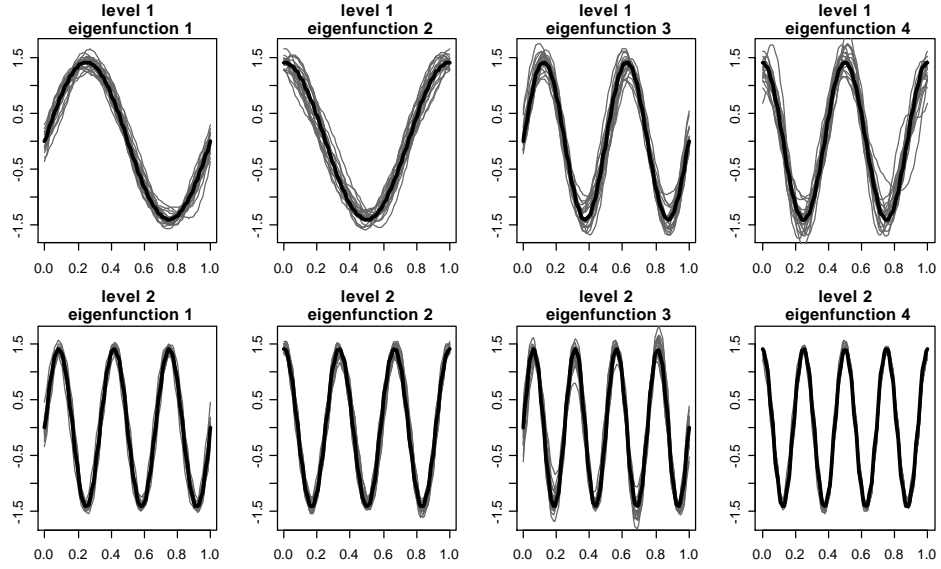
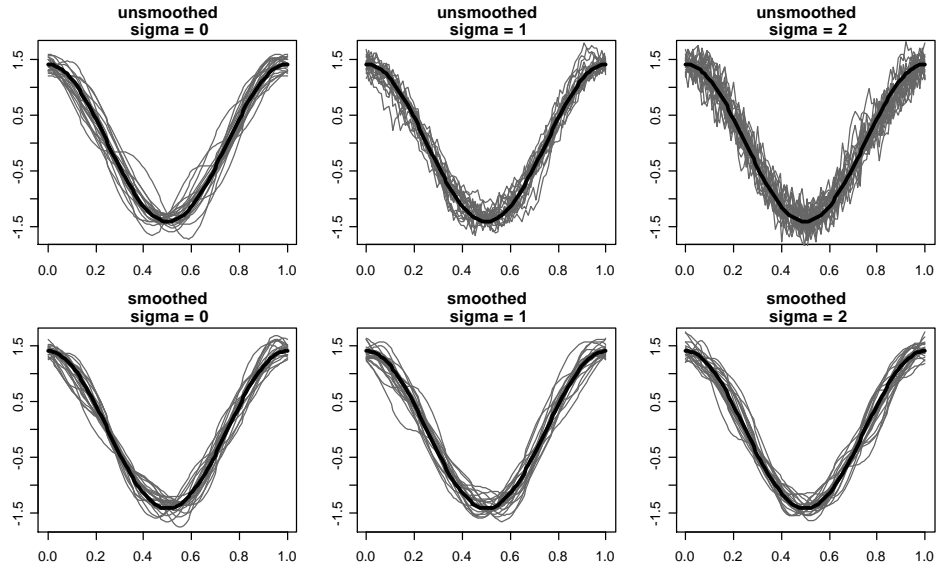FIG 12. *Estimated eigenfunctions when the functions are observed without noise, i.e, $\sigma = 0$.*

FIG 13. *Unsmoothed vs. smoothed MFPCA: the estimated level 1 eigenfunction 2 when $\sigma = 0, 1, 2$.*

MFPCA analysis shown in Figure 5. These results show that ignoring the functional hierarchy may mask the true signals and lead to incorrect conclusions.
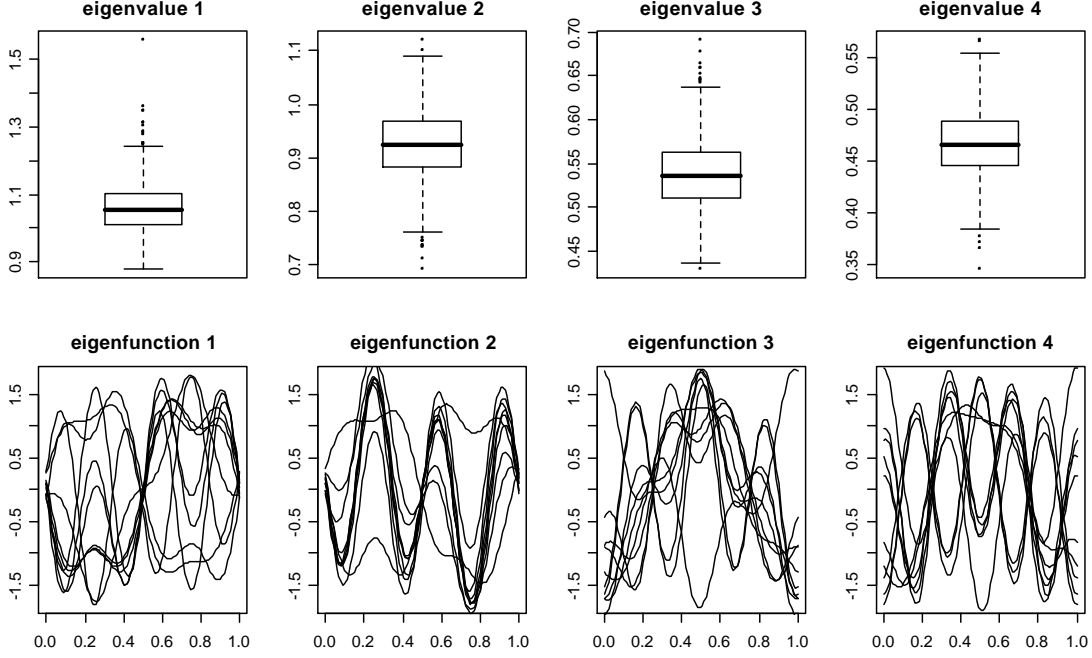

FIG 14. *Estimated eigenvalues and eigenfunctions using standard FPCA. The first row shows Boxplots of eigenvalues are from 1000 simulations. The second row displays estimated eigenfunctions from 10 randomly selected simulations.*

We also investigated the effects of using MFPCA when MFPCA is not needed, that is in the limit case when the within-subject correlation is zero. A minimal hope would be that MFPCA does not detect spurious level 1 signals and correctly estimates level 2 signals. To investigate this, in each of $1,000$ simulations, we simulate functions from model (13), by taking all level 1 eigenvalues to be 0, i.e, $\lambda_k^{(1)} = 0$, which corresponds to zero within-subject correlation. Simulated data are analyzed using MFPCA and some results are shown in Figure 15. The top left plot shows the boxplot of estimated proportion of variability explained by level 1 with a median of 4% and just a few outliers between 10 and 15%. The other three top panels show 10 randomly selected level 1 eigenfunction estimates for the largest three level 1 eigenvalues. While some patterns seem to be detected, the functional estimates are highly variable and resemble the pattern of level 2 eigenfunctions. In contrast, all level 2 eigenfunctions (bottom panels in Figure 15) are well estimated.
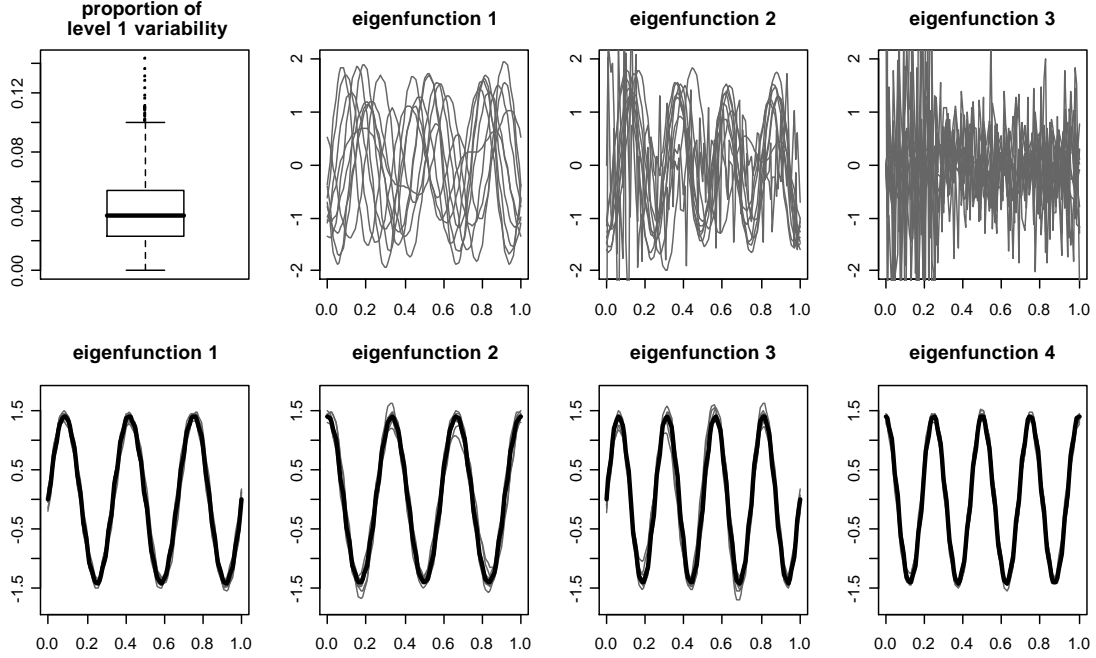
FIG 15. *Estimated eigenvalues and eigenfunctions using MFPCA, when the functions are independent. The first figure in the first row displays the boxplot of the estimated proportions of level 1 variability. The next three figures in the first row displays the three eigenfunctions at level 1 from 10 simulations. The figures in the second row show estimated level 2 eigenfunctions as well as true eigenfunctions.*

5. More results on the SHHS application. In this section, we present some additional results for the Sleep Heart Health Study (SHHS).

One might be interested in how well the MFPCA model fits the sleep data. Model checking for MFPCA is similar to that for standard FPCA. Typically, the first few dimensions capture most of the variation in the data, and finite dimensional truncations approximate raw data well. Figure 16 shows the fitted curve versus the raw data for the first two subjects, when we keep 3 dimensions at the subject level and 14 dimensions at the visit level.

To test the null hypothesis $H_0 : \rho_W = 0$ versus $H_A : \rho_W > 0$, we used a parametric bootstrap as follows. We fitted a model to the sleep data under H_0 and kept the first 14 level 2 eigenfunctions. Based on the estimated model we generated bootstrap samples and extracted eigenvalues and eigenfunctions. The upper panels in Figure 17 show the null distributions of $\hat{\rho}_W$, $\hat{N}_1$ and $\hat{N}_2$ based on 1000 bootstrap samples using the thresholds described in the paper. The 95% confidence interval for ρ is $[0.011, 0.024]$, which does not contain 0.213, indicating that there is strong evidence of within-subject correlation, even though the two visits are 5 years apart. We also conclude that the method overestimates the number of eigenvectors at level 1 and that the
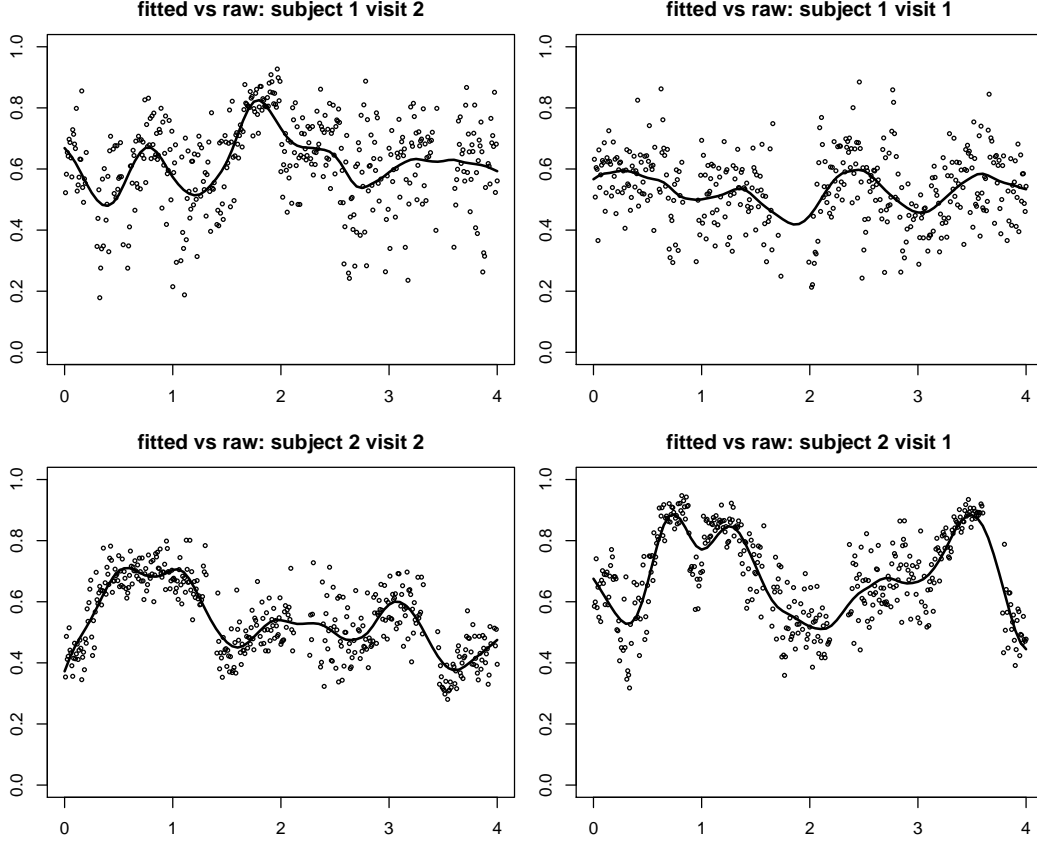
FIG 16. *Fitted versus raw functions for the first two subjects.*

number of dimensions at level 2 is well estimated. The lower panels of Figure 17 display results from a parametric bootstrap under the alternative hypothesis $H_1 : \rho_W \neq 0$. Data are simulated using 7 true level 1 and 14 level 2 eigenvalues and eigenfunctions estimated from SHHS. Under H_1, the 95% confidence interval for ρ is $[0.210, 0.236]$, covering the estimated 0.213 for the sleep data. The procedure estimated the number of dimensions on both levels well.

6. Technical details on the variance of residuals from the projection model. By definition, $\epsilon_{ijk}^{(1)} = \int_t \{ \sum_{l=N_2+1}^{\infty} \zeta_{ijl} \phi_l^{(2)}(t) + \varepsilon_{ij}(t) \} \phi_k^{(1)}(t) \, dt$ and $\epsilon_{ijl}^{(2)} = \int_t \{ \sum_{k=N_1+1}^{\infty} \xi_{ik} \phi_k^{(1)}(t) + \varepsilon_{ij}(t) \} \phi_l^{(2)}(t) \, dt$. Here are the main results and proofs.

Proposition 1 For the projection model, the following results hold:

(i) $\displaystyle\sum_{l=1}^{\infty} c_{kl}^2 = 1, \sum_{k=1}^{\infty} c_{kl}^2 = 1, \sum_{l=1}^{\infty} | c_{kl} \, c_{k'l} | \leq 1;$
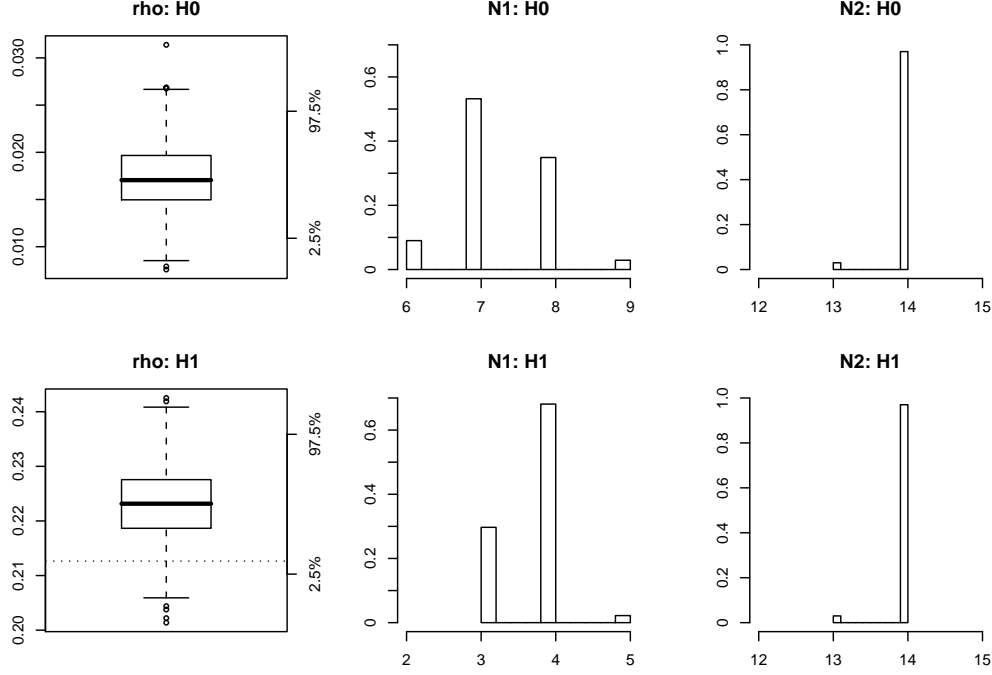
FIG 17. *Estimated within subject correlation and number of dimensions from simulations under $H_0 : \rho_W = 0$ and $H_1 : \rho_W > 0$.*

(ii) Suppose all functions are measured at common grid points $\{t_m = \frac{m-1}{T-1} : m = 1, 2, ..., T\}$,

$$\text{var}\left(\int \varepsilon_{ij}(t)\phi_k^{(1)}(t)dt\right) \approx \text{var}\left(\frac{1}{T}\sum_{r=1}^{T}\varepsilon_{ij}(t_r)\phi_k^{(1)}(t_r)\right) \approx \frac{\sigma^2}{T}$$

$$\text{cov}\left(\int \varepsilon_{ij}(t)\phi_k^{(1)}(t)dt, \int \varepsilon_{ij}(t)\phi_{k'}^{(1)}(t)dt\right) \approx \text{cov}\left(\frac{1}{T}\sum_{r=1}^{T}\varepsilon_{ij}(t_r)\phi_k^{(1)}(t_r), \frac{1}{T}\sum_{r=1}^{T}\varepsilon_{ij}(t_r)\phi_{k'}^{(1)}(t_r)\right) \approx 0$$

Proof. (i). Note that both level 1 and level 2 eigenfunctions form orthonormal bases for the functional space. Thus, level 2 eigenfunctions $\phi_l^{(2)}(t)$ can be expanded as $\sum_k c_{kl}\phi_k^{(1)}(t)$. Thus,

$$1 = \|\phi_l^{(2)}(t)\|^2 = \int \left\{\phi_l^{(2)}(t)\right\}^2 dt = \sum_k c_{kl}^2 \|\phi_k^{(1)}(t)\|^2 = \sum_k c_{kl}^2.$$

Similarly, $\sum_k c_{kl}^2 = 1$. The result $\sum_{l=1}^{\infty} |c_{kl} c_{k'l}| \leq 1$ is a consequence of the Cauchy-Schwartz inequality. The equation holds exactly for the infinite sum, but only approximately for the truncated series. However, if the true functional space is finite dimensional, this may not hold. Indeed, in Case 1 of our simulation study $c_{kl} = 0$ for $k, l \in \{1, 2, 3, 4\}$.

(ii). These two results can be obtained by straightforward calculations.

Proposition 2 Using the results from Proposition 1 and the fact that $\lambda_1^{(2)} \geq \lambda_2^{(2)} \geq \lambda_3^{(2)} \geq \cdots$, one could show the following:

(i) the variance of $\epsilon_{ijk}^{(1)}$ satisfies

$$\frac{\sigma^2}{T} \leq \text{var}\left(\epsilon_{ijk}^{(1)}\right) = \sum_{l=N_2+1}^{\infty} c_{kl}^2 \lambda_l^{(2)} + \text{var}\left(\int \varepsilon_{ij}(t)\phi_k^{(1)}(t)dt\right)$$

$$\leq \lambda_{N_2+1}^{(2)} + \frac{\sigma^2}{T};$$

(ii) the covariance between $\epsilon_{ijk}^{(1)}$ and $\epsilon_{ijk'}^{(1)}$ satisfies

$$\left|\text{cov}\left(\epsilon_{ijk}^{(1)}, \epsilon_{ijk'}^{(1)}\right)\right| = \left|\sum_{l=N_2+1}^{\infty} c_{kl}\, c_{k'l}\, \lambda_l^{(2)} + \text{cov}\left(\int \varepsilon_{ij}(t)\phi_k^{(1)}(t)dt, \int \varepsilon_{ij}(t)\phi_{k'}^{(1)}(t)dt\right)\right|$$

$$\leq \lambda_{N_2+1}^{(2)};$$

(iii) the correlation between $\epsilon_{ijk}^{(1)}$ and $\epsilon_{ijk'}^{(1)}$ satisfies

$$\left|\text{cor}\left(\epsilon_{ijk}^{(1)}, \epsilon_{ijk'}^{(1)}\right)\right| \approx \frac{\left|\sum_{l=N_2+1}^{\infty} c_{kl}\, c_{k'l}\, \lambda_l^{(2)}\right|}{\lambda_{N_2+1}^{(2)} + \sigma^2/T}$$

$$\leq \frac{\lambda_{N_2+1}^{(2)}}{\sigma^2/T};$$

The variance and covariance of $\epsilon_{ijk}^{(2)}$ could be obtained similarly.