Multivariate Functional Additive Mixed Models

Appendix

Alexander Volkmann, Almond Stöcker, Fabian Scheipl, Sonja Greven

A Derivation of Variance Decompositions

This is the derivation of the variance decompositions (3.4) and (3.5) from the model equations (2.1) and (2.2). Following the argumentation of Cederbaum et al. (2016), the variation of the response can be decomposed per dimension using iterated expectations as

$$\begin{aligned}
\int_{\mathcal{I}} \text{Var}\big(y_i^{(d)}(t)\big)dt =& \sum_{j=1}^{q}\sum_{v=1}^{V_{U_j}} z_{ijv}^2 \int_{\mathcal{I}} \text{Var}\big(U_{jv}^{(d)}(t)\big)dt + \int_{\mathcal{I}} \text{Var}\big(E_i^{(d)}(t)\big)dt + \int_{\mathcal{I}} \text{Var}\big(\epsilon_{it}^{(d)}\big)dt \\
=& \sum_{j=1}^{q}\sum_{m=1}^{\infty} \nu_{U_jm} \underbrace{\int_{\mathcal{I}} \psi_{U_jm}^{(d)}(t)\psi_{U_jm}^{(d)}(t)dt}_{=||\psi_{U_jm}^{(d)}||^2} \\
&+ \sum_{m=1}^{\infty} \nu_{Em} \underbrace{\int_{\mathcal{I}} \psi_{Em}^{(d)}(t)\psi_{Em}^{(d)}(t)dt}_{=||\psi_{Em}^{(d)}||^2} + \int_{\mathcal{I}} \sigma_d^2 dt \\
=& \sum_{j=1}^{q}\sum_{m=1}^{\infty} \nu_{U_jm}||\psi_{U_jm}^{(d)}||^2 + \sum_{m=1}^{\infty} \nu_{Em}||\psi_{Em}^{(d)}||^2 + \sigma_d^2|\mathcal{I}|
\end{aligned}$$

assuming in the second equality that each observation is in exactly one group in the jth grouping layer, i.e. exactly one indicator $z_{ijv} = z_{ijv}^2, v = 1, \ldots, V_{U_j}$ is one for each i and j, and also using in the second equality that the scores are uncorrelated across eigenfunctions, in particular $\text{Cov}(\rho_{U_jvm}, \rho_{U_jvm'}) = 0$ if $m \neq m'$.

Similarly, the total variance as measured by the sum of the univariate variances can be decomposed

as

$$
\begin{aligned}
\mathbb{E}\left(|||\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{x}_i)|||^2\right) &= \sum_{d=1}^{D} w_d \int_{\mathcal{I}} \mathrm{Var}\big(y_i^{(d)}(t)\big) dt \\
&= \sum_{j=1}^{q} \sum_{d=1}^{D} \sum_{v=1}^{V_{U_j}} z_{ijv}^2 w_d \int_{\mathcal{I}} \mathrm{Var}\big(U_{jv}^{(d)}(t)\big) dt + \sum_{d=1}^{D} w_d \int_{\mathcal{I}} \mathrm{Var}\big(E_i^{(d)}(t)\big) dt \\
&\quad + \sum_{d=1}^{D} w_d \int_{\mathcal{I}} \mathrm{Var}\big(\epsilon_{it}^{(d)}\big) dt \\
&= \sum_{j=1}^{q} \sum_{m=1}^{\infty} \nu_{U_j m} \underbrace{\sum_{d=1}^{D} w_d \int_{\mathcal{I}} \psi_{U_j m}^{(d)}(t) \psi_{U_j m}^{(d)}(t) dt}_{= |||\boldsymbol{\psi}_{U_j m}|||^2 = 1} \\
&\quad + \sum_{m=1}^{\infty} \nu_{Em} \underbrace{\sum_{d=1}^{D} w_d \int_{\mathcal{I}} \psi_{Em}^{(d)}(t) \psi_{Em}^{(d)}(t) dt}_{= |||\boldsymbol{\psi}_{Em}|||^2 = 1} + \sum_{d=1}^{D} w_d \int_{\mathcal{I}} \sigma_d^2 dt \\
&= \sum_{j=1}^{q} \sum_{m=1}^{\infty} \nu_{U_j m} + \sum_{m=1}^{\infty} \nu_{Em} + \sum_{d=1}^{D} w_d \sigma_d^2 |\mathcal{I}|.
\end{aligned}
$$

B Additional Results for the Snooker Data

B.1 Data Description

In this section, we provide additional information of the study by Enghofer (2014). The participants of the snooker training study volunteered if they wanted to take part in the training group. The training schedule for this treatment group recommended to perform a training two to three times a week which consisted of three exercises that are supposed to improve snooker specific muscular coordination. Note that these were autonomous trainings so that there is no reliable information if the participants followed the recommendations.

In order to analyse the snooker shot of maximal force, the participants were asked to place themselves in the typical playing position. Yet no exact measure of the distance between the tip of their cue and the cue ball is available and thus the time of the impulse onto the cue ball cannot be exactly determined from the data. The participants were videotaped and the open source software Kinovea (`https://www.kinovea.org`) was used to manually locate points of interest (a participant's shoulder, elbow, and hand) and track them on a two-dimensional grid over the course of the video. Figure 1 shows the univariate functions that make up the two-dimensional trajectories of the participants' elbow, hand, and shoulder. The univariate functions for the x-axis of the hand show that for most observed curves, the time of impact might lie around 0.75 of the relative time of the snooker shot. That is when the hand moves into the positive range on the x dimension. For their shot, the snooker players can either try to fix their elbow or move the elbow dynamically. In order to move the hand in a straight line (piston stroke), the elbow has to move dynamically down and up when the hand is drawn back and accelerated towards the cue ball. The latter part of the movement trajectory (after the cue hits the cue ball) does not impact the snooker shot itself. Note that a long final downwards movement of an elbow trajectory indicates that the hand is not stopped at the chest but is rather pulled through towards the shoulder. This can give insight into a player's stance at the snooker table, e.g. if the upper body is close to the table or if it is more erect.

Note that the univariate functions of Figure 1 show different characteristics over the different dimensions, and a first-order difference penalty penalizing deviations from a constant function seems to be a sensible default choice.

B.2 Data Preprocessing

In order to account for differences in body height between the participants, we first rescale the observed coordinate locations by the median distance between hand and elbow. As we are interested in the movement trajectory irrespective of phase variation, i.e. independent of the exact timing of different parts of the stroke, we apply an elastic registration approach to the data which aligns the two dimensional trajectory of the hand to its respective mean trajectory across all players and shots (Steyer et al.; 2021). We then reparameterize the time for the elbow and shoulder trajectories according to the results of the hand alignment. Other registration approaches are possible and we provide both the original and the reparameterized time of the data. Due to the high frame rate of the high speed camera and a

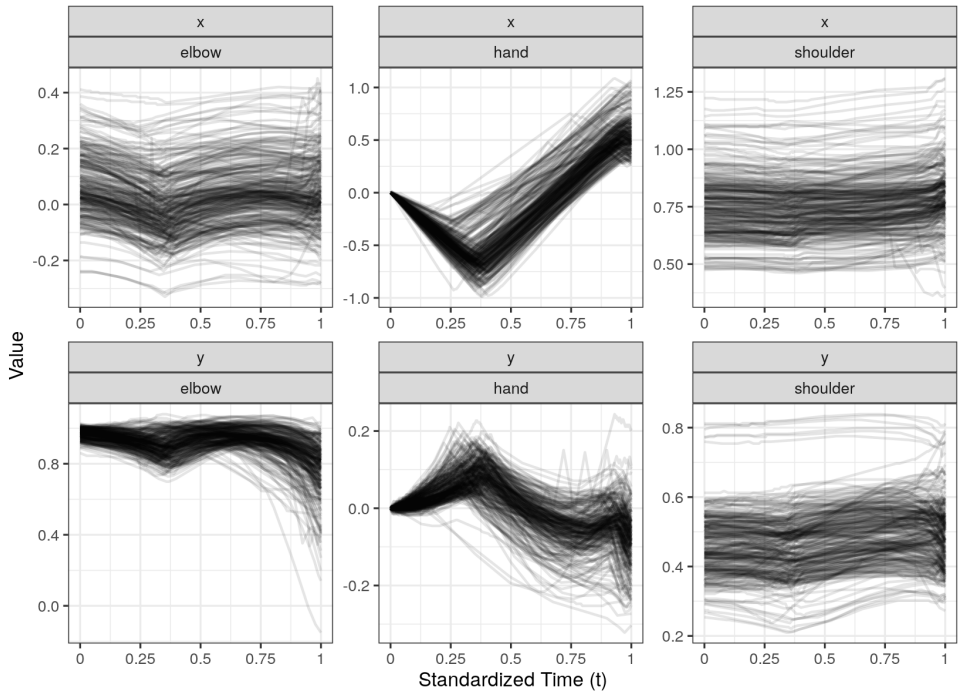

Figure 1: Univariate functions of the snooker trajectories as functions of the standardized time t. The value of the function corresponds to the position of the elbow, hand, and shoulder on the x and y axes at the associated time point.

comparatively rough grid of the tracking software, the resulting multivariate functional data are dense with redundant information. The following subsection describes the coarsening method in detail. The coarsening reduces the data set from roughly 400,000 to only 56,910 scalar observations. Note that 5 functional observations are lost due to technical problems during the recording. Even bivariate trajectories within one multivariate functional observation can have different numbers of scalar observations due to inconsistencies in the tracking programme.

Coarsening Method for the Snooker Data

Functional response variables as given by the snooker trajectories typically exhibit very high autocorrelation. Thus, up to a certain point, removing single curve measurements will inflict only very limited information loss while reducing computational cost in the model fit at least linearly. We therefore coarsen the snooker trajectories as a preprocessing step, as they were originally sampled "overly" densely in that sense. In particular, there are also evaluation points that do not carry additional information as they measure the same location as the evaluation points right before and after (e.g. because the hand does not move noticeably over these three or more time points). The coarsening should a) retain original data points via subsampling to avoid pre-smoothing losses and b) be optimal in the sense that as much information as possible is preserved given the subsample size.

Finding an optimal subsample, with respect to a general criterion and for all possible subsamples, is however typically very computationally demanding in itself. We therefore propose a fast greedy coarsening algorithm recursively discarding single points from the sample. The ultimate aim of the algorithm is to

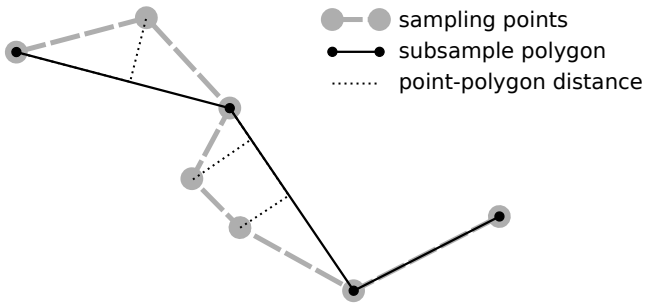

Figure 2: An example subsample polygon of length four through some curve measurements. The polygon canonically connects points in the subsample according to the order of the indices t_i of the curve measurements $y(t_i)$ in the subsample.

find a subsample polygon (Fig. 2) representing the whole data as well as possible.

One step of the coarsening algorithm described in the following is illustrated in Figure 3. Let $\mathcal{T}_0 = \{t_1, \ldots, t_m\} \subset \mathcal{I}$ be the set of evaluation points $t_1 < \cdots < t_m$ of a curve $y : \mathcal{I} \to \mathbb{R}^2$. We select a sequence of subsamples $\mathcal{T}_0 \supset \mathcal{T}_1 \supset \mathcal{T}_2 \supset \ldots$ by recursively discarding the point t from $\mathcal{T}_k$ where $y(t)$ is closest to the line segment $[y(t_-), y(t_+)]$ between its adjacent points at $t_- = \max\{s \in \mathcal{T}_k : s < t\}$ and $t_+ = \min\{s \in \mathcal{T}_k : s > t\}$. Note that this is only well-defined for $t \in \mathring{\mathcal{T}}_k = \mathcal{T}_k \setminus \{\min \mathcal{T}_k, \max \mathcal{T}_k\}$, in the "interior" of $\mathcal{T}_k$, and the first and last point t_1 and t_m are kept throughout the algorithm.

In more detail, we proceed in the kth step of the algorithm as follows

1. $\forall t \in \mathring{\mathcal{T}}_k$ compute the step-to-step quadratic error

$$\Delta^{(k)}(t) = \min_{p \in [y(t_-), y(t_+)]} (y(t), p)^2.$$

   For computing the quadratic distance $\Delta^{(k)}(t)$, we have to distinguish the case where the projection of $y(t)$ on the line through $y(t_-)$ and $y(t_+)$ lies between the two points and the case where it lies on either side of them. To do so, define $y'(t) = y(t) - y(t_-)$ and $y'_+ = y'(t_+)$, as well as the unit vector $u = \frac{y'_+}{\|y'_+\|}$ pointing from $y(t_-)$ to $y(t_+)$. Then we can compute

$$\Delta^{(k)}(t) = \begin{cases} \|y'(t)\|^2 & \text{if } \langle y'(t), u\rangle \leq 0 \\ \|y'(t) - \langle y'(t), u\rangle u\|^2 & \text{if } \|y'_+\| \geq \langle y'(t), u\rangle > 0 \\ \|y'(t) - y'_+\|^2 & \text{if } \langle y'(t), u\rangle > \|y'_+\| \end{cases}.$$

2. Choose $t^{(k)} \in \operatorname{argmin}_{t \in \mathring{\mathcal{T}}_k} \Delta^{(k)}(t)$ and set $\mathcal{T}_{k+1} = \mathcal{T}_k \setminus \{t^{(k)}\}$.

This procedure may be repeated until $\mathcal{T}_{k+1}$ is of the desired sample size. However, we consider an error based stopping criterion more convenient in most cases. We suggest to specify a threshold for the cumulative step-to-step error $S_k = \sum_{\kappa=1}^{k} \Delta^{(\kappa)}(t^{(\kappa)})$ or a relative version $R_k = S_k/\bar{\Delta}$ of it, relative to $\bar{\Delta} = \frac{1}{m-2} \sum_{\mathring{\mathcal{T}}} \Delta^{(\infty)}(t)$ the mean quadratic distance to the line segment $[y(t_1), y(t_m)]$ between the two only points left at the ultimate subsample $\mathcal{T}_\infty$.

In our experience, a threshold $R^* = 10^{-4}$ for R_k yields a close to lossless subsample from a visual point of view. For the model, $R^* = 0.003$ was chosen to limit computational demands. While the multivariate

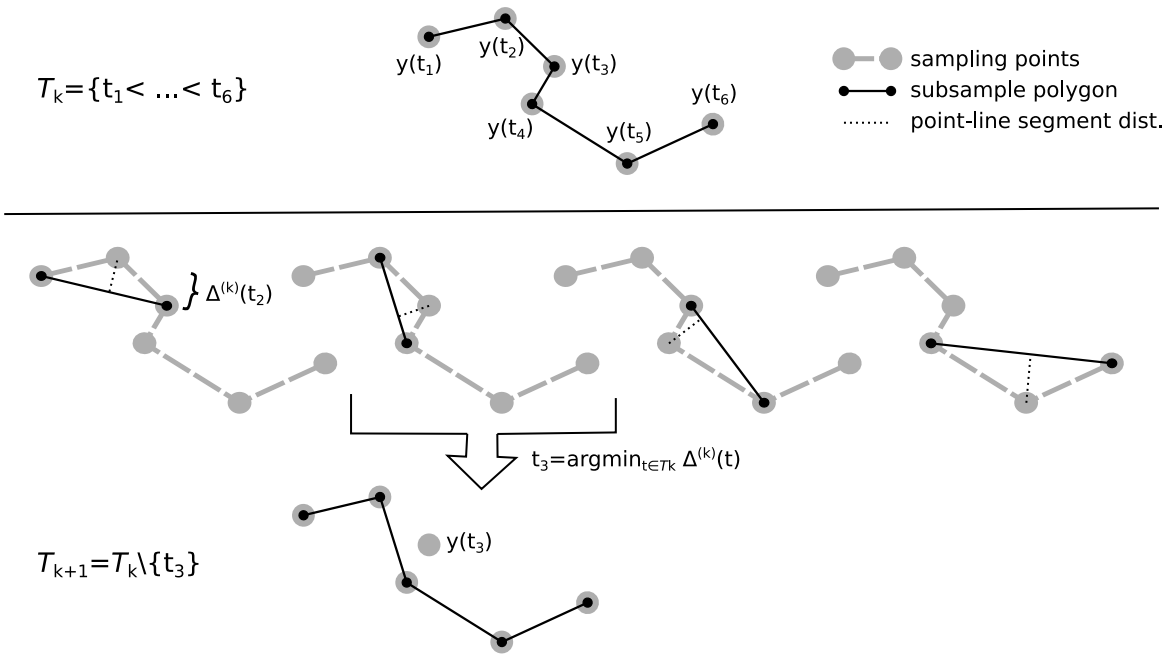

Figure 3: One iteration of the proposed fast greedy coarsening algorithm: Given the current subsample polygon (top), the loss $\Delta^{(k)}(t)$ inflicted when approximating $y(t)$ by the line segment between its neighbors is computed for each index $t_2, \ldots, t_5$ (middle). The data point with the least approximation loss is discarded (bottom). Note that $\Delta^{(k+1)}(t)$ has to be re-computed only for the two adjacent points. All others can be passed forward.

functional additive mixed model (multiFAMM) would allow for dimension specific evaluations of the curves, the warping algorithm applied as part of the preprocessing does not. Thus, we decided to apply the coarsening algorithm to the hand trajectories, which we consider the most informative component with the best signal to noise ratio. Unselected observed time points for each hand trajectory are then also dropped for the corresponding trajectories of the shoulder and elbow. As a result, the coarsened data retain the characteristic of the snooker trajectory data such that for a given time point, we observe the location of (almost) all points of interest (i.e. the grid of evaluation points is identical over the dimensions for a given observation).

B.3 Additional Results of the Analysis

Analysis of the Variance Components

The multiFAMM estimates a total of 61 univariate functional principal components (FPCs) (20 each for B and C, and 21 for E), where each process is represented by three to five univariate FPCs on each dimension. Table 1 shows that the three points of interest (hand, elbow, and shoulder) all show a similar amount of variation. The source of variation, however, differs in interpretation. While the variation in the hand trajectories stems from differing movements, the variation in the shoulder mainly reflects different positioning. Applying higher weights to dimensions associated with the hand, for example, can shift the focus of the multivariate functional principal component analysis (MFPCA) to favor movement based variation of the hand trajectories in the analysis. From Table 1 it is also apparent that movements or positions along the horizontal x-axis contribute more to the variation in the data than movements or positions along the vertical y-axis. This also suggests that variation in the x direction is the driving factor of the MFPCA. By assigning higher weights in the scalar product to dimensions associated with the y-axis, the analyst can "distort" the natural grid of observations to balance out the different variation in the axes. Note that we use an unweighted scalar product in the analysis. Also keep in mind that the variation in Table 1 is calculated based on estimation step 1.

Based on the cut-off criterion (3.4), 16 multivariate FPCs are chosen to explain 95% of total variation. For a similar amount of variance explained, 23 univariate FPCs would be needed, which would give a more complex model that contains redundancies and ignores the multivariate nature of the data. Figures 4 and 5 show i.a. the two most prominant modes of variation in the snooker training data. As described in the main part, the FPC $\boldsymbol{\psi}_{C1}$ seems to represent variation in the starting position (explaining about 27% of total variation. The second leading FPC $\boldsymbol{\psi}_{B1}$ (explaining about 15% of total variation) shows subject-specific variation related to the two snooker techniques identified by Enghofer (2014): The red trajectory (+) shows that the elbow moves first down and then up to draw the hand back and accelerate it towards the cue ball. This then vertically contracts the hand trajectory compared to fixing the elbow during the acceleration phase (blue −), which allows the hand to swing more freely (pendulum stroke). We also find that players with a personal tendency towards the pendulum stroke seem to not stop their hands at their chest. Note that for $\boldsymbol{\psi}_{B1}$ in Figure 4, the red trajectory is overlapping and thus masks the down-up-and-down movement of the elbow.

Table 1: Total variation of the centred responses per univariate dimension and overall. The variation is calculated as the sum of non-negative univariate eigenvalues and the dimension-specific measurement error variance as given from Step 1 in the estimation of the multiFAMM as described in Section 3.1

	elbow.x	elbow.y	hand.x	hand.y	shoulder.x	shoulder.y	Total
Variation	0.012	0.004	0.015	0.001	0.014	0.008	0.055
Proportion	22%	7%	28%	2%	26%	14%	100%

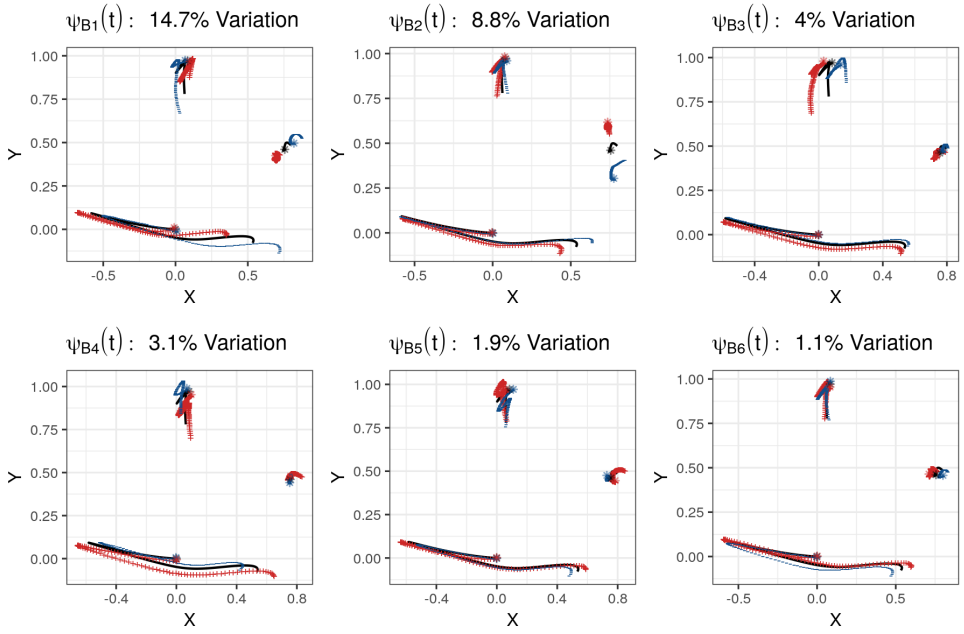

Figure 4: FPCs for the subject-specific functional random effect $\boldsymbol{B}_i(t)$ with the respective proportion of variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{B.}}$) of the FPC is added (red +) and subtracted (blue −). The start of the trajectories are marked with an asterisk.

Figure 4 shows the multivariate FPCs of the subject-specific functional random intercept included in the multiFAMM. $\boldsymbol{\psi}_{B2}(t)$ (explaining about 9% of the total variation) seems to capture variation in the positioning of the shoulder, size differences of the upper arm, and the final part of the movement trajectories after the cue ball is hit. Figure 5 shows the multivariate FPCs of the subject-and-session-specific random intercept. The second FPC (explaining about 6% of the total variation) shifts the relative positioning of the elbow and shoulder. Figure 6 shows the multivariate FPCs of the curve-specific random intercept. The leading FPC (explaining about 7% of total variation) captures variation in the starting position of elbow and shoulder, and in the final hand movement.

Analysis of the Estimated Effect Functions

The left panel of Figure 7 shows the estimated covariate effect functions for the covariate `skill`. In addition to the effect described in the main part, we point out that for skilled players, the shoulder

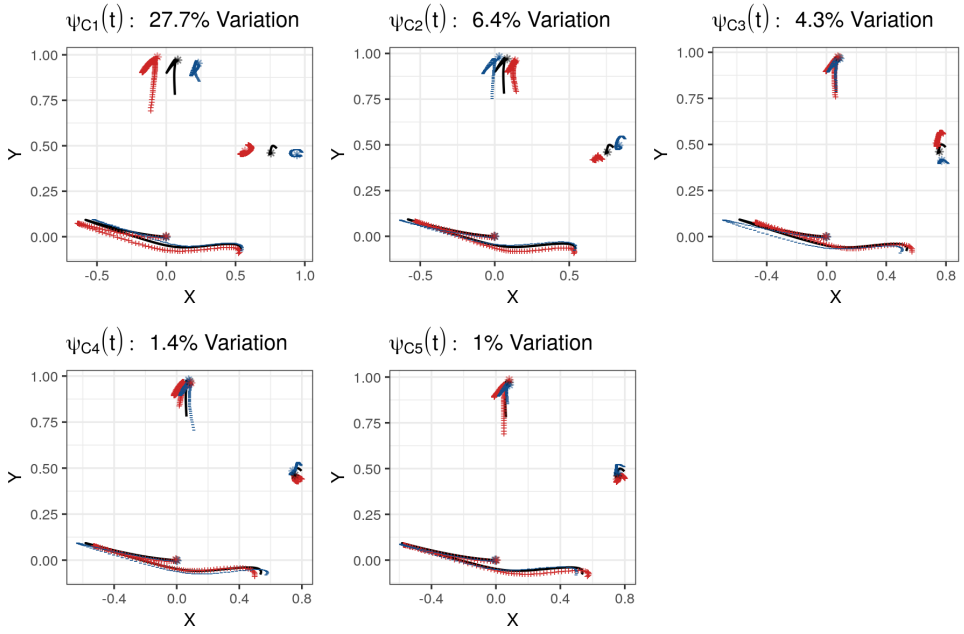

Figure 5: FPCs for the subject-and-session-specific functional random effect $\boldsymbol{C}_{ij}(t)$ with the respective proportion of variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{C\cdot}}$) of the FPC is added (red +) and subtracted (blue −). The start of the trajectories are marked with an asterisk.

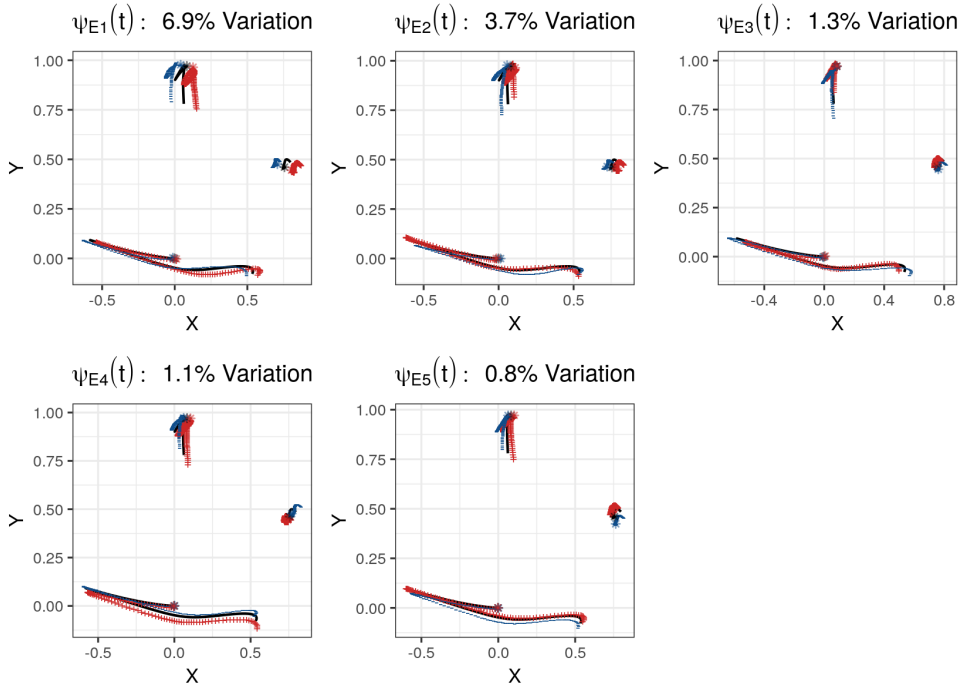

Figure 6: FPCs for the curve-specific functional random effect $\boldsymbol{E}_{ijh}(t)$ with the respective proportion of variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{E\cdot}}$) of the FPC is added (red +) and subtracted (blue −). The start of the trajectories are marked with an asterisk.

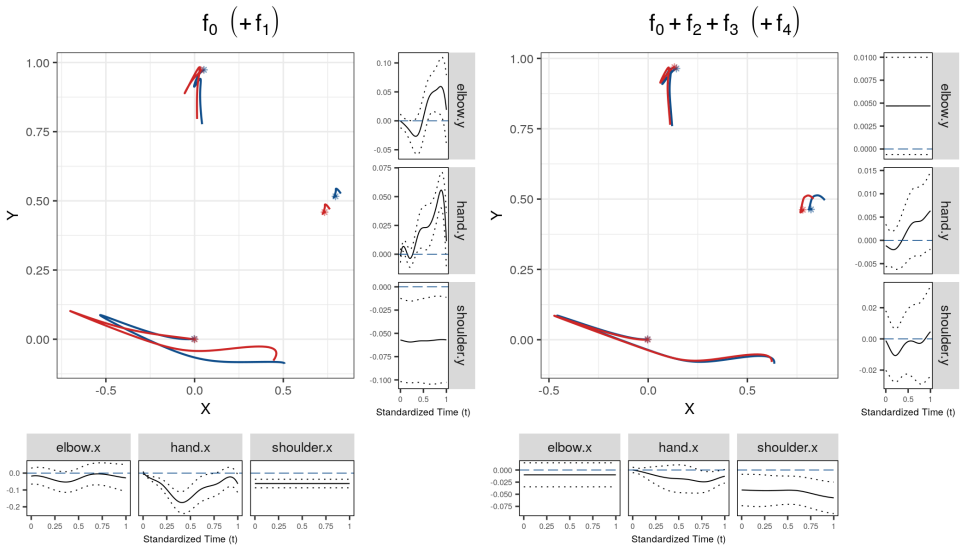

Figure 7: Estimated covariate effect functions for skill (left) and treatment effect (right). The central plot shows the effect of the coefficient function in parenthesis in the title (red) on the two dimensional trajectories for the reference (without paranthesis, blue). The start of the trajectories are marked with an asterisk. The marginal plots show the estimated univariate effect functions (black) with pointwise 95% confidence intervals (dotted) and the reference (blue dashed).

is positioned closer to the table as well as to the body all other things being equal. The movement trajectories in the right panel of Figure 7 are composed of the estimated effect functions of intercept, treatment group, and session (blue trajectories), so that the interaction effect can be separated and interpreted (red trajectories). We do not find strong evidence for differences in the displayed mean trajectories, nor in the univariate effects (marginal plots). This suggests that the training programme did not considerably change the participants' mean movement trajectories.

Figure 8 shows the estimated intercept (left) and effect functions of the covariates for treatment group (center) and session (right). The intercept (scalar plus functional) gives the mean movement trajectories (dark blue) in the reference group, i.e. an unskilled snooker player in the control group with a shot in the first session. In the middle panel, the red trajectories show the estimated effect of the treatment group added to the intercept. The marginal plots around the movement trajectories show the estimated univariate effects and their pointwise 95% confidence intervals. We find only minor differences between movement trajectories in treatment and control group. The hand trajectories of the treatment group seem to be slightly higher and further along the x-axis than the control group, all other things being equal. The right panel compares the mean trajectory for the reference group of players in the first (blue) to the second (red) session. The univariate estimated covariate effects on the y-axis seem to indicate a slight shift in the vertical direction of the trajectories. Keep in mind that the trajectories of the hand have been centred to the origin before the analysis.

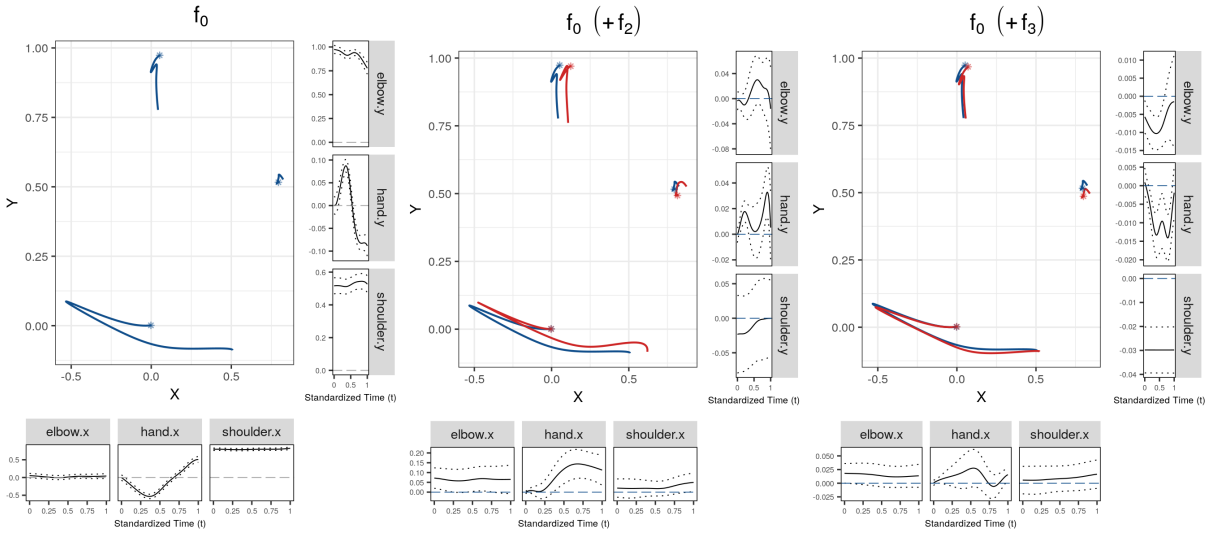

Figure 8: Estimated functional fixed effects of intercept (left), covariate treatment group (center), and covariate session (right). The main plots show the effect of the coefficient function in parenthesis (red) on the two dimensional trajectories for the intercept (blue). The start of the trajectories are marked with an asterisk. The marginal plots show the estimated univariate effect functions (black) with pointwise 95% confidence intervals (dotted) and the reference (grey/blue dashed).

Model Diagnostics and Sensitivity Analysis

We use the univariate root relative mean squared error (urrMSE) defined in Section 5 as a criterion for model fit between the different dimensions. Table 2 shows that the model fits comparatively well on all dimensions except for the x-axis of the elbow and the y-axis of the hand. In order to get an impression of the quality of the model fit, Figure 9 shows selected fitted trajectories in solid red together with the observed trajectories in dashed grey. We choose to present the quantiles from the contribution of each observation to the multivariate root relative mean squared error (mrrMSE), i.e. quantiles from $|||\boldsymbol{\zeta}_i - \hat{\boldsymbol{\zeta}}_i|||^2$ of the multivariate functional trajectory $\boldsymbol{\zeta}_i, i = 1, ..., N$ and the corresponding estimate $\hat{\boldsymbol{\zeta}}_i$ (see Section 5).

The presented fitted trajectories suggest that the estimates might not always overlap with the observed trajectories, which suggests residual structure in the model residuals. The top panels of Figure 10 plot the scalar residuals over time for each of the dimensions. Especially on the y-axis, we can discern patterns in the residuals even though this type of graphic is prone to overplotting. In order to investigate residual autocorrelation, we apply the following ad-hoc method, which allows us to approximate the well-known concept of an autocorrelation function in time series analysis. We first use fast symmetric additive covariance smoothing (Cederbaum et al.; 2018) to estimate a smoothed correlation matrix for the residuals. Then we calculate the means of the off-diagonals, which corresponds to an approximated mean autocorrelation for a given time lag. Figure 11 shows the results based on the residuals for each dimension separately. Overall, we find that the model residuals (red) show clear signs of autocorrelation, in particular up to a time lag of 0.25. However, compared to the autocorrelation of the original data (solid grey line), we see a considerable reduction, which can be attributed to the functional random effects in

Table 2: urrMSE values for the model fit on the snooker training data.

	elbow.x	elbow.y	hand.x	hand.y	shoulder.x	shoulder.y
Main Model	0.172	0.028	0.088	0.370	0.025	0.039
Sensitivity Analysis	0.121	0.019	0.084	0.233	0.019	0.025

Table 3: Proportion of predictor variance for each dimension.

	$f_0^{(d)}(t)$	$f_1^{(d)}(t)$	$f_2^{(d)}(t)$	$f_3^{(d)}(t)$	$f_4^{(d)}(t)$	$B_i^{(d)}(t)$	$C_{ij}^{(d)}(t)$	$E_{ijh}^{(d)}(t)$
elbow.x	0.018	0.020	0.053	0.003	0.001	0.210	0.120	0.574
elbow.y	0.381	0.048	0.006	0.002	0.000	0.144	0.261	0.157
hand.x	0.862	0.022	0.014	0.000	0.000	0.022	0.050	0.030
hand.y	0.755	0.037	0.012	0.004	0.000	0.103	0.076	0.012
shoulder.x	0.007	0.047	0.013	0.001	0.019	0.184	0.144	0.585
shoulder.y	0.009	0.068	0.007	0.018	0.001	0.105	0.512	0.280

the model.

Table 3 also underlines the importance of the random effects in modeling the snooker trajectories. We use the predictor variance as an indicator for quantifying the effect of the fixed and random effects on the model fit. Separately on each dimension, we calculate the variance of the partial predictors on the data and give the respective proportions in Table 3. We find that with exception of the reference mean ($\boldsymbol{f}_0(t)$), the partial predictors of the fixed effects contribute generally little to the overall predictor variance (highest proportions for $\boldsymbol{f}_1(t)$ with around 5% on most dimensions). Even the reference mean contributes little to the predictor variance on the shoulder and the x-axis of the elbow. Consequently, we find that most variance in the partial predictors can be assigned to the random effects. This suggests that the snooker training data might be dominated by random processes with little explanatory power of the available covariates.

We conduct a sensitivity analysis, where we increase the prespecified proportion of variance to 0.99. Additionally, we assume heteroscedasticity in the model given the residual plot in Figure 10. The model then estimates different measurement error variances with a predominantly large variance on dimension *hand.x*. The model almost doubles the number of FPCs and includes 30 multivariate FPCs (B: 9, C: 10, E:11). The estimation of fixed effects hardly change (results not shown) but we find a better fit to the data (compare Table 2 and Figure 9). The scalar residuals in the lower panels of Figure 10 seem to exhibit less structure for the model of the sensitivity analysis but we can still discern patterns in the residuals. Considering the residual autocorrelation, Figure 11 overall shows a reduction of autocorrelation for small time lags compared to the main model. For larger time lags we find a somewhat surprising increase of the approximated autocorrelation function. Including a large number of comparatively trivial FPCs might thus overall improve model fit while introducing new dependencies in the residuals. Given that our main interest lies in the leading FPCs and the fixed effects, we conclude that the sensitivity analysis yields similar results with added computational complexity.

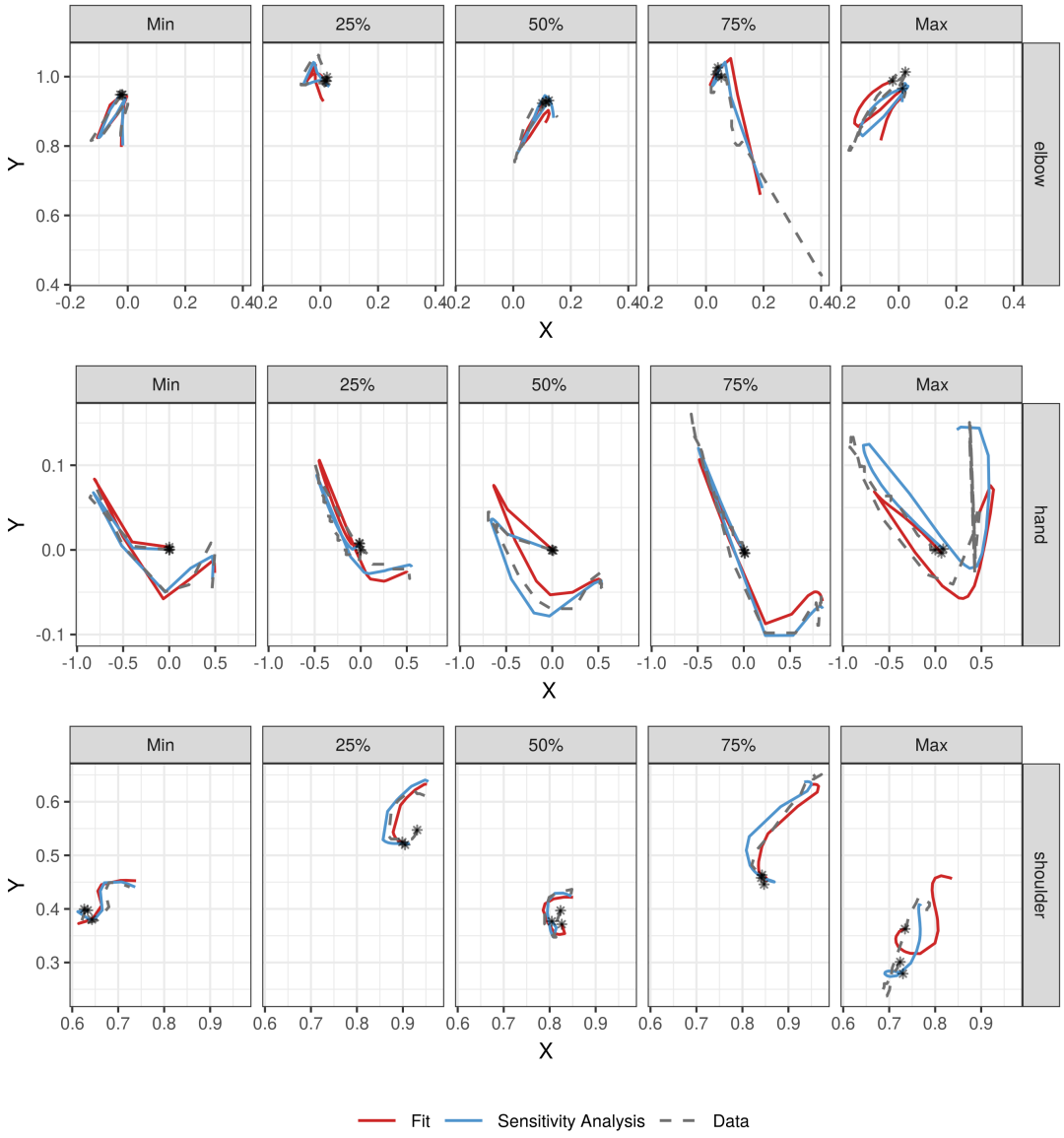

Figure 9: Fitted (red) and observed (blue) snooker training trajectories for selected observations (grey dashed). The qantiles given in the subtitles correspond to the contribution to the mrrMSE.

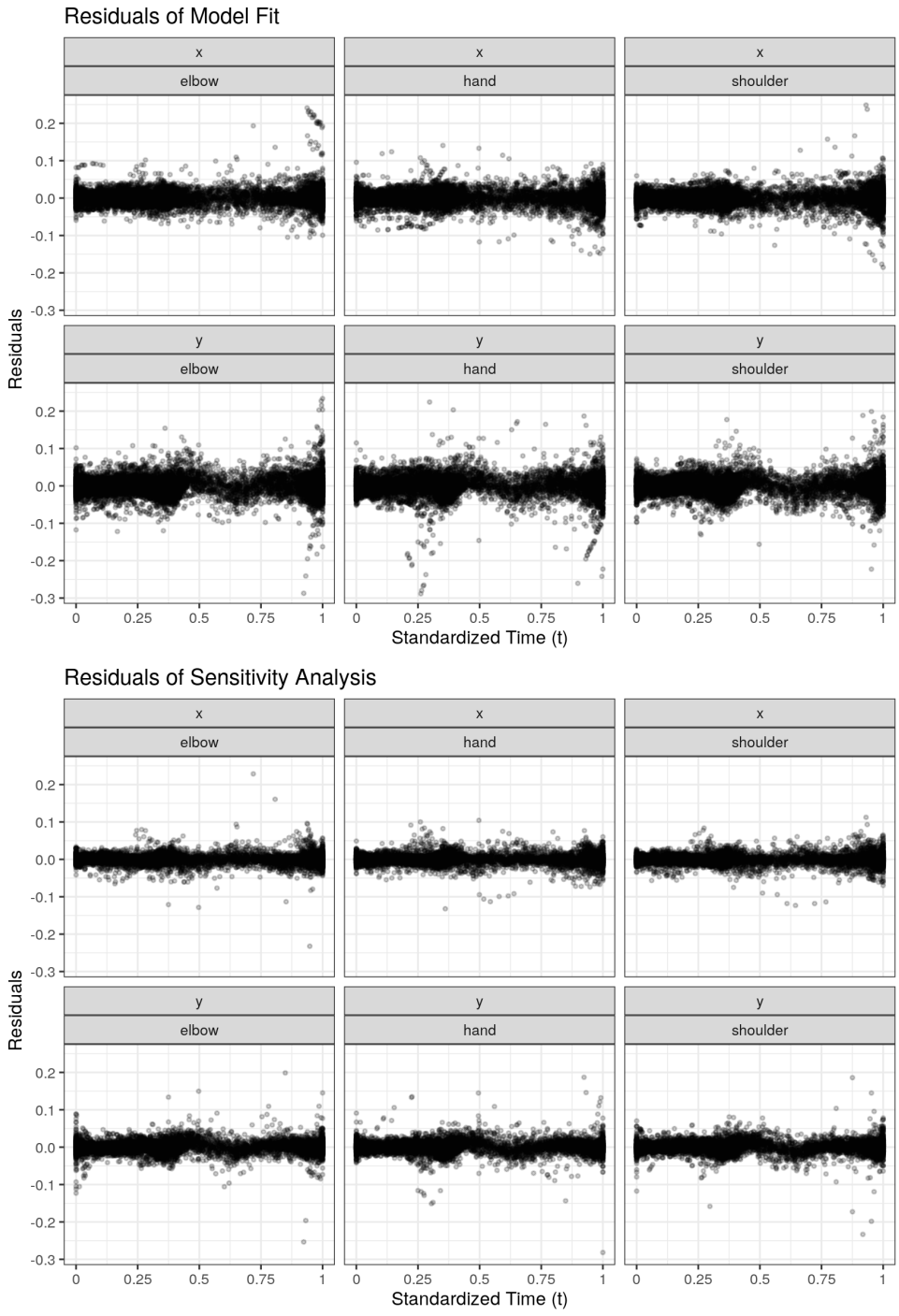

Figure 10: Model residuals per dimension over time. The top panels correspond to the model presented in the main analysis, the lower panels to the model from the sensitivity analysis. Five observations have been removed from the lower panels so that the scale of the plots is identical.

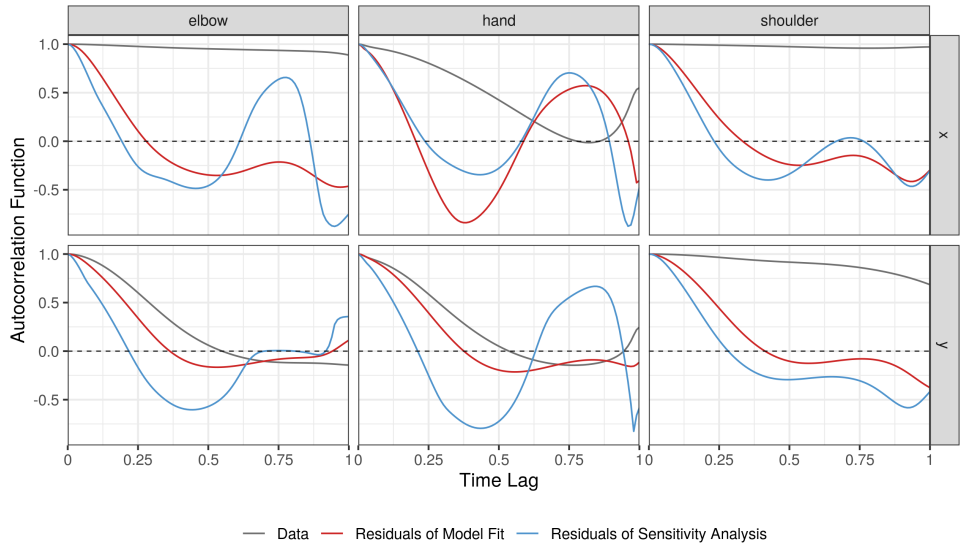

Figure 11: Approximated mean autocorrelation function over the time lag for the observed data (grey), the fit of the model presented in the main analysis (red), and the model from the sensitivity analysis (blue). The dashed line marks zero autocorrelation.

C Detailed Analysis of the Consonant Assimilation Data

C.1 Data and Model Description

Data Description

In this section we present more information on the analysis of the consonant assimilation data. The acoustic (ACO) data were recorded via microphone at 32,768 Hz. In addition, all speakers wore a custom-fit palate sensor with 62 electrodes that measured the area where the tongue touched the palate during the articulation in high temporal resolution (electropalatographic (EPG) data). These primary data were summarized and transformed into a functional index over the time of articulation for the two dimensions ACO and EPG separately. Each functional index measures how similar the articulatory or acoustic pattern is to its reference patterns for the first and second consonant at every observed time point. These similarity indices vary roughly between $+1$ and -1, where the value $+1$ indicates patterns close to the reference for the final target consonant (i.e. the consonant ending the first word) and -1 for the initial target consonant (i.e. the consonant beginning the second word). Due to the specifics of the index calculation the index values can lie slightly outside the interval $[-1, 1]$. The curves on the ACO dimension are generally smoother than on the EPG dimension, which reflects the index calculation: The ACO signal is much richer in information because 256 continuously valued points are considered in the calculation of the similarity index for every time point while only 62 binary points enter the index calculation for the EPG signal. Details on the data generation and functional index computation can be found in Pouplier and Hoole (2016) and Cederbaum et al. (2016).

Model Specification

For this application, we follow the model specification of Cederbaum et al. (2016), who analyse only the ACO dimension of the data and ignore the second mode of the consonant assimilation data. Our specified multivariate model is

$$\boldsymbol{y}_{ijh}(t) = \boldsymbol{\mu}(\boldsymbol{x}_{ij}, t) + \boldsymbol{B}_i(t) + \boldsymbol{C}_j(t) + \boldsymbol{E}_{ijh}(t) + \boldsymbol{\epsilon}_{ijht}, \tag{1}$$

with $i = 1, ..., 9$ the speaker index, $j = 1, ..., 16$ the word combination index, and $h = 1, ..., H_{ij}$ the repetition index. Again, $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ are the person-specific and curve-specific random intercepts. $\boldsymbol{C}_j(t)$ is the word combination-specific random intercept, which gives a crossed random effects structure. Note that here, the curve-specific random component $\boldsymbol{E}_{ijh}(t)$ also captures speaker and word interactions. Taking the different smoothness of the dimensions into account, the white noise measurement error $\boldsymbol{\epsilon}_{ijht}$ is assumed to follow a zero-mean bivariate normal distribution with diagonal covariance matrix $\text{diag}(\sigma^2_{\text{ACO}}, \sigma^2_{\text{EPG}})$.

The additive predictor of the multiFAMM is specified as

$$\begin{aligned}\boldsymbol{\mu}(\boldsymbol{x}_{ij}, t) = \boldsymbol{f}_0(t) &+ \texttt{order}_j \cdot \boldsymbol{f}_1(t) + \texttt{stress1}_j \cdot \boldsymbol{f}_2(t) + \texttt{stress2}_j \cdot \boldsymbol{f}_3(t) \\ &+ \texttt{vowel}_j \cdot \boldsymbol{f}_4(t) + \texttt{order}_j \cdot \texttt{stress1}_j \cdot \boldsymbol{f}_5(t) \\ &+ \texttt{order}_j \cdot \texttt{stress2}_j \cdot \boldsymbol{f}_6(t) + \texttt{order}_j \cdot \texttt{vowel}_j \cdot \boldsymbol{f}_7(t),\end{aligned}$$

with dummy covariates $\texttt{order}_j, \texttt{stress1}_j, \texttt{stress2}_j$, and $\texttt{vowel}_j$ indicating whether in the word combination /sh/ is followed by /s/ (instead of the reference /s#sh/), the final target syllable is not stressed, the initial target syllable is not stressed, and the vowel context (i. e. the adjacent vowel for each consonant sound of interest) is /a#i/ (instead of the reference /i#a/ as in "Gem**i**sch S**a**lbe"), respectively. The functional intercept $\boldsymbol{f}_0(t)$ and the effect function $\boldsymbol{f}_1(t)$ and their deviation from a sinus-like form or zero, respectively, are especially interesting for studying assimilation.

For comparability, we follow the univariate analysis in Cederbaum et al. (2016) by specifying cubic P-splines with third order difference penalty with eight and five (marginal) basis functions for the covariate effect functions and auto-covariance estimation, respectively. The MFPCA is based on an unweighted scalar product. We choose the multivariate MFPCA truncation orders so that at least 95% of the total variation in the data is explained as presented in (3.4). To handle the heteroscedasticity, we use the weighted regression approach. Given the different sampling mechanisms on the dimensions, Appendix C.3 also contains an alternative analysis based on a weighted scalar product for the MFPCA.

C.2 Results of the Model Presented in the Main Part

Analysis of the Variance Components

The univariate decomposition of the auto-covariances yields five univariate FPCs for the random components $\boldsymbol{B}$ and $\boldsymbol{E}$ on both dimensions and one and two FPCs for $\boldsymbol{C}$ on ACO and EPG, respectively. The cut-off criterion based on the sum of total variation selects a total of eight FPCs. The estimated multiFAMM then contains five multivariate FPCs for the smooth residual (representing 64% of total

Table 4: Variance components included in the multiFAMM. First row: Estimates of eigenvalues, univariate error variances, total variation. Second (third) row: Univariate norms of estimated FPCs on dimension ACO (EPG). Fourth (fifth) row: Proportion of univariate variation explained on dimension ACO (EPG) by eigenfunctions and error variance, total univariate variation explained. Sixth row: Proportion of multivariate variation explained by eigenfunctions and error variances, total multivariate variation explained.

	B_1	B_2	B_3	E_1	E_2	E_3	E_4	E_5	σ^2_{ACO}	σ^2_{EPG}	Total
Variation	0.018	0.009	0.004	0.060	0.017	0.012	0.007	0.003	0.004	0.014	0.153
$\|\|\psi^{(ACO)}\|\|^2$	0.169	0.585	0.642	0.153	0.217	0.849	0.178	0.713	$--$	$--$	$--$
$\|\|\psi^{(EPG)}\|\|^2$	0.831	0.415	0.358	0.847	0.783	0.151	0.822	0.287	$--$	$--$	$--$
$\pi^{(ACO)}$	0.068	0.126	0.052	0.210	0.082	0.226	0.026	0.056	0.094	$--$	0.940
$\pi^{(EPG)}$	0.134	0.036	0.012	0.467	0.119	0.016	0.049	0.009	$--$	0.127	0.969
π	0.115	0.062	0.023	0.393	0.108	0.077	0.042	0.023	0.027	0.091	0.961

variation) and three for the subject-specific random effects (20% of total variation) with 12% of total variation due to measurement error. With the chosen truncation order $M_C = 0$, the crossed random component $\boldsymbol{C}$ is effectively dropped from the model as a lot of variation in the data is already explained by the included fixed effects, all of which capture characteristics of the word combinations (8 fixed effects for 16 word combinations). Table 4 shows the contribution of each random process to the total variation in the consonant assimilation data according to the fitted multiFAMM. The first row gives the eigenvalues of the random components $\boldsymbol{B}$ and $\boldsymbol{E}$ as well as the estimated univariate error variances and the total amount of variation in the data as computed in (3.4). The second and third row show the univariate L^2 norm of the multivariate eigenfunctions. The proportion of explained variance π is given in the final three rows. The proportions displayed in the fourth and fifth row are computed according to (3.5) and the last row is computed according to (3.4). Note that the fitted model uses the latter multivariate cut-off criterion to determine the number of FPCs. With different sampling mechanisms for different dimensions and thus different measurement errors, it is advisable to check whether the proportion of explained variance on each dimension (3.5) is adequate. For the consonant assimilation data, the selected number of FPCs explains about 97% of variation on EPG but only 94% on ACO, which we still deem acceptable. If a greater disparity is revealed in an application, we recommend to use the proportion of explained univariate variation as a cut-off criterion (in our case this would lead to including a sixth FPC for component E).

We present the estimated multivariate FPCs for component $\boldsymbol{B}$ in Figure 12. The estimated leading multivariate FPC $\boldsymbol{\psi}_{B1}(t)$ of the subject-specific random effect $\boldsymbol{B}_i(t)$ (see left panels) shows a personal tendency to pronounce the final or initial target syllable distinctly, which explains roughly 12% of the total variation. It changes the relative time spent pronouncing each syllable and pulls the entire curve closer to the reference sound of the preferred consonant. The second FPC (accounting for 6% variation), however, shows the individual tendency for assimilation. It flattens or amplifies the sinus-like shape of

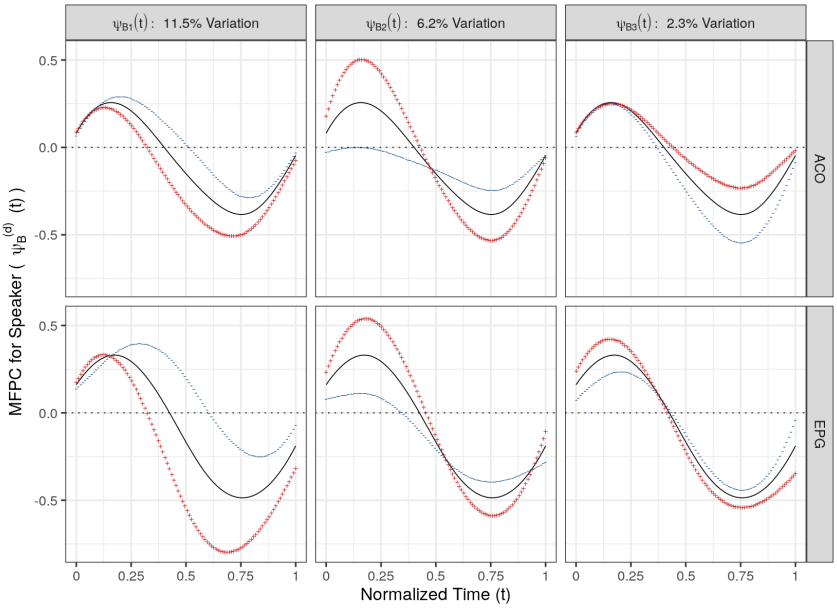

Figure 12: FPCs for the subject-specific functional random effect $\boldsymbol{B}_i(t)$ with the respective proportion of variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{B.}}$) of the FPC is added (red +) and subtracted (blue −).

the overall mean (black solid line) thus rendering the speaker more or less prone to distinguish between the two sounds. Note that in the univariate analysis for ACO presented in Cederbaum et al. (2016), this mode of variation is identified as the leading FPC of the person-specific random effect. The three leading FPCs $\boldsymbol{\psi}_{E1}(t)$, $\boldsymbol{\psi}_{B1}(t)$, and $\boldsymbol{\psi}_{E2}(t)$ all show greater univariate norms on the EPG dimension (Table 4) suggesting that this dimension is the driving source of variation in the model – but this dimension was ignored in the analysis of Cederbaum et al. (2016). Note that only $\boldsymbol{\psi}_{B3}(t)$ impacts the dimensions differently in that more assimilation on one dimension equals less assimilation on the other.

For the curve-specific random effect $\boldsymbol{E}_{ijh}(t)$ (Figure 13), the leading FPC impacts primarily the final target consonant pulling it towards or pushing it away from its reference sound. $\boldsymbol{\psi}_{E2}(t)$ has a similar shape to $\boldsymbol{\psi}_{B2}(t)$. Note that the third leading FPC of E affects the mean function in opposite directions on to the two dimensions shifting the curve up on one dimension and down on the other.

Figure 14 shows the estimated auto-and cross-covariance surface of components $\boldsymbol{B}$ and $\boldsymbol{E}$. It is evident that the cross-covariances are about as large in magnitude as the auto-covariances for the dimension ACO. In univariate analyses, however, the cross-covariance is completely ignored. Note that when no covariates are included in the model, the estimated multiFAMM contains one FPC for component C (results not shown).

Analysis of the Estimated Effect Functions

Figure 15 presents the estimated covariate effects (black, top plot for consonant order /s#sh/, bottom for /sh#s/). Overall, we find similar shapes for the estimated effects on both dimensions. In the reference group, the functional intercept $\boldsymbol{f}_0(t)$ shows signs of assimilation for consonant /s/. The effect $\boldsymbol{f}_1(t)$

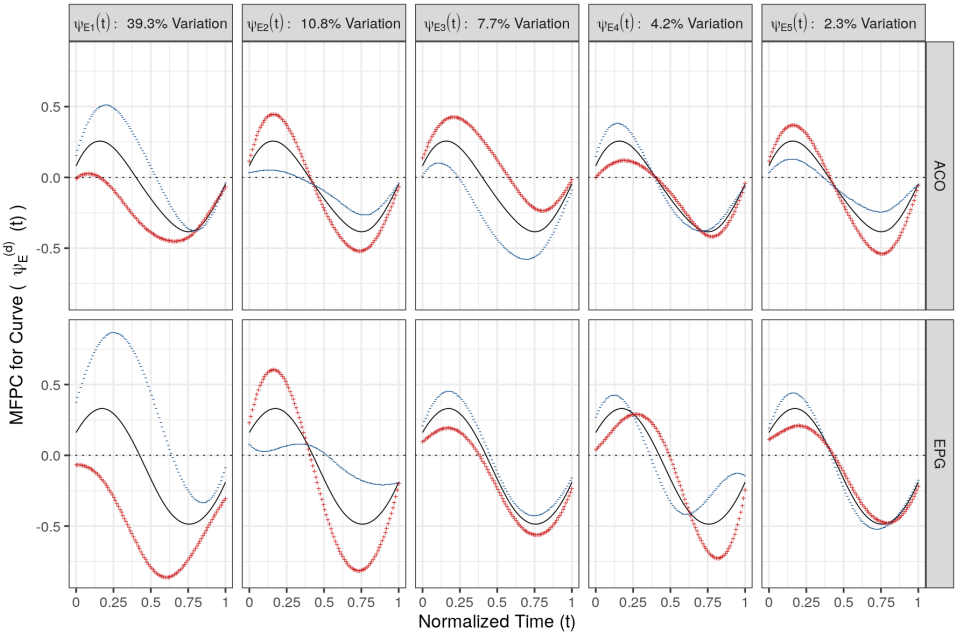

Figure 13: FPCs for the curve-specific functional random effect $\boldsymbol{E}_{ijh}(t)$ with the respective proportion of variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{E\cdot}}$) of the FPC is added (red +) and subtracted (blue −).

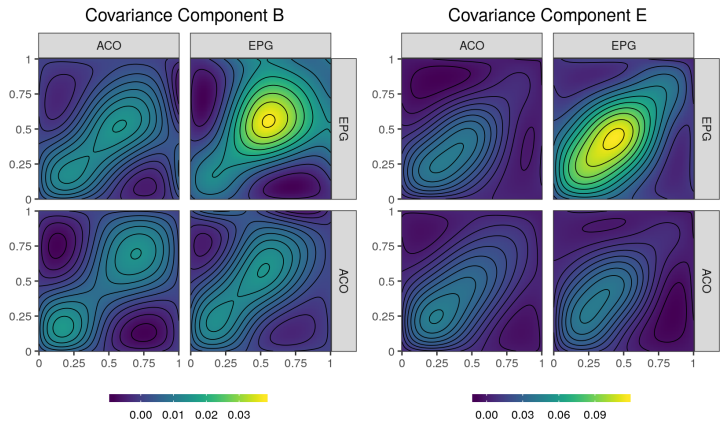

Figure 14: Auto- and cross-covariance surface of the subject-specific functional random effect $\boldsymbol{B}_i(t)$ (left) and curve-specific random effect $\boldsymbol{E}_{ijh}(t)$ (right). The bottom-left and top-right panels show the auto-covariance and the top-left and bottom-right panels show the cross-covariance function as a function of normalized time.

of covariate `order`, however, pushes the final target syllable (in this case /sh/) towards its reference, all other things being equal. The positive effect at the end of the observed interval then pushes the initial target syllable /s/ towards the center, thus indicating a more assimilated pronunciation. Shortly before, we find a negative effect on the dimension EPG which might indicate that for a brief section, the articulatory pattern of /s/ is indeed close to its reference but this does not necessarily translate to the produced sound. Thus, similar to Cederbaum et al. (2016) (red), we find that the final target syllable /sh/ is pulled towards the reference while the initial target syllable /s/ seems to be less affected. Given the similar shape on the dimension EPG, this supports their finding that assimilation is asymmetric. Since the estimated effects are similar across dimensions and similar to the univariate results, we refer to Cederbaum et al. (2016) for interpretation of the other fixed effects.

Comparison to Univariate Analysis

Compared to separate univariate functional additive mixed models (FAMMs), the multivariate model incorporates the dependency between the dimensions, thus reducing the number of FPC bases in the analysis (for a similar amount of variance explained, ten univariate FPCs would be needed). The shaded areas in Figure 15 indicate where the standard errors of the univariate analysis presented in Cederbaum et al. (2016) are smaller than the corresponding standard errors of the multiFAMM. We find that overall, the multivariate analysis seems to give smaller standard errors. In order to compare the model fit between the univariate and the multivariate analysis, we can use the urrMSE defined in Section 5. We then compare the fitted values with the observed values for the consonant assimilation data. The multivariate analysis yields urrMSE values of 0.970 (ACO) and 0.966 (EPG), whereas independent univariate FAMMs give values of 0.978 (ACO) and 0.961 (EPG), respectively. Consequently, a univariate model analogously specified to the ACO model in Cederbaum et al. (2016) gives a slightly better model fit on the EPG data as measured by the urrMSE. However, on the ACO dimension, the multivariate analysis is slightly preferred.

The added computational complexity of multivariate analyses is also negligible in our case: Fitting a univariate FAMM as proposed by Cederbaum et al. (2016) takes around 52 minutes on a 64-bit Linux platform and requires a considerable amount of RAM memory (32 GB is sufficient). The multivariate analysis maintains the requirements for internal memory, while the duration to fit the multivariate model (109 minutes on the aforementioned platform) only slightly increases compared to sequentially fitting two univariate models.

C.3 Results of the Model Using a Weighted Scalar Product

This section gives a short description of the considered multiFAMM when a weighted scalar product is used. We only present the results of estimating the eigenfunction basis as the effect on the estimated covariate effects is negligible. Table 5 shows the contribution of each random process to the total weighted variation and is structured similarly to Table 4. The number of eigenfunctions is chosen by weighted sum of total variation (3.4) which results in the same number of eigenfunctions for each random component as with the unweighted approach. However, the total share of weighted variation explained is slightly

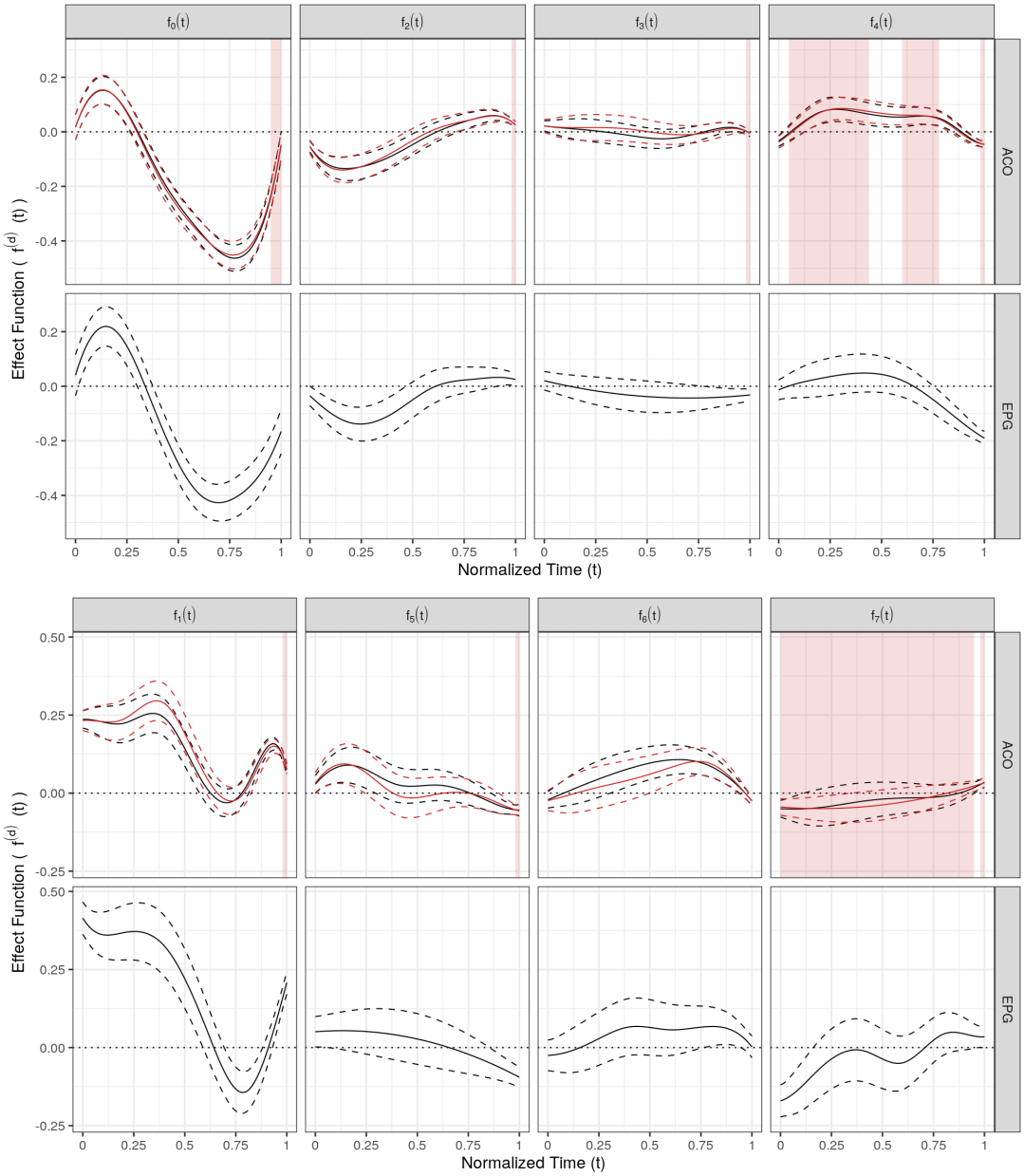

Figure 15: Estimated covariate effects (black) and comparison to univariate model (red). The corresponding 95% confidence intervals are given by the dashed line. Upper (from left to right): reference mean and covariate effects of `stress1`, `stress2`, `vowel` (/s#sh/). Lower (from left to right): covariate effect of `order` and interactions of `order` with `stress1`, `stress2`, `vowel` (/sh#s/). Areas shaded in red indicate where the standard errors of the univariate analysis are smaller.

Table 5: Variance components included in the multiFAMM using a weighted scalar product. First row: Estimates of eigenvalues, univariate error variances, total variation. Second (third) row: Univariate norms of estimated FPCs on dimension ACO (EPG). Fourth (fifth) row: Proportion of univariate variation explained on dimension ACO (EPG) by eigenfunctions and error variance, total univariate variation explained. Sixth row: Proportion of multivariate variation explained by eigenfunctions and error variances, total multivariate variation explained.

	B_1	B_2	B_3	E_1	E_2	E_3	E_4	E_5	σ^2_{ACO}	σ^2_{EPG}	Total
Variation	1.937	1.625	0.493	6.501	2.168	1.823	0.694	0.507	0.004	0.014	18.476
$\|\|\psi^{(ACO)}\|\|^2$	0.630	0.743	0.442	0.574	0.741	0.404	0.564	0.455	$--$	$--$	$--$
$\|\|\psi^{(EPG)}\|\|^2$	0.370	0.257	0.558	0.426	0.259	0.596	0.436	0.545	$--$	$--$	$--$
$\pi^{(ACO)}$	0.115	0.114	0.021	0.352	0.152	0.070	0.037	0.022	0.094	$--$	0.975
$\pi^{(EPG)}$	0.091	0.053	0.035	0.352	0.071	0.138	0.038	0.035	$--$	0.127	0.941
π	0.105	0.088	0.027	0.352	0.117	0.099	0.038	0.027	0.054	0.054	0.960

different (about 22% and 65% explained by the components $\boldsymbol{B}$ and $\boldsymbol{E}$, respectively). Consequently, the proportion of (weighted) variation assigned to the measurement error is smaller when applying the weighted scalar product. We also find that the univariate norms of the eigenfunctions are now more evenly distributed between the dimensions ACO and EPG, with most of the norms bigger on the ACO dimension. We thus see that weighting the scalar product shifts emphasis to the dimension ACO which has a lower measurement error. With this emphasis on ACO, the proportion of univariate variation explained on EPG falls now short of 95% while ACO has a proportion of univariate variation explained of about 98%. For this model, too, we have specified the cut-off using the weighted sum of total variation (3.4) and the model explains 96% of variation in the data.

Figure 16 shows the estimated multivariate FPCs of random component $\boldsymbol{B}$. We find that the leading FPC again depicts an individual tendency to spend more time pronouncing the final or the initial target consonant. For $\boldsymbol{\psi}_{B2}(t)$, almost the same mode of variation is captured on the dimension ACO as with the unweighted scalar product. On the dimension EPG, however, we find the assimilation primarily in the final target syllable with the initial target syllable rather unaffected (more or less time spent in the pronunciation). The third FPC for the subject-specific random effect is comparable to the scenario of the unweighted scalar product.

Figure 17 shows the estimated multivariate FPCs of random component $\boldsymbol{E}$. The leading FPC is somewhat unchanged by the weighting of the scalar product. While in the unweighted case the effect of $\boldsymbol{\psi}_{E2}$ was distributed equally across the two consonants, we find a stronger effect on the initial target consonant for dimension ACO and on the final target consonant for dimension EPG in the weighted case. This seems to be compensated by the third FPC, where in the weighted scenario the emphasis lies on differences in the final target consonant on ACO and in the initial target consonant on EPG. The remaining FPCs are comparable for both the weighted and unweighted scalar product.

We conclude that the model based on the weighted scalar product gives similar results. However,

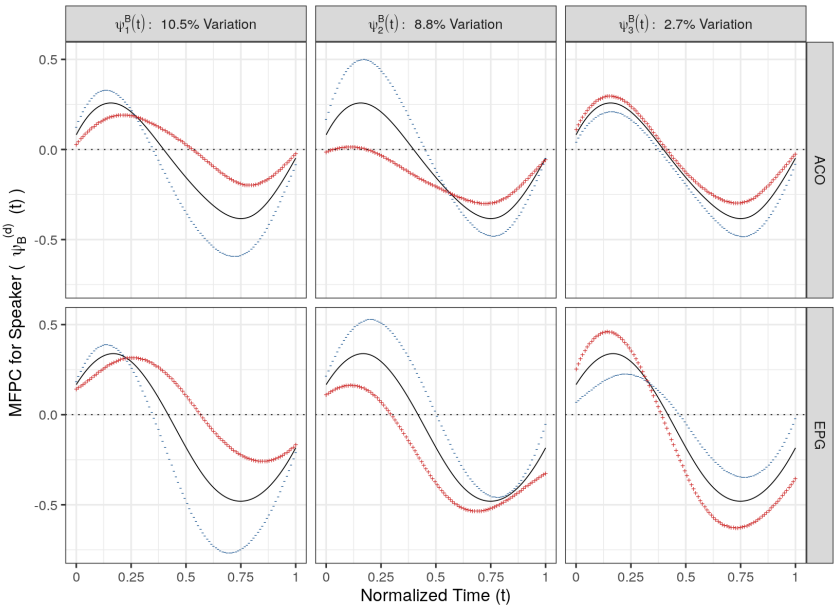

Figure 16: FPCs for the subject-specific functional random effect $\boldsymbol{B}_i(t)$ with the respective proportion of weighted variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{B\cdot}}$) of the FPC is added (red +) and subtracted (blue −).

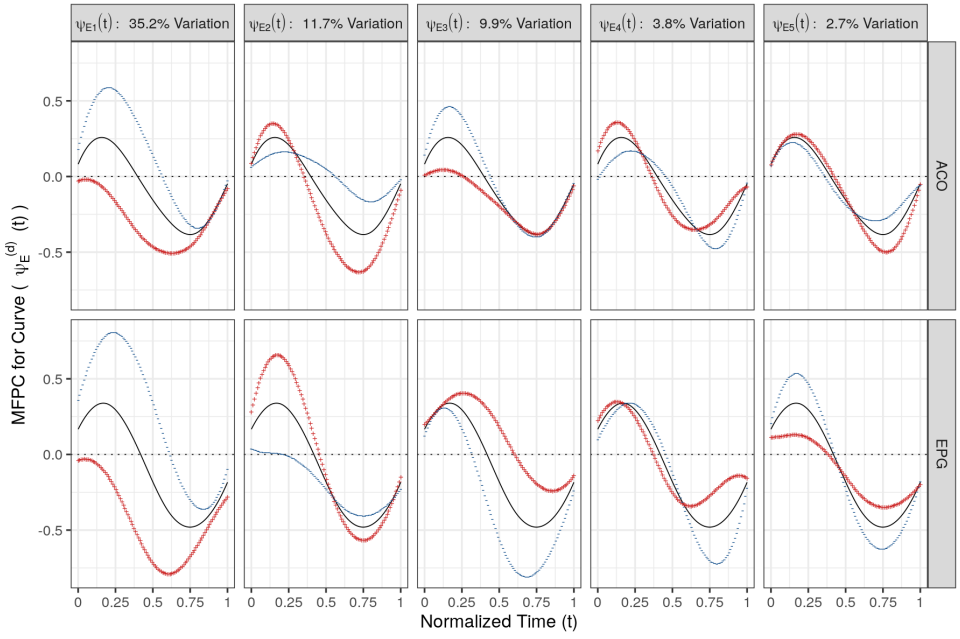

Figure 17: FPCs for the curve-specific functional random effect $\boldsymbol{E}_{ijh}(t)$ with the respective proportion of weighted variance explained. The black solid line represents the mean trajectory to which a suitable multiple ($2\sqrt{\nu_{E\cdot}}$) of the FPC is added (red +) and subtracted (blue −).

if some dimension were considerably more noisy than the rest, weighting the scalar product might be advisable (see results of the simulation).

D Detailed Analysis of the Simulation Study

We conduct an extensive simulation study in order to answer the following questions: How does the performance of the proposed multiFAMM depend on different model specifications? How does the model perform in different data settings? And how does the multiFAMM compare to a univariate modeling approach, where univariate regression models as proposed by Cederbaum et al. (2016) are estimated on each dimension independently? Additionally, we evaluate the estimation of the covariance and the fixed effects. We simulate data based on the presented model fits (4.1) of the snooker training data (main part) and (1) of the consonant assimilation data (Appendix C). Six different settings of the data generating process are analysed, where for each data setting, we additionally compare up to seven different model specifications. One modeling scenario (data setting plus model specification) consists of independent model fits based on 500 generated data sets. Table 6 provides an overview of the analysed data settings and model specifications, which are described in the following section. Note that we use data setting 1 and model specification A, respectively, to generate benchmark estimates useful for comparison to other settings and specifications.

D.1 Simulation Description

Data Settings

Table 7 provides detailed description of the data generating process for all data settings. For each data setting, 500 data sets have been simulated. For data setting 1, we generate new observations based on the model fit (1) (Appendix C) of the consonant assimilation data by randomly drawing the evaluation points, the random scores, and the measurement errors. Note that we center and decorrelate the random scores so that the empirical means and covariances coincide with the theoretical counterparts. As in the original data, each simulated data set contains the bivariate functional observations of $i = 1, .., 9$ individuals, each repeating the $j = 1, ..., 16$ different word combinations five times.

Setting 2 (strong heteroscedasticity) is analogously to setting 1 but with larger difference between the measurement error variances of the two dimensions. Setting 2 mimics an application setting with

Table 6: Summary of the different data settings and model specifications analysed in the simulation study. Each analysed combination of data setting and model specification comprises 500 model fits.

Data Setting		Model Specification	
1	Consonant assimilation data	A	True model
2	Strong heteroscedasticity	B	Truncation via total variation (TV)
3	Sparse data	C	Truncation via univariate variation (UV)
4	Uncentred scores	D	TV with alternate scalar product
5	Weighted scalar product	E	UV with alternate scalar product
6	Snooker training data	F	Scedastic misspecification
		U	Univariate models

(multimodal) multivariate data, where some dimensions are much noisier than others, in order to evaluate whether this imbalance interferes with the variance decomposition of the multiFAMM. Setting 3 (sparse data) focuses on the estimation quality for sparse functional data. Compared to setting 1, the number of evaluation points is reduced to three to ten measurements per dimension. In setting 4 (uncentred scores), the random scores are not decorrelated or centred (otherwise identical to setting 1). Especially for covariance components with few grouping levels (particularly $\boldsymbol{B}$), this can result in a considerable departure from the modeling assumptions, which is likely to occur in such settings. This setting explores the sensitivity of the multiFAMM to violations of its assumptions. Setting 5 (weighted scalar product) is identical to setting 1 but all model components are simulated based on the estimated model using a weighted scalar product for the MFPCA (see Appendix C.3). This data setting helps to understand the impact of weights in the scalar product. For setting 6 (snooker training data), we generate new trajectory data according to the model fit (4.1) of the snooker training data. This allows us to evaluate a higher dimensional multiFAMM as well as the estimation quality of nested random effects.

Model Specifications

Table 8 provides a detailed description of the different modeling scenarios used in the simulation study. We denote the most accurate approach of modeling as specification A (true model). This standard scenario mirrors the data generation so that there is no model misspecification. Most notably, we fix the number of FPCs to the number used for generating the data in order to avoid truncation effects. Though somewhat unrealistic, specification A serves to separate the impact of modeling decisions or situations with more (realistic) uncertainty for the user from the overall performance of the multiFAMM.

For model specification B (truncation via total variation (TV)), the truncation orders of the FPCs are chosen so that 95% of the total variation (3.4) are explained. In scenario C (truncation via univariate variation (UV)), we choose the number of FPCs so that on every dimension 95% of univariate variance (3.5) are explained. Specifications D (TV with alternate scalar product) and E (UV with alternate scalar product) use the cut-off criterions analogous to B and C but we alternate the scalar product on which the MFPCA is based: For data generated from a model based on an unweighted MFPCA, the scalar product used in these scenarios is weighted by $\frac{1}{\hat{\sigma}_d^2}$, and vice versa. Model specification F (scedastic misspecification) evaluates misspecifying the multiFAMM with a homoscedasticity assumption. We also contrast the multiFAMM with the univariate approach of fitting independent univariate models. In modeling scenarios denoted with U, we fit an independent FAMM on each dimension. We use the FAMMs proposed by Cederbaum et al. (2016) so that we can apply the same model specifications as for the multiFAMM (e.g. basis functions, penalties, etc.). The number of FPCs in the model is then chosen so that 95% of univariate variation is explained.

Modeling Scenarios

For each of the combinations of data settings and modeling scenarios indicated in Table 9, analyses were performed giving 500 fitted models per combination. The results presented in the main part correspond to modeling scenarios 1A (Benchmark), 1B (Cut-Off Multi), 1C (Cut-Off Uni), 3A (Sparse Data), and

Table 7: Description of data settings.

Data Setting	Description
1	Data based on consonant assimilation data (bivariate functions on $dim1, dim2$, each on $[0,1]$). 9 individuals, each repeating 16 crossed grouping levels 5 times. Number of observed time points is an independent random draw for each dimension from a uniform discrete distribution of natural numbers in $[20,50]$. Observed time point is an independent random draw from uniform distribution on $[0,1]$. Mean consists of functional intercept and 7 covariate effect functions as given in Appendix C.2. Covariates are given for each observation based on the crossed grouping level (word combination). Functional random effects $\boldsymbol{B}, \boldsymbol{E}$ are based on estimated eigenfunctions of model in Appendix C.2 (3 and 5 FPCs). Corresponding random scores are independent draws from $N(0, \hat{\nu}_{\cdot\cdot})$, then demeaned and decorrelated. Measurement errors are independent draws from $N(0, \hat{\sigma}^2_{\cdot})$ for each observed time point on the respective dimension.
2	Measurement errors are independent draws from $N(0, \hat{\sigma}^2_1)$ for each observed time point on $dim1$ and independent draws from $N(0, 16 \cdot \hat{\sigma}^2_1)$ on $dim2$. Rest as in 1.
3	Number of observed time points is a random draw for each dimension from a uniform discrete distribution of natural numbers in $[3,10]$. Rest as in 1.
4	Random scores are independent draws from $N(0, \hat{\nu}_{\cdot\cdot})$ (no centering or decorrelation). Rest as in 1.
5	All estimated components (eigenfunctions, eigenvalues, covariate effect functions, measurement error variances) stem from the model based on a weighted scalar product for the MFPCA as presented in Appendix C.3. Rest as in 1.
6	Data based on snooker training data (six-dimensional functions on $dim1, ..., dim6$, each on $[0,1]$). 25 individuals, each repeating 2 nested grouping levels 6 times. Number of observed points is an independent random draw from a uniform discrete distribution of natural numbers in $[10,50]$ per multivariate curve with observed time points identical over the dimensions. Observed time point is an independent random draw from uniform distribution on $[0,1]$. Mean consists of functional intercept and 4 covariate effect functions as given in Appendix B. Covariates `skill`, `group` are independent random draws from Bernoulli distribution with probability 0.5 per subject and `session` is given for each observation based on the nested grouping level. Functional random effects $\boldsymbol{B}, \boldsymbol{C}, \boldsymbol{E}$ are based on estimated eigenfunctions and eigenvalues of model in Appendix B ($6, 5$ and 5 FPCs). Measurement errors are independent draws from $N(0, \hat{\sigma}^2)$ for each observed time point on each dimension.

Table 8: Description of model specifications.

Model Specification	Description
A	Univariate mean, multivariate mean, univariate auto-covariance estimation based on cubic P-splines with 8, 8, 5 (marginal) basis functions and choice of penalty as in the model used for generating the data setting (third order for 1-5, first order for 6). No univariate truncation for the MFPCA except negative eigenvalues. MFPCA based on scalar product as in the model used for generating the data setting (weights of one except for 5). Number of multivariate FPCs is fixed according to the data setting (1-5: ($\boldsymbol{B}$:3, $\boldsymbol{E}$:5), 6: ($\boldsymbol{B}$:6, $\boldsymbol{C}$:5, $\boldsymbol{E}$:5)). Weighted regression approach with weights obtained from univariate variance decompositions for heteroscedastic data settings (1-5).
B	Number of FPCs chosen so that 95% of the (weighted) sum of total variation in the data is explained. Rest as in A.
C	Number of FPCs chosen so that on each dimension at least 95% of the univariate variation in the data is explained. Rest as in A.
D	Weighted scalar product for MFPCA with weights based on dimension specific measurement error variances if data setting is based on model using a scalar product with weights of one (all except 5) and vice versa (5). Number of FPCs chosen so that 95% of the (weighted) sum of total variation in the data is explained. Rest as in A.
E	Weighted scalar product for MFPCA with weights based on dimension specific measurement error variances if data setting is based on model using a scalar product with weights of one (all except 5) and vice versa (5). Number of FPCs chosen so that on each dimension at least 95% of the univariate variation in the data is explained. Rest as in A.
F	Homoscedasticity is assumed, no regression weights. Rest as in A.
U	Univariate modeling approach. Independent sparse functional linear mixed model (FLMM) models for each dimension. Number of univariate FPCs chosen so that 95% of univariate dimension is explained. Rest as in A.

Table 9: Overview of the analysed modeling scenarios.

	Data Setting					
Model Specification	1	2	3	4	5	6
A	X	X	X	X	X	X
B	X	X			X	
C	X	X			X	
D	X	X			X	
E	X	X			X	
F	X	X				
U	X	X			X	

4A (Uncentred Scores).

Evaluation Criteria

The accuracy of the estimated model components is measured using the root relative mean squared error (rrMSE). We analyse the accuracy of the covariance estimation, i.e. eigenfunctions, eigenvalues, and measurement error variances as well as the mean estimation. For evaluating the overall accuracy of the estimate $\hat{\boldsymbol{\zeta}}$ of the multivariate functional component $\boldsymbol{\zeta} = (\boldsymbol{\zeta}_1, ..., \boldsymbol{\zeta}_S)^\top$, we define the mrrMSE as

$$\text{mrrMSE}(\boldsymbol{\zeta}, \hat{\boldsymbol{\zeta}}) = \sqrt{\frac{\frac{1}{S}\sum_{s=1}^{S} |||\boldsymbol{\zeta}_s - \hat{\boldsymbol{\zeta}}_s|||^2}{\frac{1}{S}\sum_{s=1}^{S} |||\boldsymbol{\zeta}_s|||^2}}$$

with the multivariate norm based on the unweighted scalar product. Note that S depends on the model component to be evaluated, e.g. for the fitted values $\boldsymbol{y}_{ijh}(t)$, S equals N but for the subject-specific random effect $\boldsymbol{B}_i(t)$, S equals the number of individuals. The mrrMSE corresponds to the evaluation criterion used in section 5 of the main part. We also define a urrMSE

$$\text{urrMSE}\big(\boldsymbol{\zeta}^{(d)}, \hat{\boldsymbol{\zeta}}^{(d)}\big) = \sqrt{\frac{\frac{1}{S}\sum_{s=1}^{S} ||\zeta_s^{(d)} - \hat{\zeta}_s^{(d)}||^2}{\frac{1}{S}\sum_{s=1}^{S} ||\zeta_s^{(d)}||^2}}$$

for $\boldsymbol{\zeta}^{(d)} = (\zeta_1^{(d)}, ..., \zeta_S^{(d)})^\top$, which allows to evaluate the dimension specific estimation accuracy as well as a straightforward comparison to the univariate modeling approach. For scalar estimates such as eigenvalues and error variances, we define the rrMSE as

$$\text{rrMSE}(\zeta, \hat{\zeta}) = \sqrt{\frac{(\zeta - \hat{\zeta})^2}{\zeta^2}}$$

with estimate $\hat{\zeta}$ of the scalar value ζ. The rrMSE takes on (unbounded) positive values with smaller values indicating a better fit. As the rrMSE is a relative measure, small differences between estimate and true component can result in large rrMSE values for true component norms close to zero. Note that eigenfunctions are defined only up to a sign change. We thus flip the estimated eigenfunction (multiply it by (-1)) if this results in a smaller norm for the difference between the true function and its estimate.

Additionally, we evaluate the coverage of the point-wise confidence bands (CBs) of the estimated fixed effects.

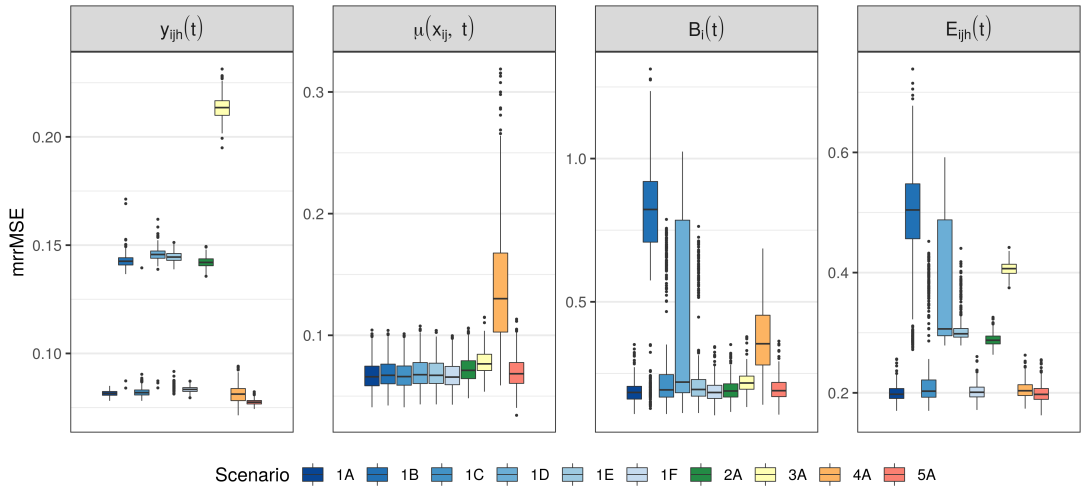

Figure 18: mrrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$, the mean $\boldsymbol{\mu}(\boldsymbol{x}_j, t)$, and the random effects $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ for all modeling scenarios in data setting 1 (blue shades) and the standard modeling scenario A in data settings 2 to 5.

D.2 Results of the Simulation Study

Impact of Model Specifications

The results for setting 1 (consonant assimilation data) demonstrate the importance of the number of FPCs in the accuracy of the estimation (see Figure 18 and Table 10). With specifications A (true model) and F (scedastic misspecification), the sets of FPCs are fixed giving overall low values for the mrrMSE (the misspecification in F yields only slightly worse results). Similarly, choosing the truncation order via the proportion of univariate variance explained in scenario C gives models with roughly the same number of FPCs as is used for the data generation. The cut-off criterion based on the total amount of variance in scenario B results in more parsimonious models and thus considerably higher mrrMSE values. The number of selected FPCs also explains the wider boxplots of the scenarios B-E compared to A: rather than a larger overall variance in estimation, we find separate clusters based on the included FPC sets (see also Figure 19). For the modeling scenarios based on a weighted scalar product (1D and 1E), the number of chosen FPCs is quite similar regardless of the cut-off criterion but the overall mrrMSE values of the fitted curves ($\boldsymbol{y}_{ijh}(t)$) are higher than for the unweighted approach. The estimate of the mean $\boldsymbol{\mu}(\boldsymbol{x}_j, t)$, however, is comparatively stable over the different model specifications.

Figure 19 shows the urrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$ for different modeling scenarios (1U, 1B, 1C, 1D, 1E) depending on the number of FPCs included in the model. For example in the univariate modeling scenario 1U, all 500 models choose two FPCs for each random effect (B2-E2) on dimension $dim1$. On $dim2$, however, different combinations of FPCs lead to considerably different urrMSE values. For the multivariate modeling scenarios, we also find that the reduction in rrMSE values depends on which additional FPC is included: $\boldsymbol{\psi}_{E5}$ reduces the values more than $\boldsymbol{\psi}_{B3}$.

Figure 20 shows the mrrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$, the mean $\boldsymbol{\mu}(\boldsymbol{x}_i, t)$, and the random effects $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ for the different modeling scenarios of data settings 1, 2, and 5. Again, we find

	1A	1B	1C	1D	1E	1U		2B	2C	2D	2E	2U	
						dim1	dim2					dim1	dim2
$\boldsymbol{B}_i(t)$	3.00	2.15	2.80	2.62	2.88	3.09		2.02	2.87	2.00	3.00	3.06	
						2.00	1.09					2.00	1.06
$\boldsymbol{E}_{ijh}(t)$	5.00	4.00	5.00	4.01	4.12	4.99		3.77	4.44	3.00	4.27	4.72	
						2.00	2.99					2.00	2.72

Table 10: Average number of eigenfunctions selected (scenario A fixed to underlying truth) in the 500 simulation iterations per modeling scenario (column) and random effect (row). For the univariate modeling approach (scenario U) we report the total amount (top) and the number per independent model (bottom).

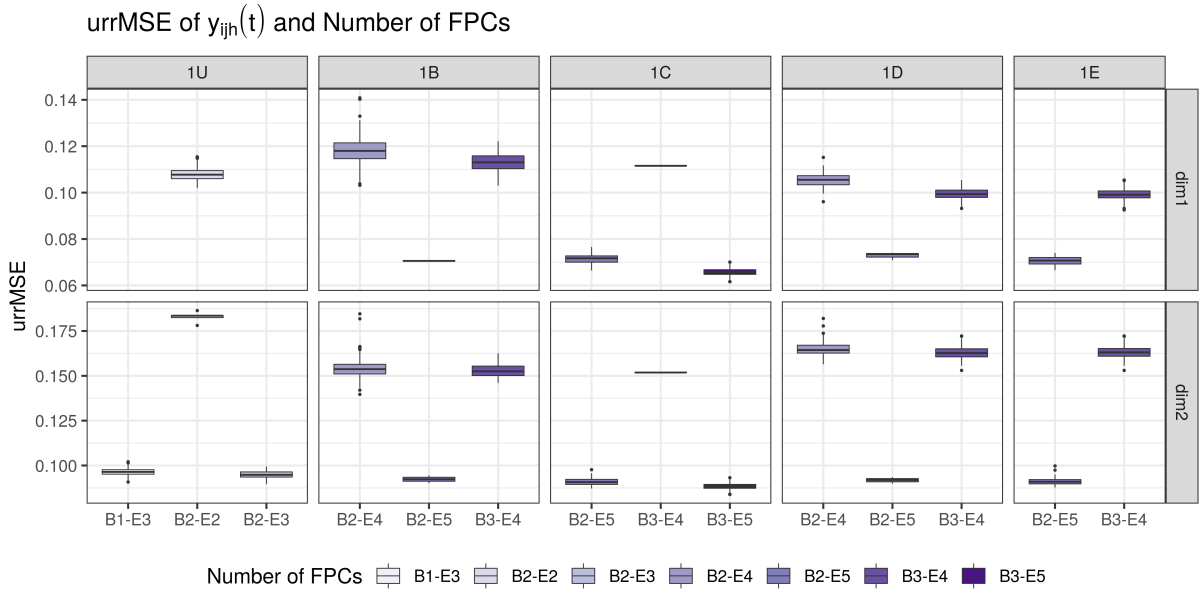

Figure 19: urrMSE values of the fitted values $\boldsymbol{y}_{ijh}(t)$ for different scenarios. The urrMSE values are grouped by the number of FPCs included in the models. Bx-Ey means the model contains x FPCs for $\boldsymbol{B}_i(t)$ and y for $\boldsymbol{E}_{ijh}(t)$.

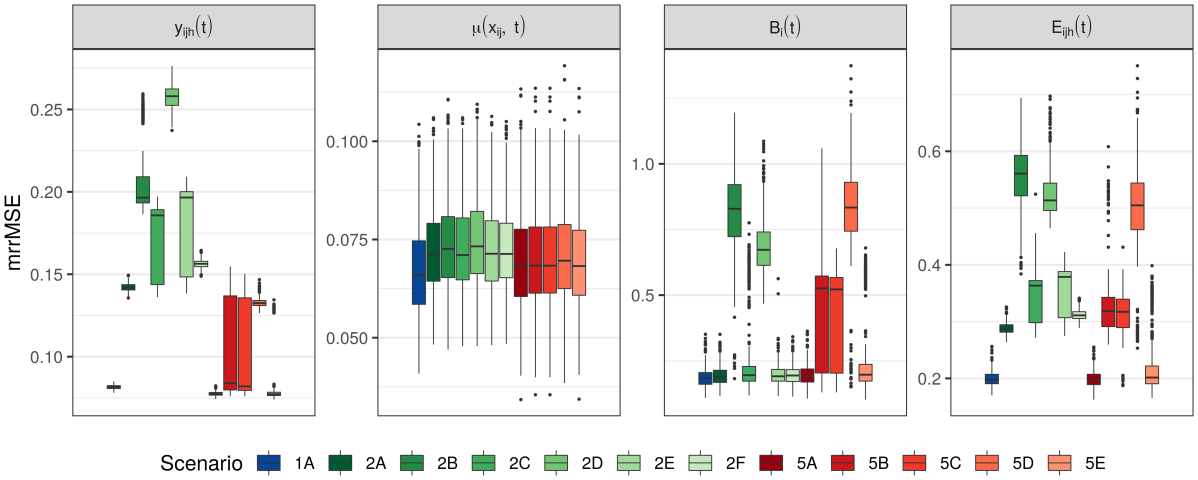

Figure 20: mrrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$, the mean $\boldsymbol{\mu}(\boldsymbol{x}_i, t)$, and the random effects $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ for different scenarios.

considerable differences for different numbers of FPCs in the models when there is strong heteroscedasticity between the different dimensions (setting 2). The overall model fit seems to be better for applying an unweighted scalar product for the MFPCA (2B, 2C compared to 2D, 2E). Later, we will see that the model fit on single dimensions can be improved by weighting the scalar product. Note that for setting 2F, misspecifying the model assumption now has a larger negative impact on the fitted curves than in setting 1F. In data setting 5, the data are generated based on the model using a weighted scalar product. Then, the number of FPCs in modeling scenarios 5B and 5C is very similar. Interestingly, basing the MFPCA on a scalar product using weights of one can lead to a similar estimation accuracy than the standard modeling scenario (5A compared to 5E).

Our simulation study thus suggests that basing the truncation orders on the proportion of explained variation on each dimension gives parsimonious and well fitting models. If interest lies mainly on the estimation of fixed effects, the alternative cut-off criterion based on the total variation in the data allows even more parsimonious models. Furthermore, an unweighted scalar product is a reasonable starting point for the multiFAMM.

Model Performance on Different Data Settings

To compare different data settings, we focus on model specification A (true model) in Figure 18. Note that the mrrMSE values of the other data settings cannot be directly compared as the denominator of the mrrMSE (slightly) changes (except for the mean of 2A, 3A, and 4A). We find that strong heteroscedasticity (2A) mainly negatively affects the fitted curves and the smooth residual. Unsurprisingly, the results for scenario 3A suggest that the estimation accuracy is lower for sparse functional data. However, the mean estimation is comparable to more densely observed data. On the other hand, the mean estimation is susceptible to violations of the assumption of uncorrelated and centred realizations of the random effects (4A). The model then has difficulties to separate intercept and random effects (see also Figure 27), which does not necessarily translate to a worse overall fit to the data. The results of scenario 5A (weighted

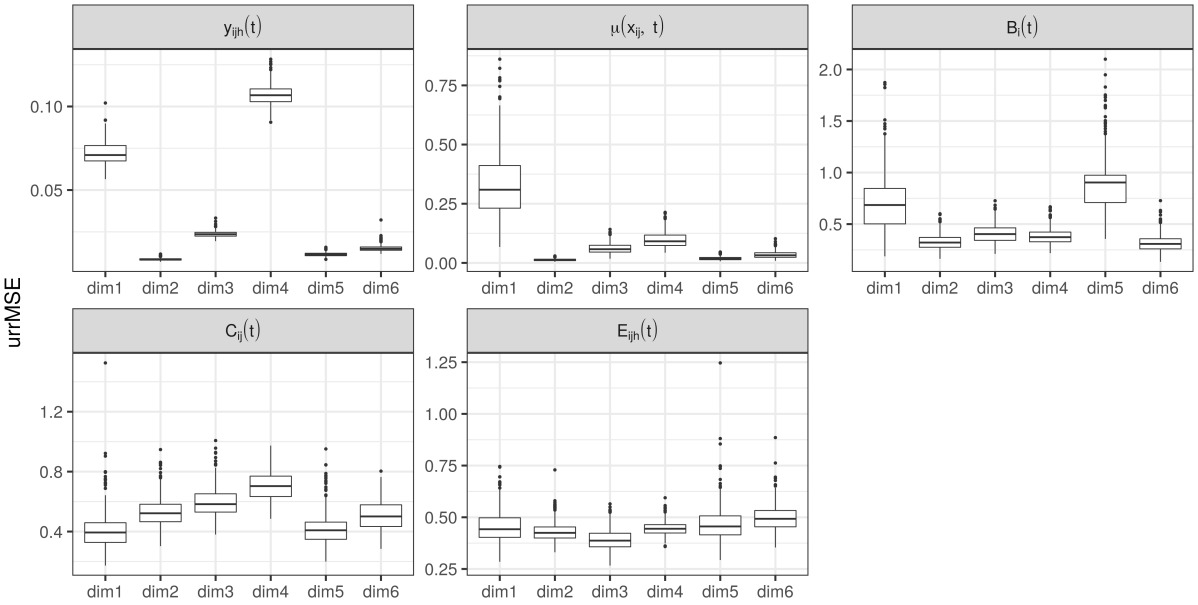

Figure 21: urrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$, the mean $\boldsymbol{\mu}(\boldsymbol{x}_i, t)$, and the random effects $\boldsymbol{B}_i(t), \boldsymbol{C}_{ij}(t)$ and $\boldsymbol{E}_{ijh}(t)$ for modeling scenario 6A. One outlying observation with high urrMSE value was removed from the boxplot for $\boldsymbol{E}_{ijh}(t)$ and $dim1$.

scalar product) suggest that the accuracy of the multiFAMM does not depend on the definition of the scalar product used for the MFPCA.

Figure 21 shows the urrMSE values for the aforementioned model components of modeling scenario 6A (snooker training data) where we compare the estimation accuracy across the dimensions. The fit of the functional curves $\boldsymbol{y}_{ijh}(t)$ suggests that there might be pronounced differences between the estimation accuracy of the dimensions, the dimensions $dim1$ (corresponding to $elbow.x$) and $dim4$ (corresponding to $hand.y$) giving high rrMSE values.

We conclude that the proposed multiFAMM performs well even in more challenging data settings such as sparse functional data or data with few grouping levels for the random effects. Especially the estimation of the fixed effects seems to be stable over the different analysed settings.

Comparison to Univariate Approach

We compare the univariate modeling approach U to multivariate modeling scenarios C (UV) and E (UV with alternate scalar product), i.e. we make sure that the proportion of explained univariate variation is also at least 95% for each model. Table 10 shows that in data settings 1 (consonant assimilation data) and 2 (strong heteroscedasticity) the number of included FPCs tends to be higher for scenario U. Yet Figure 22 indicates that the multiFAMM yields consistently lower urrMSE values on $dim1$ (smaller measurement error variance). For $dim2$, the random effects seem to be estimated more accurately and the fixed effects similarly well, but especially with the weighted scalar product (scenarios E), the overall fit for y can give higher urrMSE values. Overall, the results suggest that the unweighted scalar product is to be preferred in data situations with similar measurement error variances of the dimensions such as setting 1, where

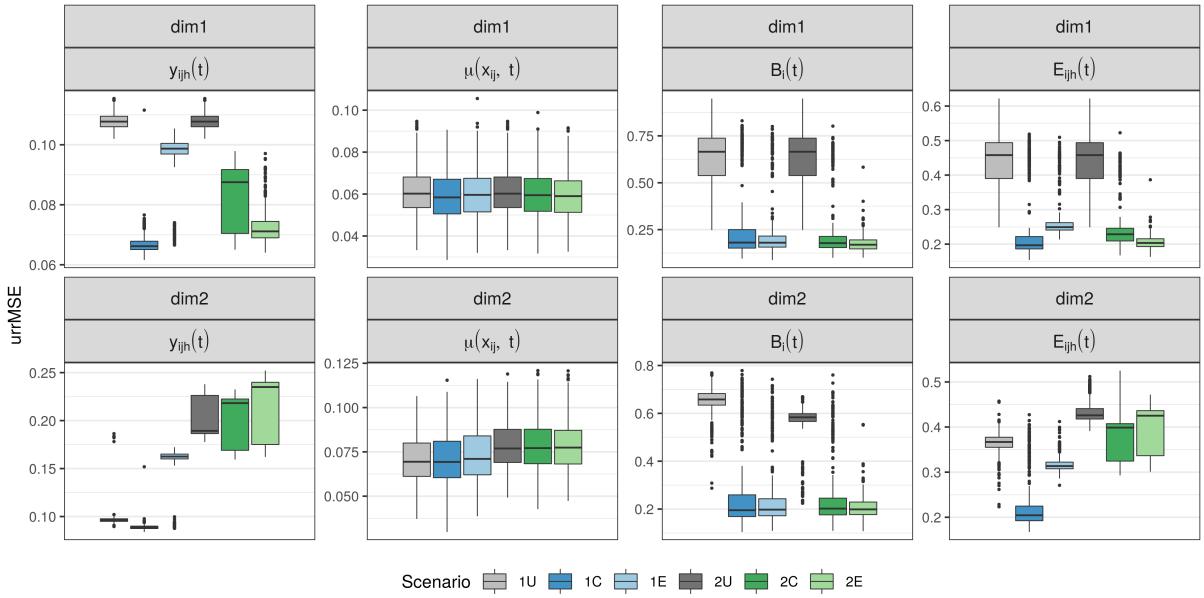

Figure 22: urrMSE values of the fitted curves $\boldsymbol{y}_{ijh}(t)$, the mean $\boldsymbol{\mu}(\boldsymbol{x}_i, t)$, and the random effects $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ for the univariate modeling scenario U (grey) and the multivariate modeling scenarios C and E in data settings 1 (blue) and 2 (green).

it gives reliably good results across all model components. However, downweighting the dimension $dim2$ with larger error seems to be a reasonable modeling decision in setting 2 if interest lies primarily on $dim1$ (lowest urrMSE values for 2E). Again, we point out that the estimation of the fixed effects is relatively stable across approaches and slightly better than for the univariate modeling approach.

Table 11 compares the standard errors of the fixed effects estimation for the multiFAMM in scenario 1C with the univariate modeling approach of scenario 1U. We look at the ratio of standard errors $\frac{se_m}{se_u}$, where se_m denotes the standard error of the multiFAMM and se_u the standard error of the corresponding univariate FAMM. For each dimension and partial predictor, we calculate the proportion of ratios smaller and larger than one over an equidistant grid of 100 time points and all simulation runs. Especially on $dim1$, we find that the multivariate approach gives smaller standard errors. On $dim2$, the proportions are more similar with slightly more proportions larger than one. Overall, we find more ratios smaller than one in our simulation study thus indicating that the multivariate approach can yield smaller standard errors. Table 12 further shows that the coverage of these two scenarios is very similar (see subsection on the estimation of fixed effects).

We conclude that the multivariate modeling approach can improve the mean estimation but is especially beneficial for the prediction of the random effects. In some cases, including weights in the multivariate scalar product might further improve the modeling.

Covariance Estimation

Figure 23 shows the mrrMSE values for the estimated eigenfunctions of the two random effects in the standard modeling scenario A of data settings 1 to 5. In general, we find that the leading eigenfunctions

Table 11: Proportion of standard error ratios of the multiFAMM (se_m) of scenario 1C and the univariate modeling approach (se_u) of scenario 1U over all simulation runs and all evaluated time points.

d	Ratio	$\beta_0^{(d)}$	$f_0^{(d)}(t)$	$f_1^{(d)}(t)$	$f_2^{(d)}(t)$	$f_3^{(d)}(t)$	$f_4^{(d)}(t)$	$f_5^{(d)}(t)$	$f_6^{(d)}(t)$	$f_7^{(d)}(t)$
$dim1$	$\frac{se_m}{se_u} < 1$	0.67	0.93	0.65	0.74	0.76	0.69	0.85	0.56	0.42
	$\frac{se_m}{se_u} > 1$	0.33	0.07	0.35	0.26	0.24	0.31	0.15	0.44	0.58
$dim2$	$\frac{se_m}{se_u} < 1$	0.28	0.48	0.45	0.36	0.48	0.41	0.31	0.54	0.62
	$\frac{se_m}{se_u} > 1$	0.72	0.52	0.55	0.64	0.52	0.59	0.69	0.46	0.38

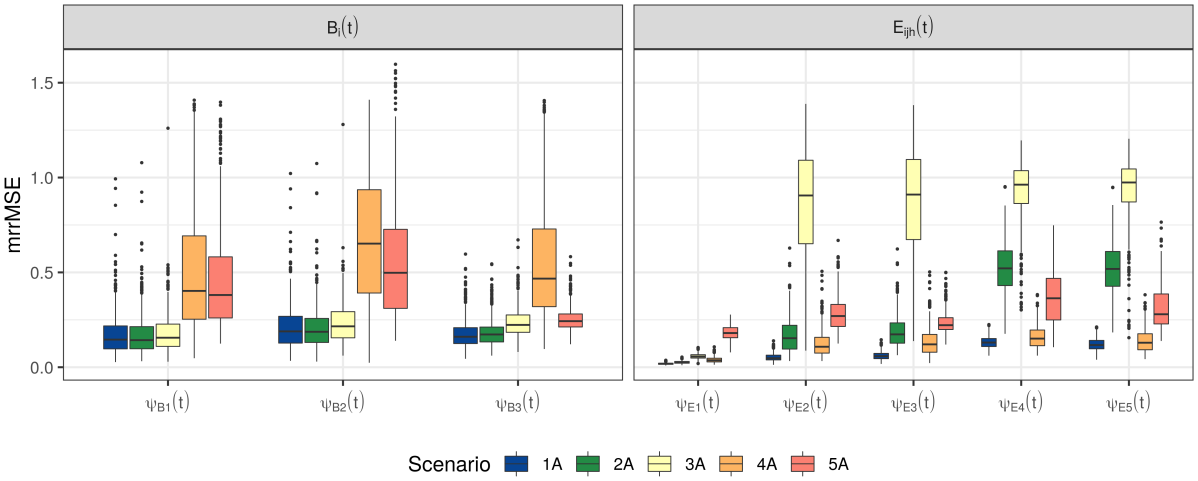

Figure 23: mrrMSE values of the eigenfunction estimation for the random effects $\boldsymbol{B}_i(t)$ and $\boldsymbol{E}_{ijh}(t)$ for different scenarios.

tend to have lower mrrMSE values. Especially the leading eigenfunctions of the smooth residual show a high accuracy. There seems to be more variance in the estimation accuracy for the subject-specific random effect. This can be confirmed with Figure 24, which contains the estimated eigenfunctions of all 500 simulation iterations for modeling scenario 1A (grey curves) and compares them to the data generating function (red curve). Overall we find that the modes can be reconstructed sufficiently well, albeit considerably better for the random smooth residual.

Figure 23 also allows to compare the eigenfunction estimation across the different data settings. Keep in mind that scenario 5A is based on a different model so the mrrMSE values are not directly comparable. We find that the eigenfunction estimation of the smooth residual suffers from strong heteroscedasticity (2A), whereas the subject-specific random effect seems unaffected. The same effect (even more pronounced) can be observed for sparse data (3A), where fewer information about the correlation within functions is available. On the other hand, $\boldsymbol{B}_i(t)$ suffers from few different individuals as this can lead to a considerable departure from the modeling assumption when the independent draws of scores are not centred and decorrelated (4A). With 720 different grouping levels of the smooth residual, we typically get an empirical covariance for the random scores that is closer to its assumption and thus smaller differences in mrrMSE values compared to 1A. The estimation of the eigenfunctions is somewhat less accurate in

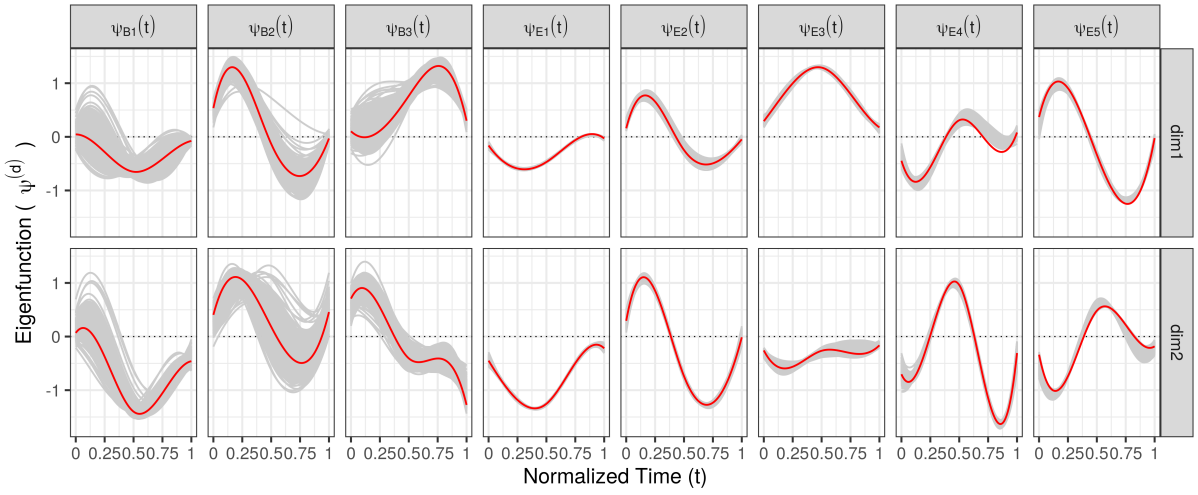

Figure 24: Estimated eigenfunctions in scenario 1A for all 500 simulation iterations (grey curves). The red curves show the data generating eigenfunctions.

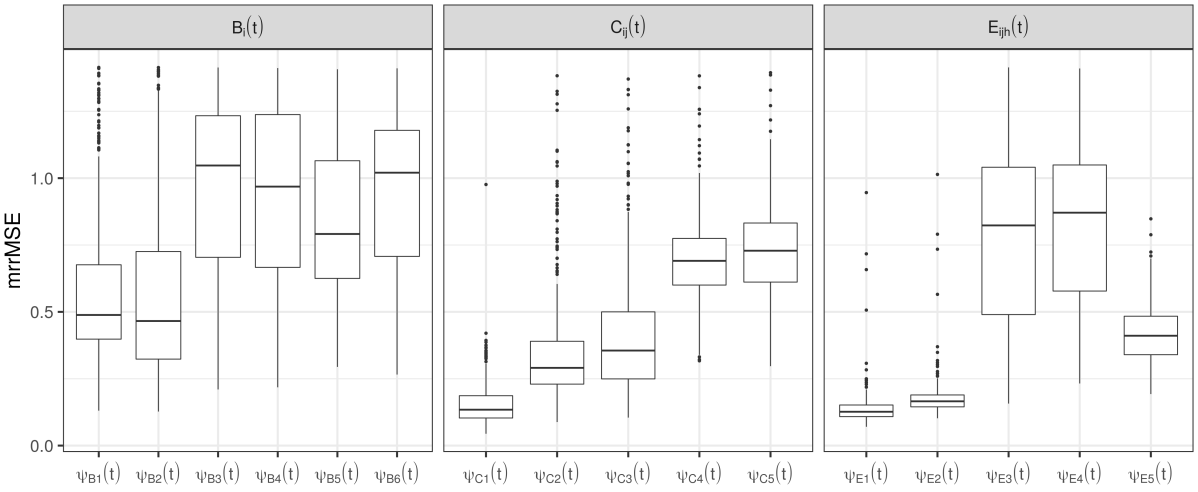

Figure 25: mrrMSE values of the eigenfunction estimation for the random effects $\boldsymbol{B}_i(t), \boldsymbol{C}_{ij}(t)$, and $\boldsymbol{E}_{ijh}(t)$ for scenario 6A.

scenario 5A as the weights of the scalar product have to be estimated as well. This additional uncertainty leads to a larger variance for the mrrMSE values and makes it harder to correctly identify the data generating modes of variation. Note that this does not affect the overall estimation of the random effects as discussed above.

We observe similar trends for the mrrMSE values of the estimated eigenfunctions in scenario 6A as presented in Figure 25. Leading eigenfunctions tend to be more accurately estimated and increasing the number of grouping levels ($\boldsymbol{B}_i(t)$: 25, $\boldsymbol{C}_{ij}(t)$: 50, $\boldsymbol{E}_{ijh}(t)$: 300) seems to have a diminishing effect on the variance of the mrrMSE values.

Figure 26 shows the rrMSE values of the estimated multivariate eigenvalues and measurement error variances. Compared to scenario 1A, the rrMSE values of the smooth residual are higher for scenarios 2A (strong heteroscedasticity) and 3A (sparse data), whereas the rrMSE values of the subject-specific

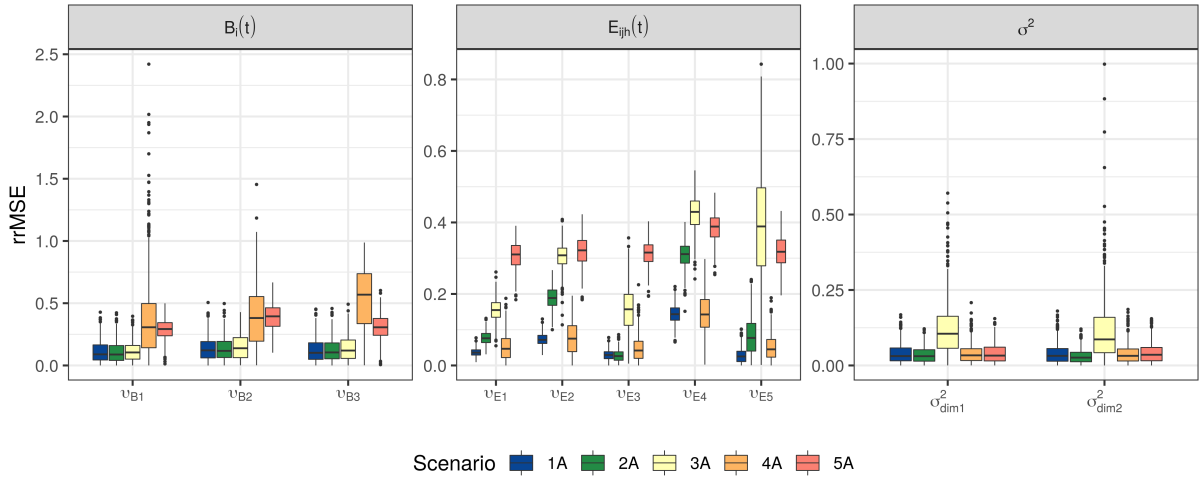

Figure 26: rrMSE values of the eigenvalue and error variance estimation for the standard scenario A of data settings 1-5.

random effect are higher for scenario 4A (correlated scores). This is along the lines of the findings for the eigenfunctions and the random effects. Scenario 5A is again not directly comparable and the overall higher rrMSE values suggest that the uncertainty in the estimation of the weights of the scalar product influences the accuracy of the estimation of the eigenvalues. With regards to the error variance, the rrMSE values are comparable over the different scenarios except for the sparse data setting, where we find higher values.

We conclude that modes of variation can be recovered well in most of the model scenarios. The nested random effect and its leading modes of variation can be well captured by the multiFAMM.

Fixed Effects Estimation

Figure 27 shows the urrMSE values of the estimated effect functions in scenario 1A (blue). We find that the estimation of the functional intercept $\boldsymbol{f}_0(t)$ and the covariate effect of `order` $\boldsymbol{f}_1(t)$ yield low urrMSE values on both dimensions. The estimation of the other covariates ($\boldsymbol{f}_2(t)$ to $\boldsymbol{f}_4(t)$) and especially the interactions of `order` with the other covariates ($\boldsymbol{f}_5(t)$ to $\boldsymbol{f}_7(t)$) give larger urrMSE values and a higher variance of these values. The yellow boxplots show the corresponding values in scenario 4A, thus indicating that only the estimation of the intercept is affected by correlated and uncentred scores (as an empirical score mean different from zero times the corresponding eigenfunctions is captured by the intercept). Figure 28 plots all estimated effect functions against the data generating effect functions in scenario 1A. This suggests that the multiFAMM can overall capture characteristics of the true underlying effect functions.

Figure 29 shows the urrMSE values for each effect function in scenario 6A. Overall, the functional intercept $\boldsymbol{f}_0(t)$ and the covariate effect of `skill` $\boldsymbol{f}_1(t)$ have a higher estimation accuracy as implied by lower urrMSE values. The other effect functions are also somewhat less pronounced, which is why the scaling with the inverse mean norm can give large urrMSE values. Figure 30 shows all estimated effect functions (grey curves) and the corresponding data generating functions in red. This plot suggests that

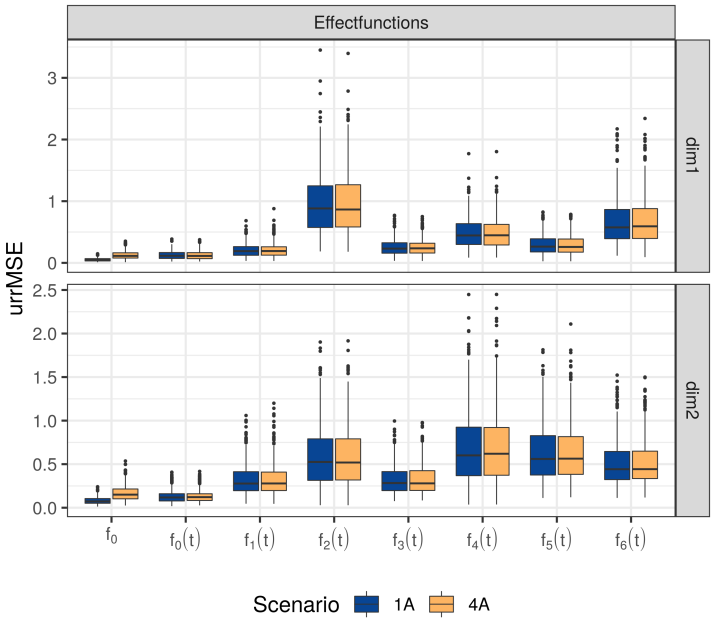

Figure 27: urrMSE values of the estimated effect functions in scenarios 1A and 4A.

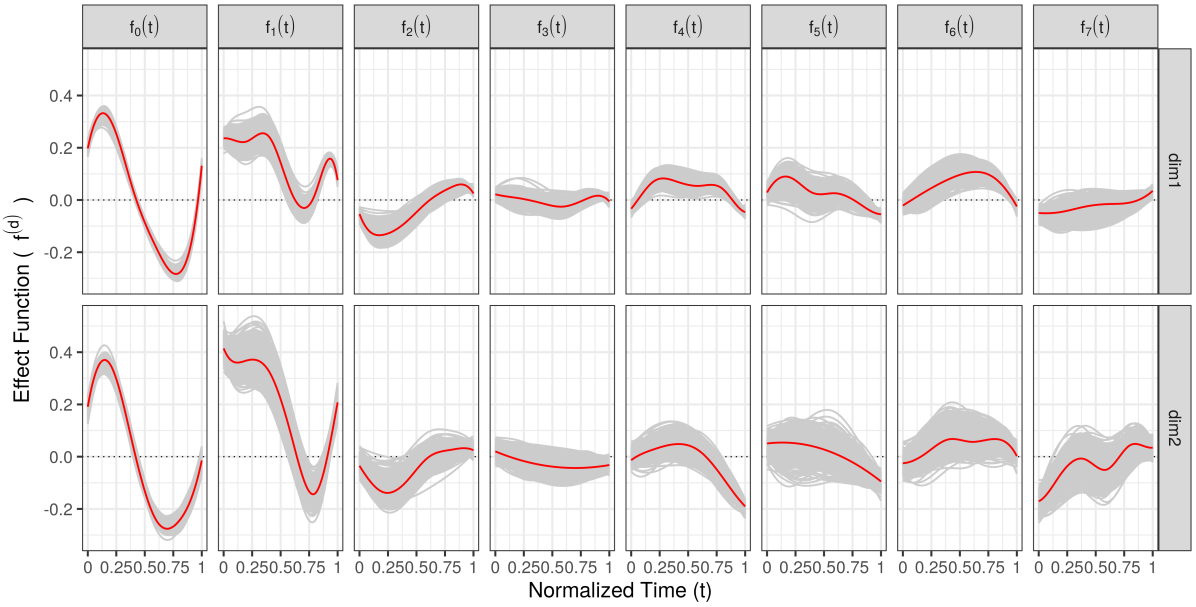

Figure 28: Estimated effect functions in scenario 1A for all 500 simulation iterations (grey curves). The red curves show the data generating effect functions.

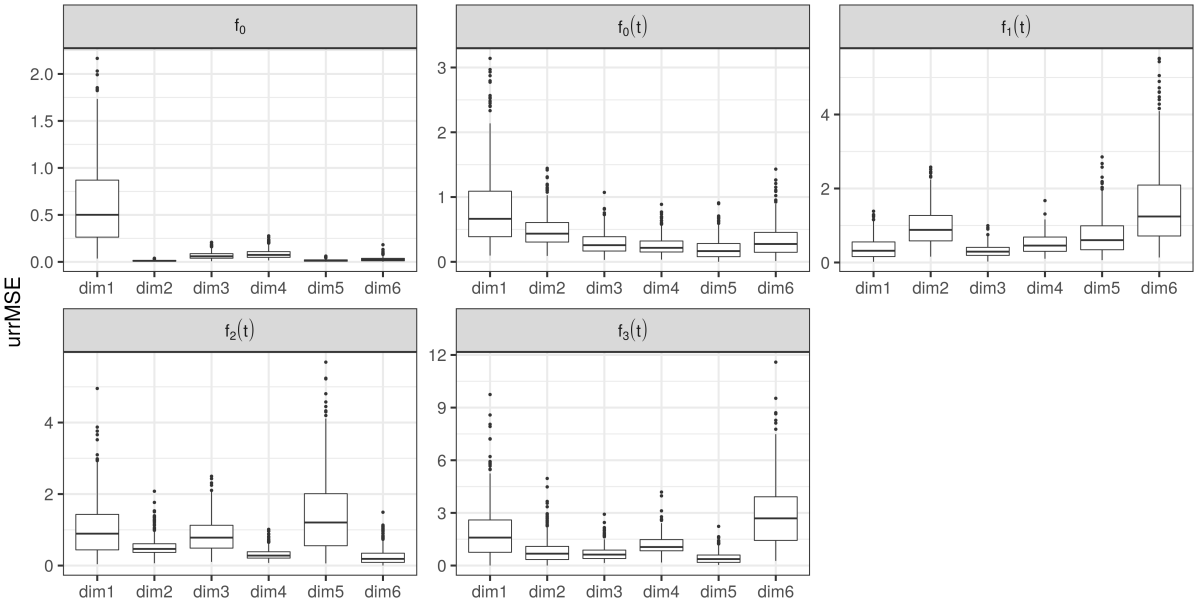

Figure 29: urrMSE values of the estimated effect functions in scenario 6A.

especially on the dimensions $dim1, dim5$, and $dim6$ the effect estimation shows a larger variance across the simulation runs. These dimensions correspond to those dimensions in the data set ($elbox.x, shoulder.x$, and $shoulder.y$ in Figure 1) where the response functions are relatively constant over t compared to the other dimensions.

For all modeling scenarios, we use the average point-wise coverage to evaluate the 95% CBs of the estimated fixed effects (see Table 12). For scenario 1A, we find the CBs to cover the true effect $88 - 95\%$ of the time. Additional uncertainty e.g. about the number of FPCs further reduces the coverage (for example 1B). Overall, the coverage of fixed effects by the corresponding CBs is comparable over the different data settings (see Appendix Table 12). In data settings 2 and 3 the uncertainty in the data is increased which allows for wider CBs and thus a slightly better coverage. The averaged pointwise coverage in 4A is comparable to 1A except for the functional intercept. Here, the averaged point-wise coverage of the scalar and functional intercept lie well below 70% as the data setting makes it particularly hard to identify the true underlying mean (the scores of the random effects are not centred or decorrelated and any mean difference to zero is absorbed by the intercept). In scenario 5A, the coverage tends to be lower than for models based on an unweighted scalar product, possibly due to the added uncertainty from estimation of the weight in the scalar problem. We find averaged point-wise coverage to be considerably lower than its nominal value in scenario 6A. In this scenario, we find that in many of the simulation runs the effects are wrongly estimated as constants (cf. also Figure 30), which can lead to undercoverage also in scalar additive mixed models (Greven and Scheipl; 2016).

Figure 31 shows the point-wise coverage average over the 500 simulation iterations in scenario 1A. It suggests that coverage tends to lie below the nominal value in areas close to the border of the functions.

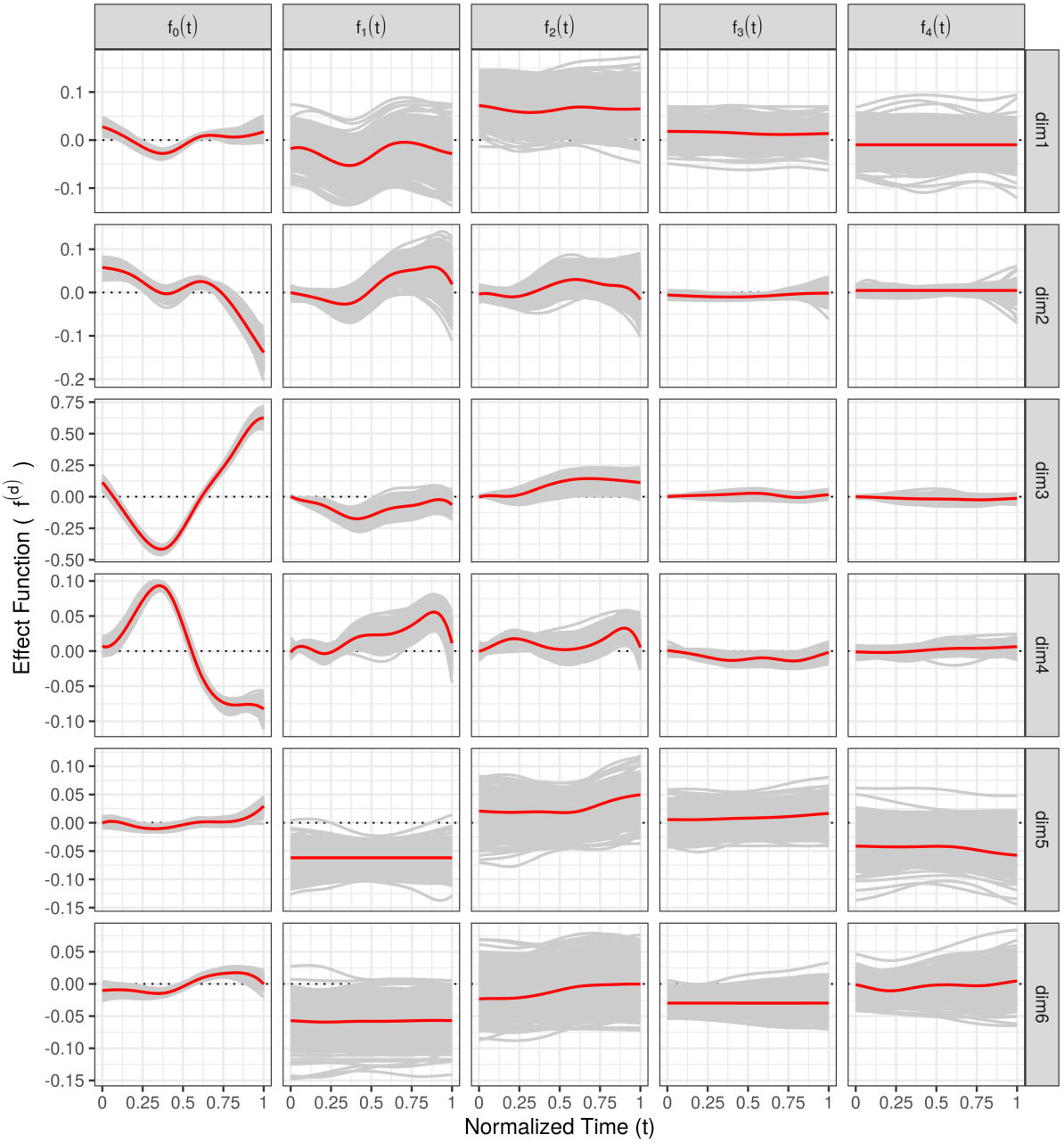

Figure 30: Estimated effect functions in scenario 6A for all 500 simulation iterations (grey curves). The red curves show the data generating effect functions.

Table 12: Averaged point-wise coverage of the point-wise CBs for the estimated effect functions and the averaged coverage of the scalar intercept $\beta_0^{(d)}$ for different scenarios. The coverage is averaged over the 500 simulation iterations and over 100 evaluation points along t.

d	Scenario	$\beta_0^{(d)}$	$f_0^{(d)}(t)$	$f_1^{(d)}(t)$	$f_2^{(d)}(t)$	$f_3^{(d)}(t)$	$f_4^{(d)}(t)$	$f_5^{(d)}(t)$	$f_6^{(d)}(t)$	$f_7^{(d)}(t)$
	1A	93.6	90.6	92.6	89.5	93.1	93.9	91.3	89.0	90.4
	1B	93.8	83.5	90.6	86.6	89.7	90.2	87.9	86.6	88.0
	1C	93.8	90.7	92.5	89.5	93.1	94.1	91.3	88.9	90.8
	1D	93.4	87.7	91.3	86.2	89.5	91.3	87.9	87.1	88.0
	1E	92.8	88.7	90.7	86.9	90.2	91.3	88.0	86.9	87.8
$dim1$	1F	94.6	92.2	93.7	90.1	94.1	94.7	91.5	90.3	93.2
	1U	92.6	90.8	92.5	90.2	93	92.6	91.7	89.5	88.2
	2A	95.4	92.8	94.2	90.6	93.9	94.8	92.9	90.5	92.0
	3A	94.4	89.2	93.0	91.2	93.0	91.7	90.8	92.4	93.5
	4A	68.4	48.7	92.7	89.7	92.7	93.4	91.4	89.4	90.2
	5A	94.0	88.4	91.4	89.8	92.8	92.7	89.7	91.4	94.2
	1A	93.4	87.6	93.3	91.8	94.6	92.9	93.3	93.5	94.1
	1B	95.0	86.1	93.1	91.9	94.0	93.3	93.6	92.2	92.5
	1C	93.8	87.5	93.2	91.9	94.9	93.0	93.4	93.5	94.0
	1D	91.2	86.5	90.7	89.4	89.4	89.5	90.0	88.2	90.0
	1E	91.0	86.6	90.8	90.0	90.0	89.3	90.2	88.5	90.0
$dim2$	1F	93.4	87.8	92.7	92.3	93.7	92.5	93.2	92.9	94.0
	1U	93.2	87.1	92.3	93.6	94.7	92.0	94.0	93.9	93.6
	2A	94.8	89.4	92.5	93.2	94.5	93.4	93.8	93.4	92.7
	3A	94.6	90.1	92.7	92.8	95.8	94.1	94.0	95.2	91.9
	4A	68.6	54.7	93.2	92.3	94.3	92.7	93.3	93.7	93.8
	5A	90.4	83.5	88.1	88.6	93.9	87.5	86.7	91.9	91.4

d	Scenario	$\beta_0^{(d)}$	$f_0^{(d)}(t)$	$f_1^{(d)}(t)$	$f_2^{(d)}(t)$	$f_3^{(d)}(t)$	$f_4^{(d)}(t)$
$dim1$		87.4	84.9	90.8	80.7	74.4	76.0
$dim2$		87.6	88.3	87.9	86.9	75.2	81.4
$dim3$	6A	82.4	85.3	87.1	84.2	81.6	71.7
$dim4$		82.4	82.2	83.6	78.5	78.3	74.1
$dim5$		85.4	82.1	84.3	84.3	78.0	76.3
$dim6$		86.2	74.1	83.0	84.2	79.3	65.7

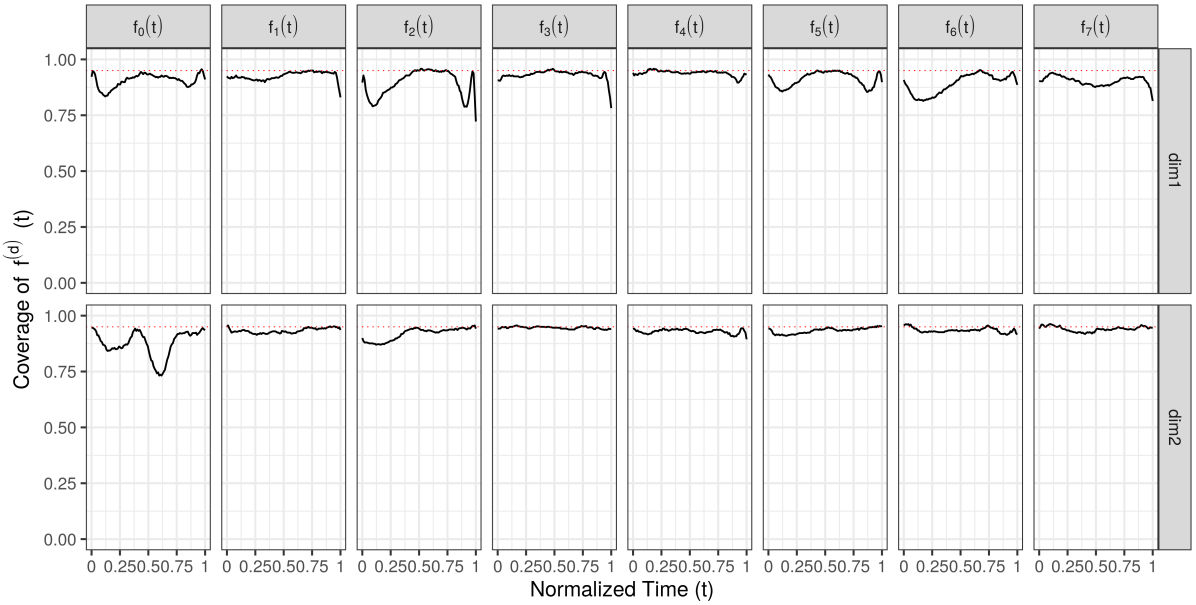

Figure 31: Point-wise averaged coverage over the functional index for the estimated effect functions in scenario 1A. The nominal value is 95% (red dotted line).

References

Cederbaum, J., Pouplier, M., Hoole, P. and Greven, S. (2016). Functional linear mixed models for irregularly or sparsely sampled data, *Statistical Modelling* **16**(1): 67–88.

Cederbaum, J., Scheipl, F. and Greven, S. (2018). Fast symmetric additive covariance smoothing, *Computational Statistics & Data Analysis* **120**: 25–41.

Enghofer, T. (2014). Überblick über die Sportart Snooker, Entwicklung eines Muskeltrainings und Untersuchung dessen Einflusses auf die Stoßtechnik. Unpublished Zulassungsarbeit, Technische Universität München.

Greven, S. and Scheipl, F. (2016). Comment, *Journal of the American Statistical Association* **111**(516): 1568–1573.

Pouplier, M. and Hoole, P. (2016). Articulatory and acoustic characteristics of german fricative clusters, *Phonetica* **73**(1): 52–78.

Steyer, L., Stöcker, A. and Greven, S. (2021). Elastic analysis of irregularly or sparsely sampled curves, *arXiv preprint arXiv:2104.11039* .