Multivariate functional additive mixed models

Alexander Volkmann[1], Almond Stöcker[1], Fabian Scheipl[2] and Sonja Greven[1]

[1]Chair of Statistics, School of Business and Economics, Humboldt-Universität zu Berlin, Germany
[2]Department of Statistics, Ludwig-Maximilians-Universität München, Germany

Abstract: Multivariate functional data can be intrinsically multivariate like movement trajectories in 2D or complementary such as precipitation, temperature and wind speeds over time at a given weather station. We propose a multivariate functional additive mixed model (multiFAMM) and show its application to both data situations using examples from sports science (movement trajectories of snooker players) and phonetic science (acoustic signals and articulation of consonants). The approach includes linear and nonlinear covariate effects and models the dependency structure between the dimensions of the responses using multivariate functional principal component analysis. Multivariate functional random intercepts capture both the auto-correlation within a given function and cross-correlations between the multivariate functional dimensions. They also allow us to model between-function correlations as induced by, for example, repeated measurements or crossed study designs. Modelling the dependency structure between the dimensions can generate additional insight into the properties of the multivariate functional process, improves the estimation of random effects, and yields corrected confidence bands for covariate effects. Extensive simulation studies indicate that a multivariate modelling approach is more parsimonious than fitting independent univariate models to the data while maintaining or improving model fit.

Key words: functional additive mixed model; multivariate functional principal components; multivariate functional data; snooker trajectories; speech production

1 Introduction

With the technological advances seen in recent years, functional datasets are increasingly multivariate. They can be multivariate with respect to the domain of a function, its codomain, or both. Here, we focus on multivariate functions with a one-dimensional domain $f = (f^{(1)}, ..., f^{(D)}) : \mathcal{I} \subset \mathbb{R} \to \mathbb{R}^D$ with square-integrable components $f^{(d)} \in L^2(\mathcal{I})$, $d = 1, ..., D$. For this type of data, we can distinguish two subclasses: One has interpretable separate dimensions and can be seen as several complementary modes of a common phenomenon ('multimodal' data, cf. Uludağ and Roebroeck, 2014) as in the analysis of acoustic signals and articulation

---

Address for correspondence: Alexander Volkmann, Humboldt-Universität zu Berlin, School of Business and Economics, Chair of Statistics, Unter den Linden 6, Berlin 10099, Germany.
E-mail: alexander.volkmann@hu-berlin.de

processes in speech production in one of our data examples. The codomain then simply is the Cartesian product $\mathcal{S} = \mathcal{S}^{(1)} \times ... \times \mathcal{S}^{(D)}$ of interpretable univariate codomains $\mathcal{S}^{(d)} \subset \mathbb{R}$. The other subclass is more 'intrinsically' multivariate insofar as univariate analyses would not yield meaningful results. Consider for example two-dimensional movement trajectories as in one of our motivating applications, where the function measures Cartesian coordinates over time: for fixed trajectories, rotation or translation of the essentially arbitrary coordinate system would change the results of univariate analyses. For intrinsically multivariate functional data a multivariate approach is the natural and preferred mode of analysis, yielding interpretable results on the observation level. Even for multimodal functional data, a joint analysis may generate additional insight by incorporating the covariance structure between the dimensions. This motivates the development of statistical methods for multivariate functional data. We here propose multivariate functional additive mixed models to model potentially sparsely observed functions with flexible covariate effects and crossed or nested study designs.

Multivariate functional data have been the interest in different statistical fields such as clustering (Jacques and Preda, 2014; Park and Ahn, 2017), functional principal component analysis (FPCAs) (Chiou et al., 2014; Happ and Greven, 2018; Backenroth et al., 2018; Li et al., 2020), and registration (Carroll et al., 2021; Steyer et al., 2021). There is also ample literature on multivariate functional data regression such as graphical models (Zhu et al., 2016), reduced rank regression (Liu et al., 2020), or varying coefficient models (Zhu et al., 2012; Li et al., 2017). Yet, so far, there are only few approaches that can handle multilevel regression when the functional response is multivariate. In particular, Goldsmith and Kitago (2016) propose a hierarchical Bayesian multivariate functional regression model that can include subject level and residual random effect functions to account for correlation between and within functions. They work with bivariate functional data observed on a regular and dense grid and assume *a priori* independence between the different dimensions of the subject-specific random effects. Thus, they model the correlation between the dimensions only in the residual function. As our approach explicitly models the dependencies between dimensions for multiple functional random effects and also handles data observed on sparse and irregular grids on more than two dimensions, the model proposed by Goldsmith and Kitago (2016) can be seen as a special case of our more general model class.

Alternatively, Zhu et al. (2017) use a two-stage transformation with basis functions for the multivariate functional mixed model. This allows the estimation of scalar regression models for the resulting basis coefficients that are argued to be approximately independent. The proposed model is part of the so-called functional mixed model (FMM) framework (Morris, 2017). While FMMs use basis transformations of functional responses (observed on equal grids) at the start of the analysis, we propose a multivariate model in the functional additive mixed model (FAMM) framework, which uses basis representations of all (effect) functions in the model (Scheipl et al., 2015). The differences between these two functional regression frameworks have been extensively discussed before (Greven and Scheipl, 2017; Morris, 2017).

The main advantages of our multivariate regression model, also compared to Goldsmith and Kitago (2016) and Zhu et al. (2017), are that it is readily available for sparse and irregular functional data and that it allows to include multiple nested or crossed random processes, both of which are required in our data examples. Another important contribution is that our approach directly models the multivariate covariance structure of all random effects included in the model using multivariate functional principal components (FPCs) and thus implicitly models the covariances between the dimensions. This makes the model representation more parsimonious, avoids assumptions difficult to verify, and allows further interpretation of the random effect processes, such as their relative importance and their dominating modes. As part of the FAMM framework, our model provides a vast toolkit of modelling options for covariate and random effects, of estimation and inference (Wood, 2017). The proposed multivariate functional additive mixed model (multiFAMM) extends the FAMM framework combining ideas from multilevel modelling (Cederbaum et al., 2016) and multivariate functional data (Happ and Greven, 2018) to account for sparse and irregular functional data and different study designs.

We illustrate the multiFAMM on two motivating examples. The first (intrinsically multivariate) data stem from a study on the effect of a training programme for snooker players with a nested study design (shots within sessions within players) (Enghofer, 2014). The movement trajectories of a player's elbow, hand, and shoulder during a snooker shot are recorded on camera, yielding six-dimensional multivariate functional data (see Figure 1). In the second data example, we analyse multimodal data from a speech production study with a crossed study design (speakers crossed with words) (Pouplier and Hoole, 2016) on so-called 'assimilation' of consonants. The two measured modes (acoustic and articulatory, see Figure 3) are expected to be closely related but joint analyses have not yet incorporated the functional nature of the data.

These two examples motivate the development of a regression model for sparse and irregularly sampled multivariate functional data that can incorporate crossed or nested functional random effects as required by the study design in addition to flexible covariate effects. The proposed approach is implemented in R (R Core Team, 2020) in package `multifamm` (Volkmann, 2021). The article is structured as follows: Section 2 specifies the multiFAMM and Section 3 its estimation process. Section 4 presents the application of the multiFAMM to the data examples and Section 5 shows the estimation performance of our proposed approach in simulations. Section 6 closes with a discussion and outlook.

2 Multivariate functional additive mixed model

2.1 General model

Let $\boldsymbol{y}_i^*(t) = (y_i^{*(1)}(t), ..., y_i^{*(D)}(t))^\top$ be the multivariate functional response of unit $i = 1, ..., N$ over $t \in \mathcal{I}$, consisting of dimensions $d = 1, ..., D$. Without loss of generality, we assume a common one-dimensional interval domain $\mathcal{I} = [0, 1]$ for all

dimensions, and square-integrable $y_i^{*(d)} \in L^2(\mathcal{I})$. Define $L_D^2(\mathcal{I}) := L^2(\mathcal{I}) \times ... \times L^2(\mathcal{I})$ so that $y_i^* \in L_D^2(\mathcal{I})$. The underlying smooth function y_i^*, however, is only evaluated at (potentially sparse or dimension specific) points $y_{it}^* = (y_{it}^{*(1)}, ..., y_{it}^{*(D)})^\top$ and the evaluation is subject to white noise, that is, $y_{it} = y_{it}^* + \epsilon_{it}$. The residual term ϵ_{it} reflects additional uncorrelated white noise measurement error, following a D-dimensional multivariate normal distribution $\mathcal{N}_D$ with zero-mean and diagonal covariance matrix $\tilde{\Sigma} = \operatorname{diag}(\sigma_1^2, ..., \sigma_D^2)$ with dimension-specific variances σ_d^2. We construct a multivariate functional mixed model as

$$
\begin{aligned}
y_{it} = y_i^*(t) + \epsilon_{it} &= \mu(x_i, t) + U(t)z_i + \epsilon_{it} \\
&= \mu(x_i, t) + \sum_{j=1}^q U_j(t)z_{ij} + E_i(t) + \epsilon_{it}, \quad t \in \mathcal{I},
\end{aligned}
\tag{2.1}
$$

where

$$
U_j(t) = (U_{j1}(t), ..., U_{jV_{U_j}}(t)); j = 1, ..., q,
$$

$$
U_{jv}(t) \overset{\text{ind.c.}}{\sim} MGP\left(0, K_{U_j}\right); v = 1, ..., V_{U_j}; \forall j = 1, ..., q,
$$

$$
E_i(t) \overset{\text{ind.c.}}{\sim} MGP\left(0, K_E\right); i = 1, ..., N, \text{ and}
$$

$$
\epsilon_{it} \overset{\text{i.i.d.}}{\sim} \mathcal{N}_D\left(0, \tilde{\Sigma} = \operatorname{diag}(\sigma_1^2, ..., \sigma_D^2)\right); i = 1, ..., N; \quad t \in \mathcal{I}.
$$

We assume an additive predictor $\mu(x_i, \cdot) = \sum_{l=1}^p f_l(x_i, \cdot)$ of fixed effects, which consists of partial predictors $f_l(x_i, \cdot) = (f_l^{(1)}(x_i, \cdot), ..., f_l^{(D)}(x_i, \cdot))^\top \in L_D^2(\mathcal{I})$, $l = 1, ..., p$, that are multivariate functions depending on a subset of the vector of scalar covariates x_i. This allows to include linear or smooth covariate effects as well as interaction effects between multiple covariates as in the univariate FAMM (Scheipl et al., 2015). Partial predictors may also depend on dimension-specific subsets of covariates.

For random effects U, we focus on model scenarios with q independent multivariate functional random intercepts for crossed and/or nested designs. For group level $v = 1, ..., V_{U_j}$ within grouping layer $j = 1, ..., q$, these take the value $U_{jv} \in L_D^2(\mathcal{I})$. For each layer, the $U_{j1}, ..., U_{jV_{U_j}}$ present independent copies of a multivariate smooth zero-mean Gaussian random process. Analogously to scalar linear mixed models, the U_{jv} model correlations between different response functions y_i^* within the same group as well as variation across groups. By arranging them in a $(D \times V_{U_j})$ matrix $U_j(t)$ per t, the jth random intercept can be expressed in the common mixed model notation in (2.1) using appropriate group indicators $z_{ij} = (z_{ij1}, ..., z_{ijV_{U_j}})^\top$ for the respective design.

Although technically a curve-specific functional random intercept, we distinguish the smooth residuals $E_i \in L_D^2(\mathcal{I})$ in the notation, as they model correlation within

rather than between response functions. We write $E_v \in L_D^2(\mathcal{I})$, $v = 1, ..., V_E$ with $V_E = N$. The E_i capture smooth deviations from the group-specific mean $\mu(x_i, \cdot) + \sum_{j=1}^q U_j(\cdot)z_{ij}$.

For a more compact representation, we can arrange all $U_j(t)$ and $E_i(t)$ together in a $(D \times (\sum_{j=1}^q V_{U_j} + N))$ matrix $U(t)$ per t, and the group indicators for all layers in a corresponding vector $z_i = (z_{i1}^\top, ..., z_{iq}^\top, e_i^\top)^\top$ with e_i the i-th unit vector. The resulting model term $U(t)z_i$ then comprises all smooth random functions, accounting for all correlation between/within response functions y_i^* given the covariates x_i as required by the respective experimental design.

E_i and U_{jv} are independent copies (ind. c.) of random processes having multivariate $D \times D$ covariance kernels K_E, K_{U_j}, with univariate covariance surfaces $K_E^{(d,e)}(t, t') = \text{Cov}\left[E_i^{(d)}(t), E_i^{(e)}(t')\right]$ and $K_{U_j}^{(d,e)}(t, t') = \text{Cov}\left[U_{jv}^{(d)}(t), U_{jv}^{(e)}(t')\right]$ reflecting the covariance between the process dimensions d and e at t and t'. We call these auto-covariances for $d = e$ and cross-covariance otherwise. The multivariate Gaussian processes are uniquely defined by their multivariate mean function, here the null function 0, and the multivariate covariance kernels K_g and we write $MGP\left(0, K_g\right)$, $g \in \{U_1, ..., U_q, E\}$. Note that vectorizing the matrix $U(t)$ allows to formulate the joint distribution assumption $\text{vec}(U(t)) \sim MGP\left(0, K_U\right)$ with $K_U(t, t')$ having a block-diagonal structure repeating each $K_{U_j}(t, t')$ for V_{U_j} times and $K_E(t, t')$ for N times.

We assume that the different sources of variation $U_j(t)$, $j = 1, ..., q$, $E_i(t)$, and ϵ_{it} are mutually uncorrelated random processes to assure model identification. Assuming smoothness of the covariance kernel K_E further guarantees that the residual process $E_i(t)$ can be separated from the white noise ϵ_{it}, removing the error variance from the diagonal of the smooth covariance kernel (e.g., Yao et al., 2005).

2.2 FPC representation of the random effects

Model (2.1) specifies a univariate functional linear mixed model (FLMM) as given in Cederbaum et al. (2016) for each dimension d. The main difference lies in the multivariate random processes that introduce dependencies between the dimensions. In order to avoid restrictive assumptions about the structure of these multivariate covariance operators, which would typically be very difficult to elicit *a priori* or verify *ex post*, we estimate them directly from the data. The main difficulty then becomes computationally efficient estimation, which is already costly in the univariate case. Especially for higher dimensional multivariate functional data, accounting for the cross-covariances can become a complex task, which we tackle with multivariate functional principal component analysis (MFPCA).

Given the covariance operators (see Section 3), we represent the multivariate random effects in Model (2.1) using truncated multivariate Karhunen-Loève (KL)

expansions

$$U_{jv}(t) \approx \sum_{m=1}^{M_{U_j}} \rho_{U_j vm} \boldsymbol{\psi}_{U_j m}(t), \ j = 1, ..., q; \ v = 1, ..., V_{U_j},$$

$$\text{(2.2)}$$

$$E_v(t) \approx \sum_{m=1}^{M_E} \rho_{Evm} \boldsymbol{\psi}_{Em}(t), \ v = 1, ..., N,$$

where the orthonormal multivariate eigenfunctions $\boldsymbol{\psi}_{gm} = (\psi_{gm}^{(1)}, ..., \psi_{gm}^{(D)})^\top \in L_D^2(\mathcal{I})$, $m = 1, ..., M_g$, $g \in \{U_1, ..., U_q, E\}$ of the corresponding covariance operators with truncation order M_g are used as basis functions and the random scores $\rho_{gvm} \sim N(0, v_{gm})$ are independent and identically distributed (i.i.d.) with ordered eigenvalues v_{gm} of the corresponding covariance operator. Note that the assumption of Gaussianity for the random processes can be relaxed. For non-Gaussian random processes, the KL expansion still gives uncorrelated (but non-normal) scores and estimation based on a penalized least squares (PLS) criterion (see Section 3.2) remains reasonable.

Using KL expansions gives a parsimonious representation of the multivariate random processes that is an optimal approximation with respect to the integrated squared error (cf. Ramsay and Silverman, 2005), as well as interpretable basis functions capturing the most prominent modes of variation of the respective process. The distinct feature of this approach is that the multivariate FPCs directly account for the dependency structure of each random process across the dimensions. If, by contrast, for example, splines were used in the basis representation of the random effects, it would be necessary to explicitly model the cross-covariances of each random process in the model (cf. Li et al., 2020). Multivariate eigenfunctions, however, are designed to incorporate the dependency structure between dimensions and allow the assumption of independent (univariate) basis coefficients ρ_{gvm} via the KL theorem (see, e.g., Happ and Greven, 2018). This leads to a parsimonious multivariate basis for the random effects, where a typically small vector of scalar scores ρ_{gvm} common to all dimensions represents nearly the entire information about these D-dimensional processes.

3 Estimation

We use a two-step approach to estimate the multiFAMM and the respective multivariate covariance operators. In a first step (Section 3.1), the D-dimensional eigenfunctions $\boldsymbol{\psi}_{gm}(t)$ with their corresponding eigenvalues v_{gm} are estimated from their univariate counterparts following Cederbaum et al. (2018) and Happ and Greven (2018). These estimates are then plugged into (2.2) and we represent the multiFAMM as part of the general FAMM framework (Section 3.2) by suitable

re-arrangement. We can view the estimated $\boldsymbol{\psi}_{gm}(t)$ simply as an empirically derived basis that parsimoniously represents the patterns in the observed data. While their estimation adds uncertainty, we are not interested in inferential statements for the variance modes and our simulations (see Section 5) suggest that the estimated eigenfunctions are reasonable approximations that work well as a basis.

3.1 Step 1: Estimation of the eigenfunction basis

3.1.1 Step 1 (i): Univariate mean estimation

In a first step, we obtain preliminary estimates of the dimension-specific means $\mu^{(d)}(\boldsymbol{x}_i, t) = \sum_{l=1}^{p} f_l^{(d)}(\boldsymbol{x}_{il}, t)$ using univariate FAMMs. We model

$$y_{it}^{(d)} = \mu^{(d)}(\boldsymbol{x}_i, t) + \epsilon_{it}^{(d)}; \ d = 1, \dots, D \tag{3.1}$$

independently for all d with i.i.d. Gaussian random variables $\epsilon_{it}^{(d)}$. The estimation of $\mu^{(d)}(\boldsymbol{x}_i, t)$ proceeds analogously to the estimation of the multiFAMM described in Section 3.2. It is based on the evaluation points of the $y_i^{*(d)}(t)$, whose locations on the interval $\mathcal{I}$ can vary across dimensions. Model (3.1) thus accommodates sparse and irregular multivariate functional data and implies a working independence assumption across scalar observations within and across functions.

3.1.2 Step 1 (ii): Univariate covariance estimation

This preliminary mean function is used to centre the data $\tilde{y}_{it}^{(d)} = y_{it}^{(d)} - \hat{\mu}^{(d)}(\boldsymbol{x}_i, t)$ in order to obtain noisy evaluations of the detrended functions $\tilde{y}_i^{*(d)}(t) = y_i^{*(d)}(t) - \mu^{(d)}(\boldsymbol{x}_i, t)$ for covariance estimation. Cederbaum et al. (2016) already find that for this purpose, the working independence assumption within functions across evaluation points in (3.1) gives reasonable results. The expectation of the crossproducts of the centred functions then coincides with the auto-covariance, that is, $\mathbb{E}\left(\tilde{y}_{it}^{(d)}\tilde{y}_{i't'}^{(d)}\right) \approx$ $\mathrm{Cov}\left[y_{it}^{(d)}, y_{i't'}^{(d)}\right]$. For the independent random components specified in Model (2.1), this overall covariance decomposes additively into contributions from each random process as

$$\mathbb{E}\left(\tilde{y}_{it}^{(d)}\tilde{y}_{i't'}^{(d)}\right) \approx \sum_{j=1}^{q} K_{U_j}^{(d,d)}(t, t')\delta_{v_j v_j'} + \left(K_E^{(d,d)}(t, t') + \sigma_d^2 \delta_{tt'}\right)\delta_{ii'}, \tag{3.2}$$

using indicators $\delta_{xx'}$ that equal one for $x = x'$ and zero otherwise. The indicator $\delta_{v_j v_j'}$ thus identifies if the curves in the crossproduct belong to the same group v_j of the jth layer. Using t, t', and the indicators $\delta_{v_j v_j'}, \delta_{tt'}, \delta_{ii'}$ as covariates and the crossproducts of the centred data as responses, we can estimate the auto-covariances $K_{U_1}^{(d,d)}, ..., K_{U_q}^{(d,d)}$,

and $K_E^{(d,d)}$ of the random processes using symmetric additive covariance smoothing (Cederbaum et al., 2018). This extends the univariate approach proposed by Cederbaum et al. (2016). In particular, we also allow a nested random effects structure as required for the snooker training application in Section 4.1 by specifying the indicator of the nested effect as the product of subject-and-session indicators. Note that estimating (3.2) also yields estimates of the dimension-specific error variances σ_d^2 as a byproduct.

3.1.3 *Step 1 (iii): Univariate eigenfunction estimation*
Based on the covariance kernel estimates, we apply separate univariate FPCAs for each random process by conducting an eigendecomposition of the respective linear integral operator. Practically, each estimated process- and dimension-specific auto-covariance is re-evaluated on a dense grid so that a univariate functional principal component analysis (FPCA) can be conducted. Alternatively, Reiss and Xu (2020) provide an explicit spline representation of the estimated eigenfunctions. Eigenfunctions with non-positive eigenvalues are removed to ensure positive definiteness, and further regularization by truncation based on the proportion of variance explained is possible (see, e.g., Di et al., 2009; Peng and Paul, 2009; Cederbaum et al., 2016). However, we suggest to keep all univariate FPCs with positive eigenvalues for the computation of the MFPCA in order to preserve all important modes of variation and cross-correlation in the data.

3.1.4 *Step 1 (iv): Multivariate eigenfunction estimation*
The estimated univariate eigenfunctions and scores are then used to conduct an MFPCA for each of the g multivariate random processes separately. The MFPCA exploits correlations between univariate FPC scores across dimensions to reduce the number of basis functions needed to sufficiently represent the random processes. We base the MFPCA on the following definition of a (weighted) scalar product

$$\langle\langle f, g\rangle\rangle := \sum_{d=1}^{D} w_d \int_{\mathcal{I}} f^{(d)}(t) g^{(d)}(t)dt, \quad f, g \in L_D^2(\mathcal{I}), \tag{3.3}$$

for the response space with positive weights w_d, $d = 1, ..., D$ and the induced norm denoted by $|||\cdot|||$. The corresponding covariance operators $\Gamma_g : L_D^2(\mathcal{I}) \to L_D^2(\mathcal{I})$ of the multivariate random processes U_{jv} and E_v are then given by $(\Gamma_g f)(t) = \langle\langle f, K_g(t, \cdot)\rangle\rangle$, $g \in \{U_1, ..., U_q, E\}$. The standard choice of weights in our applications is $w_1 = ... = w_D = 1$ (unweighted scalar product) but other choices are possible. Consider for example a scenario where dimensions are observed with different amounts of measurement error. If variation in dimensions with a large proportion of measurement error is to be downweighted, we propose to use $w_d = \frac{1}{\hat{\sigma}_d^2}$ with the dimension-specific measurement error variance estimates $\hat{\sigma}_d^2$ obtained from (3.2).

Happ and Greven (2018) show that estimates of the multivariate eigenvalues ν_{gm} of Γ_g can be obtained from an eigenanalysis of a covariance matrix of the univariate random scores. The corresponding multivariate eigenfunctions $\boldsymbol{\psi}_{gm}$ can be obtained as linear combinations of the univariate eigenfunctions with the weights given by the resulting eigenvectors. The estimates $\hat{\boldsymbol{\psi}}_{gm}$ are then substituted for the basis functions of the truncated multivariate KL expansions of the random effects $\boldsymbol{U}_{jv}$ and $\boldsymbol{E}_v$ in (2.2). Note that for each random process g, the maximum number of FPCs is given by the total number of univariate eigenfunctions included in the estimation process of the MFPCA of g. To achieve further regularization and analogously to Cederbaum et al. (2016), we propose to choose truncation orders M_g for each KL expansion of the multivariate random processes using a prespecified proportion of explained variation.

3.1.5   Step 1 (v): Multivariate truncation order
We offer two different approaches for the choice of truncation orders M_g based on different variance decompositions (derivation in Supplementary Material A):

$$\mathbb{E}\left(|||\boldsymbol{y}_i - \boldsymbol{\mu}(\boldsymbol{x}_i)|||^2\right) = \sum_{d=1}^{D} w_d \int_{\mathcal{I}} \mathrm{Var}\big(y_i^{(d)}(t)\big)dt = \sum_{g}\sum_{m=1}^{\infty} \nu_{gm} + \sum_{d=1}^{D} w_d \sigma_d^2 |\mathcal{I}|, \quad (3.4)$$

$$\text{and} \quad \int_{\mathcal{I}} \mathrm{Var}\big(y_i^{(d)}(t)\big)dt = \sum_{g}\sum_{m=1}^{\infty} \nu_{gm} ||\psi_{gm}^{(d)}||^2 + \sigma_d^2 |\mathcal{I}| \quad (3.5)$$

with $|\mathcal{I}|$ the length of the interval $\mathcal{I}$ (here equal to one) and $||\cdot||$ the L^2 norm. Multivariate variance decomposition (3.4) uses the (weighted) sum of total variation in the data across dimensions. We select the FPCs with highest associated eigenvalues ν_{gm} over all random processes g until their sum reaches a prespecified proportion (e.g., 0.95) of the total variation, thus approximating the infinite sums in (3.4) with M_g summands. For the approach based on the univariate variance (3.5), we require M_g to be the smallest truncation order for which at least a prespecified proportion of variance is explained on every dimension d. This second choice of M_g might be preferable in situations where the variation is considerably different (in amount or structure) across dimensions, whereas the first approach gives a more parsimonious representation of the random effects. Note that both approaches can lead to a simplification of the multiFAMM if $M_g = 0$ is chosen for some g. The simulation results of Section 5 suggest that increasing the number of FPCs improves model accuracy which is why sensitivity analyses with regard to the truncation order are recommended.

3.2 Step 2: Estimation of the multiFAMM

In the following, we discuss estimating the multiFAMM given the estimated multivariate FPCs. We base the proposed model on the general FAMM framework of Scheipl et al. (2015), which models functional responses using basis representations. To make the extension of the FAMM framework to multivariate functional data more apparent, the multivariate response vectors and the respective model matrices are stacked over dimensions, so that every block has the structure of a univariate FAMM over all observations i. This gives concatenated basis functions with discontinuities between the dimensions. The fixed effects are modelled analogously to the univariate case by interacting all covariate effects with a dimension indicator. The random effects are based on the parsimonious, concatenated multivariate FPC basis.

3.2.1 Matrix representation

For notational simplicity we assume that the functions are evaluated on a fixed grid of time points $t = (t^{(1)\top}, ..., t^{(D)\top})^\top$ with $t^{(d)\top} = (t_1^{(d)}, ..., t_N^{(d)\top})$ and identical $t_i^{(d)} \equiv (t_1, ..., t_T)^\top$ over all N individuals and D dimensions. However, our framework allows for sparse functional data using different grids per dimension and per observed function as in the two applications (Section 4). Correspondingly, $y = (y^{(1)\top}, ..., y^{(D)\top})^\top$ is the DNT-vector of stacked evaluation points with $y^{(d)} = (y_1^{(d)\top}, ..., y_N^{(d)\top})^\top$ and $y_i^{(d)} = (y_{i1}^{(d)}, ..., y_{iT}^{(d)})^\top$. Model (2.1) on this grid can be written as

$$y = \Phi\theta + \Psi\rho + \epsilon \qquad (3.6)$$

with Φ, Ψ the model matrices for the fixed and random effects, respectively, θ, ρ the vectors of coefficients and random effect scores to be estimated, and $\epsilon = (\epsilon^{(1)\top}, ..., \epsilon^{(D)\top})^\top$, $\epsilon^{(d)} = (\epsilon_{11}^{(d)}, ..., \epsilon_{1T}^{(d)}, ..., \epsilon_{NT}^{(d)})^\top$ the vector of residuals. We have $\epsilon \sim N(0, \Sigma)$ with $\Sigma = \text{diag}(\sigma_1^2, ..., \sigma_D^2) \otimes I_{NT}$, the Kronecker product denoted by $\otimes$, and the $(NT \times NT)$ identity matrix I_{NT}.

We estimate θ and ρ by minimizing the PLS criterion

$$(y - \Phi\theta - \Psi\rho)\Sigma^{-1}(y - \Phi\theta - \Psi\rho)^\top + \sum_{l=1}^{p} \theta_l^\top P_l(\lambda_{xl}, \lambda_{tl})\theta_l + \sum_g \lambda_g \rho_g^\top P_g \rho_g \qquad (3.7)$$

using appropriate penalty matrices $P_l(\lambda_{xl}, \lambda_{tl})$ and P_g for the fixed effects and random effects, respectively, and smoothing parameters $\lambda_{xl} = \left(\lambda_{xl}^{(1)}, ..., \lambda_{xl}^{(D)}\right)$, $\lambda_{tl} = \left(\lambda_{tl}^{(1)}, ..., \lambda_{tl}^{(D)}\right)$, and λ_g. The model and penalty matrices as well as the parameter vectors of (3.6) and (3.7) are discussed in detail below.

3.2.2 Modelling of fixed effects

The block-diagonal matrix $\boldsymbol{\Phi} = \mathrm{diag}\left(\boldsymbol{\Phi}^{(1)}, ..., \boldsymbol{\Phi}^{(D)}\right)$ models the fixed effects separately on each dimension as in a FAMM (Scheipl et al., 2015). The $(DNT \times b)$ matrix $\boldsymbol{\Phi}$ consists of the design matrices $\boldsymbol{\Phi}^{(d)} = (\boldsymbol{\Phi}_1^{(d)} \mid ... \mid \boldsymbol{\Phi}_p^{(d)})$ that are constructed for the partial predictors $f_l^{(d)}(\boldsymbol{x}, \boldsymbol{t}^{(d)}), l = 1, ..., p,$ which correspond to the NT-vectors of evaluations of the effect functions $f_l^{(d)}$. The vectors of scalar covariates $\boldsymbol{x}_i$ are repeated T times to form the matrix of covariate information $\boldsymbol{x} = (\boldsymbol{x}_1, ..., \boldsymbol{x}_1, ..., \boldsymbol{x}_N)^\top$. We use the basis representations

$$f_l^{(d)}(\boldsymbol{x}, \boldsymbol{t}^{(d)}) \approx \boldsymbol{\Phi}_l^{(d)}\boldsymbol{\theta}_l^{(d)} = (\boldsymbol{\Phi}_{xl}^{(d)} \odot \boldsymbol{\Phi}_{tl}^{(d)})\boldsymbol{\theta}_l^{(d)},$$

where $\boldsymbol{A} \odot \boldsymbol{B}$ denotes the row tensor product $(\boldsymbol{A} \otimes \boldsymbol{1}_b^\top) \cdot (\boldsymbol{1}_a^\top \otimes \boldsymbol{B})$ of the $(h \times a)$ matrix $\boldsymbol{A}$ and the $(h \times b)$ matrix $\boldsymbol{B}$ with element-wise multiplication $\cdot$ and $\boldsymbol{1}_c$ the c-vector of ones. This modelling approach combines the $(NT \times b_{xl}^{(d)})$ basis matrix $\boldsymbol{\Phi}_{xl}^{(d)}$ with the $(NT \times b_{tl}^{(d)})$ basis matrix $\boldsymbol{\Phi}_{tl}^{(d)}$. These matrices contain the evaluations of suitable marginal bases in $\boldsymbol{x}$ and $\boldsymbol{t}^{(d)}$, respectively. For a linear effect, for example, the basis matrix $\boldsymbol{\Phi}_{xl}^{(d)}$ is specified as the familiar linear model design matrix $\boldsymbol{x}$ for the linear effect $f_l^{(d)}(\boldsymbol{x}, \boldsymbol{t}^{(d)}) = \boldsymbol{x}\boldsymbol{\beta}_l^{(d)}(\boldsymbol{t}^{(d)})$ with coefficient function $\boldsymbol{\beta}_l^{(d)}(\boldsymbol{t}^{(d)})$. For a nonlinear effect $f_l^{(d)}(\boldsymbol{x}, \boldsymbol{t}^{(d)}) = g_l^{(d)}(\boldsymbol{x}, \boldsymbol{t}^{(d)})$, the basis matrix $\boldsymbol{\Phi}_{xl}^{(d)}$ contains an (e.g., B-spline) basis representation analogously to a scalar additive model. For the functional intercept, $\boldsymbol{\Phi}_{xl}^{(d)}$ is a vector of ones, and we generally use a spline basis for $\boldsymbol{\Phi}_{tl}^{(d)}$. For a complete list of possible effect specifications with examples, we refer to Scheipl et al. (2015). The tensor product basis is weighted by the $b_{xl}^{(d)}b_{tl}^{(d)}$ unknown basis coefficients in $\boldsymbol{\theta}_l^{(d)}$. Stacking the vectors $\boldsymbol{\theta}_l^{(d)}$ gives $\boldsymbol{\theta}^{(d)} = (\boldsymbol{\theta}_1^{(d)\top}, ..., \boldsymbol{\theta}_p^{(d)\top})^\top$ and finally the b-vector $\boldsymbol{\theta} = (\boldsymbol{\theta}^{(1)\top}, ..., \boldsymbol{\theta}^{(D)\top})^\top$ with $b = \sum_d \sum_l b_{xl}^{(d)}b_{tl}^{(d)}$.

Choosing the number of basis functions is a well known challenge in the estimation of nonlinear or functional effects. We introduce regularization by a corresponding quadratic penalty term in (3.7). Let $\boldsymbol{\theta}_l$ contain the coefficients corresponding to the partial predictor l and order it by dimensions. The penalty $\boldsymbol{P}_l(\lambda_{xl}, \lambda_{tl})$ is then constructed from the penalty on the marginal basis for the covariate effect, $\boldsymbol{P}_{xl}^{(d)}$, and the penalty on the marginal basis over the functional index, $\boldsymbol{P}_{tl}^{(d)}$. Specifically, $\boldsymbol{P}_l(\lambda_{xl}, \lambda_{tl})$ is a block-diagonal matrix with blocks for each d corresponding to the Kronecker sums of the marginal penalty matrices $\lambda_{xl}^{(d)}\boldsymbol{P}_{xl}^{(d)} \otimes \boldsymbol{I}_{b_{tl}^{(d)}} + \lambda_{tl}^{(d)}\boldsymbol{I}_{b_{xl}^{(d)}} \otimes \boldsymbol{P}_{tl}^{(d)}$ (Wood, 2017). A standard choice for these marginal penalty matrices given a B-splines basis representation are second or third order difference penalties, thus approximately penalizing squared second or third derivatives of the respective functions (Eilers and

Marx, 1996). For unpenalized effects such as a linear effect of a scalar covariate, the corresponding $\boldsymbol{P}_{xl}^{(d)}$ is simply a matrix of zeroes.

3.2.3 Modelling of random effects

We represent the DNT-vectors $\boldsymbol{U}_j(\boldsymbol{t}) = (\boldsymbol{U}_j(\boldsymbol{t}^{(1)})^\top, ..., \boldsymbol{U}_j(\boldsymbol{t}^{(D)})^\top)^\top$, $\boldsymbol{E}(\boldsymbol{t}) = (\boldsymbol{E}(\boldsymbol{t}^{(1)})^\top, ...,$ $\boldsymbol{E}(\boldsymbol{t}^{(D)})^\top)^\top$ with $\boldsymbol{U}_j(\boldsymbol{t}^{(d)})$, $\boldsymbol{E}(\boldsymbol{t}^{(d)})$ containing the evaluations of the univariate random effects for the corresponding groups and time points using the basis approximations

$$\boldsymbol{U}_j(\boldsymbol{t}) \approx \boldsymbol{\Psi}_{U_j}\boldsymbol{\rho}_{U_j} = (\boldsymbol{\delta}_{U_j} \odot \tilde{\boldsymbol{\Psi}}_{U_j})\boldsymbol{\rho}_{U_j}, \quad \boldsymbol{E}(\boldsymbol{t}) \approx \boldsymbol{\Psi}_E\boldsymbol{\rho}_E = (\boldsymbol{\delta}_E \odot \tilde{\boldsymbol{\Psi}}_E)\boldsymbol{\rho}_E.$$

The vth column in the $(DNT \times V_g)$, $g \in \{U_1, ..., U_q, E\}$ indicator matrix $\boldsymbol{\delta}_g$ indicates whether a given row is from the vth group of the corresponding grouping layer. Thus, the rows of the indicator matrix $\boldsymbol{\delta}_g$ contain repetitions of the group indicators $\boldsymbol{z}_{ij}^\top$ and $\boldsymbol{e}_i^\top$ in model (2.1). For the smooth residual, $\boldsymbol{\delta}_E$ simplifies to $\mathbf{1}_D \otimes (\boldsymbol{I}_N \otimes \mathbf{1}_T)$. The $(DNT \times M_g)$ matrix $\tilde{\boldsymbol{\Psi}}_g = (\tilde{\boldsymbol{\Psi}}_g^{(1)\top}|...|\tilde{\boldsymbol{\Psi}}_g^{(D)\top})^\top$ comprises the evaluations of the M_g multivariate eigenfunctions $\psi_{gm}^{(d)}(t)$ on dimensions $d = 1, ..., D$ for the NT time points contained in the $(NT \times M_g)$ matrix $\tilde{\boldsymbol{\Psi}}_g^{(d)}$. The M_gV_g vector $\boldsymbol{\rho}_g = (\boldsymbol{\rho}_{g1}^\top, ..., \boldsymbol{\rho}_{gV_g}^\top)^\top$ with $\boldsymbol{\rho}_{gv} = (\rho_{gv1}, ..., \rho_{gvM_g})^\top$ stacks all the unknown random scores for the functional random effect g. The $(DNT \times \sum_g M_gV_g)$ model matrix $\boldsymbol{\Psi} = (\boldsymbol{\Psi}_{U_1}|...|\boldsymbol{\Psi}_{U_q}|\boldsymbol{\Psi}_E)$ then combines all random effect design matrices. Stacking the vectors of random scores in a $\sum_g M_gV_g$ vector $\boldsymbol{\rho} = (\boldsymbol{\rho}_{U_1}^\top, ..., \boldsymbol{\rho}_{U_q}^\top, \boldsymbol{\rho}_E^\top)^\top$ lets us represent all functional random intercepts in the model via $\boldsymbol{\Psi}\boldsymbol{\rho}$.

For a given functional random effect, the penalty takes the form $\boldsymbol{\rho}_g^\top \boldsymbol{P}_g \boldsymbol{\rho}_g = \boldsymbol{\rho}_g^\top(\boldsymbol{I}_{V_g} \otimes \tilde{\boldsymbol{P}}_g)\boldsymbol{\rho}_g$, where $\boldsymbol{I}_{V_g}$ corresponds to the assumed independence between the V_g different groups. The diagonal matrix $\tilde{\boldsymbol{P}}_g = \text{diag}(\nu_{g1}, ..., \nu_{gM_g})^{-1}$ contains the (estimated) eigenvalues ν_{gm} of the associated multivariate FPCs. This quadratic penalty is mathematically equivalent to a normal distribution assumption on the scores $\boldsymbol{\rho}_{gv}$ with mean zero and covariance matrix $\tilde{\boldsymbol{P}}_g^{-1}$, as implied by the KL theorem for Gaussian random processes. Note that the smoothing parameter λ_g allows for additional scaling of the covariance of the corresponding random process.

3.2.4 Estimation

We estimate the unknown smoothing parameters in $\boldsymbol{\lambda}_{xl}$, $\boldsymbol{\lambda}_{tl}$, and $\boldsymbol{\lambda}_g$ using fast restricted maximum likelihood (REML)-estimation (Wood, 2017). The standard identifiability constraints of FAMMs are used (Scheipl et al., 2015). In particular, in addition to the constraints for the fixed effects, the multivariate random intercepts are subject to a sum-to-zero constraint over all evaluation points as given by, for example, Goldsmith et al. (2016).

We propose a weighted regression approach to handle the heteroscedasticity assumption contained in $\mathbf{\Sigma}$. We weigh each observation proportionally to the inverse of the estimated univariate measurement error variances $\hat{\sigma}_d^2$ from the estimation of the univariate covariances (3.2). Alternatively, updated measurement error variances can be obtained from fitting separate univariate FAMMs on the dimensions using the univariate components of the multivariate FPCs basis. In practice, we found that the less computationally intensive former option gives reasonable results.

As our proposed model is part of the FAMM framework, inference for the multiFAMM is readily available based on inference for scalar additive mixed models (Wood, 2017). Note, however, that all inferential statements do not incorporate uncertainty due to the estimated multivariate eigenfunction bases, nor in the chosen smoothing parameters. The estimation process readily provides, amongst other things, standard errors for the construction of point-wise univariate confidence bands (CBs).

3.3 Implementation

We provide an implementation of the estimation of the proposed multiFAMM in the `multifamm` R-package (Volkmann, 2021). It is possible to include up to two functional random intercepts in $\boldsymbol{U}(t)$, which can have a nested or crossed structure, in addition to the curve-specific random intercept $\boldsymbol{E}_i(t)$. While including, for example, functional covariates is conceptually straightforward (see Scheipl et al., 2015), our implementation is restricted to scalar covariates and interactions thereof. We provide different alternatives for specifying the multivariate scalar product, the multivariate cut-off criterion, and the covariance matrix of the white noise error term. Note that the estimated univariate error variances have been proposed as weights for two separate and independent modelling decisions: as weights in the scalar product of the MFPCA and as regression weights under heteroscedasticity across dimensions.

4 Applications

We illustrate the proposed multiFAMM for two different data applications corresponding to intrinsically multivariate and multimodal fuctional data. The presentation focuses on the first application with a detailed description of the multimodal data application in Supplementary Material C. We provide the data and the code to produce all analyses in the Supplementary Material (`http://www.statmod.org/smij/archive.html`).

4.1 Snooker training data

4.1.1 Data set and preprocessing

In a study by Enghofer (2014), 25 recreational snooker players split into two groups, one of which had instructions to follow a self-administered training schedule over
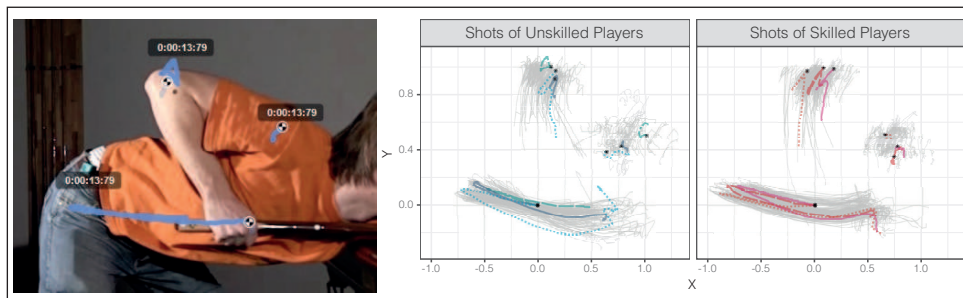
Figure 1 Screenshot of software for tracking (lines) the points of interest (circles) (left), two-dimensional trajectories of the snooker training data set (grey curves, right). For both groups of skilled and unskilled participants, three randomly selected observations are highlighted and every line type corresponds to one multivariate observation, that is, one observation consists of three trajectories: elbow (top), shoulder (right) and hand (bottom). The start of the exemplary trajectories are marked with a black asterisk with the hand trajectory centred at the origin

the next six weeks consisting of exercises aimed at improving snooker specific muscular coordination. The second was a control group. Before and after the training period, both groups were recorded on high-speed digital camera under similar conditions to investigate the effects of the training on their snooker shot of maximal force. In each of the two recording sessions, six successful shots per participant were videotaped. The recordings were then 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. This yields a six-dimensional functional observation per snooker shot $y^* = (y^{*(\text{elbow.x})}, ..., y^{*(\text{shoulder.y})}) : \mathcal{I} = [0, 1] \to \mathbb{R}^6$, that is, a two-dimensional movement trajectory for each point of interest (see Figure 1).

In their starting position (hand centred at the origin), the snooker players are positioned centrally in front of the snooker table aiming at the cue ball. From their starting position, the players draw back the cue, then accelerate it forwards and hit the cue ball shortly after their hands enter the positive range of the horizontal x-axis. After the impulse onto the cue ball, the hand movement continues until it is stopped at a player's chest. Enghofer (2014) identify two underlying techniques that a player can apply: dynamic and fixed elbow. With a dynamic elbow, the cue can be moved in an almost straight line (piston stroke) whereas additionally fixing the elbow results in a pendular motion (pendulum stroke). In both cases, the shoulder serves as a fixed point and should be positioned close to the snooker table.

We adjust the data for differences in body height and relative speed (Steyer et al., 2021) and apply a coarsening method to reduce the number of redundant data points, thereby lowering computational demands of the analysis. Supplementary Material B provides a detailed description of the data preprocessing. As some recordings and evaluations of bivariate trajectories are missing, the final dataset contains 295 functional observations with a total of 56,910 evaluation points. These multivariate

functional data are irregular and sparse, with a median of 30 evaluation points per functional observation (minimum 8, maximum 80) for each of the six dimensions.

### 4.1.2 Model specification
We estimate the following model

$$y_{ijht} = \mu(x_{ij}, t) + B_i(t) + C_{ij}(t) + E_{ijh}(t) + \epsilon_{ijht}, \tag{4.1}$$

with $i = 1, ..., 25$ the index for the snooker player, $j = 1, 2$ the index for the session, $h = 1, ..., H_{ij}$ the index for the typically six snooker shot repetitions in a session, and $t \in [0, 1]$ relative time. Correspondingly, $B_i(t)$ is a subject-specific random intercept, $C_{ij}(t)$ is a nested subject-and-session-specific random intercept, and $E_{ijh}(t)$ is the shot-specific random intercept (smooth residual). The nested random effect $C_{ij}(t)$ is supposed to capture the variation within players between sessions (e.g., differences due to players having a good or bad day). Different positioning of participants with respect to the recording equipment or the snooker table as well as shot to shot variation are captured by the smooth residual $E_{ijh}(t)$. The white noise measurement error ϵ_{ijht} is assumed to follow a zero-mean multivariate normal distribution with covariance matrix $\sigma^2 I_6$, as all six dimensions are measured with the same set-up. The additive predictor is defined as

$$\mu(x_{ij}, t) = f_0(t) + \texttt{skill}_i \cdot f_1(t) + \texttt{group}_i \cdot f_2(t) + \texttt{session}_j \cdot f_3(t)$$
$$+ \texttt{group}_i \cdot \texttt{session}_j \cdot f_4(t).$$

The dummy covariates $\texttt{skill}_i$ and $\texttt{group}_i$ indicate whether player i is an advanced snooker player and belongs to the treatment group (i.e., receives the training programme), respectively. Note that the snooker players self-select into training and control group to improve compliance with the training programme, which is why we include a group effect in the model. The dummy covariate $\texttt{session}_j$ indicates whether the shot j is recorded after the training period. The effect function $f_4(t)$ can thus be interpreted as the treatment effect of the training programme.

Cubic P-splines with first-order difference penalty, penalizing deviations from constant functions over time, with 8 basis functions are used for all effect functions in the preliminary mean estimation as well as in the final multiFAMM. For the estimation of the auto-covariances of the random processes, we use cubic P-splines with first-order difference penalty on five marginal basis functions. We use an unweighted scalar product (3.3) for the MFPCA to give equal weight to all spatial dimensions, as we can assume that the measurement error mechanism is similar across dimensions. Additionally, we find that hand, elbow, and shoulder contribute roughly the same amount of variation to the data, cf. Table 1 in Supplementary Material B.3, where we also discuss potential weighting schemes for the MFPCA. The multivariate truncation order is chosen such that 95% of the (unweighted) sum of variation (3.4) is explained.
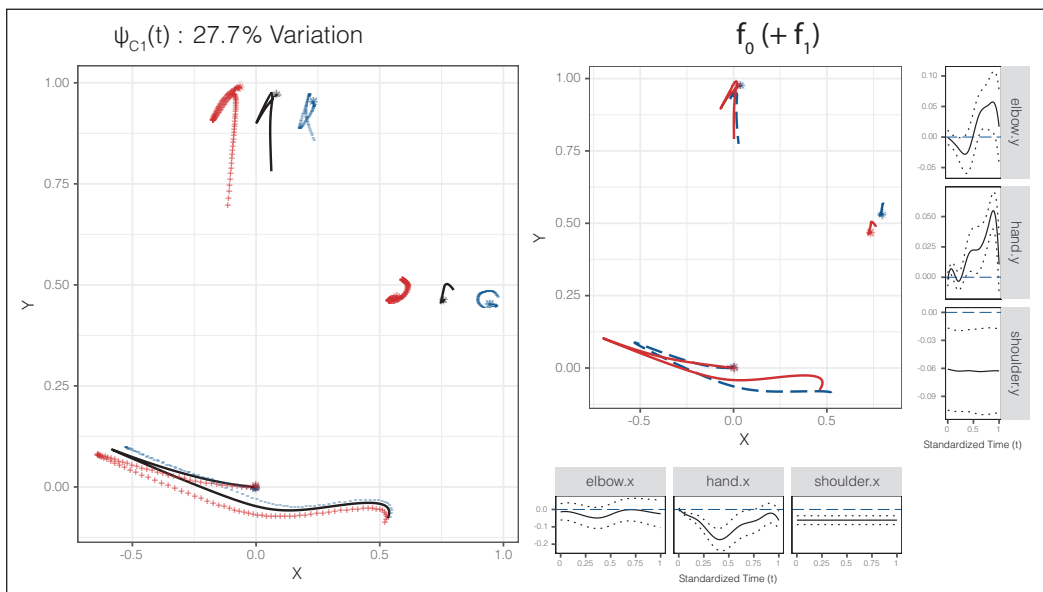
Figure 2 Dominant mode (ψ_{C1}) of the subject-and-session-specific random effect, explaining 27.7% of total variation and shown as mean trajectory (black solid) plus (+) or minus (−) $2\sqrt{\nu_{C1}}$ times the first FPC (left). An asterisk marks the start of a trajectory. Estimated covariate effect functions for skill (right). The central plot shows the effect of the coefficient function (solid) on the two-dimensional trajectories for the reference group (dashed). The marginal plots show the estimated univariate effect functions (solid) with pointwise 95% CBs (dotted) and the baseline (dashed)

4.1.3 Results

The MFPCA gives sets of five (for **C** and **E**) and six (for **B**) multivariate FPCs that explain 95% of the total variation. The estimated eigenvalues allow to quantify their relative importance. Approximately 41% of the total variation (conditional on covariates) can be attributed to the nested subject-and-session-specific random intercept $C_{ij}(t)$, 33% to the subject-specific random intercept $B_i(t)$, 14% to the shot-specific $E_{ijh}(t)$, and 7% to white noise. This suggests that day to day variation within a snooker player is larger than the variation between snooker players. Note that these proportions are based on estimation step 1 (see Section 3.1).

The left plot of Figure 2 displays the first FPC for **C**, which explains about 28% of total variation. A suitable multiple of the FPCs is added (+) to and subtracted (−) from the overall mean function (black solid line, all covariate values set to 0.5). We find that the dominant mode of the random subject-and-session-specific effect influences the relative positioning of a player's elbow, shoulder, and hand, thus suggesting a strong dependence between the dimensions. Enghofer (2014) argue from a theoretical viewpoint that the ideal starting position should place elbow and hand in a line perpendicular to the plane of the snooker table (corresponding to the *x*-axis). The most prominent mode of variation captures deviations from this ideal starting position found in the overall mean. The next most important

FPC ψ_{B1} of the subject-specific random effect, which explains about 15% of total variation, represents a subject's tendency towards the piston or pendulum stroke (see Supplementary Material Figure 4). This additional insight into the underlying structure of the variance components might be helpful for, for example, developing personalized training programmes.

The central plot on the right of Figure 2 compares the estimated mean movement trajectory for advanced snooker players (solid line) to that in the reference group (dashed). It suggests that more experienced players tend towards the dynamic elbow technique, generating a hand trajectory resembling a straight line (piston stroke). Uncertainties in the trajectory could be represented by pointwise ellipses, but inference is more straightforward to obtain from the univariate effect functions. The marginal plots display the estimated univariate effects with pointwise 95% confidence intervals. Even though we find only little statistical evidence for increased movement of the elbow (horizontal-left and vertical-top marginal panels), the hand and shoulder movements (horizontal centre and right, vertical centre and bottom) strongly suggest that the skill level indeed influences the mean movement trajectory of a snooker player. Further results indicate that the mean hand trajectories might slightly differ between treatment and control group at baseline as well as between sessions ($f_2(t)$ and $f_3(t)$, see Supplementary Material Figure 8). The estimated treatment effect $f_4(t)$ (Supplementary Material Figure 7), however, suggests that the training programme did not change the participants' mean movement trajectories substantially. Supplementary Material B.3 contains a detailed discussion of all model terms as well as some model diagnostics and sensitivity analyses.

4.2 Consonant assimilation data

4.2.1 Data set and model specification

Pouplier and Hoole (2016) study the assimilation of the German /s/ and /sh/ sounds such as the final consonant sounds in 'Kürbis' (English example: 'haggis') and 'Gemisch' (English example: 'dish'), respectively. The research question is how these sounds assimilate in fluent speech when combined across words such as in 'Kürbis-Schale' or 'Gemisch-Salbe', denoted as /s#sh/ and /sh#s/ with # the word boundary. The 9 native German speakers in the study repeated a set of 16 selected word combinations five times. Two different types of functional data, that is, acoustic (ACO) and electropalatographic (EPG) data, were recorded for each repetition to capture the acoustic (produced sound) and articulatory (tongue movements) aspects of assimilation over (relative) time t within the consonant combination.

Each functional index varies roughly between +1 and −1 and measures how similar the articulatory or acoustic pattern is to its reference patterns for the first (+1) and second (−1) consonant at every observed time point (Cederbaum et al., 2016). Without assimilation, the data are thus expected to shift from positive to negative values in a sinus-like form (see Figure 3). The dataset contains 707 bivariate functional observations with differently spaced grids of evaluation points per curve and dimension, with the number of evaluation points ranging from 22 to 59 with a median of 35. Note that the consonant assimilation data are unaligned as registration
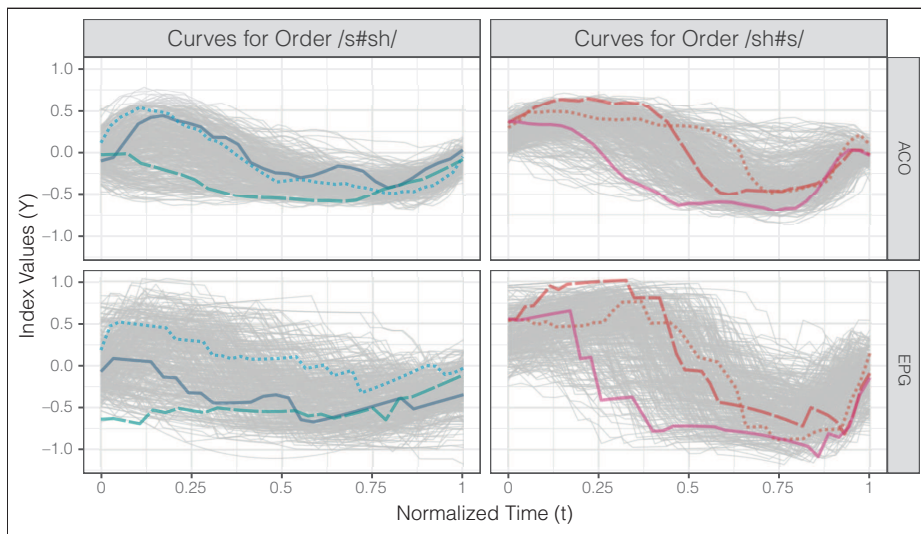
Figure 3 Index curves of the consonant assimilation dataset for both ACO and EPG data as a function of standardized time *t* (grey curves). For every consonant order, three randomly selected observations have been highlighted and every line type corresponds to one multivariate observation, that is, one observation consists of two index curves

of the time domain would mask transition speeds between the consonants, which are an interesting part of assimilation.

For comparability, we follow the model specification of Cederbaum et al. (2016), who analyse only the ACO dimension and ignore the second mode EPG. Our specified multivariate model is similar to (4.1) with $i = 1, ..., 9$ the speaker index, $j = 1, ..., 16$ the word combination index, $h = 1, ..., H_{ij}$ the repetition index and $t \in [0, 1]$ relative time. Note that the nested effect $C_{ij}(t)$ is replaced by the crossed random effect $C_j(t)$ specific to the word combinations. The additive predictor $\mu(x_j, t)$ now contains eight partial effects: the functional intercept plus main and interaction effects of scalar covariates describing characteristics of the word combination such as the order of the consonants /s/ and /sh/. The white noise measurement error ϵ_{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 basis and penalty specifications follow the univariate analysis in Cederbaum et al. (2016). Given different sampling mechanisms, we also compare the multiFAMM based on weighted and unweighted scalar products for the MFPCA.

4.2.2 Results
The multivariate analysis supports the findings of Cederbaum et al. (2016) that assimilation is asymmetric (different mean patterns for /s#sh/ and /sh#s/). Overall, the estimated fixed effects are similar across dimensions as well as comparable to the univariate analysis. Hence, the multivariate analysis indicates that previous results for the acoustics are consistently found also for the articulation. Compared to univariate

analyses, our approach reduces the number of FPC basis functions and thus the number of parameters in the analysis. The multiFAMM can improve the model fit and can provide smaller CBs for the ACO dimension compared to the univariate model in Cederbaum et al. (2016) due to the strong cross-correlation between the dimensions. We find similar modes of variation for the multivariate and the univariate analysis as well as across dimensions. In particular, the word combination-specific random effect $C_j(t)$ is dropped from the model as much of the between-word variation is already explained by the included fixed effects. The definition of the scalar product has little effect on the estimated fixed effects but changes the interpretation of the FPCs. Supplementary Material C contains a more in-depth description of this application.

5 Simulations

5.1 Simulation set-up

We conduct an extensive simulation study to investigate the performance of the multiFAMM depending on different model specifications and data settings (over 20 scenarios total), and to compare it to univariate regression models as proposed by Cederbaum et al. (2016), estimated on each dimension independently. Given the broad scope of analysed model scenarios, we refer the interested reader to Supplementary Material D for a detailed report and restrict the presentation here to the main results.

We mimic our two presented data examples (Section 4) and simulate new data based on the respective multiFAMM-fit. Each scenario consists of model fits to 500 generated datasets, where we randomly draw the number and location of the evaluation points, the random scores, and the measurement errors according to different data settings. The accuracy of the estimated model components is measured by the root relative mean squared error (rrMSE) based on the unweighted multivariate norm but otherwise as defined by Cederbaum et al. (2016), see Supplementary Material D.1. The rrMSE takes on (unbounded) positive values with smaller values indicating a better fit.

5.2 Simulation results

Figure 4 compares the rrMSE values over selected modelling scenarios based on the consonant assimilation data. We generate a benchmark scenario (far left boxplots), which imitates the original data without misspecification of any model component. In particular, the number of FPCs is fixed to avoid truncation effects. Comparing this scenario to the two scenarios left and centre illustrates the importance of the number of FPCs in the accuracy of the estimation. Choosing the truncation order via the proportion of univariate variance explained (Cut-Off Uni) as in (3.5) gives models with roughly the same number of FPCs (mean $B : 2.8, E : 5$) as is used for the data generation ($B : 3, E : 5$). The cut-off criterion based on the multivariate variance (Cut-Off Mul) given by (3.4) results in more parsimonious models (mean
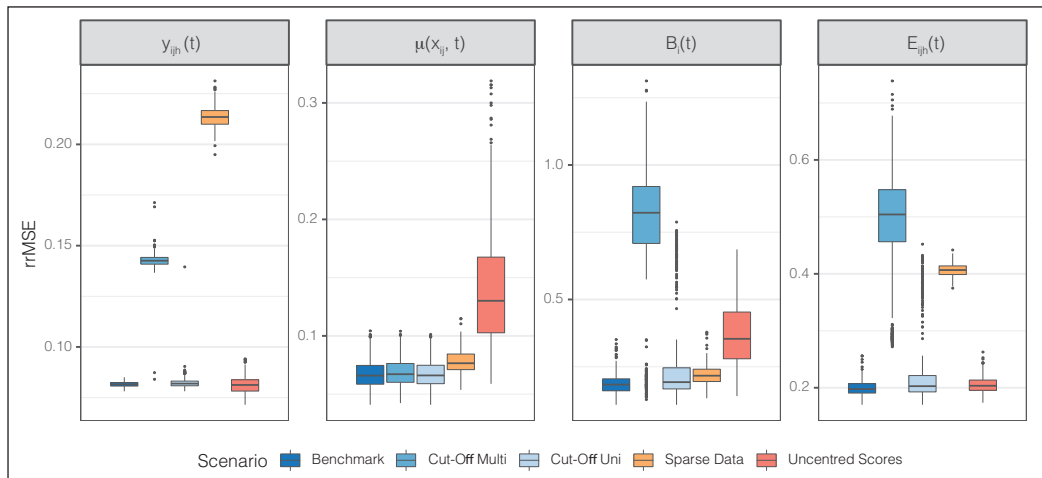
Figure 4 rrMSE values of the fitted curves $y_{ijh}(t)$, the mean $\mu(x_{ij}, t)$, and the random effects $B_i(t)$ and $E_{ijh}(t)$ for different modelling scenarios. The three leftmost scenarios correspond to different model specifications in the same data setting

$B : 2.15, E : 4)$ and thus considerably higher rrMSE values. The increased variation in the rrMSE values can also be attributed to variability in the truncation orders (cf. Supplementary Material Figure 19), leading to a mixture distribution. Comparing the benchmark scenario to more sparsely observed functional data (ceteris paribus) suggests a lower estimation accuracy for the Sparse Data scenario (right), especially for the curve-specific random effect $E_{ijh}(t)$ and resultingly the fitted curves $y_{ijh}(t)$, but pooling the information across functions helps the estimation of $\mu(x_{ij}, t)$ and $B_i(t)$. In particular, the estimation of the mean $\mu(x_{ij}, t)$ is quite robust against the increased uncertainty of these three scenarios. Only when the random scores are not centred and decorrelated as in the benchmark scenario do we find an increase in rrMSE values for the mean (Uncentred Scores, far right). This corresponds to a departure from the modelling assumptions likely to occur in practice when only few levels of a random effect are available (here for the subject-specific $B_i(t)$). The model then has difficulties to correctly separate the intercept in $\mu(x_{ij}, t)$ and the random effects $B_i(t)$. The empirical (non-zero) mean of the $B_i(t)$ is then absorbed by the intercept in $\mu(x_{ij}, t)$, resulting in higher rrMSE values for both of these model terms. However, this shift does not affect the overall fit to the data $y_{ijh}(t)$ nor the estimation of the other fixed effects (cf. Supplementary Material Figure 27). Note that the rrMSE values of the Sparse Data and Uncentred Scores scenarios are based on slightly different normalizing constants (i.e., different true data) and cannot be directly compared except for the mean.

Our simulation study thus suggests that basing the truncation orders on the proportion of explained variation on each dimension (3.5) gives parsimonious and well-fitting models. If interest lies mainly in the estimation of fixed effects, the alternative cut-off criterion based on the total variation in the data (3.4)

allows even more parsimonious models at the cost of a less accurate estimation of the random effects and overall model fit. Furthermore, the results presented in Supplementary Material D show that the mean estimation is relatively stable over different model scenarios including misspecification of the measurement error variance structure or of the multivariate scalar product, as well as in scenarios with strong heteroscedasticity across dimensions. In our benchmark scenario, the CBs cover the true effect $89 - 94\%$ of the time but coverage can further decrease with additional uncertainty, for example, about the number of FPCs. Overall, the covariance structure such as the leading FPCs can be recovered well, also for a nested random effect such as in the snooker training application. The comparison to the univariate modelling approach suggests that the multiFAMM can improve the mean estimation but is especially beneficial for the prediction of the random effects while reducing the number of parameters to estimate. In some cases like strong heteroscedasticity, including weights in the multivariate scalar product might further improve the modelling.

6   Discussion

The proposed multivariate functional regression model is an additive mixed model, which allows to model flexible covariate effects for sparse or irregular multivariate functional data. It uses FPC based functional random effects to model complex correlations within and between functions and dimensions. An important contribution of our approach is estimating the parsimonious multivariate FPC basis from the data. This allows us to account not only for auto-covariances, but also for non-trivial cross-covariances over dimensions, which are difficult to adequately model using alternative approaches such as parametric covariance functions like the Matèrn family or penalized splines, which imply a parsimonious covariance only within but not necessarily between functions. As a FAMM-type regression model, a wide range of covariate effect types is available, also providing pointwise CBs. Our applications show that the multiFAMMs can give valuable insight into the multivariate correlation structure of the functions in addition to the mean structure.

An apparent benefit of multivariate modelling is that it allows to answer research questions simultaneously relating to different dimensions. In addition, using multivariate FPCs reduces the number of parameters compared to fitting comparable univariate models while improving the random effects estimation by incorporating the cross-covariance in the multivariate analysis. The added computational costs are small: For our multimodal application, the multivariate approach prolongs the computation time by only 5% (104 vs. 109 minutes on a 64-bit Linux platform).

We find that the average point-wise coverage of the point-wise CBs can in some cases lie considerably below the nominal value. There are two main reasons for this: One, the CBs presented here do not incorporate the uncertainty of the eigenfunction estimation nor of the smoothing parameter selection. Two, coverage issues can arise in (scalar) mixed models, if effect functions are estimated as constant when in truth they are not (e.g., Wood, 2017; Greven and Scheipl, 2016). To resolve these issues,

further research on the level of scalar mixed models might be needed. A large body of research covering CB estimation for functional data (e.g., Goldsmith et al., 2013; Choi and Reimherr, 2018; Liebl and Reimherr, 2019) suggests that the construction of CBs is an interesting and complex problem, also outside of the FAMM framework.

It would be interesting to extend the multiFAMM to more general scenarios of multivariate functional data such as observations consisting of functions with different dimensional domains, for example, functions over time and images as in Happ and Greven (2018). This would require adapting the estimation of the univariate auto-covariances for spatial arguments t, t'. Exploiting properties of dense functional data, such as the block structure of design matrices for functions observed on a grid, could help to reduce computational cost in this case. Future research could further generalize the covariance structure of the multiFAMM by allowing for additional covariate effects. In our snooker training application, for example, a treatment effect of the snooker training might show itself in the form of reduced intra-player variance (cf. Backenroth et al., 2018). Ideas from distributional regression could be incorporated to jointly model the mean trajectories and covariance structure conditional on covariates.

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship and/or publication of this article.

Supplemental material

Supplementary materials for this article are available from
http://www.statmod.org/smij/archive.html

References

Backenroth D, Goldsmith J, Harran MD, Cortes JC, Krakauer JW and Kitago T (2018) Modelling motor learning using heteroscedastic functional principal components analysis. *Journal of the American Statistical Association*, **113**, 1003–1015.

Carroll C, Müller H-G and Kneip A (2021) Cross-component registration for multivariate functional data, with application to growth curves. *Biometrics*, **77**, 839–51.

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

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

Chiou J-M, Chen Y-T and Yang Y-F (2014) Multivariate functional principal component analysis: A normalization approach. *Statistica Sinica*, **24**, 1571–96.

Choi H and Reimherr M (2018) A geometric approach to confidence regions and bands for functional parameters. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **80**, 239–60.

Di C-Z, Crainiceanu CM, Caffo BS and Punjabi NM (2009) Multilevel functional principal component analysis. *The Annals of Applied Statistics*, **3**, 458.

Eilers PH and Marx BD (1996) Flexible smoothing with B-splines and penalties. *Statistical Science*, **11**, 89–102.

Enghofer T (2014) *Überblick über die Sportart snooker, Entwicklung eines Muskeltraining und Untersuchung dessen Einflusses auf die Stoßtechnik* [Overview of snooker as a sport, development of a muscular training programme, and analysis of this training programme's influence on the Snooker shot]. Unpublished thesis. Technische Universität München.

Goldsmith J and Kitago T (2016) Assessing systematic effects of stroke on motor control by using hierarchical function-on-scalar regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, **65**, 215–36.

Goldsmith J, Greven S and Crainiceanu C (2013) Corrected confidence bands for functional data using principal components. *Biometrics*, **69**, 41–51.

Goldsmith J, Scheipl F, Huang L, Wrobel J, Gellar J, Harezlak J, McLean MW, Swihart B, Xiao L, Crainiceanu C and Reiss, PT (2016) *refund: Regression with Functional Data*. URL https://cran.r-project.org/web/packages/refund/refund.pdf (last accessed 25 October 2021).

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

Greven S and Scheipl F (2017) A general framework for functional regression modelling. *Statistical Modelling*, **17**, 1–35, 100–115.

Happ C and Greven S (2018) Multivariate functional principal component analysis for data observed on different (dimensional) domains. *Journal of the American Statistical Association*, **113**, 649–59.

Jacques J and Preda C (2014) Model-based clustering for multivariate functional data. *Computational Statistics & Data Analysis*, **71**, 92–106.

Li C, Xiao L and Luo S (2020) Fast covariance estimation for multivariate sparse functional data. *Stat*, **9**, e245.

Li J, Huang C, Hongtu Z and Alzheimer's Disease Neuroimaging Initiative (2017) A functional varying-coefficient single-index model for functional response data. *Journal of the American Statistical Association*, **112**, 1169–81.

Liebl D and Reimherr M (2019) Fast and fair simultaneous confidence bands for functional parameters. *arXiv preprint arXiv:1910.00131*.

Liu Y, Yan B, Merikangas K and Shou H (2020) Graph-fused multivariate regression via total variation regularization. *arXiv preprint arXiv:2001.04968*.

Morris JS (2017) Comparison and contrast of two general functional regression modelling

frameworks. *Statistical Modelling*, **17**, 59–85.

Park J and Ahn J (2017) Clustering multivariate functional data with phase variation. *Biometrics*, **73**, 324–33.

Peng J and Paul D (2009) A geometric approach to maximum likelihood estimation of the functional principal components from sparse longitudinal data. *Journal of Computational and Graphical Statistics*, **18**, 995–1015.

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

R Core Team (2020) *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. URL https://www.R-project.org (last accessed 25 October 2021).

Ramsay JO and Silverman BW (2005) *Functional data analysis, 2nd edition*. Springer Science & Business Media.

Reiss PT and Xu M (2020) Tensor product splines and functional principal components. *Journal of Statistical Planning and Inference*, **208**, 1–12.

Scheipl F, Staicu A-M and Greven S (2015) Functional additive mixed models. *Journal of Computational and Graphical Statistics*, **24**, 477–501.

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

Uludağ K and Roebroeck A (2014) General overview on the merits of multimodal neuroimaging data fusion. *Neuroimage*, **102**, 3–10.

Volkmann A (2021) *multifamm: Multivariate Functional Additive Mixed Models*. URL https://cran.r-project.org/web/packages/multifamm (last accessed 25 October 2021).

Wood SN (2017) *Generalized additive models: An introduction with R, 2nd edition*. Chapman and Hall/CRC Press.

Yao F, Müller H-G and Wang J-L (2005) Functional data analysis for sparse longitudinal data. *Journal of the American Statistical Association*, **100**, 577–90.

Zhu H, Li R and Kong L (2012) Multivariate varying coefficient model for functional responses. *Annals of Statistics*, **40**, 2634.

Zhu H, Morris JS, Wei F and Cox DD (2017) Multivariate functional response regression, with application to fluorescence spectroscopy in a cervical pre- cancer study. *Computational Statistics & Data Analysis*, **111**, 88–101.

Zhu H, Strawn N and Dunson DB (2016) Bayesian graphical models for multivariate functional data. *The Journal of Machine Learning Research*, **17**, 7157–83.