The Annals of Applied Statistics
2020, Vol. 14, No. 3, 1164–1181
https://doi.org/10.1214/20-AOAS1324

OPTIMAL EMG PLACEMENT FOR A ROBOTIC PROSTHESIS CONTROLLER WITH SEQUENTIAL, ADAPTIVE FUNCTIONAL ESTIMATION (SAFE)

BY JONATHAN STALLRICH[1,*], MD NAZMUL ISLAM[1,†], ANA-MARIA STAICU[1,‡], DUSTIN CROUCH[2], LIZHI PAN[3] AND HE HUANG[4]

[1]*Department of Statistics, North Carolina State University,* [*]*jwstalli@ncsu.edu;* [†]*mnislam@ncsu.edu;* [‡]*astaicu@ncsu.edu*
[2]*Department of Mechanical, Aerospace, and Biomedical Engineering, University of Tennessee, Knoxville, dcrouch3@utk.edu*
[3]*School of Mechanical Engineering, Tianjin University, melzpan@tju.edu.cn*
[4]*Joint Department of Biomedical Engineering, University of North Carolina and North Carolina State University, helen-huang@unc.edu*

Robotic hand prostheses require a controller to decode muscle contraction information, such as electromyogram (EMG) signals, into the user's desired hand movement. State-of-the-art decoders demand extensive training, require data from a large number of EMG sensors and are prone to poor predictions. Biomechanical models of a single movement degree-of-freedom tell us that relatively few muscles, and, hence, fewer EMG sensors are needed to predict movement. We propose a novel decoder based on a dynamic, functional linear model with velocity or acceleration as its response and the recent past EMG signals as functional covariates. The effect of each EMG signal varies with the recent position to account for biomechanical features of hand movement, increasing the predictive capability of a single EMG signal compared to existing decoders. The effects are estimated with a multistage, adaptive estimation procedure that we call Sequential Adaptive Functional Estimation (SAFE). Starting with 16 potential EMG sensors, our method correctly identifies the few EMG signals that are known to be important for an able-bodied subject. Furthermore, the estimated effects are interpretable and can significantly improve understanding and development of robotic hand prostheses.

1. Introduction. Over 160,000 Americans are transradial (i.e., below-elbow) amputees, henceforth TRAs, and must learn how to perform tasks without their intact hand (Ziegler-Graham et al. (2008)). Passive hand prostheses are useful but cannot fully emulate the functionality of an intact hand. State-of-the-art multifunctional robotic prostheses, such as the FDA-approved DEKA arm system (Resnik (2011), Resnik et al. (2011)), have more advanced mechanical systems that move according to a prosthesis controller (PC) that decodes user inputs into movement. These inputs are often electromyogram (EMG) signals because they measure activation of muscles due to contractions which are known to cause hand movement for an able-bodied (AB) subject.

Figure 1 summarizes the movement process for an AB subject and a TRA equipped with a robotic prosthesis. For both AB subjects and TRAs, intended hand movement begins with initiation of action potentials that are conducted along motor neurons from the central nervous system. The neural signals travel along the motor neural pathway and propagate onto the muscle fibers, causing the muscle to contract. The action potential measured from the muscle fibers is called motor unit action potential (MUAP). An EMG signal taken from a surface electrode placed on an intact or residual forearm represents the sum of MUAPs across one or

Received April 2018; revised September 2019.
Key words and phrases. Electromyography signal, varying functional regression, functional variable selection, adaptive group LASSO, correlated functional predictors, sequential adaptive functional estimation.

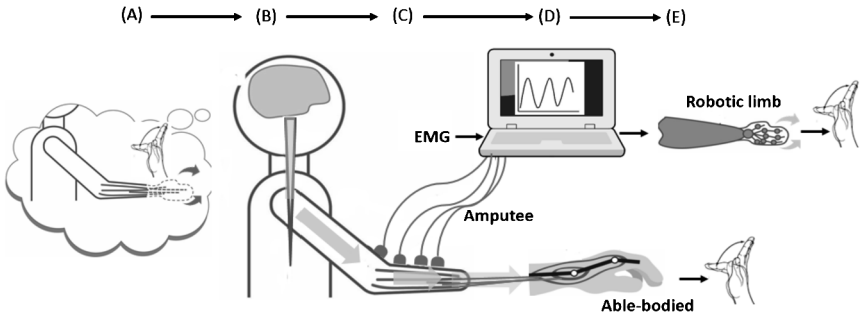

FIG. 1. *The biomechanical process for hand movement for an AB subject and TRA. A motor command starts as an internal limb representation of the intended hand movement* (*A*) *followed by neural signals sent from the motor cortex* (*B*) *to forearm muscles* (*C*) *causing them to contract. For an AB subject*, *tendons connecting the muscles to joints in the hand* (*D*, *bottom*) *are stretched*, *producing hand movement* (*E*, *bottom*). *For a TRA equipped with a robotic prosthesis*, *their muscle contractions are measured with EMG sensors* (*C*) *and the PC* (*D*, *top*) *decodes the EMG signals into robotic hand movements* (*E*, *top*).

more forearm muscles and conveys information about the magnitude and duration of muscle contractions. For an AB subject the contractions stretch tendons connected to bones in the hand, leading to hand movement. These physical connections no longer exist for a TRA, but TRAs can still activate their residual muscles and may sense movement in their missing limb (Mercier et al. (2006)).

Two important and related questions are: (1) how many EMG sensors are needed to capture relevant muscle contraction information?, and (2) how should the PC decode the EMG data to produce movement? A traditional direct myoelectric PC uses two EMG sensors corresponding to antagonistic muscles for control of a single movement along a degree-of-freedom (DOF), such as wrist flexion/extension. To determine the sensor locations, prosthetists first palpate a muscle while the TRA contracts their muscles to attempt movement along a single DOF. The EMG signal on the targeted skin surface areas is visually checked by probing different locations with a bipolar electrode. A chosen EMG sensor location should be independent of a coactivating antagonistic muscle. Unfortunately, it is difficult to use these PCs to operate multiple DOFs intuitively (Resnik et al. (2018)).

EMG pattern recognition (PR) is the state-of-the-art decoding method for prosthesis control (Scheme and Englehart (2011)). PR assumes that the varying patterns of multichannel EMG signals can be decoded to different hand/wrist motions. In both clinical practice and research settings, the placement of EMG sensors is not muscle specific because EMG cross talks (i.e., signals detected from nearby muscles) do not significantly affect decoding performance (Resnik et al. (2018)). Rather, the amount and variety of neural information recorded in EMG signals determines the accuracy for classifying motions. To ensure that "sufficient" neural information is captured, studies have used redundant (Huang et al. (2010)) or high-density electrodes (Zhou et al. (2007)) to saturate the surface of the residual limb. However, this electrode saturation comes with costs: (1) it requires frequent, lengthy calibration, (2) its implementation increases the challenges in computer hardware design for data streaming and real-time EMG processing and (3) its prediction performance may degrade due to noise introduced by additional sensor information.

Unlike EMG PR, an EMG-driven musculoskeletal model decodes the EMG input into movement based on human physiology (Crouch and Huang (2016)). For these models, saturation of the EMG recording surface does not necessarily increase the neural information captured. In fact, there are only 20 muscles in the human forearm (O'Rahilly and Müller (1983)). If muscle-specific EMG recordings can be captured, the PC decoder needs fewer electrodes. Crouch and Huang (2017) proposed a low-dimensional PC decoder based on a

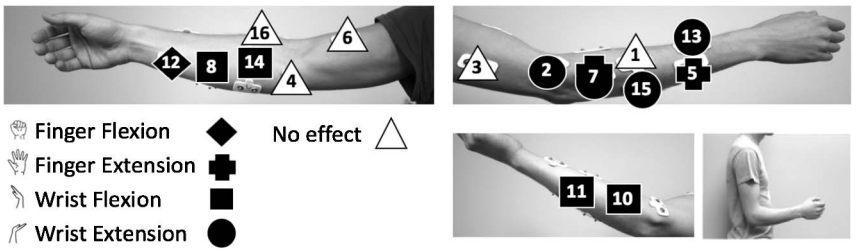

FIG. 2. *EMG sensor labels differentiated by movement DOF contribution. EMG label* 7 *measures a muscle that contributes to both finger and wrist extension and so has a combined symbol. The posture used for data collection is shown in the bottom right.*

planar link-segment dynamic model that includes only four virtual muscles while still incorporating relevant knowledge of the biomechanical system. Their model accurately predicted continuous wrist and finger movement with only four independent EMG signals, one for each virtual muscle. However, selection of the relevant EMG sensors remains an important and critical problem for TRAs, as many muscles are inaccessible to external EMG sensors due to muscle loss or depth of the muscle in the arm.

This paper introduces a novel EMG-based decoder and functional estimation procedure that is selective of the EMG without sacrificing prediction performance. The velocity or acceleration of a given movement DOF is predicted by multiple functional covariates corresponding to the recent past behavior EMG signals, whose effects can vary with the recent DOF's position. EMG selection and estimation is done simultaneously through our proposed Sequential Adaptive Functional Estimation (SAFE) procedure that uses multiple, adaptive stages of a penalized fitting criterion.

The paper is organized as follows. Section 2 describes the collection and processing of the EMG and movement data and provides motivation for our proposed decoder. Section 3 briefly reviews current EMG-based PCs and highlights their deficiencies. Section 4 details the dynamic, functional linear model that makes up our decoder, and Section 5 details the SAFE approach. Section 6 presents the analysis results from data collected from an AB subject, including EMG selection, effect interpretations and prediction performance. A simulation study is performed in Section 7, showing the robustness of our method to varying covariance assumptions. Section 8 concludes the paper with a discussion of impacts and extensions.

2. Data collection and processing. EMG and movement data were collected from an AB subject's right limb. Figure 2 shows the placement of 15 EMG surface electrodes on the subject's limb and labels those electrodes placed near muscles known to contribute to the movement DOFs of interest, metacarpophalangeal (MCP) flexion/extension and wrist flexion/extension. For simplicity, we refer to MCP flexion/extension as finger flexion/extension, because it involves simultaneous movement of all fingers excluding the thumb. To provide a baseline for selection performance, we also generated an external signal unrelated to movement, giving a total of 16 signals, denoted by $X_1, \ldots, X_{16}$ with X_9 being the external signal.

The EMG signals were normalized between zero (no contraction) and one (maximal contraction) based on previous training. They then were high-pass filtered at 40 Hz, rectified and low-pass filtered at six Hz using a 4th order Butterworth zero-phase filter (Butterworth (1930)). Movement data were collected by reflective markers placed on nine anatomical locations on the forearm, wrist and hand, and were recorded at 120 Hz using an infrared motion capture system (Vicon Motion Systems Ltd., UK). The movement data were filtered at six Hz using a 4th order Butterworth filter and the joint angles (in degrees) were calculated from a musculoskeletal model (Holzbaur, Murray and Delp (2005)) in OpenSim (Delp et al. (2007)).

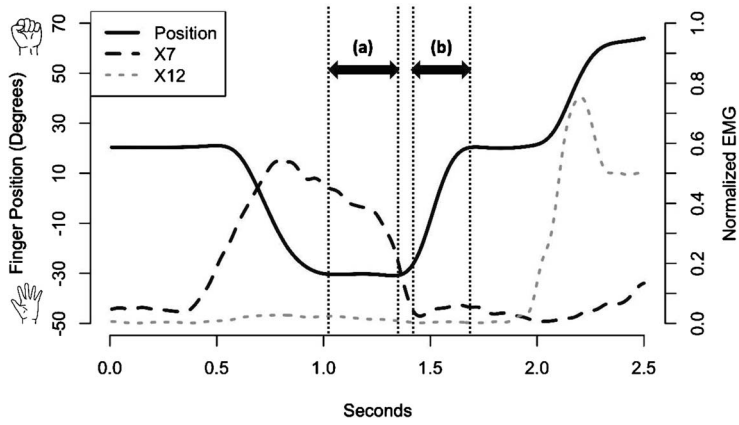

FIG. 3. *Joint finger position* (*degrees*) *for extension and flexion movements and normalized EMG signals for* X_7 *and* X_{12}. *Event* (*a*) *demonstrates restricted movement due to maximal finger extension. Event* (*b*) *demonstrates movement due to passive forces, lacking concurrently active EMG signals.*

EMG and joint angle data were collected synchronously for approximately 30 seconds while the subject performed basic hand movements in a fixed arm posture shown in Figure 2. Isolated, single DOF movements were performed following either a consistent pattern or a random pattern. The movements during a random pattern were determined by the AB subject. The consistent pattern had the subject alternate between performing maximal finger flexion and extension. Before performing the opposite movement, they allowed their hand to return to a neutral position by relaxing their muscles. There were six independent datasets for isolated finger movement, three observing consistent movement patterns and three observing random movement patterns. The same was done for isolated wrist movement.

Figure 3 is a snapshot of synchronous data for a consistent finger movement pattern. Prior to Figure 3(a), X_7's signal concurrently increases as the fingers extend. In Figure 3(a), the fingers are at maximal extension while X_7's signal decreases, meaning the subject is relaxing the corresponding muscle. Figure 3(b) has no concurrent muscle contractions while the fingers flex toward a neutral state of 20 degrees. This type of movement is attributable to passive forces generated by muscle relaxation. Note the time delay between the muscle's relaxation in Figure 3(a) and the resulting passive force movement in Figure 3(b).

Figure 3 is convincing that EMG data can predict movement but omits all the potential EMG signals that complicate EMG selection. The left plot in Figure 4 shows all 15 EMG signals (ignoring X_9) for the same time period as in Figure 3, and the right plot shows the concurrent correlations among all 16 signals, calculated across the entire 30-second data window. Between 0.5 and 1.5 seconds, the dominant signal belongs to X_7, but many other signals are active as well, such as X_5. The high concurrent correlations indicate redundant EMG information and the presence of latent signals that complicate EMG selection. The next section discusses how decoders in current EMG-based PCs address these issues.

3. Current EMG-based control strategies. The ith instance ($i = 1, \ldots, N$) of data collection includes K measured and processed EMG signals, x_{ik} ($k = 1, \ldots, K$), and a measure for each movement DOF of interest. For example, we let z_i denote the movement DOF position. It is common to divide the data into time windows and summarize the EMG and movement data within each window. In some cases movement in each window is assigned to a movement category, such as flexion or extension. The x_{ik}'s are often converted into a feature set summarizing its behavior in the time window, such as the moving average or slope (Hargrove, Englehart and Hudgins (2007)), and these feature sets are the decoder's inputs instead of the original x_{ik}'s.

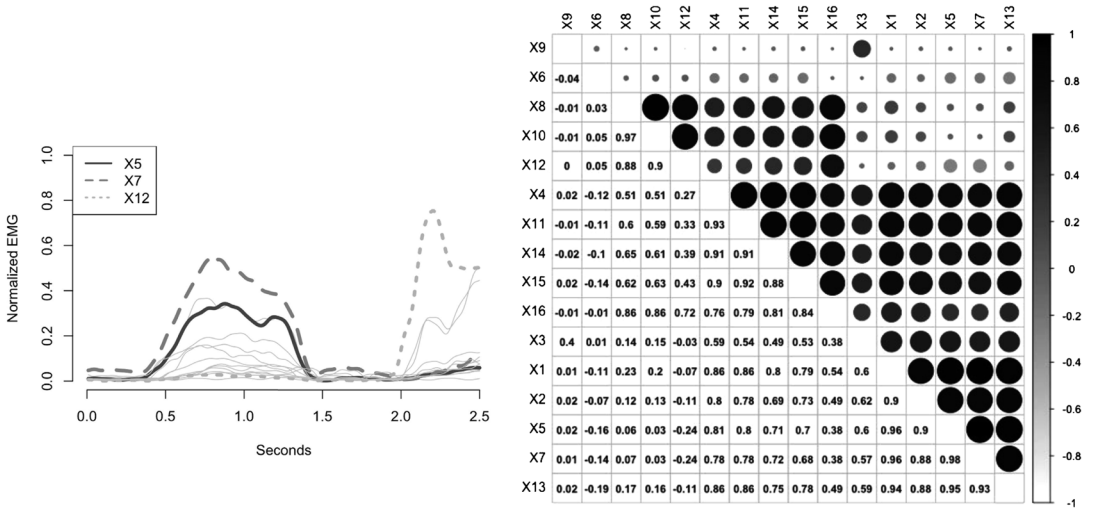

FIG. 4. *Overlay plot (left) of* 15 *EMG signals during one instance of finger extension and flexion observed and concurrent correlation plot (right) between all* 16 *signals for the entire dataset.*

Hahne et al. (2014) simultaneously predicted multiple wrist movement DOFs with linear and nonlinear regression (e.g., mixture of linear experts and kernel ridge regression) based on feature sets from 192 EMG signals. They compared performance across different subsets of the EMG sensors and determined fewer signals produced comparable predictions to the full set, but no formal variable selection was performed. Scheme and Englehart (2011) summarized and compared several statistical PR classifiers (e.g., K-nearest-neighbors, support vector machine, quadratic discriminant analysis) on EMG/movement data and discussed challenges faced by these classifiers. Both papers noted the importance of the amount and the type of EMG information in the feature set for a decoder's prediction performance.

Jiang, Englehart and Parker (2009) and Jiang et al. (2014) proposed a generative model to incorporate underlying neurophysiological processes largely ignored by PR approaches. The jth movement DOF ($j = 1, \ldots, M$) was assigned two opposing latent control signals representing positive and negative directions (e.g., extension/flexion), denoted by c_{ij}^p and c_{ij}^n, respectively. Let $\mathbf{x}_i = (x_{i1}, \ldots, x_{iK})^T$ be the vector of K EMG signals at instance i. Equating $\mathbf{x}_i = \boldsymbol{W}\mathbf{c}_i$ where $\boldsymbol{W}$ is the $K \times 2M$ synergy matrix and $\mathbf{c}_i = (c_{i1}^p, c_{i1}^n, \ldots, c_{iM}^p, c_{iM}^n)^T$, they solved for $\boldsymbol{W}$ and $\mathbf{c}_i$ requiring both to have nonnegative elements. Their approach does not require any measured movement data but restricts the DOF of movements the subject performs to help identify the control signals during the calibration session.

Crouch and Huang (2016) developed a decoder based on a planar lumped-parameter musculoskeletal model for the two DOFs considered in this paper. The decoder reliably predicted continuous movement with only four EMG signals corresponding to muscles known to contribute to these movements for AB subjects. The results relied on their prior knowledge of the important muscles and their ability to target them with the four surface electrodes. This knowledge is unknown for a given TRA whose musculoskeletal structure has been significantly altered, and it is unclear how the important EMG signals would be selected for a TRA. We next describe a new class of EMG-based decoders based on a flexible statistical model that can recover many of the biomechanical features directly incorporated in the musculoskeletal model in Crouch and Huang (2016) and lends itself to an EMG selection method.

4. Proposed decoder and penalized estimation. As demonstrated in Figure 3, movement can be influenced by concurrent and recent past changes in X_k. The effect may also

depend on the position, as observed in Figure 3(a) in which there is no change in position despite a decreasing EMG signal. Hence, our proposed decoder is based on a functional linear model with a scalar response, y_i, and K functional predictors defined on $\mathcal{S}$ whose effects can vary with covariate, $z_i \in \mathcal{Z}$,

$$E[y_i|X_{i1},\ldots,X_{iK},z_i] = \sum_{k=1}^{K}\int_{\mathcal{S}} X_{ik}(s)\gamma_k(s,z_i)\,ds. \tag{4.1}$$

Here, $\gamma_k(\cdot,\cdot)$ is an unknown bivariate function defined on $\mathcal{S}\times\mathcal{Z}$ that quantifies the effect of the kth functional covariate, briefly written as X_{ik}, on the mean of y_i conditional on z_i. When possible, we shorten $\gamma_k(\cdot,\cdot)$ as simply γ_k.

For model (4.1) to be used as an EMG-based decoder, y_i is set to either movement velocity or acceleration, z_i is the movement position and X_{ik} is a curve describing the recent past behavior of the kth EMG signal. Each X_{ik} is defined on $\mathcal{S}=[-\delta,0]$, where $s=0$ refers to the concurrent time relative to the velocity y_i and δ is a predetermined historical length. Consistent with our application, each X_{ik} is measured at $\delta+1$ equally-spaced time points, $X_{ik}(j)=x_{(i+j)k}$ for $j=-\delta,\ldots,0$. As in Section 3, we window the data; however, we do not reduce the EMG data into a feature set. Section 6.1 and Section A of the Supplementary Material (Stallrich et al. (2020)) detail how the y_i values were estimated from the observed z_i and how the X_{ik} were created from the EMG data for this application.

With y_i being the first or second derivative of z_i, model (4.1) is a forced differential equation system with nonlinear forcing functions for each EMG signal (Ramsay and Hooker (2017)). The covariate-varying effects in the forcing functions are a direct extension of the functional linear model described in Cardot, Ferraty and Sarda (2003), Ferraty et al. (2012), Goldsmith et al. (2011), McLean et al. (2014), Ramsay and Silverman (2005) and have been recently applied to functional data in Cardot and Sarda (2008), Davenport (2013), Wu, Fan and Müller (2010). In their Section 1.1.6, Ramsay and Hooker (2017) analyzed handwriting movements of Chinese characters by modeling the acceleration with a harmonic oscillator, but they did not use EMG-based forcing functions. Both velocity and acceleration models are considered in Section 6 as well as forcing functions that do not depend on position.

4.1. *Model approximation.* Following Wood (2006) and Eilers and Marx (2003), we approximate γ_k with a tensor product of two finite dimensional univariate basis functions, $\{\omega_l(\cdot)\}$ and $\{\tau_m(\cdot)\}$, defined on $\mathcal{S}$ and $\mathcal{Z}$, respectively. We use the same bases for all γ_k; so, $\gamma_k(s,z)=\sum_{l=1}^{L}\sum_{m=1}^{M}\omega_l(s)\tau_m(z)\beta_{klm}$, where the β_{klm}'s are basis coefficients for the kth functional covariate's effect that need to be estimated. In matrix form, $\gamma_k(s,z)\approx \boldsymbol{\omega}^T(s)\boldsymbol{B}_k\boldsymbol{\tau}(z)$ where $\boldsymbol{\omega}^T(s)=(\omega_1(s),\ldots,\omega_L(s))$, $\boldsymbol{\tau}^T(z)=(\tau_1(z),\ldots,\tau_M(z))$ and $\boldsymbol{B}_k=(\beta_{klm})$ is an $L\times M$ coefficient matrix.

Approximate the integrals in (4.1) by

$$\int_{\mathcal{S}} X_{ik}(s)\gamma_k(s,z_i)\,ds \approx \left\{\sum_{j=-\delta}^{0} x_{(i+j)k}\boldsymbol{\omega}(j)^T\right\}\boldsymbol{B}_k\boldsymbol{\tau}(z_i) = \boldsymbol{X}_{ik\omega}^T\boldsymbol{B}_k\boldsymbol{\tau}(z_i). \tag{4.2}$$

Define $\boldsymbol{X}_{ik\omega\tau}^T=\boldsymbol{X}_{ik\omega}^T\otimes\boldsymbol{\tau}(z_i)$, where $\otimes$ is the Kronecker product operator. For notational simplicity denote $\boldsymbol{X}_{ik\omega\tau}^T$ by $\widetilde{\boldsymbol{X}}_{ik}^T$, and let $\boldsymbol{\beta}_k$ denote the vectorized version of $\boldsymbol{B}_k$. Then, we approximate model (4.1) by

$$E[y_i|\widetilde{\boldsymbol{X}}_{i1},\ldots,\widetilde{\boldsymbol{X}}_{iK}] \approx \sum_{k=1}^{K}\widetilde{\boldsymbol{X}}_{ik}^T\boldsymbol{\beta}_k, \tag{4.3}$$

so that estimation of γ_k is done through estimation of $\boldsymbol{\beta}_k$.

4.2. *Penalized estimation.* It is necessary to regularize estimation of $\boldsymbol{\beta}_k$ to prevent overfitting and to encourage interpretable $\hat{\gamma}_k$. We further desire the estimation procedure to force some estimates to zero to determine a subset of important functional covariates. Fan, James and Radchenko (2015), Gertheiss, Maity and Staicu (2013), Matsui and Konishi (2011), Pannu and Billor (2017) have discussed functional variable selection with scalar-on-functional regression models. Our proposed approach is inspired by these ideas and accommodates more complex regression coefficients.

Following Gertheiss, Maity and Staicu (2013), we adopt the penalized least squares approach for estimation that simultaneously induces sparsity and controls smoothness of the γ_k's. For simplicity, assume $\{\omega_l(\cdot)\}_{l=1}^{L}$ and $\{\tau_m(\cdot)\}_{m=1}^{M}$ are orthogonal B-spline bases. We allow L and M to be sufficiently large to capture the complexity of the regression surfaces and penalize the degree of smoothness. We measure smoothness through the total curvature in each direction with $\|\gamma''_{k,s}\|$ and $\|\gamma''_{k,z}\|$, where $\|\gamma''_{k,s}\|^2 = \int_{\mathcal{S}}\int_{\mathcal{Z}}\{\gamma''_{k,s}(s,z)\}^2\,dz\,ds$ and $\gamma''_{k,s} = \partial^2\gamma_k/\partial s^2$. Sparsity results from penalizing the effect's total magnitude, measured by $\|\gamma_k\|$. By orthogonality of the bases,

$$\begin{aligned} \|\gamma_k\|^2 &= \boldsymbol{\beta}_k^T\boldsymbol{\beta}_k, \qquad \|\gamma''_{k,s}\|^2 = \boldsymbol{\beta}_k^T(\boldsymbol{\Omega}_s \otimes \boldsymbol{I}_M)\boldsymbol{\beta}_k, \\ \|\gamma''_{k,z}\|^2 &= \boldsymbol{\beta}_k^T(\boldsymbol{I}_L \otimes \boldsymbol{\Omega}_z)\boldsymbol{\beta}_k, \end{aligned} \tag{4.4}$$

where $\boldsymbol{\Omega}_s = \int \boldsymbol{\omega}''(s)\boldsymbol{\omega}''(s)^T\,ds$ and $\boldsymbol{\omega}''(\cdot)$ is the vector of second derivatives $\omega_l''(\cdot)$ with a similar definition for $\boldsymbol{\Omega}_z$. The three norms in (4.4) are combined into the penalty term

$$P_\phi(\gamma_k) = \big(\|\gamma_k\|^2 + \phi_s\|\gamma''_{k,s}\|^2 + \phi_z\|\gamma''_{k,z}\|^2\big)^{1/2} = \big(\boldsymbol{\beta}_k^T\boldsymbol{Q}_\phi\boldsymbol{\beta}_k\big)^{1/2}, \tag{4.5}$$

where $\boldsymbol{Q}_\phi = \boldsymbol{I}_{\mathrm{LM}} + \phi_s(\boldsymbol{\Omega}_s \otimes \boldsymbol{I}_M) + \phi_z(\boldsymbol{I}_L \otimes \boldsymbol{\Omega}_z)$ is positive-definite and the tuning parameters $\phi = (\phi_s, \phi_z) > 0$ control the smoothness of γ_k in the direction s and z. For instance, large values of ϕ_s force $\hat{\gamma}_k$ to be marginally linear in $\mathcal{S}$. We estimate $\boldsymbol{\beta}_k$ by minimizing the criterion

$$\sum_{i=1}^{N}\Bigg(y_i - \sum_{k=1}^{K}\widetilde{\boldsymbol{X}}_{ik}^T\boldsymbol{\beta}_k\Bigg)^2 + \lambda\sum_{k=1}^{K}\big(\boldsymbol{\beta}_k^T\boldsymbol{Q}_\phi\boldsymbol{\beta}_k\big)^{1/2}, \tag{4.6}$$

where $\lambda > 0$ controls the sparsity of the γ_k's.

Let $\boldsymbol{R}_\phi$ be the Cholesky factor of $\boldsymbol{Q}_\phi$, that is, $\boldsymbol{Q}_\phi = \boldsymbol{R}_\phi\boldsymbol{R}_\phi^T$, and set $\widetilde{\boldsymbol{\beta}}_k = \boldsymbol{R}_\phi^T\boldsymbol{\beta}_k$ and $\boldsymbol{W}_{ik} = \boldsymbol{R}_\phi^{-1}\widetilde{\boldsymbol{X}}_{ik}$. The criterion (4.6) can be written as

$$\sum_{i=1}^{N}\Bigg(y_i - \sum_{k=1}^{K}\boldsymbol{W}_{ik}^T\widetilde{\boldsymbol{\beta}}_k\Bigg)^2 + \lambda\sum_{k=1}^{K}\|\widetilde{\boldsymbol{\beta}}_k\|_2, \tag{4.7}$$

which, for a fixed ϕ, is the group LASSO criterion described in Yuan and Lin (2006) and Yang and Zou (2013), Yang and Zou (2015). Minimizing (4.7) can be done efficiently using the groupwise-majorization-descent algorithm (Yang and Zou (2013)) for given λ and ϕ. Rebuilding $\widehat{\boldsymbol{B}}_k$ from $\widehat{\boldsymbol{\beta}}_k = (\boldsymbol{R}_\phi^T)^{-1}\widehat{\widetilde{\boldsymbol{\beta}}}$, we estimate γ_k with $\widehat{\gamma}_k(s,z) = \boldsymbol{\omega}(s)^T\widehat{\boldsymbol{B}}_{\boldsymbol{k}}\boldsymbol{\tau}(z)$.

4.3. *Selection of λ and ϕ.* The optimal tuning parameters $\pi^* = (\lambda^*, \phi_s^*, \phi_z^*)$ are chosen empirically by comparing the prediction performance of the minimizer of (4.7) across multiple settings in the tuning parameter space, for example, K-fold cross-validation (CV). For longitudinal data, like this data application, K-fold block CV is recommended, in which the data are not partitioned randomly but rather into K equally-sized sequential sections (Roberts et al. (2017)). CV is performed across a set of π's, each producing a set of estimates, $\{\hat{\gamma}_{k,\pi}\}$ and an average prediction error (APE_π) across the folds. Let $\pi_{\min}$ be the π

having $\text{APE}_{\min}$, the smallest APE_π. The $\pi_{\min}$ is one choice for the optimal π, but we recommend a version of the one-standard-error rule (Friedman, Hastie and Tibshirani (2001), Krstajic et al. (2014), Yang and Zou (2015)). First, gather the considered π's whose corresponding $\text{APE}_\pi \leq \text{APE}_{\min} + \text{SE}_{\min}$, where $\text{SE}_{\min}$ is the estimated standard error of $\text{APE}_{\min}$. From this subset choose the π whose estimates minimize $\sum_k P_\phi(\hat{\gamma}_{k,\pi})$ with $\phi = (1, 1)$, being a measure of model complexity balancing sparsity and smoothness of the $\hat{\gamma}_k$'s.

5. Sequential adaptive functional estimation (SAFE). The fitting criterion (4.6) does not allow for different shrinkage and smoothness for the functional predictors which may inflate the number of false positives. Adaptive estimation is one way to improve variable selection performance (Ciuperca (2019), Gertheiss, Maity and Staicu (2013), Guo et al. (2015), Ivanoff, Picard and Rivoirard (2016), Meier (2009), Zou (2006)). This section describes a sequential, adaptive fitting procedure that is generally applicable to linear models fit using a group LASSO-type penalty.

5.1. *Adaptive weight penalty.* Adaptive LASSO is an iterative fitting procedure that translates preliminary estimates $\tilde{\gamma}_k$ into weights that are incorporated into penalty (4.5) to improve variable selection performance. Here, the kth functional covariate is assigned the positive weight vector $w_k = (f_k, g_k, h_k)$, where f_k weights the sparsity penalty $\|\gamma_k\|^2$ and g_k and h_k weight the smoothing penalties $\|\gamma''_{k,s}\|^2$ and $\|\gamma''_{k,z}\|^2$, respectively. We set the weights to $f_k = 1/\|\tilde{\gamma}_k\|$, $g_k = 1/\|\tilde{\gamma}''_{k,s}\|$ and $h_k = 1/\|\tilde{\gamma}''_{k,z}\|$. The adaptive penalty function is denoted

$$P_{\phi,w}(\gamma_k) = \left(f_k\|\gamma_k\|^2 + g_k\phi_s\|\gamma''_{k,s}\|^2 + h_k\phi_z\|\gamma''_{k,z}\|^2\right)^{1/2}. \tag{5.1}$$

The weighted penalty for γ_k can be efficiently calculated by $P_{\phi,w}(\gamma_k) = (\boldsymbol{\beta}_k^T \boldsymbol{Q}_{\phi,w}\boldsymbol{\beta}_k)^{1/2}$, where $\boldsymbol{Q}_{\phi,w} = f_k\boldsymbol{I}_{\text{LM}} + g_k\phi_s(\boldsymbol{\Omega}_s \otimes \boldsymbol{I}_M) + h_k\phi_z(\boldsymbol{I}_L \otimes \boldsymbol{\Omega}_z)$. Constructing (4.7) with $\boldsymbol{Q}_{\phi,w}$ instead of $\boldsymbol{Q}_\phi$ incorporates these weights in the estimation procedure.

Selection performance of the adaptive group LASSO procedure depends on the accuracy of the $\tilde{\gamma}_k$'s (Zou (2006)). Initial estimation by minimizing (4.7) without penalizing $\|\gamma_k\|^2$ (Gertheiss, Maity and Staicu (2013)) may produce poor estimates for this data application due to high correlation between the functional predictors (see Figure 4) and dependency of the X_{ik} distributions on z_i; see Section A of the Supplementary Material (Stallrich et al. (2020)). Instead, we recommend initial estimates by optimizing (4.7) which includes the sparsity penalty. This is essentially a two-stage, adaptive estimation procedure where the first stage sets $w_k = (1, 1, 1)$ for all covariates to generate initial estimates. The weights are updated, and a second stage of adaptive estimation is performed. Next, we formalize and generalize this sequential, adaptive procedure.

5.2. *Sequential adaptive algorithm.* The first stage of Sequential, Adaptive Functional Estimation (SAFE) performs adaptive group LASSO with $w_k^1 = (1, 1, 1)$, producing the stage 1 estimates $\{\hat{\gamma}_k^1\}_{k=1}^K$ for optimal tuning parameters π^{1*}. Let $\mathcal{K}^1$ be the active variable subset of $\{1, \ldots, K\}$ where $\hat{\gamma}_k^1 \neq 0$, and let $\Gamma^1 = \{\hat{\gamma}_k^1; k \in \mathcal{K}^1\}$. All $k \notin \mathcal{K}^1$ are removed from consideration for the remaining stages of the estimation procedure.

The second stage of SAFE starts by calculating weights w_k^2 for $k \in \mathcal{K}^1$ from their corresponding $\hat{\gamma}_k^1 \in \Gamma^1$. Adaptive group LASSO is performed with these w_k^2 producing the stage 2 estimates $\{\hat{\gamma}_k^2\}_{k\in\mathcal{K}^1}$ for optimal tuning parameters π^{2*}. Note, the π^{2*} are found from a new implementation of K-fold CV (or block CV). This gives the second stage active variable set $\mathcal{K}^2$ and active variable estimates Γ^2. This two-stage approach is similar to the relaxed LASSO (Meinshausen (2007)) and the work by Wei and Huang (2010) and Guo et al. (2015).

SAFE is easily generalized to more than two stages and, as shown in Section 6, may perform better with more stages. Let $r = 1, \ldots, R$ index the stages of SAFE where stage r

uses weights, w_k^r, calculated from Γ^{r-1} for $r \geq 2$ and $w_k^1 = (1, 1, 1)$. This gives R active variable sets, $\mathcal{K}^1, \ldots, \mathcal{K}^R$, and active variable estimates, $\Gamma^1, \ldots, \Gamma^R$. The number of stages could be preset or chosen based on a stopping criterion that assesses the consistency of the $\mathcal{K}^r$ and/or Γ^r for increasing r. For example, SAFE could stop if $\mathcal{K}^r = \mathcal{K}^{r-1}$ or when changes in the effect estimates in Γ^r compared to Γ^{r-1} are below some threshold. In Section 6 we fixed $R = 5$ and assessed convergence by plotting the norms in (4.4) across the stages to check whether any of the nonzero $\hat{\gamma}_k$ approached 0. An example is given in Section C.4 of the Supplementary Material (Stallrich et al. (2020)).

If the tuning parameter space is not explored well, it is possible that the APE_{min} could increase significantly for increasing stages. To prevent this, we retain the APE_{min} and its SE_{min} from the first stage and include these values in the second stage CV. Denote these retained values by $\text{APE}^*_{\text{min}}$ and SE^*_{min}. If the second stage $\text{APE}_{\text{min}} \leq \text{APE}^*_{\text{min}}$, we continue with CV as usual and updated $\text{APE}^*_{\text{min}} = \text{APE}_{\text{min}}$ and $\text{SE}^*_{\text{min}} = \text{SE}_{\text{min}}$. Otherwise, we collect those second stage π's with $\text{APE}_\pi \leq \text{APE}^*_{\text{min}} + \text{SE}^*_{\text{min}}$ and select π^{2*} as the one that minimizes $\sum_k P_\phi(\hat{\gamma}_{k,\pi})$ with $\phi = (1, 1)$. If $\text{APE}_\pi > \text{APE}^*_{\text{min}} + \text{SE}^*_{\text{min}}$ for all second stage π's, we restart SAFE with a finer grid of tuning parameter candidates. The process for more than two stages is straightforward.

In a similar spirit to Leeb, Pötscher and Ewald (2015) and Zhao, Shojaie and Witten (2017), after the last SAFE stage we may re-estimate the γ_k for $k \in \mathcal{K}^R$ using only a smooth regularization to potentially reduce the estimation bias caused by the sparsity penalty. We recommend these estimates be compared to those in Γ^R, as the latter estimates may be more interpretable due to their penalization of $\|\gamma_k\|^2$.

SAFE's approach is easily generalized to models other than (4.1). The model need not even be functional. In the next section we implement variations of SAFE with three other functional variable selection algorithms and show how it can also improve their selection performance.

6. EMG selection and prediction for hand movement. We applied the proposed methods to the AB subject's data described in Section 2. Based on the structure of a biological limb, at least one signal is needed to control flexion and extension of a given movement DOF. However, the AB musculoskeletal system has muscle redundancies, meaning multiple muscles can contribute to a given movement DOF (see Figure 2). The ideal decoder would need only one EMG signal for extension and another for flexion without sacrificing prediction performance.

We expected estimation of γ_k to be challenging for this data application for two reasons. Figure 4 suggested the presence of latent factors among the 16 measured signals, which can be explained by the EMG sensors being indirect measures of redundant muscle contractions. Moreover, Section A of the Supplementary Material (Stallrich et al. (2020)) demonstrates how the distribution of the X_{ik} can vary with position, with some positions having $X_{ik} \approx 0$. The latter can cause some inconsistent behavior of the estimated effects, as we will see in Section 6.4.

We considered three competitors that perform functional variable selection and estimation on a model with a smooth, position effect and covariate-invariant effects for the EMG

$$E[y_i | X_{i1}, \ldots, X_{iK}, z_i] = \alpha(z_i) + \sum_k \int_{-\delta}^{0} X_{ik}(s)\gamma_k(s)\, ds. \tag{6.1}$$

Since y_i is either the first or second derivative of z_i, model (6.1) resembles a forced harmonic oscillator (Ramsay and Hooker (2017)), $D^2 z = -\beta z + u$, where u is a forcing input. Model (6.1) sets $u = \sum_k \int_{-\delta}^{0} X_{ik}(s)\gamma_k(s)\, ds$ and replaces $-\beta z$ with a smooth effect $\alpha(z_i)$,

due to the positional constraints on extension/flexion of the fingers and wrist. The three competitors represent current functional variable selection methods and represent the approaches described in Section 3. All four methods are summarized here:

- SAFE(z): Model (4.1) with SAFE.
- AGL: Model (6.1) with SAFE, essentially adaptive group LASSO from Gertheiss, Maity and Staicu (2013).
- LAD: Model (6.1) with SAFE minimizing least absolute deviation (Pannu and Billor (2017)).
- FAR: Model (6.1) with group smoothly clipped absolute deviation penalty, that is, linear functional additive regression from Fan, James and Radchenko (2015).

Adaptive weighting was incorporated into AGL and LAD but not FAR because it was unclear how weighting should be performed. FAR was performed for $R = 2$ stages due to its lack of weighting, while the other three methods used $R = 5$ stages. The position effect was approximated by $\alpha(z_i) = \boldsymbol{\tau}(z_i)^T \boldsymbol{\beta}_z$, and its weighted sparse/smooth measures were included in penalty (5.1).

6.1. *Computational details*. The estimation procedure is similar to gradient matching (Ramsay and Hooker (2017), Chapter 8) which uses initial estimates of the response (i.e., velocity or acceleration) from the observed position data. The velocity values, y_i, were estimated from a smoother of the position data, z_i, based on a large set of sixth-order B-spline basis functions fit with a third-order penalty using `fda` (Ramsay and Silverman (2005), Ramsey et al. (2014)) in `R` ((2017)). The acceleration values were similarly estimated but with a fourth-order penalty. The $\hat{y}_i$ were centered and scaled to have mean 0 and variance 1, and thinned by selecting every 20th observation in order to reduce the high temporal correlation, leaving approximately $N = 200$ responses. For each $\hat{y}_i$, we extracted the previous $\delta + 1$ EMG observations, ending with the ith EMG value. We chose a past time window of approximately 300 ms ($\delta = 40$) based on observed passive force movement from Figure 3. All X_{ik} were previously normalized between 0 and 1, so we only centered each curve at time points $s = -\delta, \ldots, 0$. More details may be found in Section A of the Supplementary Material (Stallrich et al. (2020)).

For SAFE(z), $\gamma_k(\cdot, \cdot)$ was modeled with orthogonal cubic B-splines of dimension $L = 10$ and $M = 10$ in the s and z directions, respectively. Both alternatives AGL and LAD set $L = 10$ basis functions for $\gamma_k(\cdot)$, while FAR used $L = 5$ because it does not penalize smoothness. The optimal tuning parameters were found using five-fold block CV, as described in Section 4.3, with tuning parameter values $\log(\phi_s), \log(\phi_z) \in \{-10, -5, 0, 5, 10\}$ and $\log(\lambda)$ from -20 to 0 in increments of 0.25. We required $g_k, h_k \leq \exp(10)$ to circumvent computational issues that arose when $\widehat{\gamma}_k$ was approximately linear in either the s or z direction. We used the `R` package `gglasso` (Yang and Zou (2013)) for AGL and SAFE(z), `rqPen` (Sherwood and Maidman (2017)) for LAD and `R` code from the corresponding author of Fan, James and Radchenko (2015) for FAR.

6.2. *Performance metrics*. For the AB subject, the muscles driving the movement of interest (finger or wrist extension/flexion) and the subset of EMG sensors that best measure these muscles' contractions were known. Let $\mathcal{K} = \mathcal{K}_F \cup \mathcal{K}_E$ denote the index set of these important EMG signals where $\mathcal{K}_F$ and $\mathcal{K}_E$ partition $\mathcal{K}$ into index sets corresponding to flexion and extension, respectively. From Figure 2, these sets for finger and wrist movements were:

- Finger: $\mathcal{K}_F = \{X_{12}\}$, $\mathcal{K}_E = \{X_5, X_7\}$.
- Wrist: $\mathcal{K}_F = \{X_8, X_{10}, X_{11}, X_{14}\}$, $\mathcal{K}_E = \{X_2, X_7, X_{13}, X_{15}\}$.

TABLE 1

Variable selection performance metrics for consistent (top three rows) and random (bottom three row) finger movement patterns with velocity as the response. The top and bottom row in each cell represents the results of the initial and final stage, respectively. TPR = 1 *for all scenarios*

	AGL		LAD		FAR		SAFE(z)	
Pattern	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]
FC1	0.21 [13]	0.77 [10]	0.21 [13]	0.77 [10]	0.93 [3]	0.00 [0]	0.86 [4]	0.15 [2]
	0.50 [9]	0.54 [7]	0.71 [6]	0.31 [4]	0.93 [3]	0.00 [0]	1.00 [2]	0.00 [0]
FC2	0.43 [10]	0.54 [7]	0.29 [12]	0.69 [9]	0.86 [4]	0.08 [1]	0.93 [3]	0.08 [1]
	0.57 [8]	0.39 [5]	1.00 [2]	0.00 [0]	0.86 [4]	0.08 [1]	1.00 [2]	0.00 [0]
FC3	0.29 [12]	0.69 [9]	0.79 [5]	0.15 [2]	0.93 [3]	0.00 [0]	0.86 [4]	0.15 [2]
	0.57 [8]	0.46 [6]	1.00 [2]	0.00 [0]	0.93 [3]	0.00 [0]	1.00 [2]	0.00 [0]
FR1	0.43 [10]	0.54 [7]	0.86 [4]	0.08 [1]	0.93 [3]	0.08 [1]	0.79 [5]	0.15 [2]
	0.86 [4]	0.15 [2]	1.00 [2]	0.00 [0]	0.93 [3]	0.08 [1]	1.00 [2]	0.00 [0]
FR2	0.86 [4]	0.08 [1]	0.86 [4]	0.077 [1]	0.86 [4]	0.08 [1]	0.86 [4]	0.08 [1]
	1.00 [2]	0.00 [0]	0.93 [3]	0.00 [0]	0.86 [4]	0.08 [1]	0.93 [3]	0.00 [0]
FR3	0.93 [3]	0.08 [1]	0.43 [10]	0.54 [7]	0.64 [7]	0.31 [4]	0.86 [4]	0.08 [1]
	1.00 [2]	0.00 [0]	1.00 [2]	0.00 [0]	0.64 [7]	0.31 [4]	1.00 [2]	0.00 [0]

We compared the four methods with metrics focusing on identification of $\mathcal{K}_F$ and $\mathcal{K}_E$ as well as prediction accuracy:

- Size $= |\widehat{\mathcal{K}}|$, ideal size $= 2$ with $|\widehat{\mathcal{K}} \cap \mathcal{K}_F| = 1$ and $|\widehat{\mathcal{K}} \cap \mathcal{K}_E| = 1$.
- Sparsity (SP) $= 1 - |\widehat{\mathcal{K}}|/K$, ideal sparsity $= 14/16 = 0.875$. Also, relative sparsity (RSP) $=$ SP/0.875.
- False positive rate (FPR) $= FP/|\mathcal{K}^c|$, where $\mathcal{K}^c$ is the complement of $\mathcal{K}$ and $FP = |\widehat{\mathcal{K}} \cap \mathcal{K}^c|$.
- True positive rate (TPR) $= TP/2$, where $TP = \mathbb{1}(|\widehat{\mathcal{K}} \cap \mathcal{K}_F| \geq 1) + \mathbb{1}(|\widehat{\mathcal{K}} \cap \mathcal{K}_E| \geq 1)$ focuses on whether we find a member of $\mathcal{K}_F$ and $\mathcal{K}_E$.
- Average prediction error (APE) $= \sum_{i=1}^{N_0} (y_{0i} - \hat{y}_{0i})^2/N_0$ for N_0 holdout observations, y_{0i}.

6.3. *Variable selection results*. Table 1 shows the results of the variable selection performance for the six finger movement datasets with velocity as the response; acceleration and wrist movement results may be found in the Supplementary Material (Stallrich et al. (2020)). The datasets are referenced by movement DOF (F = Finger, W = Wrist), movement pattern (C = Consistent, R = Random) and replicate $(1, 2, 3)$, for example, FC1 = Finger movement, Consistent pattern, replicate 1. In Table 1 all methods tended to overselect in the initial stage. The final iteration of SAFE(z) selected signals $\{X_7, X_{12}\}$ in all six datasets; for FR2 it also selected X_5. Both AGL and LAD benefited from adopting the SAFE approach, generally having better selection properties in the final stage than the initial stage. LAD's final selection iteration was similar to SAFE(z), except for FC1 selecting six total EMG signals. AGL overselected for all datasets, except FR2 and FR3. FAR generally performed well but overselected for FC2 and FR1-FR3.

Selection results for the acceleration-based models were generally worse than those shown in Table 1; see Table 1 in the Supplementary Material (Stallrich et al. (2020)). SAFE(z) identified the correct signals for all datasets but dropped X_{12} for FR3. The competing methods had more erratic results, both over- and under-selecting. For these reasons, we will focus our attention only on the velocity-based model.

For wrist movement, AGL, LAD and FAR selected many EMG signals from $\mathcal{K}_F$ and $\mathcal{K}_E$ and often had nonzero FPRs; see Table 2 in the Supplementary Material (Stallrich et al.

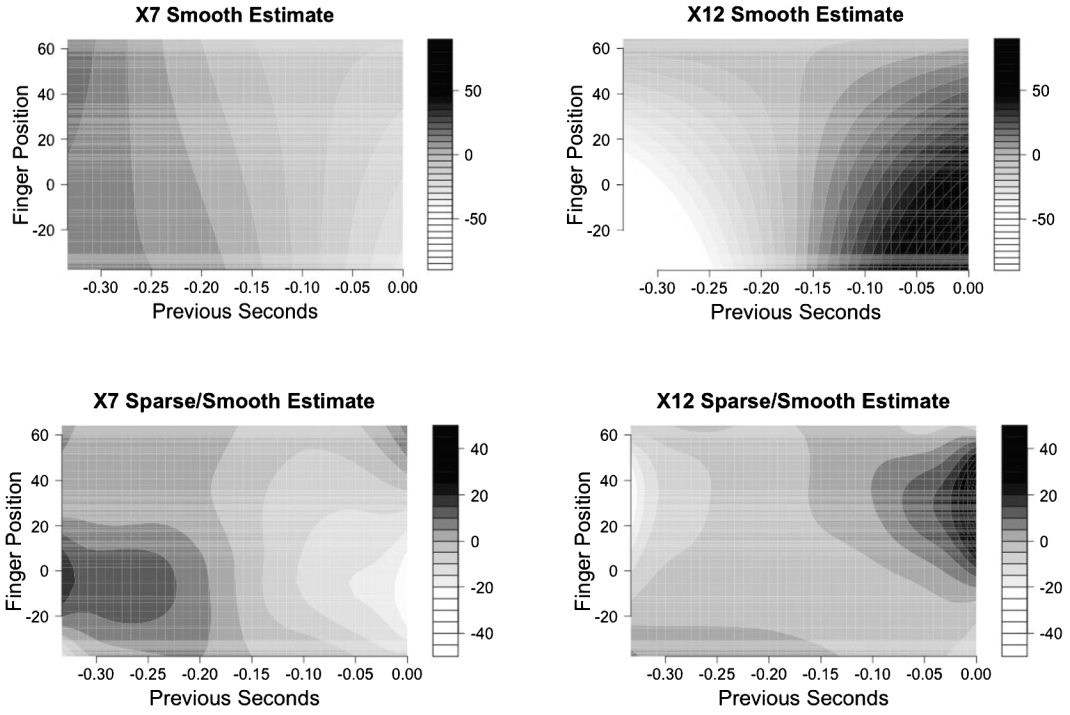

FIG. 5. *Estimated coefficients for X_7 (left) and X_{12} (right) under a smoothness-only penalty (top) and the sparse/smooth estimates from SAFE(z) (bottom). To help interpretation, the x-axis is in terms of previous seconds from the concurrent time, indicated by* 0.00. *Darker colors indicate positive velocity (flexion), and lighter colors indicate negative velocity (extension).*

(2020)). SAFE(z) always selected one EMG sensor from each $\mathcal{K}_F$ and $\mathcal{K}_E$, except for WC1 when it selected two signals from each group. SAFE(z) also maintained FPR $= 0$ for all datasets.

6.4. *EMG effect interpretations.* The SAFE(z) estimated effects for all finger/wrist datasets can be found in Sections B.3 and C.3 of the Supplementary Material (Stallrich et al. (2020)), respectively. Figure 5 here shows two sets of SAFE(z) estimates $\hat{\gamma}_7$ and $\hat{\gamma}_{12}$ for FC3, one using only a smoothness penalty (top panels) and another from the final SAFE stage (bottom panels). Fixing $s = 0$, concurrent activation of X_7 led to finger extension (negative velocity), while concurrent activation of X_{12} led to finger flexion (positive velocity). The top panels suggest the effect of X_{12} was strongest when the hand was fully extended (negative position). This is impossible for the consistent movement pattern because the flexor digitorum, measured by X_{12}, never contracted during maximal finger extension, meaning $X_{i,12}(s) \approx 0$ at this position. By not penalizing $\|\gamma_{12}\|^2$, the estimated effect deviated from 0 while maintaining its smoothness and without degrading the model's fit. When this penalty was included, the estimated effects approached 0 at these positions, providing clearer interpretations of the EMG's movement contributions. For this reason we use the final stage estimates for interpretation and prediction.

The effect of varying X_{12} on finger movement was concentrated between $0 < z_i < 55$, during which the fingers were in a neutral or flexed position. For these z_i, concurrent activation of X_{12} produced finger flexion, while past activation of X_{12} contributed to finger extension. The impact of X_7 was most relevant between angles $-30 < z_i < 20$, a range of extended and neutral finger positions. For these z_i, concurrent activation of X_7 led to finger extension, while past activation led to finger flexion. This indicates our model successfully

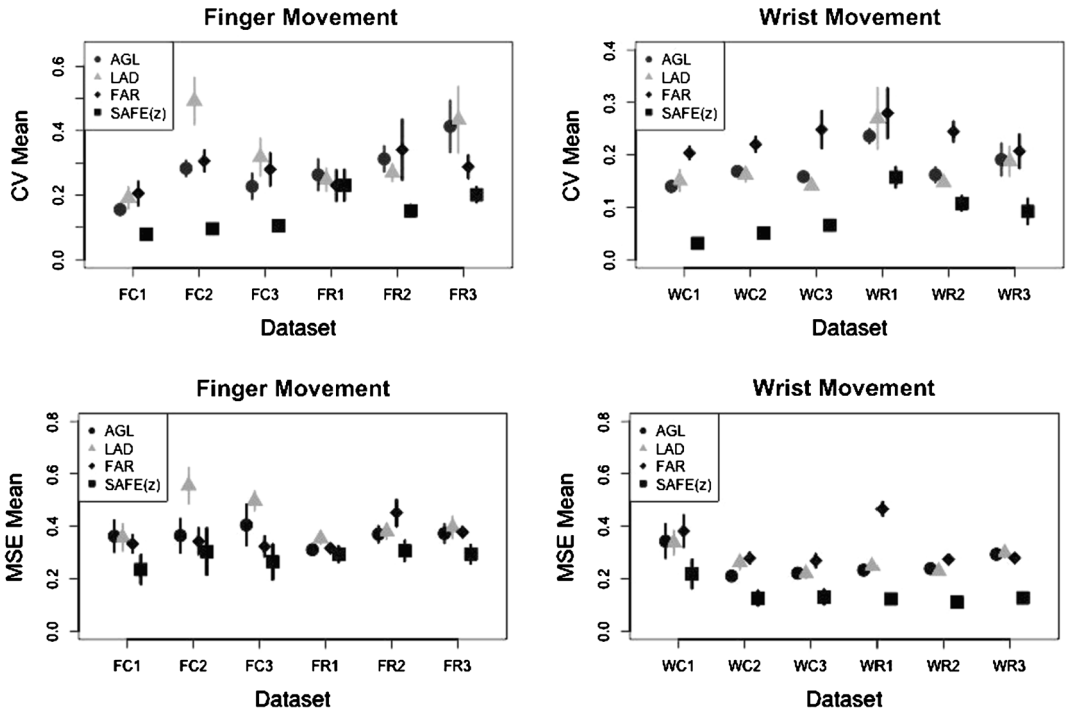

FIG. 6. *In-sample CV MSE means and standard errors for the optimal tuning parameters of the last selection stage* (*top panel*) *and out-sample prediction MSE means and standard errors* (*bottom panel*) *using the remaining five data sets*.

captured the passive force effects from past EMG behavior. Section B.4 of the Supplementary Material (Stallrich et al. (2020)) gives the interested reader additional interpretations of these effects.

6.5. *Prediction performance*. Each model for a given dataset was used to predict movement for the remaining five data sets. We calculated the mean and standard error of APEs across these five holdout samples for each data set and refer to this as out-sample prediction. Prediction was challenging across datasets, particularly when the consistent-movement models were asked to predict random movement. In-sample predictions were also considered, referring to the mean and standard error of the CV APEs across the five in-sample folds under the optimal π^*. Figure 6 shows that across nearly all scenarios, SAFE(z) outperforms the competitors in mean APE using fewer EMG signals. As expected, finger out-sample APE for the consistent movement datasets had higher standard errors, likely due to the prediction of random movements. SAFE(z) APEs for the wrist movement data were consistently small across the two movement patterns.

7. Simulation study. Recall that the original data were thinned to circumvent potential issues resulting from temporal correlations between the y_i's. A simulation study was performed to investigate the impact of covariance misspecification on variable selection performance. The FC3 data and its estimated coefficients (Figure 5) were used to generate responses under various error structures. The data-generating model was $y_i = \int_{\mathcal{S}} X_{i7}(s)\hat{\gamma}_7(s, z_i)\,ds + \int_{\mathcal{S}} X_{i,12}(s)\hat{\gamma}_{12}(s, z_i)\,ds + \epsilon_i$, where ϵ_i was a zero-mean error process with isotropic covariance function

$$\operatorname{cov}(\epsilon_i, \epsilon_{i'}) = \sigma_h^2[\mathbb{1}(i = i') + \theta \exp\{-(|i - i'|/\eta)^2\}].$$

TABLE 2

Average performance metrics across 100 *data sets for six covariance settings* $\theta = 0.25, 10, 100$ *and correlation decay that is slow*($\eta = 100$) *and fast* ($\eta = 10$) *with model based on fits of dataset FC*3. *The top and bottom row in each cell represents the results of the first and second stage*, *respectively.* TPR $= 1$ *for all scenarios*

Settings		AGL		LAD		FAR		SAFE(z)	
θ	Decay	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]	RSP [Size]	FPR [FP]
0.25	Slow	0.86 [3.92]	0.13 [1.74]	0.76 [5.37]	0.23 [3.00]	0.93 [2.94]	0.04 [0.55]	0.88 [3.66]	0.13 [1.62]
		0.95 [2.66]	0.05 [0.62]	0.90 [3.36]	0.10 [1.26]	0.94 [2.88]	0.04 [0.50]	0.99 [2.06]	0.01 [0.06]
0.25	Fast	0.83 [4.32]	0.17 [2.14]	0.73 [5.85]	0.27 [3.44]	0.94 [2.90]	0.04 [0.52]	0.86 [3.93]	0.15 [1.90]
		0.95 [2.71]	0.05 [0.64]	0.89 [3.58]	0.11 [1.41]	0.94 [2.84]	0.04 [0.50]	0.99 [2.12]	0.01 [0.12]
10	Slow	0.88 [3.62]	0.12 [1.49]	0.77 [5.20]	0.22 [2.90]	0.95 [2.75]	0.03 [0.41]	0.93 [3.04]	0.08 [1.02]
		0.94 [2.81]	0.06 [0.73]	0.93 [2.98]	0.07 [0.90]	0.95 [2.71]	0.03 [0.38]	0.99 [2.02]	0.01 [0.02]
10	Fast	0.75 [5.47]	0.24 [3.17]	0.64 [7.07]	0.36 [4.63]	0.90 [3.35]	0.07 [0.84]	0.84 [4.24]	0.17 [2.19]
		0.92 [3.15]	0.08 [1.07]	0.90 [3.41]	0.10 [1.34]	0.91 [3.20]	0.06 [0.74]	0.99 [2.07]	0.01 [0.07]
100	Slow	0.87 [3.81]	0.13 [1.66]	0.76 [5.38]	0.24 [3.05]	0.95 [2.76]	0.04 [0.46]	0.92 [3.08]	0.08 [1.06]
		0.94 [2.82]	0.06 [0.75]	0.91 [3.22]	0.09 [1.13]	0.95 [2.69]	0.03 [0.42]	0.99 [2.05]	0.01 [0.05]
100	Fast	0.76 [5.39]	0.24 [3.09]	0.61 [7.48]	0.38 [4.91]	0.90 [3.39]	0.06 [0.80]	0.83 [4.41]	0.18 [2.31]
		0.94 [2.91]	0.07 [0.87]	0.87 [3.89]	0.14 [1.78]	0.90 [3.34]	0.06 [0.77]	0.99 [2.08]	0.01 [0.08]

Here, θ is related to the dominant sources of dependence; $\theta = 0$ means that the responses are uncorrelated with $\text{var}(\epsilon_i) = \sigma_h^2$ while large θ reflects a higher degree of dependency and, for fixed σ_h^2, increases variance to $\text{var}(\epsilon_i) = \sigma_h^2(1+\theta)$. In addition, $\eta > 0$ controls the correlation decay where larger values imply slower correlation decay. The simulation study varied over two factors:

- $\theta = 0.25, 10, 100$
- $\eta = 10$ (Fast Decay) and 100 (Slow Decay)

and sets σ_h^2 so that $\text{var}(\epsilon_i) = \sigma_h^2(1+\theta)$ equals the in-sample MSE from the FC3 data application. For each setting we simulated 100 independent samples with $N = 200$ observations and analyzed the data with all four competitors from Section 6 with $S = 2$ stages.

Table 2 shows the average variable selection for the six covariance scenarios. All methods had a TPR $= 1$ and a negligible FPR. For example, SAFE(z)'s FPR never exceeded 0.009. We also calculated the proportion of all 100 second-stage models with size greater than 2. For SAFE(z), the largest proportion was 10% while the other methods had proportions ranging from 15% (AGL, $\theta = 0.25$, $\eta = 10$) to 72% (FAR, $\theta = 100$, $\eta = 100$).

Figure 7 shows the in-sample MSE performance for the six covariance scenarios. As expected, the MSEs were larger for $\eta = 100$ than for $\eta = 10$, but they only slightly increased as θ increased. These results provide evidence of SAFE's robustness to covariance misspecification. Section D of the Supplementary Material (Stallrich et al. (2020)) has results from another numerical experiment that investigated the role of model misspecification on variable selection and prediction.

8. Discussion and extensions. In this paper we proposed a new PC decoder based on a covariate-varying, scalar-on-function linear model to account for the biomechanical characteristics involved in hand movement. The functional predictors were the recent past behavior of EMG signals measured across the subject's limb, and the responses were finger and wrist velocity during flexion/extension. The effects for each EMG signal were allowed to vary with the current finger or wrist position. The bivariate effects were approximated using a tensor product of basis expansions whose coefficients were estimated with a group LASSO penalty that combined smoothing and sparseness penalization. We developed an adaptive, multistage

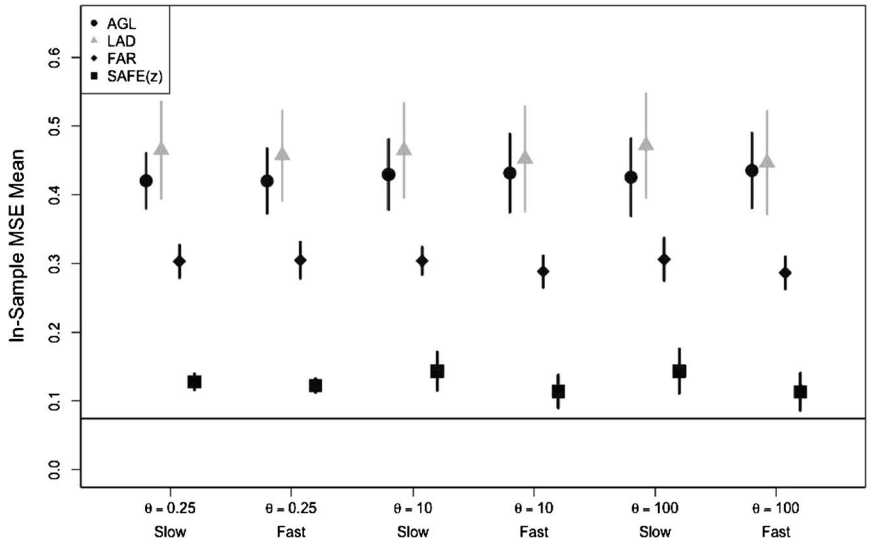

FIG. 7. *Average and standard deviation of in-sample MSEs across* 100 *datasets for six covariance settings* $\theta = 0.25, 10, 100$ *and Slow*$(\eta = 100)$*/Fast*$(\eta = 10)$ *correlation decay with model based on fits of dataset FC*3. *The bottom reference line is the in-sample MSE from the analysis of the FC*3 *data.*

estimation procedure, called Sequential, Adaptive Functional Estimation (SAFE), motivated by Meinshausen (2007), Wei and Huang (2010) and Guo et al. (2015). Section 6 showed our proposed decoder estimated with SAFE was able to recover a sparse set of important EMG signals for finger and wrist movement for an AB subject while competitors based on a more conventional model overselected and had worse prediction. A numerical experiment based on the data application also showed SAFE's robustness to covariance misspecification for data with high signal-to-noise ratios.

Our methodology provides biomedical engineers new tools for developing a TRA prosthesis controller. For example, if it is difficult to identify the function and location of residual muscles in the residual limb, engineers can first place high-density EMG electrodes over the skin surface. If the TRA performs mirrored hand movements with both their intact and residual limb, movement data can be captured from their intact limb that corresponds to the EMG patterns collected from the residual limb. Using our decoder and SAFE, engineers can remove redundant EMG signals and examine the estimated effects to interpret the role each EMG signal has on movement. This reduces the computational burden associated with streaming and processing large amounts of EMG data that hinder state-of-the-art PR methods and provides a unique way to determine EMG electrode location for musculoskeletal model-based PCs.

The proposed SAFE procedure opens several research questions. This paper focuses on approximating the bivariate coefficients with tensor products of finite dimensional basis functions, which can cause the number of parameters to increase exponentially. Other economical basis functions should be explored, such as splines over a triangulation or radial basis functions, although these involve computational challenges for calculating the penalty. Another avenue is to extend it to functional linear models where the functional measurements are perturbed by error or are sparsely measured. In this case, one could smooth the functional covariates using existing approaches (Xiao et al. (2016), Yao, Müller and Wang (2005)) and employ them in a SAFE selection algorithm. We have also investigated constructing prediction intervals based on the distribution-free split-conformal prediction sets (see Section E of the Supplementary Material (Stallrich et al. (2020)) for a short demonstration), which as far as the authors are aware has not been applied to functional regression.

We conjecture that SAFE requires more stages when the true model is driven by few latent variables. To see this, let $L_1(s)$ and $L_2(s)$ be two latent factors where $E(y_i|L_1(s), L_2(s)) = \int_{\mathcal{S}} L_1(s)\beta_1(s)\,ds + \int_{\mathcal{S}} L_2(s)\beta_2(s)\,ds$. Suppose we perform variable selection across K observed covariates where, for simplicity, $X_k(s) = \alpha_{k1}L_1(s) + \alpha_{k2}L_2(s)$. If we attempt to fit a model with these covariates, we get the following model equivalency:

$$\begin{aligned}\sum_{k=1}^{K}\int_{\mathcal{S}} X_k(s)\gamma_k(s)\,ds &= \sum_{k=1}^{K}\int_{\mathcal{S}}\big(\alpha_{k1}L_1(s) + \alpha_{k2}L_2(s)\big)\gamma_k(s)\,ds \\ &= \int_{\mathcal{S}} L_1(s)\sum_{k}\alpha_{k1}\gamma_k(s)\,ds + \int_{\mathcal{S}} L_2(s)\sum_{k}\alpha_{k2}\gamma_k(s)\,ds.\end{aligned}$$

For the fitted model to approximate the true model, the estimates, $\hat{\gamma}_k(s)$, should satisfy $\sum_k \alpha_{k1}\hat{\gamma}_k(s) \approx \beta_1(s)$ and $\sum_k \alpha_{k2}\hat{\gamma}_k(s) \approx \beta_2(s)$. That is, the true effects $\beta_1(s)$ and $\beta_2(s)$ will be partitioned across the $\hat{\gamma}_k(s)$ and variable selection across the $X_k(s)$ implies we want a sparse partition. Across the multiple stages of SAFE, these partitioned effects will be reflected in the updated adaptive weights and performing an additional fitting stage will eventually cause the predictors with larger weight to be removed from consideration. The number of required stages to arrive at a sparse model is then likely driven by the α_{k1} and α_{k2}. Paul et al. (2008) investigated variable selection for predictors with this latent factor structure, and many of their ideas could be extended to the functional variable problem. Their technique may be combined with SAFE selection and would allow us to efficiently screen a large number of EMG signals.

Acknowledgments. The project described was supported by NSF grant number DIIS1527202. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NSF. A.-M. Staicu's contributions were supported by NSF grant number DMS1454942 and NIH grant number P01CA142538. We also thank Rebecca North at North Carolina State University for her help with the simulation studies.

SUPPLEMENTARY MATERIAL

Additional numerical results (DOI: 10.1214/20-AOAS1324SUPPA; .pdf). The Supplementary Material (Stallrich et al. (2020)) contains details for data re-construction, additional results for finger and wrist movement analysis, simulation results, and demonstration of constructing distribution-free prediction intervals using split conformal inference.

R code, and data (DOI: 10.1214/20-AOAS1324SUPPB; .zip). R code and data (Stallrich et al. (2020)) are provided in a zip file.

REFERENCES

BUTTERWORTH, S. (1930). On the theory of filter amplifiers. *Wireless Engineer* **7** 536–541.

CARDOT, H., FERRATY, F. and SARDA, P. (2003). Spline estimators for the functional linear model. *Statist. Sinica* **13** 571–591. MR1997162

CARDOT, H. and SARDA, P. (2008). Varying-coefficient functional linear regression models. *Comm. Statist. Theory Methods* **37** 3186–3203. MR2467760 https://doi.org/10.1080/03610920802105176

CIUPERCA, G. (2019). Adaptive group LASSO selection in quantile models. *Statist. Papers* **60** 173–197. MR3905445 https://doi.org/10.1007/s00362-016-0832-1

CROUCH, D. L. and HUANG, H. (2016). Lumped-parameter electromyogram-driven musculoskeletal hand model: A potential platform for real-time prosthesis control. *Journal of Biomechanics* **49** 3901–3907.

CROUCH, D. L. and HUANG, H. H. (2017). Musculoskeletal model-based control interface mimics physiologic hand dynamics during path tracing task. *J. Neural. Eng.* **14** 036008. https://doi.org/10.1088/1741-2552/aa61bc

DAVENPORT, C. A. (2013). Semiparametric regression models for interacting covariates. Ph.D. thesis, North Carolina State University, Raleigh, NC. MR3218062

DELP, S. L., ANDERSON, F. C., ARNOLD, A. S., LOAN, P., HABIB, A., JOHN, C. T., GUENDELMAN, E. and THELEN, D. G. (2007). OpenSim: Open-source software to create and analyze dynamic simulations of movement. *IEEE Trans. Biomed. Eng.* **54** 1940–1950.

EILERS, P. H. and MARX, B. D. (2003). Multivariate calibration with temperature interaction using two-dimensional penalized signal regression. *Chemom. Intell. Lab. Syst.* **66** 159–174.

FAN, Y., JAMES, G. M. and RADCHENKO, P. (2015). Functional additive regression. *Ann. Statist.* **43** 2296–2325. MR3396986 https://doi.org/10.1214/15-AOS1346

FERRATY, F., GONZÁLEZ-MANTEIGA, W., MARTÍNEZ-CALVO, A. and VIEU, P. (2012). Presmoothing in functional linear regression. *Statist. Sinica* **22** 69–94. MR2933168 https://doi.org/10.5705/ss.2010.085

FRIEDMAN, J., HASTIE, T. and TIBSHIRANI, R. (2001). *The Elements of Statistical Learning*: *Data Mining*, *Inference*, *and Prediction*. Springer, New York.

GERTHEISS, J., MAITY, A. and STAICU, A.-M. (2013). Variable selection in generalized functional linear models. *Stat.* **2** 86–101. MR4027303 https://doi.org/10.1002/sta4.20

GOLDSMITH, J., CRAINICEANU, C. M., CAFFO, B. S. and REICH, D. S. (2011). Penalized functional regression analysis of white-matter tract profiles in multiple sclerosis. *NeuroImage* **57** 431–439.

GUO, P., ZENG, F., HU, X., ZHANG, D., ZHU, S., DENG, Y. and HAO, Y. (2015). Improved variable selection algorithm using a LASSO-type penalty, with an application to assessing hepatitis B infection relevant factors in community residents. *PLoS ONE* **10** e0134151. https://doi.org/10.1371/journal.pone.0134151

HAHNE, J. M., BIESSMANN, F., JIANG, N., REHBAUM, H., FARINA, D., MEINECKE, K.-R. and PARRA, L. C. (2014). Linear and nonlinear regression techniques for simultaneous and proportional myoelectric control. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 269–279.

HARGROVE, L. J., ENGLEHART, K. B. and HUDGINS, B. (2007). A comparison of surface and intramuscular myoelectric signal classification. *IEEE Trans. Biomed. Eng.* **54** 847–853.

HOLZBAUR, K. R., MURRAY, W. M. and DELP, S. L. (2005). A model of the upper extremity for simulating musculoskeletal surgery and analyzing neuromuscular control. *Ann. Biomed. Eng.* **33** 829–840.

HUANG, H. H., ZHANG, F., SUN, Y. L. and HE, H. (2010). Design of a robust EMG sensing interface for pattern classification. *Journal of Neural Engineering* **7**.

IVANOFF, S., PICARD, F. and RIVOIRARD, V. (2016). Adaptive Lasso and group-Lasso for functional Poisson regression. *J. Mach. Learn. Res.* **17** 1-46. MR3504615

JIANG, N., ENGLEHART, K. B. and PARKER, P. A. (2009). Extracting simultaneous and proportional neural control information for multiple-DOF prostheses from the surface electromyographic signal. *IEEE Trans. Biomed. Eng.* **56** 1070–1080.

JIANG, N., REHBAUM, H., VUJAKLIJA, I., GRAIMANN, B. and FARINA, D. (2014). Intuitive, online, simultaneous, and proportional myoelectric control over two degrees-of-freedom in upper limb amputees. *IEEE Trans. Neural Syst. Rehabil. Eng.* **22** 501–509.

KRSTAJIC, D., BUTUROVIC, L. J., LEAHY, D. E. and THOMAS, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics* **6**. https://doi.org/10.1186/1758--2946--6--10

LEEB, H., PÖTSCHER, B. M. and EWALD, K. (2015). On various confidence intervals post-model-selection. *Statist. Sci.* **30** 216–227. MR3353104 https://doi.org/10.1214/14-STS507

MATSUI, H. and KONISHI, S. (2011). Variable selection for functional regression models via the L_1 regularization. *Comput. Statist. Data Anal.* **55** 3304–3310. MR2825412 https://doi.org/10.1016/j.csda.2011.06.016

MCLEAN, M. W., HOOKER, G., STAICU, A.-M., SCHEIPL, F. and RUPPERT, D. (2014). Functional generalized additive models. *J. Comput. Graph. Statist.* **23** 249–269. MR3173770 https://doi.org/10.1080/10618600.2012.729985

MEIER, L. (2009). grplasso: Fitting user specified models with group lasso penalty. *R Foundation for Statistical Computing*.

MEINSHAUSEN, N. (2007). Relaxed Lasso. *Comput. Statist. Data Anal.* **52** 374–393. MR2409990 https://doi.org/10.1016/j.csda.2006.12.019

MERCIER, C., REILLY, K. T., VARGAS, C. D., ABALLEA, A. and SIRIGU, A. (2006). Mapping phantom movement representations in the motor cortex of amputees. *Brain* **129** 2202–2210.

O'RAHILLY, R. and MÜLLER, F. (1983). *Basic Human Anatomy*: *A Regional Study of Human Structure*. Saunders, Philadelphia.

PANNU, J. and BILLOR, N. (2017). Robust group-LASSO for functional regression model. *Comm. Statist. Simulation Comput.* **46** 3356–3374. MR3656106 https://doi.org/10.1080/03610918.2015.1096375

PAUL, D., BAIR, E., HASTIE, T. and TIBSHIRANI, R. (2008). "Preconditioning" for feature selection and regression in high-dimensional problems. *Ann. Statist.* **36** 1595–1618. MR2435449 https://doi.org/10.1214/009053607000000578

R CORE TEAM (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria.

RAMSAY, J. and HOOKER, G. (2017). *Dynamic Data Analysis*: *Modeling Data With Differential Equations*. *Springer Series in Statistics*. Springer, New York. MR3645102 https://doi.org/10.1007/978-1-4939-7190-9

RAMSAY, J. O. and SILVERMAN, B. W. (2005). *Functional Data Analysis*, 2nd ed. *Springer Series in Statistics*. Springer, New York. MR2168993

RAMSEY, J. O., WICKHAM, H., GRAVES, S. and HOOKER, G. (2014). Package fda: Functional data analysis. *R Foundation for Statistical Computing*.

RESNIK, L. (2011). Development and testing of new upper-limb prosthetic devices: Research designs for usability testing. *Journal of Rehabilitation Research and Development* **48** 697–706.

RESNIK, L., ETTER, K., KLINGER, S. L. and KAMBE, C. (2011). Using virtual reality environment to facilitate training with advanced upper-limb prosthesis. *Journal of Rehabilitation Research and Development* **48** 707–718.

RESNIK, L., HUANG, H. H., WINSLOW, A., CROUCH, D. L., ZHANG, F. and WOLK, N. (2018). Evaluation of EMG pattern recognitiion for upper limb prosthesis control: A case study in comparison with direct myoelectric control. *Journal of NeuroEngineering and Rehabilitation* **15**.

ROBERTS, D. R., BAHN, V., CIUTI, S., BOYCE, M. S., ELITH, J., GUILLERA-ARROITA, G., HAUENSTEIN, S., LAHOZ-MONFORT, J. J., SCHRÖDER, B. et al. (2017). Cross-validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure. *Ecography* **40** 913–929.

SCHEME, E. and ENGLEHART, K. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use. *Journal of Rehabilitation Research and Development* **48** 643–659.

SHERWOOD, B. and MAIDMAN, A. (2017). Package 'rqPen'. *R Foundation for Statistical Computing*.

STALLRICH, J., ISLAM, M. N., STAICU, A.-M., CROUCH, D. L., PAN, L. and HUANG, H. H. (2020). Supplement to "Optimal EMG placement for a robotic prosthesis controller with sequential, adaptive functional estimation (SAFE)." https://doi.org/10.1214/20-AOAS1324SUPPA, https://doi.org/10.1214/20-AOAS1324SUPPB.

WEI, F. and HUANG, J. (2010). Consistent group selection in high-dimensional linear regression. *Bernoulli* **16** 1369–1384. MR2759183 https://doi.org/10.3150/10-BEJ252

WOOD, S. N. (2006). Low-rank scale-invariant tensor product smooths for generalized additive mixed models. *Biometrics* **62** 1025–1036. MR2297673 https://doi.org/10.1111/j.1541-0420.2006.00574.x

WU, Y., FAN, J. and MÜLLER, H.-G. (2010). Varying-coefficient functional linear regression. *Bernoulli* **16** 730–758. MR2730646 https://doi.org/10.3150/09-BEJ231

XIAO, L., ZIPUNNIKOV, V., RUPPERT, D. and CRAINICEANU, C. (2016). Fast covariance estimation for high-dimensional functional data. *Stat. Comput.* **26** 409–421. MR3439382 https://doi.org/10.1007/s11222-014-9485-x

YANG, Y. and ZOU, H. (2013). gglasso: Group lasso penalized learning using a unified BMD algorithm. *R Foundation for Statistical Computing*.

YANG, Y. and ZOU, H. (2015). A fast unified algorithm for solving group-lasso penalize learning problems. *Stat. Comput.* **25** 1129–1141. MR3401877 https://doi.org/10.1007/s11222-014-9498-5

YAO, F., MÜLLER, H.-G. and WANG, J.-L. (2005). Functional data analysis for sparse longitudinal data. *J. Amer. Statist. Assoc.* **100** 577–590. MR2160561 https://doi.org/10.1198/016214504000001745

YUAN, M. and LIN, Y. (2006). Model selection and estimation in regression with grouped variables. *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **68** 49–67. MR2212574 https://doi.org/10.1111/j.1467-9868.2005.00532.x

ZHAO, S., SHOJAIE, A. and WITTEN, D. (2017). In defense of the indefensible: A very naive approach to high-dimensional inference. Preprint. Available at arXiv:1705.05543.

ZHOU, P., LOWERY, M. M., ENGLEHART, K. B., HUANG, H., LI, G., HARGROVE, L., DEWALD, J. P. A. and KUIKEN, T. A. (2007). Decoding a new neural-machine interface for control of artificial limbs. *Journal of Neuralphysiology* **98** 2974–2982.

ZIEGLER-GRAHAM, K., MACKENZIE, E. J., EPHRAIM, P. L., TRAVISON, T. G. and BROOKMEYER, R. (2008). Estimating the prevalence of limb loss in the United States: 2005 to 2050. *Archives of Physical Medicine and Rehabilitation* **89** 422–429.

ZOU, H. (2006). The adaptive lasso and its oracle properties. *J. Amer. Statist. Assoc.* **101** 1418–1429. MR2279469 https://doi.org/10.1198/016214506000000735