Model Form Identification in High-dimensional Functional Linear Regressions

Xingche Guo

Department of Statistics, University of Connecticut

Sep 29, 2026

Presented at the Statistics Seminar
Department of Mathematics, IU Indianapolis

The talk is based on:

- Xingche Guo, Yehua Li, and Pang Du. "Model Form Identification in High-Dimensional Functional Linear Regressions." arXiv preprint arXiv:2605.05458 (2026).

A related prior work:

- Xingche Guo, Yehua Li, and Tailen Hsing. "Variable Selection and Minimax Prediction in High-dimensional Functional Linear Model." Statistica Sinica (2026+)

Functional linear regression

Consider the functional linear regression (FLR):

$$Y_i = \sum_{j=1}^{p} \langle X_{ij}, \beta_{0j} \rangle_2 + \varepsilon_i, \quad i = 1, \ldots, n$$

- $X_{ij}(\cdot)$ are functional predictors in $\mathbb{L}_2 := \mathbb{L}_2[0,1]$ with mean zero
- $\beta_{0j}(\cdot)$ are unknown coefficient (i.e., slope) functions in $\mathbb{L}_2$
- $\langle \cdot, \cdot \rangle_2$ is inner product in $\mathbb{L}_2$
- $\| \cdot \|_2$ is the norm in $\mathbb{L}_2$
- $\varepsilon_i \overset{i.i.d.}{\sim} N(0, \sigma^2)$ are errors in $\mathbb{R}$
- p may be large relative to n

Reproducing kernel Hilbert space (RKHS)

Key assumption: $\beta_j(\cdot) \in \mathcal{H}_j = \mathcal{H}(K_j)$, which is the reproducing kernel Hilbert space (RKHS) with kernel K_j (Wahba, 1990).

Definition: A real, symmetric, square-integrable, and nonnegative definite function $K(\cdot, \cdot)$ on $[0, 1]^2$ is called a reproducing kernel for a Hilbert space of functions $\mathcal{H}(K)$ on $[0, 1]$ if

- $K(\cdot, t) \in \mathcal{H}(K)$ for any $t \in [0, 1]$, and
- $\mathcal{H}(K)$ is equipped with the inner product such that

$$\langle \beta, K(\cdot, t) \rangle_{\mathcal{H}(K)} = \beta(t), \ \beta \in \mathcal{H}(K), t \in [0, 1].$$

The Hilbert space $\mathcal{H}(K)$ is then called an RKHS (Aronszajn, 1950).

Variable Selection

In high-dimensional FLR, a natural first question is:

$$\beta_{0j}(t) \equiv 0 \quad \text{or} \quad \beta_{0j}(t) \not\equiv 0, \qquad j = 1, \ldots, p.$$

This gives an active set of relevant functional predictors:

$$\mathcal{S} = \{j : \beta_j(t) \not\equiv 0\}.$$

Define $\widehat{\mathcal{S}} = \{j : \widehat{\beta}_j(t) \not\equiv 0\}$. Variable selection is consistent when

$$\Pr(\widehat{\mathcal{S}} = \mathcal{S}) \to 1.$$

Why does variable selection matter?

Variable Selection

In high-dimensional FLR, a natural first question is:

$$\beta_{0j}(t) \equiv 0 \quad \text{or} \quad \beta_{0j}(t) \not\equiv 0, \qquad j = 1, \ldots, p.$$

This gives an active set of relevant functional predictors:

$$\mathcal{S} = \{j : \beta_j(t) \not\equiv 0\}.$$

Define $\widehat{\mathcal{S}} = \{j : \widehat{\beta}_j(t) \not\equiv 0\}$. Variable selection is consistent when

$$\Pr(\widehat{\mathcal{S}} = \mathcal{S}) \to 1.$$

Why does variable selection matter?

- Reduces model complexity when the number of predictors p is large.
- Improves interpretability by identifying which functional predictors are associated with the response.

Variable Selection

In high-dimensional FLR, a natural first question is:

$$\beta_{0j}(t) \equiv 0 \quad \text{or} \quad \beta_{0j}(t) \not\equiv 0, \qquad j = 1, \ldots, p.$$

This gives an active set of relevant functional predictors:

$$\mathcal{S} = \{j : \beta_j(t) \not\equiv 0\}.$$

Define $\widehat{\mathcal{S}} = \{j : \widehat{\beta}_j(t) \not\equiv 0\}$. Variable selection is consistent when

$$\Pr(\widehat{\mathcal{S}} = \mathcal{S}) \to 1.$$

Why does variable selection matter?

- Reduces model complexity when the number of predictors p is large.
- Improves interpretability by identifying which functional predictors are associated with the response.

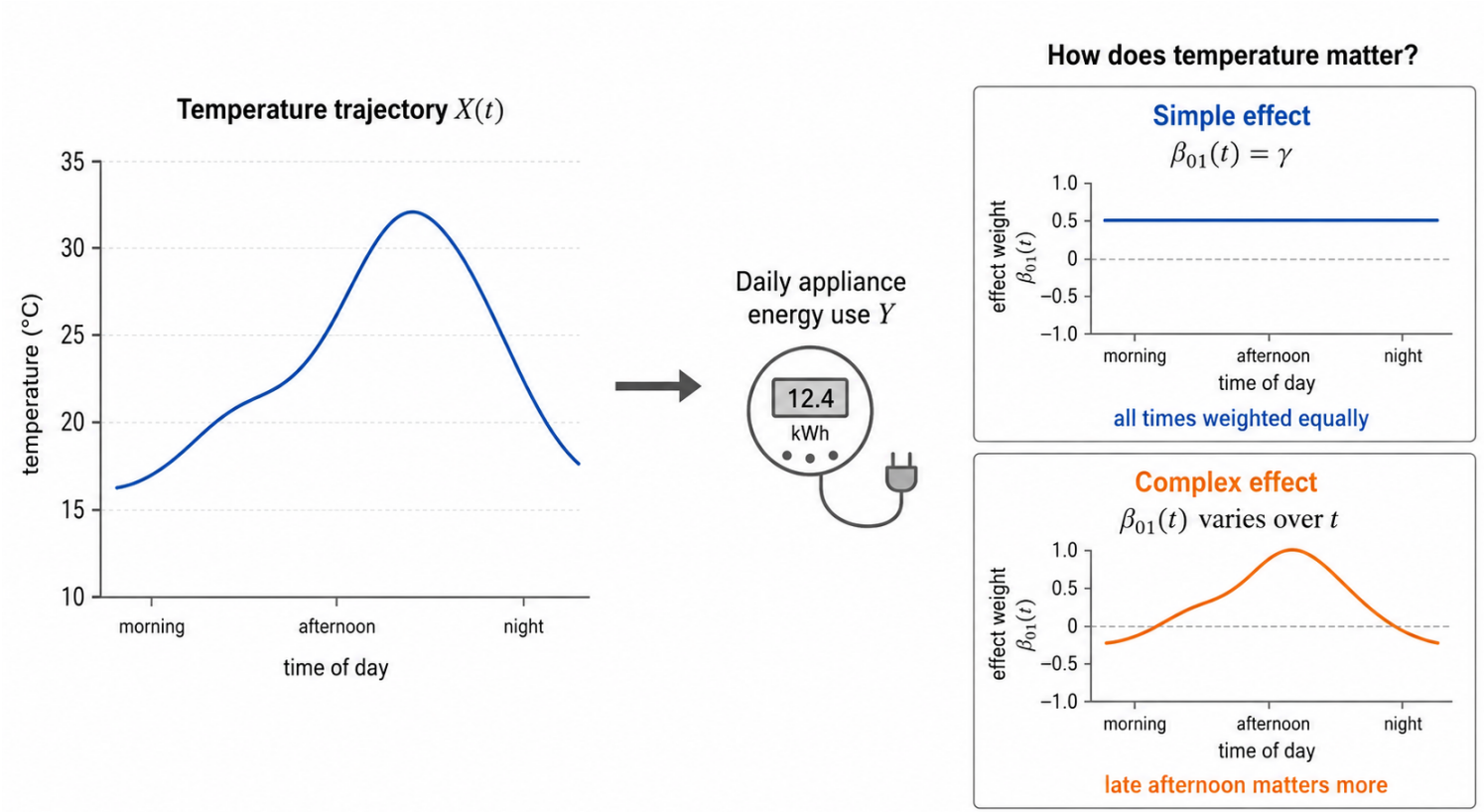
But is this enough?

Motivating example 1: temperature and energy use

Suppose

- $X_{i1}(t)$ is the house temperature trajectory during day i.
- Y_i is the total appliances energy consumption on day i.
- After variable selection, we find $\beta_{01} \neq 0$.

So temperature matters. But what kind of effect is it?

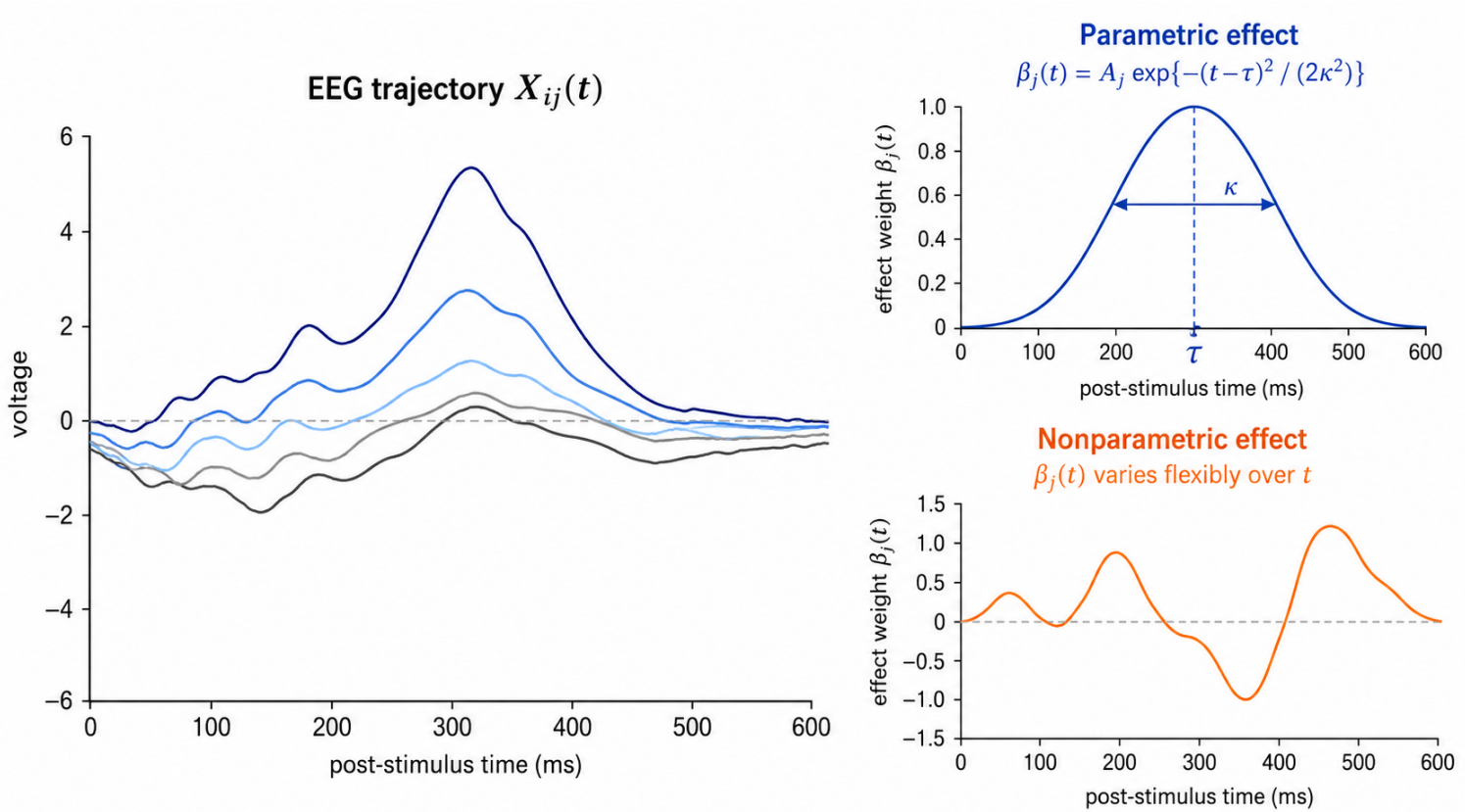


Motivating example 2: EEG–behavior latency

Suppose

- $X_{ij}(t)$ is the EEG voltage trajectory for trial i at electrode j.
- Y_i is the behavioral outcome for trial i (e.g., reaction time).
- After variable selection, we find $\beta_j \neq 0$ for some electrode j.

So neural activity matters. But what kind of effect is it?



Model Form Identification (MoFI)

For an active predictor $j \in \widehat{\mathcal{S}}$, knowing $\beta_{0j} \neq 0$ does not tell us the form of the effect.

MoFI decomposes the coefficient space into

$$\mathcal{H}_j = \mathcal{H}_{0j} \oplus \mathcal{H}_{1j},$$

where $\mathcal{H}_{0j}$ contains **simple components** and $\mathcal{H}_{1j}$ contains **complementary components**.

Equivalently,

$$\beta_{0j} = \beta_{0j}^{(0)} + \beta_{0j}^{(1)}.$$

MoFI asks:

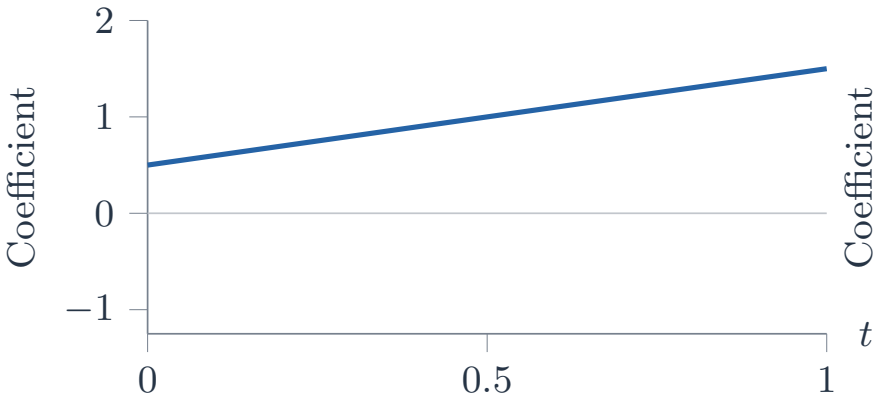
$$\beta_{0j}^{(1)} = 0 \quad \text{or} \quad \beta_{0j}^{(1)} \neq 0?$$

- **Simple effect:** $\beta_{0j}^{(1)} = 0$.
- **Complex effect:** $\beta_{0j}^{(1)} \neq 0$.

Model Form Identification (MoFI)

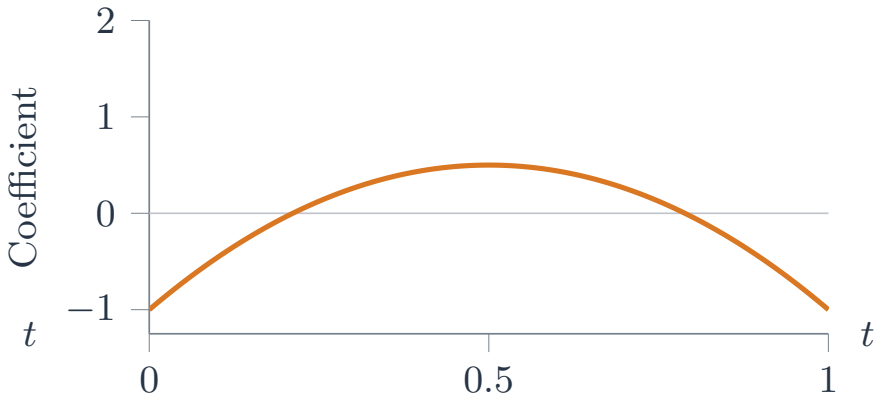
Linear component

$\mathcal{H}_0 = \text{span}\{1, t\}, \quad \beta^{(0)}(t) = 0.5 + t$



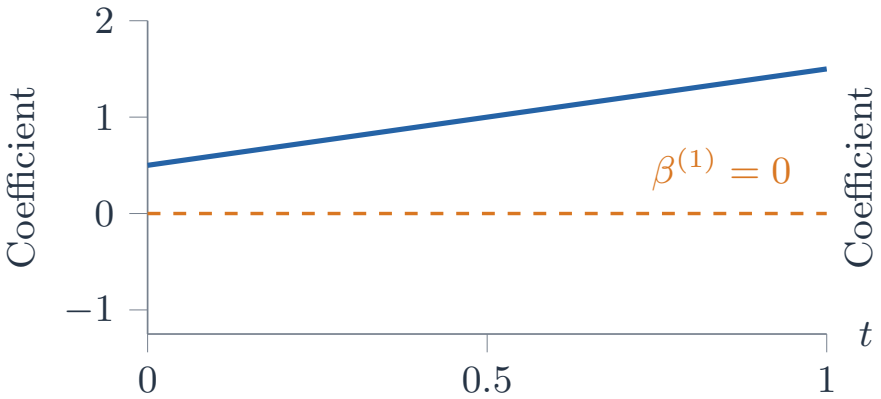
Complementary component

$\mathcal{H}_1 = \mathcal{H}_0^{\perp}, \quad \beta^{(1)}(t) = -6t^2 + 6t - 1$



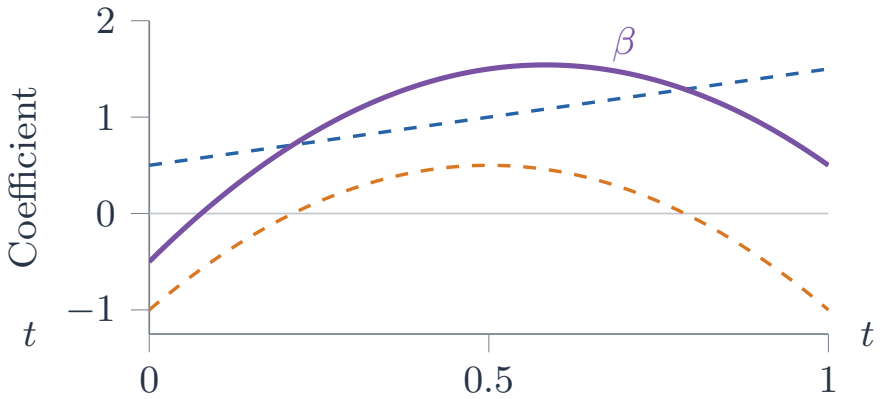
Simple effect

$\beta(t) = \beta^{(0)}(t), \qquad \beta^{(1)}(t) \equiv 0$



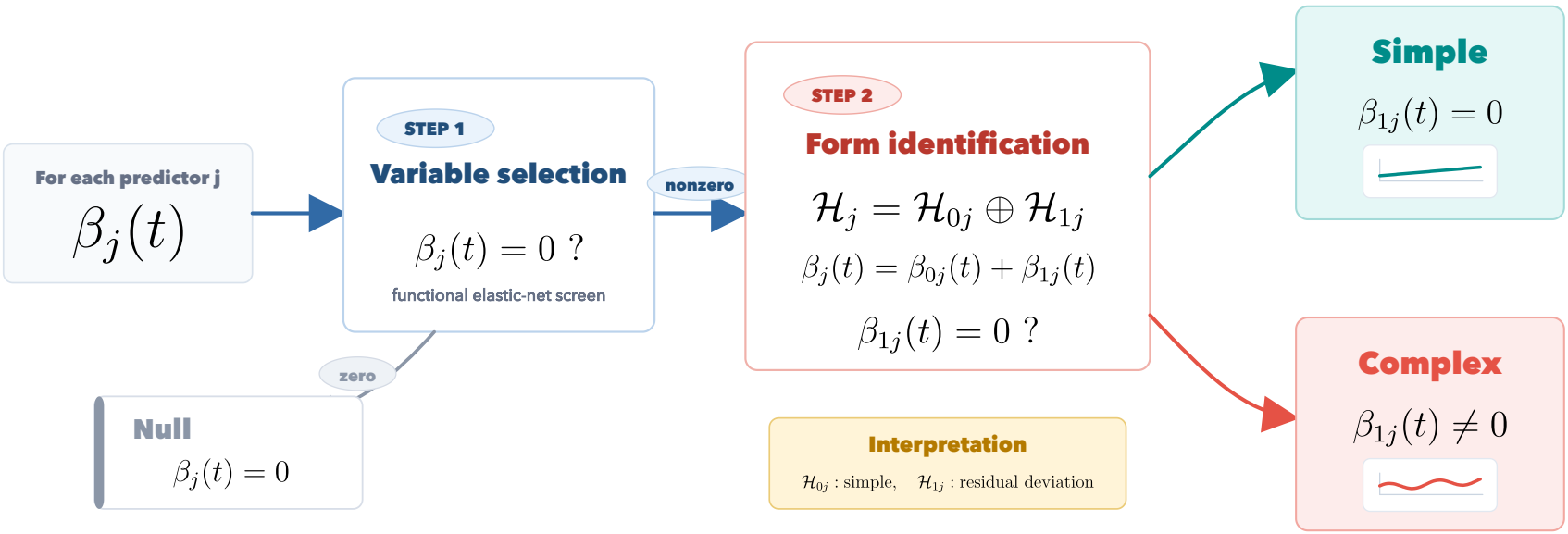
Complex effect

$\beta(t) = \beta^{(0)}(t) + \beta^{(1)}(t), \qquad \beta^{(1)} \not\equiv 0$



MoFI-FLR: two-step procedure

1. **Variable selection:** use a functional elastic-net to screen out null functional predictors.
2. **Model form identification:** for each selected predictor, distinguish simple effects from complex effects.



Step 1: Functional elastic-net (fEnet) screening

Goal: Identify active functional predictors by encouraging entire coefficient functions β_j to be zero.

Penalized estimation:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathcal{H}}{\arg\min} \left\{ \underbrace{\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\langle X_{ij}, \beta_j\rangle_2\right)^2}_{\text{Data fit}} + \underbrace{\sum_{j=1}^{p}\text{Pen}(\beta_j; \boldsymbol{\lambda})}_{\text{Sparsity and regularization}} \right\},$$

where $\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_p$ and $\boldsymbol{\lambda}$ controls the penalty strength.

Step 1: Functional elastic-net (fEnet) screening

Goal: Identify active functional predictors by encouraging entire coefficient functions β_j to be zero.

Penalized estimation:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta}\in\mathcal{H}}{\arg\min} \left\{ \underbrace{\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j=1}^{p}\langle X_{ij}, \beta_j\rangle_2\right)^2}_{\text{Data fit}} + \underbrace{\sum_{j=1}^{p}\text{Pen}(\beta_j;\boldsymbol{\lambda})}_{\text{Sparsity and regularization}} \right\},$$

where $\mathcal{H} = \mathcal{H}_1 \times \cdots \times \mathcal{H}_p$ and $\boldsymbol{\lambda}$ controls the penalty strength.

Computational challenge: The penalty involves the RKHS norm $\|\beta_j\|_{\mathcal{H}_j}$, which complicates direct optimization.

Key question: Can we reparameterize the model so that the penalty involves only $\mathbb{L}_2$ norms?

Model reparameterization

Idea: Represent the RKHS coefficient β through an $\mathbb{L}_2$ function f.

Kernel and its square root:

$$K(s,t) = \sum_{k=1}^{\infty} \theta_k \, \varphi_k(s)\varphi_k(t), \qquad K^{1/2}(s,t) = \sum_{k=1}^{\infty} \sqrt{\theta_k} \, \varphi_k(s)\varphi_k(t).$$

The corresponding square-root operator is

$$(\mathscr{K}^{1/2} f)(t) = \int_0^1 K^{1/2}(s,t) f(s) \, ds.$$

Key identity Cai and Yuan (2012)

For every $\beta \in \mathcal{H}(K)$, there exists $f \in \mathbb{L}_2[0,1]$ such that

$$\boxed{\beta = \mathscr{K}^{1/2} f} \qquad \text{and} \qquad \boxed{\|\beta\|_{\mathcal{H}(K)} = \|f\|_2}.$$

Benefit: Optimize over f with an $\mathbb{L}_2$-**norm penalty** instead of over β with an RKHS-norm penalty.

Reparameterize coefficients and predictors:

$$\beta_j = \mathscr{K}_j^{1/2} f_j, \qquad \widetilde{X}_{ij} = \mathscr{K}_j^{1/2} X_{ij}.$$

Estimate the transformed coefficients:

$$\widehat{\boldsymbol{f}} = \underset{\boldsymbol{f} \in \mathbb{L}_2^p}{\arg\min} \left\{ \frac{1}{2n} \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{p} \langle \widetilde{X}_{ij}, f_j \rangle_2 \right)^2 + \underbrace{\lambda_1 \sum_{j=1}^{p} \|\Psi_j f_j\|_2}_{\text{Group sparsity}} + \underbrace{\frac{\lambda_2}{2} \sum_{j=1}^{p} \|f_j\|_2^2}_{\text{Estimation stability}} \right\}.$$

A practical choice of penalty scaling is

$$\|\Psi_j f_j\|_2 = \left\{ \frac{1}{n} \sum_{i=1}^{n} \langle \widetilde{X}_{ij}, f_j \rangle_2^2 + \theta \|f_j\|_2^2 \right\}^{1/2}, \qquad \theta > 0.$$

Variable selection: Retain predictors whose estimated coefficient functions are nonzero:

$$\widehat{\mathcal{S}} = \{ j : \widehat{f}_j \not\equiv 0 \} = \{ j : \|\widehat{f}_j\|_2 > 0 \}.$$

Step 2: Model Form Identification

Goal: For each selected predictor $j \in \widehat{\mathcal{S}}$, determine whether its effect is simple or complex.

Decompose the coefficient function:

$$\beta_j = \underbrace{\beta_j^{(0)}}_{\text{Simple component}} + \underbrace{\beta_j^{(1)}}_{\text{Complementary component}}, \qquad \beta_j^{(r)} \in \mathcal{H}_{rj}.$$

Step 2: Model Form Identification

Goal: For each selected predictor $j \in \widehat{\mathcal{S}}$, determine whether its effect is simple or complex.

Decompose the coefficient function:

$$\beta_j = \underbrace{\beta_j^{(0)}}_{\text{Simple component}} + \underbrace{\beta_j^{(1)}}_{\text{Complementary component}}, \qquad \beta_j^{(r)} \in \mathcal{H}_{rj}.$$

Reparameterize each component: For $r = 0, 1$,

$$\beta_j^{(r)} = (\mathscr{K}_j^{(r)})^{1/2} f_j^{(r)}, \qquad \widetilde{X}_{ij}^{(r)} = (\mathscr{K}_j^{(r)})^{1/2} X_{ij}.$$

The contribution of predictor j becomes

$$\langle X_{ij}, \beta_j \rangle_2 = \langle \widetilde{X}_{ij}^{(0)}, f_j^{(0)} \rangle_2 + \langle \widetilde{X}_{ij}^{(1)}, f_j^{(1)} \rangle_2.$$

Step 2: Model Form Identification

Goal: For each selected predictor $j \in \widehat{\mathcal{S}}$, determine whether its effect is simple or complex.

Decompose the coefficient function:

$$\beta_j = \underbrace{\beta_j^{(0)}}_{\text{Simple component}} + \underbrace{\beta_j^{(1)}}_{\text{Complementary component}}, \qquad \beta_j^{(r)} \in \mathcal{H}_{rj}.$$

Reparameterize each component: For $r = 0, 1$,

$$\beta_j^{(r)} = (\mathscr{K}_j^{(r)})^{1/2} f_j^{(r)}, \qquad \widetilde{X}_{ij}^{(r)} = (\mathscr{K}_j^{(r)})^{1/2} X_{ij}.$$

The contribution of predictor j becomes

$$\langle X_{ij}, \beta_j \rangle_2 = \langle \widetilde{X}_{ij}^{(0)}, f_j^{(0)} \rangle_2 + \langle \widetilde{X}_{ij}^{(1)}, f_j^{(1)} \rangle_2.$$

Key question: Is the complementary component needed?

$$f_j^{(1)} \equiv 0 \qquad \text{or} \qquad f_j^{(1)} \not\equiv 0?$$

Jointly estimate both components by minimizing

$$\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(0)}, f_j^{(0)}\rangle_2 - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(1)}, f_j^{(1)}\rangle_2\right)^2 + \sum_{j\in\widehat{\mathcal{S}}}\mathrm{Pen}\left(f_j^{(1)}\right),$$

where

$$\mathrm{Pen}\left(f_j^{(1)}\right) = \underbrace{\bar{\lambda}_1\|\Psi_j^{(1)} f_j^{(1)}\|_2}_{\text{Select model form}} + \underbrace{\frac{\bar{\lambda}_2}{2}\|f_j^{(1)}\|_2^2}_{\text{Stabilize estimation}}.$$

Jointly estimate both components by minimizing

$$\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(0)}, f_j^{(0)}\rangle_2 - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(1)}, f_j^{(1)}\rangle_2\right)^2 + \sum_{j\in\widehat{\mathcal{S}}}\mathrm{Pen}\left(f_j^{(1)}\right),$$

where

$$\mathrm{Pen}\left(f_j^{(1)}\right) = \underbrace{\bar{\lambda}_1\|\Psi_j^{(1)} f_j^{(1)}\|_2}_{\text{Select model form}} + \underbrace{\frac{\bar{\lambda}_2}{2}\|f_j^{(1)}\|_2^2}_{\text{Stabilize estimation}}.$$

Only the complementary component is penalized:

- The simple component is estimated without penalization.
- The complementary component can shrink entirely to zero.

Jointly estimate both components by minimizing

$$\frac{1}{2n}\sum_{i=1}^{n}\left(Y_i - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(0)}, f_j^{(0)}\rangle_2 - \sum_{j\in\widehat{\mathcal{S}}}\langle \widetilde{X}_{ij}^{(1)}, f_j^{(1)}\rangle_2\right)^2 + \sum_{j\in\widehat{\mathcal{S}}}\mathrm{Pen}\left(f_j^{(1)}\right),$$

where

$$\mathrm{Pen}\left(f_j^{(1)}\right) = \underbrace{\bar{\lambda}_1\|\Psi_j^{(1)} f_j^{(1)}\|_2}_{\text{Select model form}} + \underbrace{\frac{\bar{\lambda}_2}{2}\|f_j^{(1)}\|_2^2}_{\text{Stabilize estimation}}.$$

Only the complementary component is penalized:

- The simple component is estimated without penalization.
- The complementary component can shrink entirely to zero.

Identify the effect form:

$$\widehat{f}_j^{(1)} \equiv 0 \quad \Longrightarrow \quad \text{Simple effect},$$
$$\widehat{f}_j^{(1)} \not\equiv 0 \quad \Longrightarrow \quad \text{Complex effect}.$$

Model-Form Selection Consistency

Let $\mathcal{S} = \{j : \beta_j \not\equiv 0\}$ be the true active set.

True effect forms:

$$\mathcal{S}_1 = \{j \in \mathcal{S} : \beta_j^{(1)} \equiv 0\} \quad \text{(simple effects)},$$
$$\mathcal{S}_2 = \{j \in \mathcal{S} : \beta_j^{(1)} \not\equiv 0\} \quad \text{(complex effects)}.$$

Estimated effect forms:

$$\widehat{\mathcal{S}}_1 = \{j \in \widehat{\mathcal{S}} : \widehat{f}_j^{(1)} \equiv 0\},$$
$$\widehat{\mathcal{S}}_2 = \{j \in \widehat{\mathcal{S}} : \widehat{f}_j^{(1)} \not\equiv 0\}.$$

MoFI consistency

As the sample size increases, recover both sets with probability approaching one:

$$\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2\right) \longrightarrow 1 \qquad \text{as } n \to \infty.$$

Why Does the Two-Step Procedure Work?

1. **Step 1: Recover the active set** (Guo et al., 2026)

$$\Pr\left(\widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_1.$$

2. **Step 2: Recover the effect forms, given correct screening**

$$\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2 \,\middle|\, \widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_2.$$

Why Does the Two-Step Procedure Work?

1. **Step 1: Recover the active set** (Guo et al., 2026)

$$\Pr\left(\widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_1.$$

2. **Step 2: Recover the effect forms, given correct screening**

$$\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2 \,\middle|\, \widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_2.$$

Combine the two guarantees:

$$\begin{aligned}\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2\right) &\geq (1 - P_1)(1 - P_2) \\ &\geq 1 - P_1 - P_2.\end{aligned}$$

Thus, $P_1 \to 0$ and $P_2 \to 0$ imply **MoFI consistency**.

Why Does the Two-Step Procedure Work?

1. **Step 1: Recover the active set** (Guo et al., 2026)

$$\Pr\left(\widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_1.$$

2. **Step 2: Recover the effect forms, given correct screening**

$$\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2 \,\middle|\, \widehat{\mathcal{S}} = \mathcal{S}\right) \geq 1 - P_2.$$

Combine the two guarantees:

$$\begin{aligned}\Pr\left(\widehat{\mathcal{S}}_1 = \mathcal{S}_1, \widehat{\mathcal{S}}_2 = \mathcal{S}_2\right) &\geq (1 - P_1)(1 - P_2) \\ &\geq 1 - P_1 - P_2.\end{aligned}$$

Thus, $P_1 \to 0$ and $P_2 \to 0$ imply **MoFI consistency**.

Remaining challenge: controlling the second-stage error P_2.

Theory: Model-Form Selection Consistency

Theorem (informal)

Under regularity conditions and suitable tuning parameters, the two-stage MoFI-FLR procedure satisfies

$$\Pr\left(\widehat{\mathcal{S}}_0 = \mathcal{S}_0,\, \widehat{\mathcal{S}}_1 = \mathcal{S}_1\right) \geq 1 - \underbrace{\exp\left(-D\frac{n\lambda_2^2}{q}\right)}_{\text{Step 1: variable-selection error}} - \underbrace{\exp\left(-\bar{D}\frac{n\bar{\lambda}_2^2}{q_1}\right)}_{\text{Step 2: form-selection error}},$$

- $q = |\mathcal{S}|$: number of active predictors.
- $q_1 = |\mathcal{S}_1|$: number of complex effects.

Theory: Model-Form Selection Consistency

Theorem (informal)

Under regularity conditions and suitable tuning parameters, the two-stage MoFI-FLR procedure satisfies

$$\Pr\left(\widehat{\mathcal{S}}_0 = \mathcal{S}_0,\, \widehat{\mathcal{S}}_1 = \mathcal{S}_1\right) \geq 1 - \underbrace{\exp\left(-D\frac{n\lambda_2^2}{q}\right)}_{\text{Step 1: variable-selection error}} - \underbrace{\exp\left(-\bar{D}\frac{n\bar{\lambda}_2^2}{q_1}\right)}_{\text{Step 2: form-selection error}},$$

- $q = |\mathcal{S}|$: number of active predictors.
- $q_1 = |\mathcal{S}_1|$: number of complex effects.

Consistency: Under the stated conditions, if

$$\frac{n\lambda_2^2}{q} \longrightarrow \infty \qquad \text{and} \qquad \frac{n\bar{\lambda}_2^2}{q_1} \longrightarrow \infty,$$

both error bounds vanish, yielding **correct recovery of all effect forms with probability approaching one**.

Restrictions on $\lambda_1, \lambda_2, \bar{\lambda}_1, \bar{\lambda}_2$

$$\lambda_1/\lambda_2 > \left(\frac{3}{\gamma} - 2\right) C_{\max}^{-1},$$

$$D_{1,1}^* > \lambda_1 > D_{1,2}^* \frac{\tau^{1/2}(1+\sigma)}{C_{\min}\gamma} \left\{\frac{\log(p-q)}{n}\right\}^{1/2},$$

$$D_{2,1}^* > \lambda_2 > D_{2,2}^* \frac{\tau(1+\sigma)(\rho_1+1)}{(C_{\min}/C_{\max})^2\gamma^2} \max\left\{\frac{q\log(p-q)}{n}, \left(\frac{q^2}{n}\right)^{1/2}\right\}.$$

$$\bar{\lambda}_1/\bar{\lambda}_2 > \left(\frac{3}{\bar{\gamma}} - 2\right) C_{\max}^{-1},$$

$$\bar{D}_{1,1}^* > \bar{\lambda}_1 > \bar{D}_{1,2}^* \frac{\tau^{1/2}(1+\sigma)}{C_{\min}\bar{\gamma}} \left\{\frac{\log(q_0)}{n}\right\}^{1/2},$$

$$\bar{D}_{2,1}^* > \bar{\lambda}_2 > \bar{D}_{2,2}^* \frac{\tau(1+\sigma)(\bar{\rho}_1+1)}{(C_{\min}/C_{\max})^2\bar{\gamma}^2} \max\left\{\frac{q_1\log(q_0)}{n}, \left(\frac{q_1^2}{n}\right)^{1/2}\right\}.$$

Key Regularity Conditions

1. **Limited functional complexity**

   Eigenvalues of functional predictors decay sufficiently fast.

2. **Limited predictor dependence**

   - Active and inactive predictors are not too strongly dependent.
   - Predictors with simple and complex effects are not too strongly dependent in the complementary space.
   - Predictors within each group are not nearly collinear.

3. **Sufficient signal strength**

   The smallest nonzero signal is large enough to overcome noise and regularization.

Prediction Performance

Excess prediction risk:

$$\mathcal{R}(\widehat{\boldsymbol{\beta}}) = \mathbb{E}^* \left[\left\{ \sum_{j=1}^{p} \left\langle X_j^*, \beta_{0j} - \widehat{\beta}_j \right\rangle_2 \right\}^2 \right],$$

where $\mathbb{E}^*$ averages over an independent test observation.

Step 1 (fEnet):

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\text{fenet})}\right) = O_p\left(\frac{q\lambda_1^2}{\lambda_2}\right).$$

Step 2 (MoFI): Under a partially separable covariance structure,

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\text{MoFI})}\right) = O_p\left[\frac{qM_0}{n}\left(1 + \frac{q_1\bar{\lambda}_1^2}{\bar{\lambda}_2^2}\right) + \frac{q_1\bar{\lambda}_1^2}{\bar{\lambda}_2}\right].$$

- q: number of active predictors; q_1: number of complex effects.
- M_0: dimension of the simple-component space.

High-Dimensional Regime: Polynomial Growth

Dimension and sparsity:

$$p \asymp n^{\alpha}, \qquad q \asymp n^{\varsigma}, \qquad \alpha > \varsigma, \qquad 0 < \varsigma < \frac{1}{4}.$$

Tuning rates:

$$\lambda_1 \asymp \lambda_2 \asymp \bar{\lambda}_1 \asymp \bar{\lambda}_2 \asymp n^{-(1/2-\varsigma)}.$$

Selection and prediction guarantees

MoFI achieves **model-form selection consistency**, and

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\mathrm{fenet})}\right) = O_p\left(n^{-(1/2-2\varsigma)}\right) = o_p(1),$$

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\mathrm{MoFI})}\right) = O_p\left(n^{-(1/2-2\varsigma)}\right) = o_p(1).$$

Takeaway: Both estimators have vanishing excess prediction risk, with the same upper-bound rate.

Ultra-High-Dimensional Regime

Dimension and sparsity:

$$\log p = O\left(n^{1-2\varsigma}\right), \qquad q \asymp n^{\varsigma}, \qquad 0 < \varsigma < \frac{1}{4}.$$

Stage-specific tuning rates:

$$\lambda_1 \asymp \lambda_2 \asymp n^{-\varsigma}, \qquad \bar{\lambda}_1 \asymp \bar{\lambda}_2 \asymp n^{-(1/2-\varsigma)}.$$

Selection and prediction guarantees

MoFI achieves **model-form selection consistency**, and

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\mathrm{fenet})}\right) = O_p(1),$$

$$\mathcal{R}\left(\widehat{\boldsymbol{\beta}}^{(\mathrm{MoFI})}\right) = O_p\left(n^{-(1/2-2\varsigma)}\right) = o_p(1).$$

Takeaway: MoFI guarantees vanishing excess prediction risk; the fEnet bound alone does not establish this.

Simulation design

Functional predictors. Generate

$$X_{ij}(t) = \sum_{k=1}^{30} z_{ijk} \nu_k^{1/2} \phi_k(t), \qquad \nu_k = \exp(-k/4),$$

using the cosine basis

$$\phi_1(t) = 1, \qquad \phi_k(t) = \sqrt{2}\cos\{(k-1)\pi t\}, \quad k > 1.$$

The scores are correlated across predictors:

$$\boldsymbol{z}_{i\cdot k} = (z_{i1k}, \ldots, z_{ipk})^\top \sim N(\mathbf{0}, \boldsymbol{\Sigma}_p), \qquad (\boldsymbol{\Sigma}_p)_{j\ell} = 0.5^{|j-\ell|}.$$

Signal structure. We set

$$(p, q, q_0) = (100, 10, 5),$$

so 10 predictors are active and 5 active effects are simple.

- Scenario I: simple effects are constant.
- Scenario II: simple effects are low-frequency periodic.
- Scenario III: simple effects combine constant and periodic components.

Simulation results (Prediction)

Table: Mean and $(5\%, 95\%)$ quantile of the relative excess risk (RER) computed over 200 simulation replicates under three scenarios.

Scenario	n	σ	fENet	fENet-refine	MoFI-fix	MoFI-optim
I	125	1	0.0194 (0.0111, 0.0271)	0.0176 (0.0106, 0.0229)	0.0115 (0.0054, 0.0144)	**0.0112** (0.0053, 0.0134)
		2	0.0536 (0.0279, 0.1199)	0.0519 (0.0239, 0.1374)	0.0355 (0.0146, 0.0918)	**0.0349** (0.0144, 0.1230)
		4	0.1895 (0.1031, 0.3297)	0.1957 (0.1086, 0.3373)	0.1858 (0.0718, 0.3418)	**0.1740** (0.0675, 0.3284)
		8	0.5413 (0.3194, 0.8469)	**0.5229** (0.3126, 0.7761)	0.5806 (0.3580, 0.9797)	0.6049 (0.3210, 0.9853)
	250	1	0.0064 (0.0049, 0.0086)	0.0056 (0.0042, 0.0073)	0.0036 (0.0026, 0.0049)	**0.0033** (0.0023, 0.0047)
		2	0.0178 (0.0126, 0.0239)	0.0176 (0.0123, 0.0241)	0.0114 (0.0075, 0.0158)	**0.0112** (0.0077, 0.0159)
		4	0.0501 (0.0317, 0.0830)	0.0537 (0.0345, 0.0849)	0.0385 (0.0213, 0.0863)	**0.0363** (0.0213, 0.0678)
		8	**0.2471** (0.1531, 0.3674)	0.2639 (0.1629, 0.3929)	0.2799 (0.1615, 0.4270)	0.2551 (0.1489, 0.4019)
II	125	1	0.0190 (0.0093, 0.0606)	0.0197 (0.0097, 0.0599)	0.0120 (0.0045, 0.0433)	**0.0116** (0.0047, 0.0435)
		2	0.0507 (0.0232, 0.0898)	0.0533 (0.0246, 0.0906)	**0.0395** (0.0140, 0.0785)	0.0398 (0.0137, 0.0820)
		4	0.1430 (0.0893, 0.2440)	0.1523 (0.0964, 0.2712)	0.1417 (0.0624, 0.2558)	**0.1401** (0.0768, 0.2567)
		8	0.4386 (0.2197, 0.7432)	**0.4202** (0.2274, 0.6710)	0.4934 (0.2462, 0.8986)	0.4880 (0.2459, 0.8932)
	250	1	0.0054 (0.0040, 0.0073)	0.0050 (0.0037, 0.0065)	0.0032 (0.0023, 0.0043)	**0.0029** (0.0019, 0.0041)
		2	0.0152 (0.0106, 0.0209)	0.0157 (0.0110, 0.0211)	**0.0103** (0.0073, 0.0142)	**0.0103** (0.0070, 0.0155)
		4	0.0541 (0.0282, 0.0893)	0.0586 (0.0320, 0.0901)	0.0449 (0.0194, 0.0832)	**0.0448** (0.0198, 0.0846)
		8	**0.1803** (0.1089, 0.2969)	0.1920 (0.1109, 0.3193)	0.2067 (0.1058, 0.3478)	0.1840 (0.1023, 0.3261)

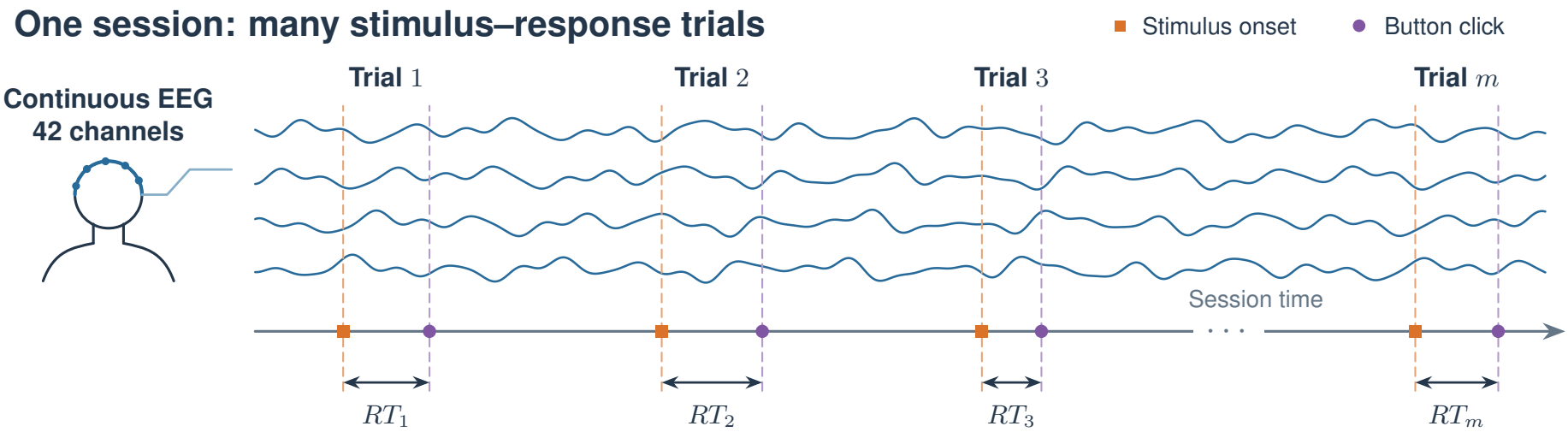
Simulation results (Selection)

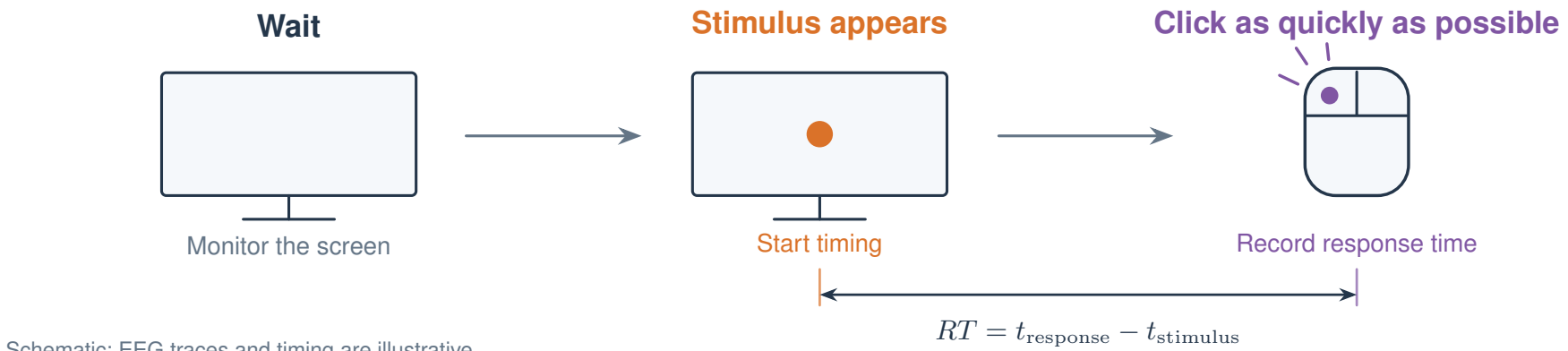
Table: Variable-selection and form-identification metrics over 200 simulation replicates under three scenarios.

Scenario	n	σ	FPR$\times 100\%$	FNR$\times 100\%$	$r(0\to 1)\times 100\%$		$r(1\to 0)\times 100\%$	
					fix	optim	fix	optim
I	125	1	0.19	0.20	0.10	9.00	0.00	0.00
		2	0.81	1.40	2.50	9.70	0.80	0.10
		4	2.67	13.30	8.80	30.90	15.50	2.20
		8	4.55	39.58	14.79	29.94	29.09	7.64
	250	1	0.00	0.00	0.00	2.60	0.00	0.00
		2	0.18	0.00	0.30	3.70	0.00	0.00
		4	0.18	0.50	0.10	6.00	1.40	0.00
		8	1.16	27.80	1.50	25.10	30.20	3.20
II	125	1	0.41	1.55	0.80	8.20	0.00	0.00
		2	0.60	8.85	2.30	9.40	0.10	0.00
		4	1.74	27.30	5.90	22.20	4.50	0.40
		8	4.96	45.62	13.37	23.03	13.37	2.25
	250	1	0.01	0.00	0.00	2.50	0.00	0.00
		2	0.18	0.05	0.00	6.20	0.00	0.00
		4	0.24	9.20	0.20	5.40	0.20	0.00
		8	0.69	41.15	0.50	10.70	12.70	0.20

Psychomotor vigilance task (PVT; Hinss et al., 2023)

Participants monitor a screen for a visual stimulus, and are instructed to press a button as quickly as possible upon stimulus onset.





Schematic; EEG traces and timing are illustrative.

Real Data: Analysis Setup

Data and variables

- **Design:** 29 participants, 3 sessions per participant, and 90 trials per session.
- **Observation unit:** One trial.
- **Response:** Trial-level reaction time.
- **Predictors:** 42 EEG channel curves per trial, with time rescaled to $[0, 1]$.

Simple-component space

$$\beta_j^{(0)}(t) = \sum_{k=1}^{3} a_{jk} g_k(t),$$

where g_k are Gaussian basis functions with centers motivated by sensory and cognitive processing windows:

Scaled time	Latency	ERP window
0.3	180 ms	P1/N1
0.5	300 ms	P2/N2
0.7	420 ms	P3b

Three Gaussian basis functions

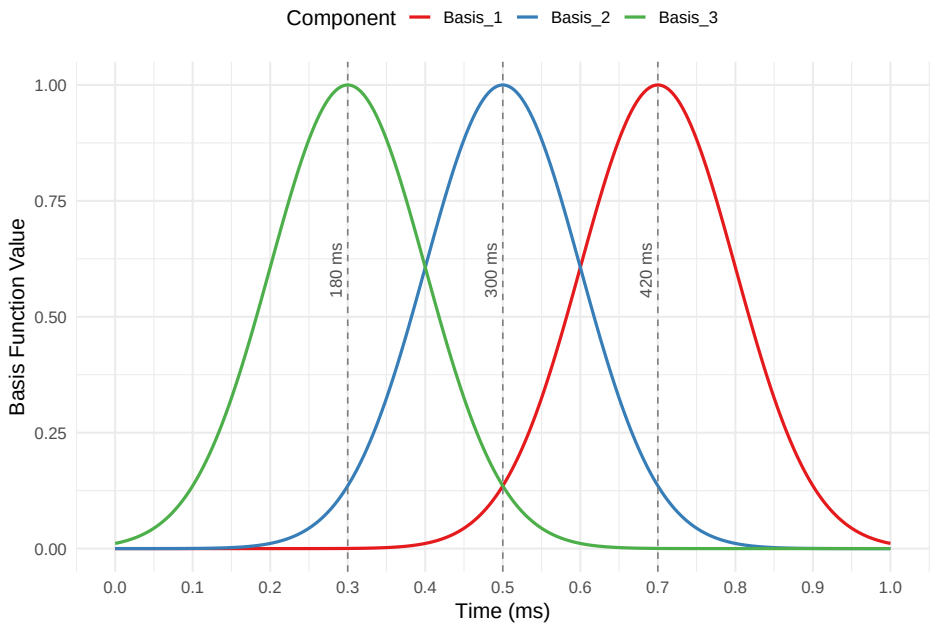


Figure: Three Gaussian basis functions

- $K(s,t) = \exp\left\{-50(s-t)^2\right\}$ and $K_0(s,t) = \Phi_0^\top(s) G^{-1} \Phi_0(t)$.
- $\Phi_0(t) = \left(K(t_1,t), K(t_2,t), K(t_3,t)\right)^\top$ and $G = \left(K(t_{r_1}, t_{r_2})\right)_{1 \le r_1, r_2 \le 3}$ to be the 3×3 Gram matrix.

EEG channels on scalp surface

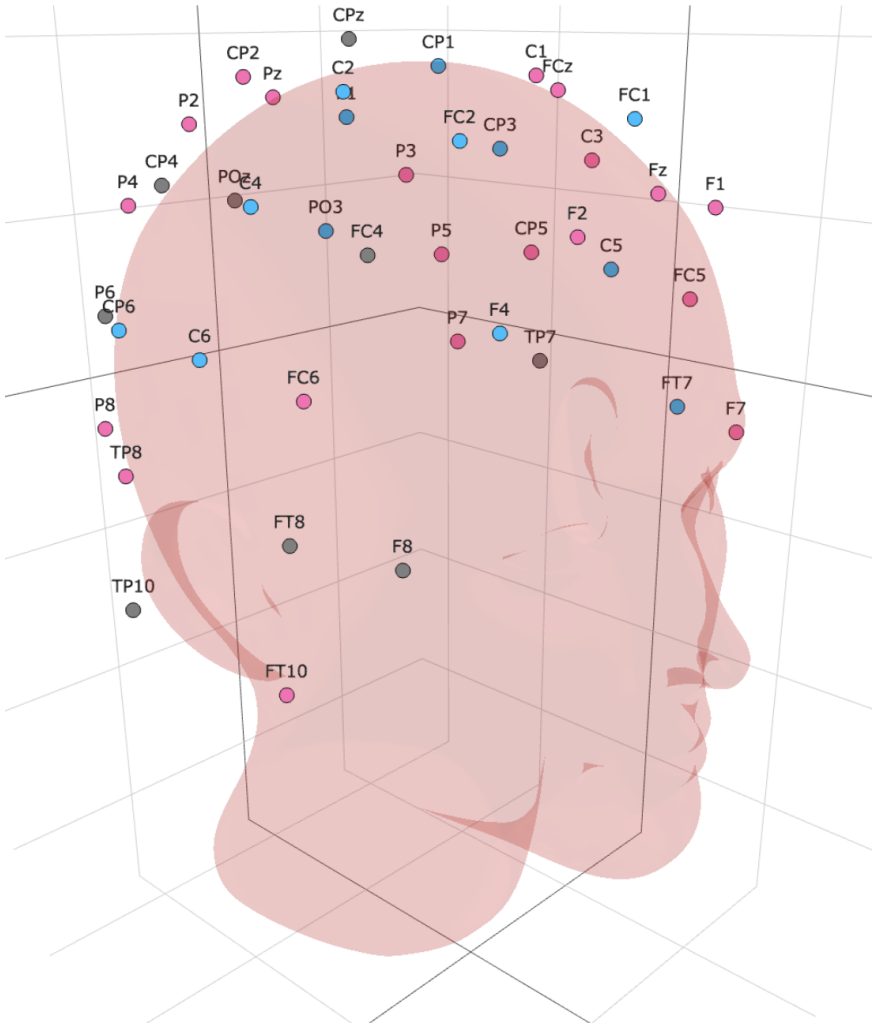
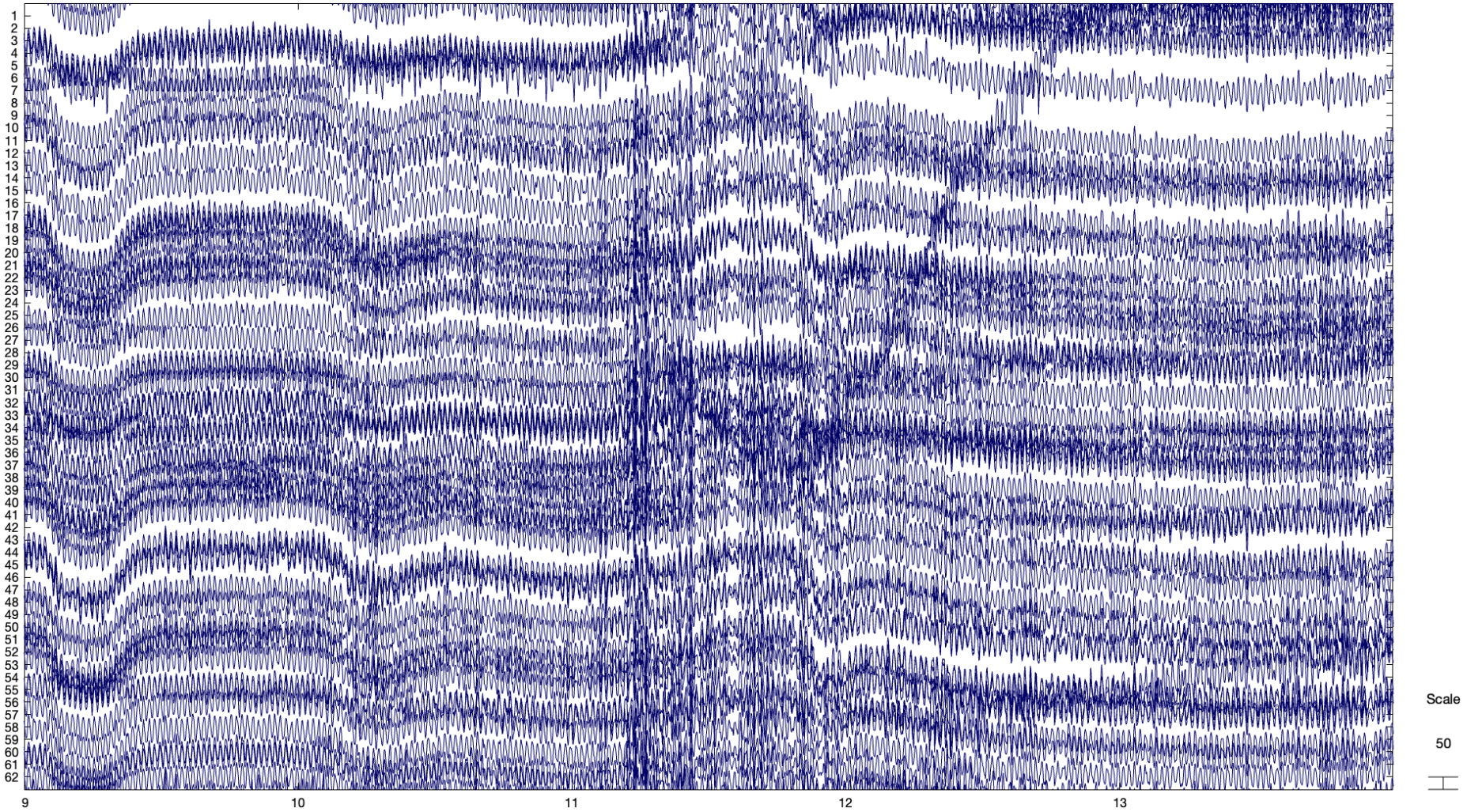
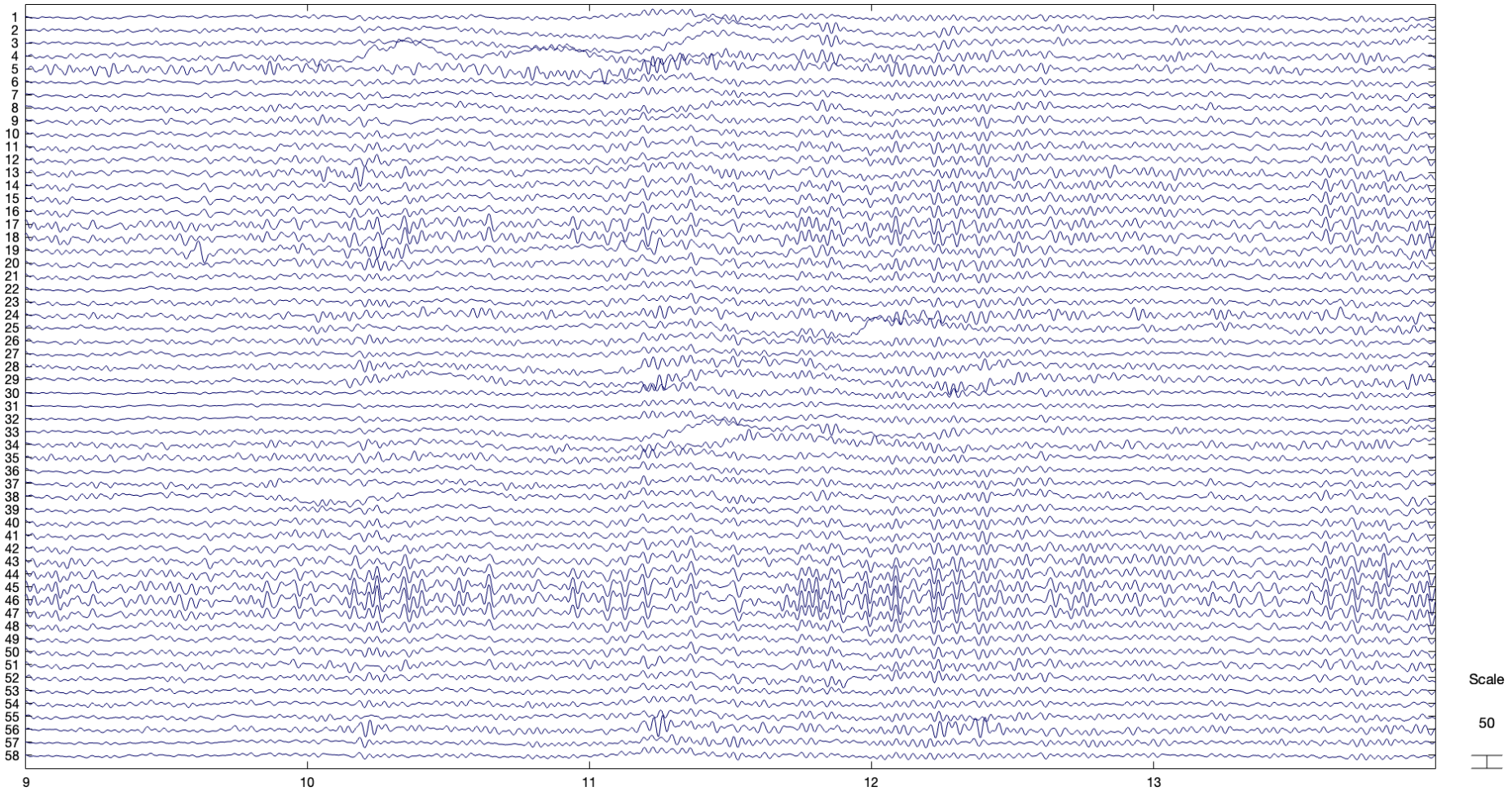


Figure: EEG channels on scalp surface

Raw EEG data


1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
9
10
11
12
13
Scale
50

Preprocessed EEG data



Results

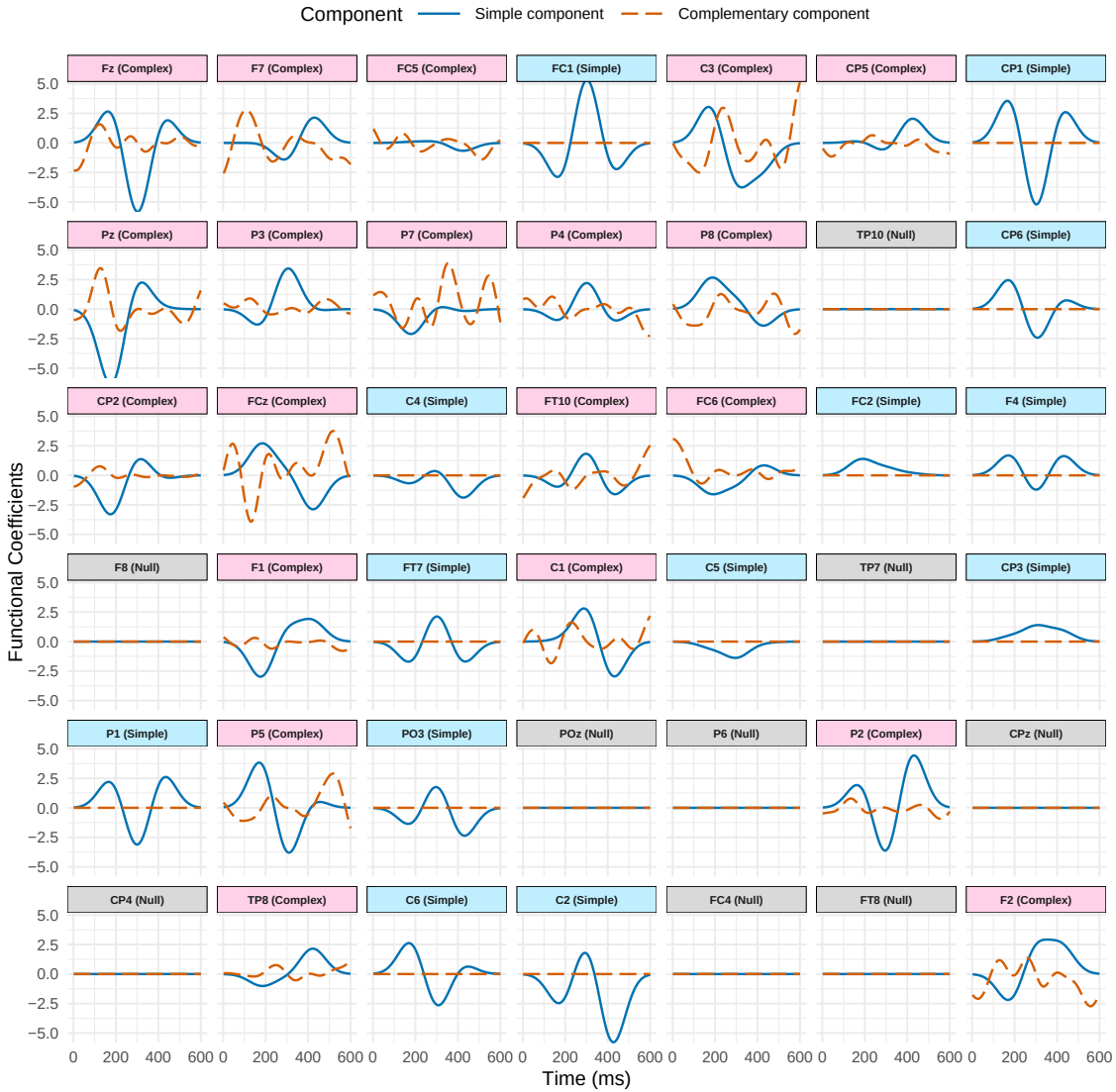


Figure: Estimated functional coefficients

Limitations and future directions

MoFI-FLR provides a first step toward model-form identification in high-dimensional FLR, but several important questions remain.

- **Choice of simple structure.**

  The current method requires the user to specify the null space $\mathcal{H}_{0j}$, which defines what counts as a simple effect.

- **Statistical inference.**

  The present work focuses on selection and estimation consistency. Inference tools such as confidence bands and formal tests remain to be developed.

- **Observation scheme.**

  The theory assumes functional predictors are fully observed. Extending MoFI-FLR to discretely observed or sparsely sampled functional data is an important next step.

Thank You!

References

Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3):337–404.

Cai, T. T. and Yuan, M. (2012). Minimax and adaptive prediction for functional linear regression. *Journal of the American Statistical Association*, 107:1201–1216.

Guo, X., Li, Y., and Hsing, T. (2026+). Variable selection and minimax prediction in high-dimensional functional linear model. *arXiv preprint arXiv:2310.14419*.

Hinss, M. F., Jahanpour, E. S., Somon, B., Pluchon, L., Dehais, F., and Roy, R. N. (2023). Open multi-session and multi-task eeg cognitive dataset for passive brain-computer interface applications. *Scientific Data*, 10(1):85.

Wahba, G. (1990). *Spline Models for Observational Data*. SIAM, Philadelphia.