Biostatistics, 2025, **26**(1), kxaf050
https://doi.org/10.1093/biostatistics/kxaf050
Article

High-dimensional inference for functional regression with an application to the Alzheimer's disease magnetoencephalography study

Huaqing Jin [1] and Fei Jiang [2,*]

[1]Department of Statistics and Data Science, Tsinghua University, Qinghuayuan, Beijing 100084, China
[2]Department of Epidemiology and Biostatistics, University of California, San Francisco, 16th Street, San Francisco, CA 94158, United States

*Corresponding author: Department of Epidemiology and Biostatistics, University of California, San Francisco, 16th Street, San Francisco, CA 94158, United States. Email: fei.jiang@ucsf.edu

SUMMARY

Alzheimer's disease (AD) is a progressive, chronic neurodegenerative disorder affecting millions worldwide. A new clinical magnetoencephalography (MEG) study was conducted to identify neural activity biomarkers and key brain regions in AD. Traditional methods for analyzing MEG data, which typically extract features from power spectral density, suffer from information loss. Furthermore, functional regression with variable selection tends to produce non-robust results, making it less ideal for drawing reliable scientific conclusions. To address these challenges, we propose a high-dimensional hypothesis testing (HDHT) framework for functional covariates and introduce a rigorous inference process to support scientific conclusions. We establish the theoretical properties of the HDHT framework and validate its performance through simulation studies. Applying the HDHT framework to the AD MEG data, we identify 19 important regions associated with cognitive functions that align with established AD pathophysiology. These findings suggest that the non-invasive MEG can be a potential low-risk and low-toxicity modality for monitoring neurodegenerative progression.

KEYWORDS: Alzheimer's disease; B-spline; generalized functional linear model; neuroimaging data.

1. INTRODUCTION

Alzheimer's disease (AD) is a progressive, chronic neurodegenerative disorder affecting over 6.5 million people in the United States and more than 30 million worldwide (World Health Organization 2023). Growing evidence has demonstrated that abnormal neural activities, such as subclinical epileptiform activity, are frequently found in people with AD and can accelerate the disease's progression. These findings underscore the importance of identifying crucial neural activity biomarkers associated with AD. Studying these biomarkers can help determine whether medications designed to suppress abnormal neural activities would slow neurodegeneration. Furthermore, neural activity biomarkers play an increasingly important role in developing neuromodulation treatments for AD and related diseases. For example, beta band oscillations from the subthalamic nucleus can be a therapeutic target for Parkinson's disease (Weinberger et al. 2006). Neuromodulation treatments that improve frontal gamma oscillatory responses are a promising

Received: June 26, 2025. **Revised:** October 15, 2025. **Accepted:** November 24, 2025

approach to delay cognitive decline (Shu et al. 2024). These studies highlight the significance of identifying neural activity biomarkers for AD.

To understand the neurodegeneration process, UCSF researchers conducted a clinical MEG study in 2022 to identify neural activity biomarkers for AD. The study recruited 176 patients aged 30 to 65 years, including 104 females and 72 males. Each patient underwent a resting-state magnetoencephalography (MEG) test and a cognitive assessment using the mini-mental state examination (MMSE).

After excluding outliers, we assemble a cohort consisting of 82 AD samples and 61 control samples. The objective is to investigate the relationship between the power spectrum density (PSD) of MEG brain signals and MMSE scores. It can help us understand the connection between the brain activities measured by MEG and cognitive function. A PSD of a brain signal represents how the signal's power is distributed across different frequencies. It can be estimated by taking the Fourier transform of the brain signal and then computing the squared magnitude of that transform. The scalp MEG signals are mapped onto the brain using the MEG source localization techniques (Michel and He 2019; Jin et al. 2023). The PSD is then calculated from these source brain signals.

As a preliminary study, we group the whole-brain PSD into delta, theta, alpha, and beta bands, corresponding to the PSD within the frequency ranges of 2 to 4 Hz, 4 to 8 Hz, 8 to 12 Hz, and 12 to 35 Hz, respectively. We then adopt a linear regression model to examine the relationship between MMSE scores and the average PSD in each frequency band for different brain regions of interest (ROIs). We perform hypothesis testing for each ROI $j = 1, \ldots, 68$ and each frequency band $b = 1, \ldots, 4$. The null hypothesis $\mathbf{H}_0^{(j,b)}$ is that the average PSD at ROI j in frequency band b has no effect on the MMSE score. Since there are a total of 68×4 hypothesis tests, the significance level is adjusted to $0.05/68/4$ to account for the multiple testing. The significant ROIs where PSDs exhibit a strong association with MMSE scores are presented in Fig. 1. This analysis reveals that significant ROIs vary across frequency bands, indicating that the relationship between PSD and MMSE scores is a function of frequencies. While this band-specific analysis provides initial insights, it treats each frequency band as an independent feature, potentially overlooking finer details within the continuous spectrum of PSD data. To address this limitation and enable a more refined analysis, we adopt the functional regression model, which considers the PSD as a continuous function across frequencies. Furthermore, given the high dimensionality of the PSD data across numerous ROIs, we propose a high-dimensional functional regression model to study the relationship. To ensure interpretable scientific conclusions, we also develop a hypothesis testing procedure to identify key predictors associated with MMSE scores within this high-dimensional framework.

Functional regression is a widely used method for modeling the relationship between a scalar response and low-dimensional functional predictors (Yuan and Cai 2010; Cai and Yuan 2012; Lian 2013; Kong et al. 2016) as well as high-dimensional predictors (Matsui and Konishi 2011; Aneiros et al. 2022). Various functional regression methods (Gertheiss et al. 2013; Fan et al. 2015; Zhang et al. 2022; Xue et al. 2024) have also been proposed to identify important predictors associated with the outcomes. However, a significant limitation of these methods is their high sensitivity to the choice of tuning parameters. Furthermore, they fail to provide uncertainty measures for the selected predictors. Consequently, these methods are often inadequate for drawing scientific conclusions. In contrast, hypothesis testing is specifically designed to address such issues by offering uncertainty quantification. Nevertheless, options for hypothesis testing under the high-dimensional generalized functional linear model framework remain scarce. To the best of our knowledge, the bootstrap method proposed by Xue and Yao (2021) stands out as the sole known testing approach in this area. However, this method has several limitations. Firstly, it is restricted to linear regression models and cannot handle binary outcomes under the generalized functional linear model framework. Handling binary outcomes is important, because these outcomes, such as whether a subject has AD, are the most reliable endpoints used in clinical diagnosis. Secondly, the method by Xue and Yao (2021) cannot incorporate baseline covariates. Incorporating baseline covariates such as age and education level is essential, as these are well-established factors that strongly affect cognitive

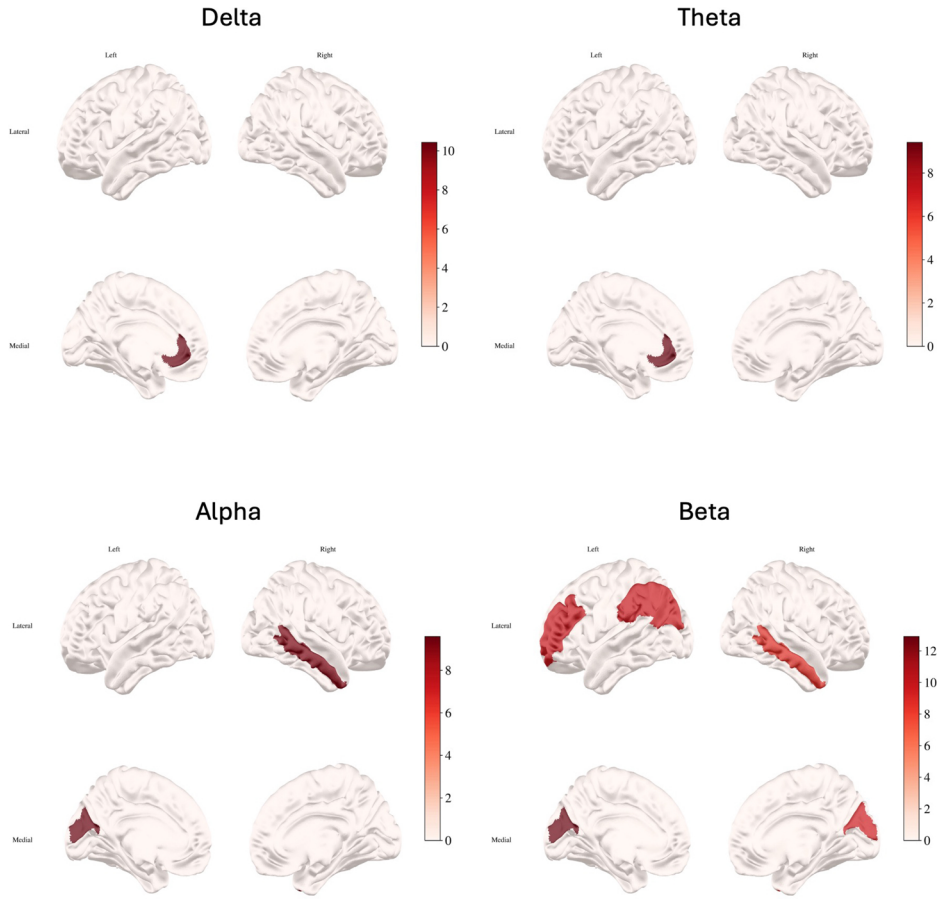

Fig. 1. Negative log *P*-values from the significant ROIs selected via fitting MMSE score on each ROI's average PSD at delta, theta, alpha, and beta bands, adjusted for age, gender, and education level. We conduct hypothesis testing for each ROI $j = 1, \ldots, 68$ and each frequency band $b = 1, \ldots, 4$. The null hypothesis $\mathbf{H}_0^{(j,b)}$ is that the average PSD at ROI j in frequency band b has no effect on the MMSE score. Since there are a total of 68×4 hypothesis tests, the significance level is adjusted to $0.05/68/4$ to account for the multiple testing.

function. Ignoring such covariates can lead to model misspecification, reduce interpretability, and potentially bias the estimated effects of other variables. Lastly, Xue and Yao (2021)'s method lacks the capability to test for linear hypotheses. Testing linear hypotheses of the functional parameters is particularly valuable for our study, as it allows us to determine whether the effects on cognitive function differ across distinct brain regions.

To overcome these challenges, we propose a novel high-dimensional hypothesis testing (HDHT) procedure for high-dimensional functional predictors. Our approach is built on the generalized functional linear model framework. It is able to test linear hypotheses and allows for incorporating baseline covariates. Furthermore, it also maintains power under the local alternative hypotheses (Lee and Young 2016; Shi et al. 2019). The local alternative hypothesis is an alternative hypothesis that is infinitesimally close to the null hypothesis and moves closer to it as the sample size (n) increases. By applying the HDHT to the AD MEG data, our approach identifies clinically meaningful neurophysiological markers for neurodegeneration. These markers can be used to track the neurodegenerative process during the pre-symptomatic stage, when positron emission tomography

(PET) and cerebrospinal fluid (CSF) tests are not recommended due to their associated toxicities and risks.

The rest of the paper is organized as follows. Section 2 presents the model, details the estimation procedure and introduces the test statistic. We apply the proposed method to analyze the data from the AD MEG study in Section 3. In Section 4, we validate the theoretical properties of the proposed method, and perform the benchmark comparison through simulation studies. We conclude with the findings in Section 5.

2. MODEL, ESTIMATION, AND TEST STATISTIC

2.1. Problem formulation

Consider a high-dimensional functional covariate $\mathbf{X}(s) = \{X_1(s), \ldots, X_d(s)\}^{\mathrm{T}} \in \mathbb{R}^d, s \in [0, \tau]$, where d is the number of functional predictors. Without loss of generality, we assume $\tau = 1$, because we can always project any compact intervals to $[0, 1]$. In our application, we consider $X_j(s)$ to be the PSD (in $\log_{10}$ scale) of the brain activity from the jth ROI. Therefore, s represents a specific frequency. Note that while we focus on the PSD in this paper, our method can also be applied to the original time series data. Let $\mathbf{Z} = (Z_1, \ldots, Z_q)^{\mathrm{T}}$ be a vector of baseline covariates with $q < \infty$. Let $p = d + q$ and Y be the outcome of interest, such as cognitive scores or a disease indicator. We consider the conditional distribution of Y_i given $\mathbf{X}_i, \mathbf{Z}_i$ for the independent and identically distributed $\{(Y_i, \mathbf{X}_i, \mathbf{Z}_i), i = 1, \ldots, n\}$ follows the exponential family distribution. The conditional density function of Y_i given $\mathbf{X}_i, \mathbf{Z}_i$ is given by

$$f\left\{Y_i, \boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{Z}_i + \int_0^1 \boldsymbol{\beta}^{\mathrm{T}}(s)\mathbf{X}_i(s)ds\right\} \tag{1}$$
$$\propto \exp\left[\frac{Y\left\{\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{Z}_i + \int_0^1 \boldsymbol{\beta}^{\mathrm{T}}(s)\mathbf{X}(s)ds\right\} - \psi\left\{\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{Z}_i + \int_0^1 \boldsymbol{\beta}^{\mathrm{T}}(s)\mathbf{X}(s)ds\right\}}{c(\sigma)}\right],$$

where $\boldsymbol{\beta}(s) = \{\beta_1(s), \ldots, \beta_d(s)\}^{\mathrm{T}}$ represents the functional effects, $\boldsymbol{\alpha}_0$ is an unknown true effect of the baseline covariates, $c(\sigma)$ is a scale parameter, and $\psi(\cdot)$ is a known function defined by the exponential family distribution. Since we are interested in $\boldsymbol{\alpha}_0$ and $\boldsymbol{\beta}(s)$, we rescale the loss function and assume $c(\sigma) = 1$.

For a vector $\mathbf{v} \in \mathbb{R}^p$ and a subset $S \subseteq (1, \ldots, p)$, we use $\mathbf{v}_S \in \mathbb{R}^{|S|}$ to denote the vector obtained by restricting $\mathbf{v}$ on the set S. We aim to test the hypothesis that

$$\mathbf{H}_0 : \mathbf{C}\boldsymbol{\beta}_{\mathcal{M}}(s) = \mathbf{t}(s) \text{ vs. } \mathbf{H}_1 : \mathbf{C}\boldsymbol{\beta}_{\mathcal{M}}(s) = \mathbf{t}(s) + \mathbf{h}_n(s), \tag{2}$$

where $\mathcal{M} \subset (1, \ldots, d)$ with $|\mathcal{M}| = m$ is the subset of indices for the functional coefficients of interest in the hypothesis. Here $\mathbf{C}$ is an $r \times m$ matrix and $\mathbf{t}(s)$, $\mathbf{h}_n(s)$ are the known functions, where $\mathbf{h}_n(s)$ quantifies the effect size under the alternative hypothesis. We allow the alternative hypothesis to be local, that is, $\|\mathbf{h}_n(s)\|_2$ can vanish when $n \to \infty$. For example, if we are interested in testing the difference of the effects from the first two functional predictors, we can set $\mathbf{C} = [1, -1]$, $\mathbf{t}(s) \equiv \mathbf{0}$ and $\mathbf{h}_n(s)$ to be some nonzero function. In summary, by varying $\mathbf{C}$, $\mathbf{t}$, $\mathbf{h}_n$, and $\mathcal{M}$, we can generate different linear hypotheses to test.

2.2. Parameter estimation

Assume that $\beta_j(s) \in C^{\omega}([0, 1]), j = 1, \ldots, d$, with ωth order derivatives and we approximate $\beta_j(s)$ by a B-spline function $\mathbf{B}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0j}$, where $\mathbf{B}(s)$ is the N dimensional B-spline basis function and $\boldsymbol{\gamma}_{0j}$ $j = 1, \ldots, d$ are the oracle projection coefficients. These coefficients ensure that $\mathbf{B}(s)^{\mathrm{T}}\boldsymbol{\gamma}_{0j}$ is the orthogonal projection of $\beta_j(s)$ onto the functional space spanned by the B-spline basis $\mathbf{B}(s)$. Let $\mathcal{S} = \{j : \beta_j(s) \neq 0, \exists s, j \in \mathcal{M}^c\}$ and $k_0 = |\mathcal{S}|$, and $\boldsymbol{\Gamma}_0 \equiv (\boldsymbol{\gamma}_{0j}, j = 1, \ldots, d)$.

Furthermore, we consider the loss function as the negative of the logarithm of the likelihood function

$$\mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\Gamma}) = -n^{-1} \sum_{i=1}^{n} \log \left[f \left\{ Y_i, \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{Z}_i + \int_0^1 \mathbf{B}(s)^{\mathrm{T}} \boldsymbol{\Gamma} \mathbf{X}_i(s) ds \right\} \right],$$

where $\boldsymbol{\Gamma} = (\boldsymbol{\gamma}_j, j = 1, \ldots, d)$. Let $\psi_i'(\boldsymbol{\alpha}, \boldsymbol{\Gamma})$, $\psi_i''(\boldsymbol{\alpha}, \boldsymbol{\Gamma})$ be the first and second derivatives of $\psi(t)$ evaluated at $t = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{Z}_i + \int_0^1 \mathbf{B}(s)^{\mathrm{T}} \boldsymbol{\Gamma} \mathbf{X}_i(s) ds$. Slightly abusing notation, we define $\psi_i\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}$, $\psi_i'\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}$, $\psi_i''\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}$ be the cumulant function and its first and second derivatives of $\psi_i(t)$ evaluated at $t = \boldsymbol{\alpha}^{\mathrm{T}} \mathbf{Z}_i + \int_0^1 \boldsymbol{\beta}(s)^{\mathrm{T}} \mathbf{X}_i(s) ds$. In a generalized linear model $\psi_i''(\boldsymbol{\alpha}, \boldsymbol{\Gamma}) > 0$ and we assume that there is a constant C_ψ such that $|\psi_i''(\cdot, \cdot)| \leq C_\psi$ uniformly. These conditions are often used in the generalized linear model to ensure the estimation consistency (Loh and Wainwright 2015).

Suppose we are interested in testing whether $\mathbf{C}\boldsymbol{\beta}_{\mathcal{M}}(s) = \mathbf{t}(s)$ or not. For a matrix and a vector, let $\| \cdot \|_p$ denote the L_p norm and l_p norm, respectively. When estimating $\boldsymbol{\alpha}, \boldsymbol{\Gamma}$, we minimize the regularized loss function

$$\widehat{\boldsymbol{\alpha}}, \widehat{\boldsymbol{\Gamma}} = \mathrm{argmin}_{\boldsymbol{\alpha}, \boldsymbol{\Gamma}} \left\{ \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\Gamma}) + \sum_{j \in \mathcal{M}^c} \rho_\lambda(\|N^{-1/2} \boldsymbol{\gamma}_j\|_2) \right\}, \text{ s. t.} \|\boldsymbol{\alpha}\|_1 + \sum_{j=1}^{d} \|N^{-1/2} \boldsymbol{\gamma}_j\|_2 \leq R, \tag{3}$$

where ρ_λ is some regularized function and λ is the tuning parameter, and R is a constant greater than $2\|\boldsymbol{\alpha}_0\|_1 + 2\sum_{j=1}^{d} \|N^{-1/2} \boldsymbol{\gamma}_{0j}\|_2$, $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\Gamma}} = (\widehat{\boldsymbol{\gamma}}_j, j = 1, \ldots, d)$ are the estimators of $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$, respectively. The additional constraints $\|\boldsymbol{\alpha}_0\|_1 + 2\sum_{j=1}^{d} \|N^{-1/2} \boldsymbol{\gamma}_{0j}\|_2$ are only used to facilitate the theoretical derivation to control the magnitude of the estimator and have little empirical effect. Note that in (3), we do not impose the L_2 penalty on $\boldsymbol{\gamma}_j$ for $j \in \mathcal{M}$, and hence do not shrink the estimators corresponding to the functional coefficients involved in the hypothesis. Since we do not shrink the estimators, the procedure can effectively test local alternative hypotheses.

We demonstrate that the estimators from (3) have asymptotic normality, which is the key to constructing the confidence intervals for $\boldsymbol{\alpha}_0$ and the point-wise confidence intervals for $\beta_j(s)$. Following the asymptotic normality, the statistical consistency of $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\boldsymbol{\Gamma}}$ is established. All the theoretical results are deferred in Appendix SG and Appendix SI of the Supplementary Material.

2.3. Computational algorithm

For notation simplicity, we define a generic $q + dN$ dimensional vector $\boldsymbol{\theta} \equiv (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_{2j}^{\mathrm{T}}, j = 1, \ldots, d)^{\mathrm{T}} \equiv \{\boldsymbol{\alpha}^{\mathrm{T}}, N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma})^{\mathrm{T}}\}^{\mathrm{T}}$, $\boldsymbol{\theta}_0 \equiv (\boldsymbol{\theta}_{01}^{\mathrm{T}}, \boldsymbol{\theta}_{02j}^{\mathrm{T}}, j = 1, \ldots, d)^{\mathrm{T}} \equiv \{\boldsymbol{\alpha}_0^{\mathrm{T}}, N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}_0)^{\mathrm{T}}\}^{\mathrm{T}}$ and $\widehat{\boldsymbol{\theta}} = (\widehat{\boldsymbol{\theta}}_1^{\mathrm{T}}, \widehat{\boldsymbol{\theta}}_{2j}^{\mathrm{T}}, j = 1, \ldots, d)^{\mathrm{T}} \equiv \{\widehat{\boldsymbol{\alpha}}^{\mathrm{T}}, N^{-1/2}\mathrm{vec}(\widehat{\boldsymbol{\Gamma}})^{\mathrm{T}}\}^{\mathrm{T}}$. Recall that $p - q$ is the total number of functional coefficients, $m = |\mathcal{M}|$ is the number of coefficients involved in the hypothesis, and $k_0 = |\mathcal{S}|$ is the number of nonzero coefficients not in $\mathcal{M}$. Without loss of generality, we assume that the B-spline coefficients corresponding to the $p - q - m - k_0$ zero $\beta_j(s)$'s are placed at the end of $\boldsymbol{\theta}_0$. That is, the last $(p - q - m - k_0)N$ elements of $\boldsymbol{\theta}_0$ are zero. We define $\boldsymbol{\theta}_{\mathcal{F}} = \{\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_{2j}^{\mathrm{T}}, j \in \mathcal{F}\}$ where $\mathcal{F}$ is a subset of $\{1, \ldots, d\}$ and define $\boldsymbol{\theta}_{0\mathcal{F}}$ and $\widehat{\boldsymbol{\theta}}_{\mathcal{F}}$ in a similar way and denote $\mathcal{L}(\boldsymbol{\theta}) \equiv \mathcal{L}(\boldsymbol{\alpha}, \boldsymbol{\Gamma})$.

We utilize the contractive Peaceman-Rachford splitting method (CPRSM) (He et al. 2014) to solve for $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ in (3). CPRSM splits the minimization of the objective function into two separate convex optimization steps, thereby bypassing the non-convexity issue of the original objective function. More specifically, let $\mathbf{D} = [\mathbf{0}_{dN,q}, \mathbf{I}_{dN \times dN}]$ and let $f_1(\boldsymbol{\theta}) = \mathcal{L}(\boldsymbol{\theta})$, where

$\boldsymbol{\theta} = (\boldsymbol{\theta}_1^{\mathrm{T}}, \boldsymbol{\theta}_{2j}^{\mathrm{T}}, j = 1, \ldots, d)$, and let $f_2(\boldsymbol{\Gamma}) = \sum_{j \in \mathcal{M}^c} \rho_\lambda(\|N^{-1/2}\boldsymbol{\gamma}_j\|_2)$. We estimate $\boldsymbol{\alpha}$ and $\boldsymbol{\Gamma}$ by minimizing

$$f_1(\boldsymbol{\theta}) + f_2(\boldsymbol{\Gamma}) \text{ subject to } \mathbf{D}\boldsymbol{\theta} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}) = \mathbf{0} \text{ and } \|\boldsymbol{\theta}_1\|_1 + \sum_{j=1}^{d} \|\boldsymbol{\theta}_{2j}\|_2 \le R,$$

where the constraints $\mathbf{D}\boldsymbol{\theta} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}) = \mathbf{0}$ is due to the fact that $\mathbf{D}\boldsymbol{\theta} = N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma})$. At the kth iteration, CPRSM updates $\boldsymbol{\theta}^{(k)}$ as

$$\begin{aligned}
\boldsymbol{\theta}^{k+1/2} &= \mathrm{argmin}_{\boldsymbol{\theta}} f_1(\boldsymbol{\theta}) - \boldsymbol{\rho}^{(k)\mathrm{T}}\{\mathbf{D}\boldsymbol{\theta} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}^{(k)})\} \\
&\quad + \frac{\kappa}{2}\|\mathbf{D}\boldsymbol{\theta} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}^{(k)})\|_2^2, \qquad (4)\\
\check{\boldsymbol{\theta}}_1, \|\check{\boldsymbol{\theta}}_{2j}\|_2 &= \mathrm{Proj}_{\left\{\boldsymbol{\theta}: \|\boldsymbol{\theta}_1\|_1 + \sum_{j=1}^{d}\|\boldsymbol{\theta}_{2j}\|_2 \le R\right\}}([\boldsymbol{\theta}_1^{(k+1/2)\mathrm{T}}, \|\boldsymbol{\theta}_{2j}^{(k+1/2)}\|_2, j = 1, \ldots, d]^{\mathrm{T}}), \qquad (5)\\
\boldsymbol{\theta}^{(k+1)} &= (\check{\boldsymbol{\theta}}_1^{\mathrm{T}}, \boldsymbol{\theta}_{2j}^{(k+1/2)\mathrm{T}}\|\check{\boldsymbol{\theta}}_{2j}\|_2/\|\boldsymbol{\theta}_{2j}^{(k+1/2)}\|_2, j = 1, \ldots, d)^{\mathrm{T}}, \qquad (6)\\
\boldsymbol{\rho}^{k+1/2} &= \boldsymbol{\rho}^{(k)} - \zeta\kappa\{\mathbf{D}\boldsymbol{\theta}^{(k+1)} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}^{(k)})\}, \qquad (7)\\
\boldsymbol{\Gamma}^{(k+1)} &= \mathrm{argmin}_{\boldsymbol{\Gamma}} f_2(\boldsymbol{\Gamma}) - \boldsymbol{\rho}^{(k+1/2)\mathrm{T}}\{\mathbf{D}\boldsymbol{\theta}^{(k+1)} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma})\} \qquad (8)\\
&\quad + \frac{\kappa}{2}\|\mathbf{D}\boldsymbol{\theta}^{(k+1)} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma})\|_2^2, \\
\boldsymbol{\rho}^{(k+1)} &= \boldsymbol{\rho}^{(k+1/2)} - \zeta\kappa\{\mathbf{D}\boldsymbol{\theta}^{(k+1)} - N^{-1/2}\mathrm{vec}(\boldsymbol{\Gamma}^{(k+1)})\}, \qquad (9)
\end{aligned}$$

where (ζ, κ) are the relaxation and penalty parameters in CPRSM and $\mathrm{Proj}_C(\cdot)$ projects a vector to the feasible set C. When $\mathcal{L}(\boldsymbol{\theta})$ is the logarithm of a Gaussian likelihood, (4) reduces to a quadratic minimization which can be solved analytically. When $\mathcal{L}(\boldsymbol{\theta})$ is the log-likelihood from the other generalized linear models, such as the logistic model, (4) can be solved using Newton-Raphson method iteratively. The projection step (5) can be achieved by adopting the linear programming method (Duchi et al. 2008). (8) is a standard penalized quadratic minimization step. Throughout the paper, we choose $\rho_\lambda(\cdot)$ as the smoothly clipped absolute deviation (SCAD) penalty (Fan and Li 2001), which is an amenable penalty satisfying Conditions (A1) to (A6) in Appendix SF of the supplementary material (Loh and Wainwright 2015, 2017). When using the SCAD penalty, the minimization in (8) can be addressed by first calculating the analytical solution of the minimization without the penalty term, and then applying SCAD shrinkage to the resulting solution.

We summarize the CPRSM algorithm in Algorithm S1 in Appendix SE of the Supplementary Material and more details of the computational algorithm are also provided there.

2.4. Test statistic

We propose a test statistic to test the hypothesis (2). Let $\otimes$ denote the Kronecker product. Define

$$\mathbf{Q}\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\} \equiv E\left\{\psi_i''(\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot))\left(\left[\mathbf{Z}_i^{\mathrm{T}}, N^{1/2}\left\{\int_0^1 \mathbf{X}_i(s) \otimes \mathbf{B}(s)ds\right\}^{\mathrm{T}}\right]^{\mathrm{T}}\right)^{\otimes 2}\right\}.$$

The covariance matrix of the residuals is

$$\Sigma\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\} \equiv E\left\{\left([\psi_i'\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\} - Y_i]\left[\mathbf{Z}_i^{\mathrm{T}}, N^{1/2}\left\{\int_0^1 \mathbf{X}_i(s) \otimes \mathbf{B}(s)ds\right\}^{\mathrm{T}}\right]^{\mathrm{T}}\right)^{\otimes 2}\right\}.$$

Further, define

$$\mathbf{A} = (\mathbf{C} \otimes \mathbf{I}_{N\times N})[\mathbf{0}_{mN\times q}, \mathbf{I}_{mN\times mN}, \mathbf{0}_{mN\times k_0 N}],$$
$$\boldsymbol{\Psi}(\Sigma, \mathbf{Q}, \boldsymbol{\beta}) = \mathbf{A}\mathbf{Q}\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}(\cdot)\}^{-1}_{\mathcal{M}\cup S,\mathcal{M}\cup S}\Sigma_{\mathcal{M}\cup S,\mathcal{M}\cup S}\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}\mathbf{Q}\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}(\cdot)\}^{-1}_{\mathcal{M}\cup S,\mathcal{M}\cup S}\mathbf{A}^{\mathrm{T}},$$

where $\mathbf{M}_{\mathcal{M}\cup S,\mathcal{M}\cup S}$ is the top left $\{q + (m + k_0)N\} \times \{q + (m + k_0)N\}$ of an arbitrary matrix $\mathbf{M}$.

The sample versions of $\mathbf{Q}\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}$, $\Sigma\{\boldsymbol{\alpha}, \boldsymbol{\beta}(\cdot)\}$ and $\boldsymbol{\Psi}(\Sigma, \mathbf{Q}, \boldsymbol{\beta})$ based on the estimator $\widehat{\boldsymbol{\theta}}$ are denoted by $\widehat{\mathbf{Q}}(\widehat{\boldsymbol{\theta}})$, $\widehat{\Sigma}(\widehat{\boldsymbol{\theta}})$ and $\boldsymbol{\Psi}(\widehat{\Sigma}, \widehat{\mathbf{Q}}, \widehat{\boldsymbol{\theta}})$, respectively. Then the test statistic is defined as

$$T = n\left(\mathbf{A}\widehat{\boldsymbol{\theta}}_{\mathcal{M}\cup S} - N^{-1/2}\mathrm{vec}\left[\left\{\int_0^1 \mathbf{B}(s)\mathbf{B}(s)^{\mathrm{T}}ds\right\}^{-1}\int_0^1 \mathbf{B}(s)\mathbf{t}^{\mathrm{T}}(s)ds\right]\right)^{\mathrm{T}}\boldsymbol{\Psi}(\widehat{\Sigma}, \widehat{\mathbf{Q}}, \widehat{\boldsymbol{\theta}})^{-1} \tag{10}$$
$$\left(\mathbf{A}\widehat{\boldsymbol{\theta}}_{\mathcal{M}\cup S} - N^{-1/2}\mathrm{vec}\left[\left\{\int_0^1 \mathbf{B}(s)\mathbf{B}(s)^{\mathrm{T}}ds\right\}^{-1}\int_0^1 \mathbf{B}(s)\mathbf{t}^{\mathrm{T}}(s)ds\right]\right).$$

Our test statistic follows the Wald-type framework and is derived by expressing the functional parameters in the null hypothesis through their spline approximations. Details of the construction of the test statistic (10) are provided after Remark A.1 in the supplementary material.

In Theorem 1, we show that the test statistic T asymptotically follows a central chi-squared distribution with degree of freedom rN under the null hypothesis.

Theorem 1. *Assume Conditions (A1) to (A6) and (C1) to (C5) in the supplementary material hold. Furthermore, assume the conditions in Theorem A.2 of the supplementary material hold. In addition, assume* $(rN)^{1/4}n^{-1/2} \to 0$ *and*

$$E\left[\|\boldsymbol{\Psi}^{-1/2}(\Sigma, \mathbf{Q}, \boldsymbol{\beta})\mathbf{A}\mathbf{Q}^{-1}_{\mathcal{M}\cup S,\mathcal{M}\cup S}\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}(\cdot)\}^{-1}_{\mathcal{M}\cup S,\mathcal{M}\cup S}\right.$$
$$\left.\times\{\psi_i'\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}(s)\} - Y_i\}\left[\mathbf{Z}_i^{\mathrm{T}}, N^{1/2}\left\{\int_0^1 \mathbf{X}_i(s)\otimes\mathbf{B}(s)ds\right\}^{\mathrm{T}}\right]^{\mathrm{T}}_{\mathcal{M}\cup S}\|_2^3\right] = O(1).$$

Then $\lim_{n\to\infty}|\Pr(T \le x) - \Pr\left\{\chi^2(rN, \mathbf{c}_n^{\mathrm{T}}\boldsymbol{\Psi}^{-1}(\Sigma, \mathbf{Q}, \boldsymbol{\beta})\mathbf{c}_n) \le x\right\}| = 0$, *where* $\chi^2(r, \gamma)$ *is an* r *degree of freedom non-central chi-square distribution, with non-centrality parameter* γ *and* $\mathbf{c}_n = vec\left[\sqrt{n/N}\left\{\int_0^1 \mathbf{B}(s)\mathbf{B}(s)^{\mathrm{T}}ds\right\}^{-1}\int_0^1 \mathbf{B}(s)\mathbf{h}_n^{\mathrm{T}}(s)ds\right]$.

The proof of Theorem 1 is deferred to Appendix SI of the Supplementary Material. Based on Theorem 1, to conduct the hypothesis test (2) with significance level α, we reject the null hypothesis if $T > \chi^2_{1-\alpha,rN}$, where $\chi^2_{1-\alpha,rN}$ is the $(1-\alpha)$th quantile of the central chi-squared distribution with degree of freedom rN. Furthermore, the asymptotic distribution results also hold under alternative hypotheses, making it possible to estimate the sample sizes needed to achieve a given statistical power.

3. REAL DATA ANALYSIS

3.1. Data description and preprocessing

We apply HDHT to analyze our AD data, which includes 82 AD patients and 61 healthy subjects. The baseline characteristics are shown in Table SA.1 in the Supplementary Material. To improve interpretability, we map the raw neural activity data, which is measured as time series signals

recorded from the scalp, onto the brain surface (Jin et al. 2023). We then consolidate the voxel-level time series into region-level aggregates, which leads to 68 functional predictors based on the Desikan-Killiany (DK) brain atlas (Desikan et al. 2006).

A Fourier transform is applied independently to each brain signal to generate region-wise PSDs over the whole brain. For each individual, the final brain signal data consist of the PSDs, evaluated at 2 to 35 Hz, from the MEG source time series across 68 ROIs. Detailed preprocessing procedures are presented in Appendix SA of the Supplementary Material. The outcome of interest is the MMSE score that assesses cognitive impairment. The lower the score, the worse the cognitive function (Dellasega and Morris 1993). In addition, we consider gender, age, and the number of years of education as the baseline covariates.

3.2. HDHT identifies 19 significant ROIs with effects that vary across different frequency bands

We apply the HDHT method to study the association between the PSDs and the MMSE score adjusted for the baseline covariates and test the null hypothesis that $\mathbf{H}_{0j} : \beta_j(s) = 0, \forall s$ for $j = 1, \ldots, 68$, where s represents the frequency point. Because there are 68 ROIs, the significance level is set as $\alpha = 0.05/68$ to account for the multiple testing. The number of B-spline bases N and the penalty parameter λ are optimized using a 10-fold cross-validation (CV) process aimed at minimizing the prediction error. Parameters N and λ are chosen from the set $\{4, 6, 8, 10, 12\}$ and the set $\{0.1, 0.3, 0.5, 0.7, 0.9, 1.1, 1.3\}$, respectively. The parameter a in the SCAD penalty is set to be 3.7 as recommended by Fan and Song (2010). We set $R = 2 \times 10^5$ to guarantee it to be larger than $\|\boldsymbol{\alpha}_0\|_1 + 2\sum_{j=1}^{d} \|N^{-1/2}\boldsymbol{\gamma}_{0j}\|_2$. In the CPRSM algorithm, we set $\zeta = 0.9$, $\kappa = 1$ to ensure rapid convergence according to He et al. (2014). The algorithm terminates when the norm of the difference between the parameters in two consecutive iterations is less than $5e-4$ or when the number of iterations reaches 5000.

We include the high dimensional non-functional Wald test proposed by Shi et al. (2019) and the bootstrap method proposed by Xue and Yao (2021) for comparison. To make the real data compatible with the non-functional Wald test, we use the mean and standard deviation of the functional predictors as the predictors in the non-functional test. For each ROI, we test the null hypothesis that the coefficients of the mean and standard deviation of that ROI are zero. Since the bootstrap method does not account for the baseline covariates, we employ a projection step to eliminate the influence of baseline covariates. Specifically, consider $\widetilde{\mathbf{Z}}$ as an $n \times (p+1)$ matrix where the first column consists of ones and the remaining columns represent the baseline covariates. Let $\mathbf{P} = \mathbf{I} - \widetilde{\mathbf{Z}}(\widetilde{\mathbf{Z}}^{\mathrm{T}}\widetilde{\mathbf{Z}})^{-1}\widetilde{\mathbf{Z}}^{\mathrm{T}}$ and $\mathbf{Y} = (Y_i, i = 1, \ldots, n)^{\mathrm{T}}$. We generate the new outcomes as $\mathbf{PY}$ and the new functional predictors as $\mathbf{P}\{\mathbf{X}_1(t), \ldots, \mathbf{X}_n(t)\}^{\mathrm{T}}$. Then we implement the bootstrap method considering the new outcomes and the new functional predictors. Here we remove the effects from the baseline covariates because $\mathbf{P}\widetilde{\mathbf{Z}} = \mathbf{0}$.

After applying the HDHT method, we identify 19 significant ROIs with p-values less than $0.05/68$. We demonstrate the significant ROIs with p-values less than $0.05/68$ in Fig. 2B. The non-functional Wald test identifies 17 significant ROIs which are presented in Fig. SA.2 in Appendix SB.2. We do not present the results from the bootstrap method of Xue and Yao (2021) as no significant ROIs are identified after applying the Bonferroni correction. Overall, the significant ROIs identified by HDHT exhibit a symmetrical pattern between the left and right hemispheres, which is intuitive and consistent with the existing literature in AD pathophysiology (Jagust 2018). Most of the significant ROIs are located in the inferior and posterior temporal cortices and the posterior parietal-occipital cortices, which reflects the similar regional distribution of the glucose hypometabolism and the high tau accumulation in the AD brain (Jagust 2018; Jin et al. 2023). These findings suggest that our method identifies clinically meaningful neurophysiological markers for neurodegeneration. These markers can be used to monitor the neurodegenerative process during the pre-symptomatic screening stage, when PET and CSF tests are not recommended due to their toxicities and risks.

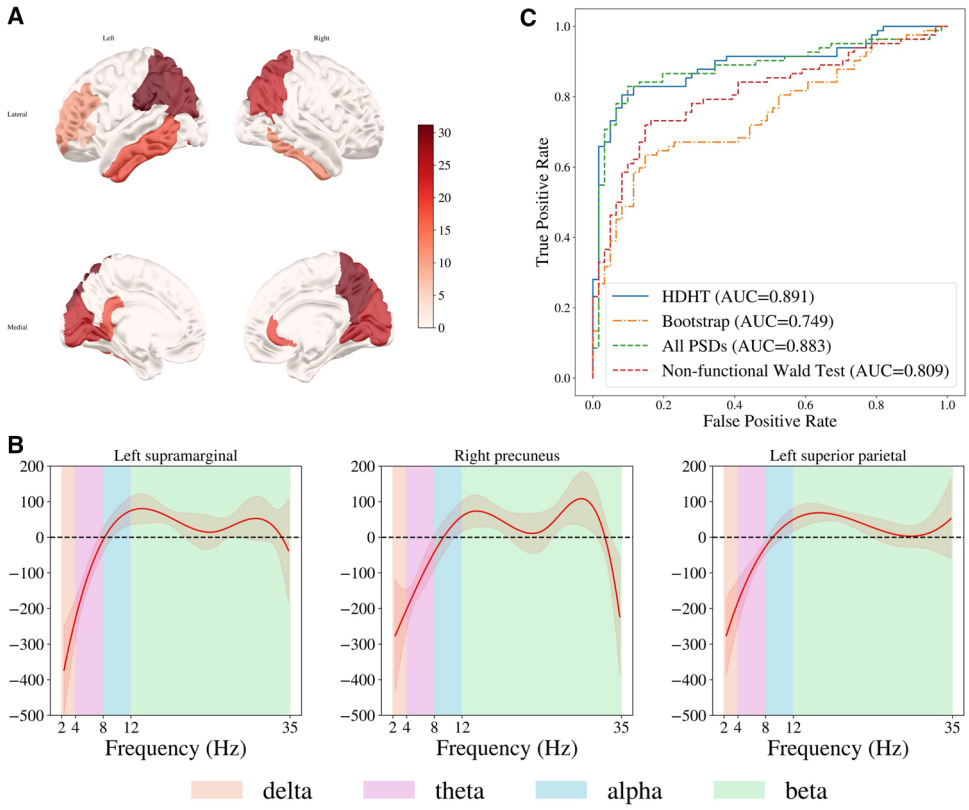

Fig. 2. A) Negative log *P*-values from the significant ROIs selected via HDHT. Only the significant regions with *P*-value less than 0.05/68 are shown. B) The functional effects of the top three significant ROIs and their $100(1-0.05/19)\%$ confidence intervals (red shading). The functional effects are estimated by the HDHT method and the order is based on the *P*-values with the smallest *P*-value on the left. From left to right, the four frequency bands are the delta, theta, alpha, and beta bands. C) The receiver operating characteristic curves and corresponding AUCs for HDHT and other compared methods in distinguishing healthy subjects from AD patients.

The subsequent analysis is focused on the 19 significant ROIs. In Fig. 2B, we further present the estimated functional effects from the top three significant ROIs identified by the HDHT method. The remaining significant ROIs are shown in Fig. SA.3 of the Supplementary Material. Since the analysis focuses on the 19 significant ROIs, we show the $100(1-0.05/19)\%$ confidence intervals for the functional effects of these ROIs after adjusting for the multiple comparisons. The results show that across the top 3 significant ROIs, the functional effects in delta (2 to 4 Hz) and theta bands (4 to 8 Hz) are significantly negative, while the functional effects in the low beta (12 to 20 Hz) band are significantly positive and the effects are not significant in the alpha band.

We also investigate the effects of the baseline covariates on the MMSE score and construct the 95% CIs. The effects and corresponding 95% CIs for gender, age and years of education are -0.502 ($[-1.920, 0.916]$), 0.388 ($[-0.241, 1.017]$) and 0.417 ($[0.155, 0.679]$), respectively. Among the three covariates, education level is a strong predictor of cognitive function, with higher education implying higher MMSE scores on average. It indicates that an additional year of education is associated with an increase in the MMSE score, thereby indicating better cognitive function. This finding is consistent with the existing literature (Sando et al. 2008).

3.3. HDHT has high sensitivity and specificity in classifying AD and control samples

We further evaluate the accuracy of HDHT, the non-functional Wald test, and the bootstrap method in classifying AD versus healthy controls. For HDHT, we construct a logistic model relating the disease indicator to the PSDs from the significant ROIs, adjusting for baseline covariates. We estimate the parameters by minimizing the penalized objective similar to (3), where penalties are applied to all functional coefficients from the significant ROIs. For the non-functional Wald test, we build a logistic model using the means and standard deviations of the PSDs from the 17 significant ROIs and baseline covariates. Parameters are estimated by maximizing the penalized likelihood function with penalties applied to all coefficients. Since the bootstrap method fails to identify any significant PSDs, we fit a standard logistic regression using only the baseline covariates and estimate parameters via maximum likelihood estimation. For comprehensive evaluation, we also fit a logistic regression using all PSDs along with baseline covariates. We apply penalties to all 68 functional coefficients and minimize the penalized objective function.

We evaluate classification accuracy using leave-one-out cross-validation, fitting the model with all samples except one and predicting the disease status of the excluded sample. We then compare the area under the receiver operating characteristic (ROC) curve (AUC) for both methods. Figure 2C illustrates that HDHT achieves highest classification accuracy, indicating that the PSDs from selected ROIs provide valuable information for distinguishing AD from control samples. It is worth mentioning that the AUC obtained using covariates from all ROIs (All PSDs in Fig. 2C) is similar to that achieved using covariates from the significant ROIs identified by HDHT. This suggests that the significant ROIs selected by HDHT capture all the essential information in the covariates for distinguishing between the disease and control groups.

3.4. The important ROIs vary across frequency bands with the low-beta band strongly associated with cognitive function

We further show that important brain ROIs vary across the frequency bands. We conduct the hypothesis test at each frequency for the 19 significant ROIs identified by the HDHT method. The null hypothesis is the effect of PSD at a specific frequency on MMSE score is zero for each significant ROI, so the significance level is set as $0.05/19$. Specifically, we first obtain the point-wise $100(1-0.05/19)\%$ CIs for the functional coefficients. A functional effect is significant at a specific frequency if the corresponding CI does not cover zero. We then plot the percentage of the significant ROIs across frequencies in the top panel of Fig. 3. Figure 3 shows that the theta and low-beta (12 to 20 Hz) frequency bands have the largest number of significant ROIs. This indicates that the theta and low-beta bands are important frequency bands in neural activity data for explaining cognitive function. Similar findings are also reported in Ranasinghe et al. (2022)'s study.

To further explore band-specific effects, we examine association of the delta, theta, alpha, and beta frequency bands with cognitive function for each significant ROI. The null hypothesis is for each significant ROI, the effect of PSD on MMSE score at each frequency band is zero. Since there are 19 ROIs and four frequency bands, we set the significance level to $0.05/19/4$ to account for multiple testing. For each hypothesis test, we calculate the $100(1-0.05/19/4)\%$ CI for the average functional coefficient within the frequency band of interest, defined as $\int_{f_1}^{f_2}\beta_j(s)ds/(f_2-f_1)$ where f_1 and f_2 are the lower and upper bounds of the frequency band, respectively. If the CI does not cover zero, the ROI is considered significant for that frequency band. The bottom panel of Fig. 3 shows that more significant ROIs are identified in delta, theta and beta bands, which is consistent with the observation in Ranasinghe et al. (2022) and Jin et al. (2023).

3.5. The significant brain ROIs identified by HDHT are not symmetric between the left and right hemispheres

Furthermore, as shown in Fig. 2B, the significant regions are not completely symmetric. This raises a question of whether the functional pattern asymmetrically affects the cognitive function. To explore this, we further examine the hypothesis regarding whether the effects from the PSDs derived from

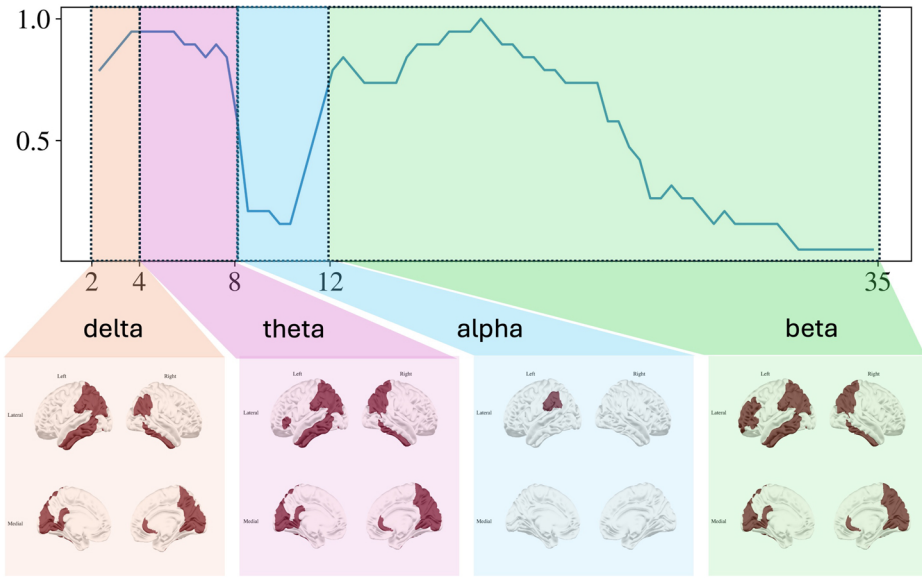

Fig. 3. Top: The percentage of the significant ROIs based on the point-wise $100(1-0.05/19)\%$ confidence intervals for each frequency point among the 19 significant ROIs identified by the HDHT method. Bottom: The significant ROIs at each frequency band on brain based on the hypothesis testing at the $0.05/19/4$ significance level. The highlighted regions represent the significant ROIs.

the left and right hemispheres are equal or not. We use $\mathcal{G}$ to represent the set of indices of the ROIs that are significant in only one hemisphere. If $\beta_j(s)$ is the functional coefficient corresponding to the jth ROI in one hemisphere, we denote $\tilde{\beta}_j(s)$ as the functional coefficient corresponding to the jth ROI in the other hemisphere.

We then test the null hypothesis that $\mathbf{H}_0 : \tilde{\beta}_j(s) - \beta_j(s) = 0, \forall s$ for all $j \in \mathcal{G}$. All the implementation details are the same as in the Section 3.2. We obtain the test statistic $T = 60.248$ with a p-value of 3.786×10^{-4}, indicating a significant difference between the left and right hemispheres in at least one ROI. This analysis suggests that there is indeed an asymmetry in the PSD effects from a subset of ROIs.

4. SIMULATION STUDIES

4.1. Simulation setup

We carry out comprehensive numerical analyses to assess both the theoretical properties and the performance of the proposed HDHT method. Data are generated using the model specified in (1), with the linear model for continuous outcomes and the logistic model for binary outcomes. For the linear model, random errors are drawn from either a standard normal distribution or a t-distribution with three degrees of freedom, and these errors are subsequently standardized to have a variance of one. In the logistic model, we model Y given $\boldsymbol{\alpha}_0^{\mathrm{T}}\mathbf{Z}_i + \int_0^1 \boldsymbol{\beta}^{\mathrm{T}}(s)\mathbf{X}(s)ds$ as following a Bernoulli distribution.

The $\boldsymbol{\beta}(s)$ is constructed using a Fourier basis following Xue and Yao (2021). Specifically, we set

$$\beta_j(s) = c_j \sum_{k=1}^{50} \eta_k \phi_k(s), \eta_k = I_{\{k\leq 4\}}(1.2 - 0.2k) + 0.4 I_{\{5\leq k\leq 50\}}(k-3)^{-4}, \text{for } k = 1, \ldots, 50,$$

where $I_{\{\cdot\}}$ is an indicator function, $c_j = 0$ for $j = 3, \ldots, d-1$ and c_j is chosen differently for each j with $j = 1, 2, d$ under various settings to control the signal-to-noise ratio and $\phi_k(s)$'s are the Fourier basis functions. Specifically, $\phi_1(s) = 1$, $\phi_{2k}(s) = \sqrt{2}\cos\{k\pi(2s-1)\}$ for $1 \le k \le 25$ and $\phi_{2k-1}(s) = \sqrt{2}\sin\{(k-1)\pi(2s-1)\}$ for $2 \le k \le 25$.

We consider two types of functional predictors $\mathbf{X}(s)$, the uncorrelated and highly-correlated functional predictors. The uncorrelated predictors $\mathbf{X}(s)$ are generated using a B-spline basis. Specifically, we generate the functional predictors as

$$X_j(s) = \sum_{k=1}^{10} \xi_{jk} B_k(s) + 0.5\epsilon_j(s), \tag{11}$$

where ξ_{jk} is drawn from $N(0, 5^2)$, $B_k(s)$ is the kth B-spline basis function and $\epsilon_j(s)$ is a standard Gaussian process. The terms $\epsilon_j(s)$ and ξ_{jk} are independent for different j, ensuring zero correlation between different functional predictors.

The highly-correlated predictors $\mathbf{X}(s)$ are generated with the spectral graph model (SGM) (Raj et al. 2020; Verma et al. 2022), a generative model that can mimic the PSD of the human brain activity at the resting state. A brief introduction to the SGM and examples of simulated PSDs are provided in Appendix SC of the Supplementary Material. After generating the functional predictors with SGM, we standardize them to have a mean of zero and a variance of one at each ROI. We then add independent Gaussian noises with a mean of 0 and a variance of 100 to the functional predictors at each ROI, where this variance is estimated from the real data. More details of generating the functional predictors using SGM can be found in Verma et al. (2022). The resulting functional predictors exhibit strong correlation between different ROIs, with an average pair-wise correlation coefficient of approximately 0.6.

Unless stated otherwise, $\boldsymbol{\alpha}_0 = [\alpha_{00}, -1, 2]$ under both the linear and logistic models. For the linear model, the intercept α_{00} is 5. For the logistic model, α_{00} is adaptively adjusted to ensure the probability of success remains at 0.5. The two baseline covariates in $\mathbf{Z}$ are derived respectively from the standard normal distribution and a Bernoulli distribution with a probability of 0.5.

In the implementation, the number of B-spline bases N and the penalty parameter λ are optimized using a 10-fold CV process aimed at minimizing the prediction error. The number of B-spline bases is chosen from the set $\{4, 6, 8, 10, 12\}$. We select λ from $[0, 1.5]$ with increments of 0.1 under the linear model and from $\{0, 0.05, 0.1, 0.2, 0.3, 0.4\}$ under the logistic model. In the CPRSM, we set $\zeta = 0.9$, $\kappa = 10$ for the linear model, and $\kappa = 1.5$ for the logistic model to ensure rapid convergence according to He et al. (2014). All other implementation details are the same as in Section 3.2. The results presented in this section are based on 1,000 repetitions.

4.2. Evaluate the asymptotic property of $\boldsymbol{\alpha}$

We use the asymptotic normality outlined in Remark A.1 in the supplementary material to construct 95% CIs for the coefficients of the baseline covariates $\boldsymbol{\alpha}$. We validate the empirical coverage of these theoretical CIs for the coefficients of the baseline covariates $\boldsymbol{\alpha}$ and confirm their consistency. We choose $d = 200$ and vary the sample size n from 100 to 3,200. The data are derived from a linear model with standard normal random errors and a logistic model. The uncorrelated functional predictors are generated as specified in (11). Additionally, we set $c_j = 2$ for $j = d$ and $c_j = 0$ for all other values in (11) to generate the functional coefficients. In this scenario, only the last functional covariate is associated with the outcome. Since only $\boldsymbol{\alpha}$ is the parameter of interest, we apply penalties to all functional coefficients in the estimation.

For each given sample size, we repeat the simulation 1,000 times and the results are shown in Fig. 4, where we report both the average estimation error $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|_2$ and the empirical coverage of the 95% CI for $\boldsymbol{\alpha}_0$ over the 1,000 simulations. Figure 4A shows that the average estimation error decreases when the sample size increases, which validates the consistency of the estimator. Panels B and C of Fig. 4 illustrate that the CI coverage probabilities for α_{01} and α_{02} in the linear

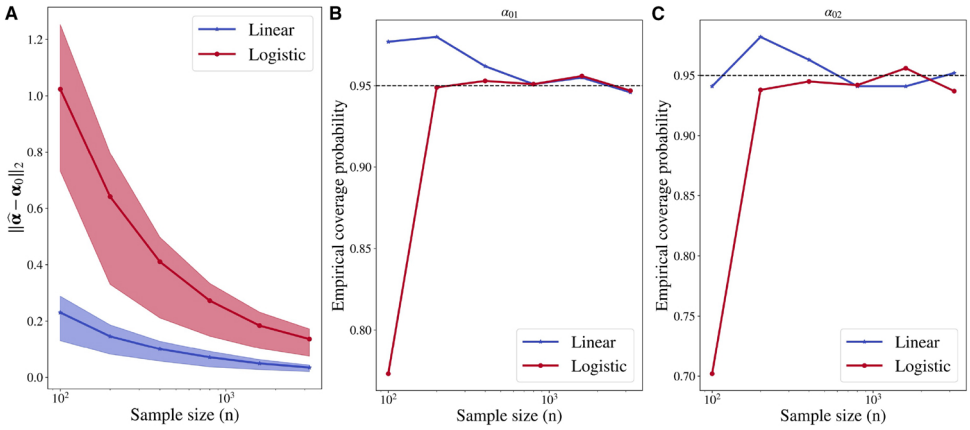

Fig. 4. A) $\|\widehat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}_0\|_2$ for the coefficients of the baseline covariates when the sample size n varies under the linear and logistic scenarios. The shaded area represents the 25% and 75% quantiles of the estimated error over 1000 simulations. B) and C) Empirical coverage probabilities of the 95% confidence interval of α_{01} and α_{02}, respectively under linear and logistic models.

model are close to 95% across all evaluated sample sizes. Additionally, the coverage probability approaches 95% more closely as the sample size increases. A similar trend is observed under the logistic model, although the coverage probability only reaches 95% when the sample size exceeds 300. This suggests that the estimation is more efficient under the linear model.

4.3. Evaluate the hypothesis testing procedure

We then evaluate the performance of HDHT in testing linear functional hypotheses. We compare our method with the bootstrap method proposed by Xue and Yao (2021). Note that the bootstrap method is developed under the linear model with no baseline covariate, and therefore, our comparative study is conducted without baseline covariates. All the tuning parameters in the bootstrap method are chosen through the 10-fold CV.

We choose $d = 200$ and sample size $n = 100$. We generate the uncorrelated functional predictor from (11) and correlated predictor from the SGM as described in Section 4.1. We choose $c_d = 1$ and varying c_1 and c_2 in (11) under different simulation settings. Specifically, we aim to test the null hypotheses that $\mathbf{H}_{0,1} : \beta_1(s) = 0, \forall s$ and $\mathbf{H}_{0,2} : \beta_1(s) = \beta_2(s) = 0, \forall s$. To assess the performance for testing $\mathbf{H}_{0,1}$, we generate functional parameters according to (11) with $c_1 = c$ and $c_2 = 0$, where we vary c from 0 to 0.4 corresponding to the null hypothesis and the alternative hypotheses with increasing effect size settings. Similarly, to evaluate the performance for testing $\mathbf{H}_{0,2}$, we generate functional parameters from (11) with both $c_1 = c_2 = c$, varying c from 0 to 0.4.

We show the size and power of testing $\mathbf{H}_{0,1}$ and $\mathbf{H}_{0,2}$ in Table 1 under various settings. Table 1 shows that when the functional predictors are uncorrelated, both HDHT and the bootstrap method control the type I error $(c = 0)$ when testing $\mathbf{H}_{0,1}$, while HDHT and the bootstrap yield similar powers under the alternative hypotheses when $c \neq 0$. Furthermore, when the hypotheses involve two functional parameters as those in $\mathbf{H}_{0,2}$, HDHT yields greater power than the bootstrap method does. When the functional predictors are highly-correlated, both methods control the type I error under the nominal level. However, the bootstrap method only achieves a power of less than 0.15 in all scenarios. The results indicate that the HDHT method outperforms the bootstrap method, particularly when the functional predictors are highly correlated.

Table 1. The empirical size and power of hypothesis testing for the comparison between HDHT and bootstrap methods under the linear model with normal noise and t noise based on 1,000 repetitions. Significance level is 0.05 and type I errors are in bold.

Type of $\mathbf{X}(s)$	Err. dist.	c	HDHT	Bootstrap	HDHT	Bootstrap
			$\beta_1(s)=0$		$\beta_1(s)=\beta_2(s)=0$	
Uncorrelated		**0.0**	**0.045**	**0.047**	**0.050**	**0.049**
	$N(0,1)$	0.1	0.115	0.115	0.167	0.099
		0.2	0.377	0.361	0.637	0.388
		0.4	0.862	0.930	0.975	0.854
		0.0	**0.038**	**0.048**	**0.044**	**0.035**
	$t(3)$	0.1	0.134	0.118	0.154	0.099
		0.2	0.403	0.370	0.615	0.379
		0.4	0.909	0.933	0.941	0.915
Highly-correlated (PSD)		**0.0**	**0.038**	**0.036**	**0.040**	**0.024**
	$N(0,1)$	0.1	0.114	0.037	0.150	0.035
		0.2	0.187	0.048	0.496	0.063
		0.4	0.624	0.051	0.935	0.055
		0.0	**0.026**	**0.027**	**0.038**	**0.019**
	$t(3)$	0.1	0.066	0.030	0.130	0.027
		0.2	0.176	0.051	0.304	0.020
		0.4	0.628	0.144	0.944	0.052

We choose number of functional covariates $d=200$, sample size $n=100$ and significance level $\alpha=0.05$. $c=0$ indicates the null hypothesis is true. Err. dist. stands for the distribution of the random error. Type I errors are in **bold**.

4.4. Evaluate the testing procedure under the real data setting

We assess the performance of the HDHT method in a more realistic setting when the distributions of the functional predictors mimic those in the real data. Therefore, we choose $d=68$ and consider $n=200$ and $n=500$. Furthermore, we generate a continuous and a binary baseline covariate from the standard normal distribution and the Bernoulli distribution with probability 0.5, respectively.

The baseline effect is $\boldsymbol{\alpha}_0=[5,-1,2]$ under the linear model. For the logistic model, the baseline effect is $\boldsymbol{\alpha}_0=[\alpha_{00},-1,2]$ where α_{00} is adaptively chosen in each scenario to ensure that the positive and negative samples remain balanced. The functional predictors are generated by the SGM (Verma et al. 2022), and the functional coefficients $\boldsymbol{\beta}(s)$ are generated from the Fourier basis as described in (11) with $c_d=2$, $c_j=0$ for $j\neq 1,2,d$. We vary (c_1,c_2) under different simulation settings.

We consider two sets of hypothesis tests, where we aim to test the null hypotheses that for all s in the domain,

$$\mathbf{H}^1_{0,1}:\beta_1(s)=2c\sum_{k=1}^{50}\eta_k\phi_k(s),c=0;\quad \mathbf{H}^2_{0,1}:\beta_1(s)-\beta_2(s)=2c\sum_{k=1}^{50}\eta_k\phi_k(s),c=0.$$

$$\mathbf{H}^1_{0,2}:\beta_1(s)=2c\sum_{k=1}^{50}\eta_k\phi_k(s),c=0.1;\quad \mathbf{H}^2_{0,2}:\beta_1(s)-0.8\beta_2(s)=(0.2c+0.2)\sum_{k=1}^{50}\eta_k\phi_k(s),c=0.$$

$$\mathbf{H}^1_{0,3}:\beta_1(s)=2c\sum_{k=1}^{50}\eta_k\phi_k(s),c=0.2;\quad \mathbf{H}^2_{0,3}:4\beta_1(s)-3\beta_2(s)=(8c+1)\sum_{k=1}^{50}\eta_k\phi_k(s),c=0,$$

against the alternative hypotheses in the same forms with c being different from the ones in the corresponding null hypotheses.

We set the significance level $\alpha=0.05$ and repeat the simulation 1,000 times. The top panel of Table 2 shows the results of testing hypotheses $\mathbf{H}^1_{0,1}$ to $\mathbf{H}^1_{0,3}$, where only one parameter is involved

Table 2. The empirical size and power of hypothesis testing for the HDHT under the linear and logistic models for different null hypotheses based on 1,000 repetitions. Significance level is 0.05 and type I errors are in bold.

$\mathbf{H}_0$	c	$n = 200$			$n = 500$		
		$N(0,1)$	$t(3)$	Logistic	$N(0,1)$	$t(3)$	Logistic
		Interested functional coefficients: $\{\beta_1(s)\}$					
	0.0	**0.054**	**0.051**	**0.046**	**0.044**	**0.048**	**0.032**
$\mathbf{H}^1_{0,1}$	0.1	0.649	0.732	0.149	0.998	0.999	0.180
	0.2	0.997	1.000	0.314	1.000	1.000	0.594
	0.4	1.000	1.000	0.794	1.000	1.000	1.000
	0.0	0.555	0.645	0.116	0.997	0.996	0.197
$\mathbf{H}^1_{0,2}$	**0.1**	**0.054**	**0.053**	**0.064**	**0.041**	**0.049**	**0.038**
	0.2	0.640	0.730	0.103	0.997	0.999	0.138
	0.4	1.000	1.000	0.486	1.000	1.000	0.884
	0.0	0.996	1.000	0.383	1.000	1.000	0.659
$\mathbf{H}^1_{0,3}$	0.1	0.573	0.638	0.155	0.997	0.996	0.169
	0.2	**0.053**	**0.049**	**0.054**	**0.043**	**0.043**	**0.059**
	0.4	0.998	1.000	0.209	1.000	1.000	0.449
		Interested functional coefficients: $\{\beta_1(s), \beta_2(s)\}$					
	0.0	**0.054**	**0.063**	**0.041**	**0.046**	**0.049**	**0.051**
$\mathbf{H}^2_{0,1}$	0.1	0.248	0.251	0.066	0.833	0.811	0.093
	0.2	0.849	0.858	0.088	1.000	1.000	0.202
	0.4	1.000	1.000	0.227	1.000	1.000	0.608
	0.0	**0.052**	**0.063**	**0.041**	**0.048**	**0.047**	**0.051**
$\mathbf{H}^2_{0,2}$	0.1	0.311	0.323	0.059	0.916	0.893	0.080
	0.2	0.921	0.923	0.062	1.000	1.000	0.180
	0.4	1.000	1.000	0.161	1.000	1.000	0.607
	0.0	**0.053**	**0.062**	**0.040**	**0.047**	**0.049**	**0.050**
$\mathbf{H}^2_{0,3}$	0.1	0.336	0.345	0.057	0.930	0.919	0.075
	0.2	0.929	0.934	0.063	1.000	1.000	0.178
	0.4	1.000	1.000	0.149	1.000	1.000	0.600

Under the linear model, random errors are drawn from either a standard normal distribution or a $t(3)$-distribution. The results are based on 1000 repetitions. We choose number of functional covariates $d = 68$, sample size $n = 200$ and significance level $\alpha = 0.05$. The type I errors are in **bold**.

in the test. HDHT effectively controls the type I errors under the linear model. However, it shows slightly inflated type I error rates when testing $\mathbf{H}^1_{0,3}$ under the logistic model, which is likely a finite sample phenomenon. Furthermore, the power increases as the discrepancy between the true value of c and its value under the null hypothesis increases. Moreover, a larger sample size also results in a higher power. The bottom panel of Table 2 shows the results from testing the hypotheses $\mathbf{H}^2_{0,1}$ to $\mathbf{H}^2_{0,3}$ that involve the linear combinations of two parameters. The type I errors are well controlled when the sample size reaches 500 in all scenarios. Furthermore, the power increases when c deviates from 0, and when the sample size increases. In summary, the simulation studies demonstrate that HDHT has satisfactory performance in testing linear hypotheses under both linear and logistic models.

5. CONCLUSION AND DISCUSSION

This study investigates the relationship between high-dimensional functional brain activity data obtained through MEG and cognitive function in individuals with AD. We introduce the HDHT method within a generalized functional linear model framework. This method incorporates a regularization technique for estimating functional parameters, with validated asymptotic properties,

and a hypothesis testing procedure sensitive to local alternative hypotheses. Applying the HDHT method, we identified 19 ROIs significantly associated with cognitive function, which is consistent with findings from molecular imaging studies (Jagust 2018; Villemagne et al. 2021; Wang et al. 2023), fortifying the crucial role of these regions in cognitive processes. Further analysis shows that the association between cognitive function and MEG neural activities changes over frequencies. Specifically, MEG neural activities in the theta and low beta bands exhibit the strongest association with cognitive function, while those in the alpha band show the weakest association. Our method utilizes MEG to track neurodegeneration. This offers a new approach to monitor the condition during the pre-symptomatic stage when PET and CSF tests are not recommended due to their toxicities and risks.

We also show that our method yields significantly higher out-of-sample sensitivity and specificity in distinguishing AD samples from the control samples. The application shows that our method achieves higher interpretability and prediction accuracy than existing methods. HDHT is generalizable to study other functional data, such as fMRI data and EEG data. Further research along this line is ongoing in our group.

When analyzing a large number of functional predictors, one can reduce the computational complexity by preprocessing the data using a screening method (Fan and Li 2001; Fan and Lv 2008). We have conducted preliminary studies on adopting a screening preprocessing step prior to implementing HDHT. These studies suggest that such preprocessing not only significantly reduces computational complexity but also preserves the error rates in the hypothesis testing.

ACKNOWLEDGEMENTS

We thank the Editors, Associate Editor, and two anonymous reviewers for their insightful comments, which helped us to improve the presentation of this paper.

SUPPLEMENTARY MATERIALS

Supplementary material is available at *Biostatistics Journal* online. The PDF file contains a brief description of the real data preprocessing, the additional results from the real data analysis, a brief introduction of the spectral graph model (SGM), a detailed explanation of the CPRSM algorithm, and all related the theoretical results and proofs.

FUNDING

This work was supported by NIH [K25AG071840, R01NS132766] (F.J.) and by new faculty startup funds from Department of Statistics and Data Science, Tsinghua University (H.J.).

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

DATA AVAILABILITY

The python code to reproduce the simulation studies and the real data analysis is available at https://github.com/JINhuaqing/HDHT_paper/tree/main. The dataset that supports the findings of this study is available from the corresponding author upon reasonable request.

REFERENCES

Aneiros G, Novo S, Vieu P. 2022. Variable selection in functional regression models: a review. J Multivar Anal. 188:104871.

Cai TT, Yuan M. 2012. Minimax and adaptive prediction for functional linear regression. J Am Stat Assoc. 107:1201–1216.

Dellasega C, Morris D. 1993. The MMSE to assess the cognitive state of elders. J Neurosci Nurs. 25:147–152.

Desikan RS et al. 2006. An automated labeling system for subdividing the human cerebral cortex on mri scans into gyral based regions of interest. Neuroimage. 31:968–980.
Duchi J, Shalev-Shwartz S, Singer Y, Chandra T. Efficient projections onto the L1-ball for learning in high dimensions. In: Proceedings of the 25th International Conference on Machine Learning. 2008. p. 272–279. ACM.
Fan J, Li R. 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J Am Stat Assoc. 96:1348–1360.
Fan J, Lv J. 2008. Sure independence screening for ultrahigh dimensional feature space. J R Stat Soc Ser B Stat Methodol. 70:849–911.
Fan J, Song R. 2010. Sure independence screening in generalized linear models with np-dimensionality. Ann Statist. 38:3567–3604.
Fan Y, James GM, Radchenko P. 2015. Functional additive regression. Ann Statist. 43:2296–2325.
Gertheiss J, Maity A, Staicu A-M. 2013. Variable selection in generalized functional linear models. Statistics. 2:86–103.
He B, Liu H, Wang Z, Yuan X. 2014. A strictly contractive Peaceman–Rachford splitting method for convex programming. SIAM J Optim. 24:1011–1040.
Jagust W. 2018. Imaging the evolution and pathophysiology of Alzheimer disease. Nat Rev Neurosci. 19:687–700.
Jin H et al. 2023. Dynamic functional connectivity MEG features of Alzheimer's disease. Neuroimage. 281:120358. https://doi.org/10.1016/j.neuroimage.2023.120358
Kong D, Xue K, Yao F, Zhang HH. 2016. Partially functional linear regression in high dimensions. Biometrika. 103:147–159.
Lee SM, Young GA. 2016. Distribution of likelihood-based p-values under a local alternative hypothesis. Biometrika. 103:641–652.
Lian H. 2013. Shrinkage estimation and selection for multiple functional regression. Stat Sinica. 23:51–74.
Loh P-L, Wainwright MJ. 2015. Regularized M-estimators with nonconvexity: Statistical and algorithmic theory for local optima. J Mach Learn Res. 16:559–616.
Loh P-L, Wainwright MJ. 2017. Support recovery without incoherence: a case for nonconvex regularization. Ann Statist. 45:2455–2482.
Matsui H, Konishi S. 2011. Variable selection for functional regression models via the L1 regularization. Comput Stat Data Anal. 55:3304–3310.
Michel CM, He B. 2019. EEG source localization. Handb Clin Neurol. 160:85–101.
Raj A et al. 2020. Spectral graph theory of brain oscillations. Hum Brain Mapp. 41:2980–2998.
Ranasinghe KG et al. 2022. Altered excitatory and inhibitory neuronal subpopulation parameters are distinctly associated with tau and amyloid in Alzheimer's disease. Elife. 11:e77850.
Sando SB et al. 2008. Risk-reducing effect of education in Alzheimer's disease. Int J Geriatr Psychiatry. 23:1156–1162.
Shi C, Song R, Chen Z, Li R. 2019. Linear hypothesis testing for high dimensional generalized linear models. Ann Stat. 47:2671–2703.
Shu I-W, Lin Y, Granholm EL, Singh F. 2024. A focused review of gamma neuromodulation as a therapeutic target in Alzheimer's spectrum disorders. J Psychiatry Brain Sci. 9:e240001.
Verma P, Nagarajan S, Raj A. 2022. Spectral graph theory of brain oscillations—revisited and improved. Neuroimage. 249:118919.
Villemagne VL et al. 2021. Molecular imaging approaches in dementia. Radiology. 298:517–530.
Wang J et al. 2023. PET molecular imaging for pathophysiological visualization in Alzheimer's disease. Eur J Nucl Med Mol Imaging. 50:765–783.
Weinberger M et al. 2006. Beta oscillatory activity in the subthalamic nucleus and its relation to dopaminergic response in Parkinson's disease. J Neurophysiol. 96:3248–3256.
World Health Organization. Dementia. https://www.who.int/news-room/fact-sheets/detail/dementia, 2023. https://www.who.int/news-room/fact-sheets/detail/dementia. Accessed: 2025-01-30.
Xue K, Yao F. 2021. Hypothesis testing in large-scale functional linear regression. Stat Sinica. 31:1101–1123.
Xue K, Yang J, Yao F. 2024. Optimal linear discriminant analysis for high-dimensional functional data. J Am Stat Assoc. 119:1055–1064.
Yuan M, Cai TT. 2010. A reproducing kernel Hilbert space approach to functional linear regression. Ann Statist. 38:3412–3444.
Zhang X, Zhang Q, Ma S, Fang K. 2022. Subgroup analysis for high-dimensional functional regression. J Multivar Anal. 192:105100.