Prediction of working memory ability based on EEG by functional data analysis

Yuanyuan Zhang, Chienkai Wang, Fangfang Wu, Kun Huang, Lijian Yang, Linhong Ji

Pll: S0165-0270(19)30409-1
DOI: https://doi.org/10.1016/j.jneumeth.2019.108552
Reference: NSM 108552

To appear in: Journal of Neuroscience Methods

Received Date: 25 July 2019
Revised Date: 14 December 2019
Accepted Date: 15 December 2019

Please cite this article as: Yuanyuan Zhang, Chienkai Wang, Fangfang Wu, Kun Huang, Lijian Yang, Linhong Ji, Prediction of working memory ability based on EEG by functional data analysis, Journal of Neuroscience Methods (2019), doi:https://doi.org/10.1016/j.jneumeth.2019.108552

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2019 Published by Elsevier.
Graphical Abstract

**Prediction of working memory ability based on EEG by functional data analysis**

Yuanyuan Zhang, Chienkai Wang, Fangfang Wu, Kun Huang, Lijian Yang, Linhong Ji

145 university students aged 18 to 22 were recruited, 142 with complete records. The N-back task test (N=0–2) has been recorded (6 dimension score). Reaction time and accuracy have been recorded. EEG signals recorded at 1000 Hz samplerate for 3 seconds from 32 scalp electrodes location according to international 10-20 system. PCA is used to yield subject i’s working memory ability as numerical N-back score $S_i$, $i = 1, \ldots, 142$.

N-back task test does not involve EEG and its results will be used as the variable to be predicted when using eyes-closed resting-state EEG signals as input variables.

For each subject $i$, predict score of Nback $S_{i,j}$, $1 \leq i \leq N$, $1 \leq j \leq L$, $L=32$, $N=3000$.

**Transformed score** $Y_i = \log \left( \frac{(S_i + 1)}{(17 - S_i)} \right)$, $i = 1, \ldots, 142$

**Multiple functional linear model (MFLM)**

$Y_i = \sum_{l=1}^{L} \int_0^1 \beta_l(t) X_{i,l}(t) dt + \epsilon_i$, $i = 1, \ldots, 142$

**Training set : 122 subjects (random with 50 replications)**

**Testing set : 20 subjects**

Predicted : $\sum_{l=1}^{L} \int \hat{\beta}_l(t) \hat{X}_{i,l}(t) dt$, $i = 123, \ldots, 142$

The estimated score : $\hat{S}_i = \text{round} \left( \frac{17 \exp \hat{Y}_i - 1}{\exp \hat{Y}_i + 1} \right)$

The adjusted score : $\hat{Y}_{i,\text{adj}} = \log \left( \frac{(\hat{S}_i + 1)}{(17 - \hat{S}_i)} \right)$

**Result**

R$^2$ between $\hat{Y}_{i,\text{adj}}$ and $Y_i = 0.72$
Highlights

Prediction of working memory ability based on EEG by functional data analysis
Yuanyuan Zhang, Chienkai Wang, Fangfang Wu, Kun Huang, Lijian Yang, Lihong Ji

- MFLM is built to predict working memory measured with the N-back task based on FDA of eyes-closed resting-state EEG signals from 32 scalp locations.
- MFLM quantifies the correlation between neural dynamics and cognitive performance.
- The working memory is explained by the EEG signals of eight frontal electrode positions.
- MFLM can be potentially applied to predict other behavioral responses from sensor data in cognitive neuroscience.
Prediction of working memory ability based on EEG by functional data analysis*

Yuanyuan Zhang a,1, Chienkai Wang b,1 (Co-first author), Fangfang Wu b,3, Kun Huang a,4, Lijian Yang a,*,5 and Linhong Ji b,*,6

a Center for Statistical Science and Department of Industrial Engineering, Tsinghua University, Beijing 100084, China
b Division of Intelligent and Biomechanical System, State Key Laboratory of Tribology, Department of Mechanical Engineering, Tsinghua University, Beijing 100084, China

ARTICLE INFO

Keywords:
B-spline Basis
EEG
Functional Principal Component Analysis (FPCA)
N-back
Least Squares
Working Memory

ABSTRACT

Background: There is always a demand for fast and accurate algorithms for EEG signal processing. Owing to the high sample rate, EEG signals usually come with a large number of sample points, making it difficult to predict the working memory ability in cognitive research with EEG.

New Method: Following well-designed experiments, the functional linear model provides a simple framework for regressions involving EEG signal predictors. The use of a data-driven basis in a linear structure naturally extends the standard linear regression model. The proposed approach utilizes B-spline approximation of functional principal components that greatly facilitates implementation.

Results: Using LASSO feature selection, critical features have been extracted from eight frontal electrodes, and the R-square of 0.72 indicates rather strong linear association of actual observations and out-of-sample predictions.

Comparison with Existing Methods: There does not seem to be any existing methods of predicting working memory ability from N-back task tests via EEG signals; the data-driven functional linear regression method proposed in this work is, to the best of our knowledge, the first of its kind.

Conclusions: The data analytics suggest that a multiple functional linear regression model for the predictive relationship between working memory ability and frontal activity of the brain is both feasible and accurate via EEG signal processing.

1. Introduction

Electroencephalography (EEG) is known for containing a great deal of information about the function of the brain. As EEG is a non-invasive method to monitor the electrical activity of the brain and has a high time resolution; it has become a popular technique used in cognitive science (Nie- dermeyer and da Silva, 2005). Many studies have suggested that human intelligence or cognitive function is related to EEG (Anokhin et al., 1999; Marosi et al., 1999; Thatcher et al., 2005). Working memory (WM), the ability to maintain information momentarily in an active state, plays a key role in cognition. This has made WM one of the most studied topics in cognitive research. However, due to the high sample rate, EEG signals usually end up with a large number of sample points. Studies have presented a variety of methods to extract features from EEG and make analysis more efficient.

Existing research to date has concentrated on power spectral density (PSD) computed with the Welch method, which uses a windowed Fourier transforms of signal segments to observe several frequency bands and brain oscillations. With that method, the phenomenon that the alpha rhythm decreases and the theta rhythm increases has been shown to corre-

*Corresponding author
**Principal corresponding author

ORCID(s): 0000-0003-8225-6748 (Yuanyuan Zhang)
yy-z17@mails.tsinghua.edu.cn (Yuanyuan Zhang)
wjk17@mails.tsinghua.edu.cn (Chienkai Wang)
wylijianmail@tsinghua.edu.cn (Lijian Yang)
jilh@tsinghua.edu.cn (Linhong Ji)

Preprint submitted to Elsevier
al., 2006), the detection of epileptic seizures (Hu et al., 2006), and other neuropathology research (Li et al., 2008). Sample entropy (SampEn) (Richman and Moorman, 2000), another method to improve approximate entropy (ApEn), has an excellent anti-noise and anti-interference performance on the analysis of biomedical signals (Al-Angari and Sahakian, 2007). It can also be used for emotion recognition in EEG applications (Jie et al., 2014). More recently, resting-state functional connectivity has been proposed and shown its potential use on the pathophysiological mechanism of schizophrenia and epilepsy (Micheloyannis et al., 2006; Ponten et al., 2007).

This work presents a functional data analysis (FDA) approach to extract information from EEG data that explains the variation in WM from the N-back task, which has never been reported in the WM literature. The completely new method begins with estimating an unobservable latent variable by minimizing the distance between the recorded signal and a weighted sum of B-spine basis functions. The covariance structure of the estimated variable is then decomposed into principal components and the eigenfunctions accounting for a certain percentage of variance are chosen as regressors. The model is further reduced using the LASSO method to arrive at the final prediction of cognitive performance using functional PCA-assisted regression. For detailed exposition and recent development on FDA, see (Bosq, 2000; Ramsay and Silverman, 2007; Ma et al., 2012; Zheng et al., 2014; Cao et al., 2016) and the references therein.

The rest of this paper is organized as follows. In Section 2, the pilot experiments and data collection are described. In Section 3, we briefly describe a multiple functional linear model (MFLM) with the aid of functional principal component analysis (FPCA), and propose an algorithm with full implementation details. The proposed method is applied in Section 4 to build a prediction model of the N-back task score by the EEG signals, while Section 5 presents our conclusions.

2. Materials

2.1. Subjects

There were 145 university students aged 18 to 22 who were recruited for the study. All of them had no known history of neurological or psychiatric disorders, had normal or corrected-to-normal vision, and were right-handed. Subjects were asked to stay in good condition, avoid lack of sleep, and any intake of alcohol or caffeine for the 24 h before the experiment. An informed consent form was given to each subject before participation. After EEGLAB bandpass filtering, baseline removal, and referencing to average, each of the EEG recordings was visually examined and three were removed as they had been interfered with by surrounding electrical noise. Thus, 142 subjects remained for the study without further pre-processing work. Clearly, these subjects are from a narrow sub-population of university students of a certain age, but the methodology developed in our work can easily be adapted to more inclusive populations provided one collects a great deal more records of the right kind.

2.2. Cognitive psychological measures

Our study focuses on quantifying the correlation between neural dynamics and one specific cognitive function, namely, working memory. Therefore, an appropriate working memory assessment is needed for verifying the characteristics extracted with the proposed algorithm. N-back is a popular paradigm for neuroimaging studies of working memory and 2-back is also widely used (Owen et al., 2005). Thus, we chose it as the assessment to working memory and set the N value from 0 to 2. During the N-back task, participants were successively shown a stream of English letters (randomly chosen from A to Z) on a screen and then asked to decide for each letter whether it matches the one appearing N times before. As the N value was set from 0 to 2, the difficulty increases correspondingly with the value of N. In the 0-back, participants simply decided whether the letter on the screen was ‘X’ or not by pressing key ‘J’ on the keyboard as ‘true’ and key ‘K’ as ‘false’. In the 1-back and 2-back, participants had to compare the present letter with the one or two previous letters, respectively, and decide if they matched. Participants’ WM ability was measured in terms of accuracy and reaction time. Therefore, participants were asked to ensure their accuracy and respond as quickly as possible. According to the result of the N-back task, each participant was assigned six-dimensional data with the accuracy and reaction for each 0-back, 1-back, and 2-back as criteria. The first principal component of six-dimensional data was designated as the score representing WM ability. More specifically, the N-back task test does not involve EEG data was designated as the score representing WM ability.

2.3. EEG recording

EEG signals were recorded from 32 scalp locations based on the international 10-20 system of electrode placement; see Fig. 1. The electrode positions were labeled Fp1, Fpz,
Fp2, F7, F3, Fz, F4, F8, FC5, FC1, FC2, FC6, M1, T7, T3, Cz, C4, T8, M2, CP5, CP1, CP2, CP6, CPz, P7, P3, Pz, P4, P8, P0z, O1, and O2. The ground electrode was placed at the AFz electrode and electrical reference at the Cpz electrode. An ANT Neuro built EEG amplifier, Eego mylab 88 channels version (64 EEG channels and 24 EMG channels), was used to acquire EEG at a 1000-Hz sample rate. To record an accurate detection of the voltage of the brain neuron current, the impedance of electrodes was kept below 5kΩ. The experiment required participants to go through a 5-min eyes-closed resting state while EEG was being recorded. EEGLAB was used for 0.5-45-Hz bandpass filtering, baseline removal, and referencing to average (Delorme and Makeig, 2004; Onton et al., 2005).

3. Methods

3.1. Functional data analysis

Large complex EEG data collected in modern science and technology pose tremendous challenges to traditional methods because their extremely high dimensionality is entangled with complicated structures. Functional data analysis (FDA) treats signals as model units and is designed for commonly encountered signal data, such as speech frame, EEG, and ECG; see Ramsay and Silverman (2007) for a comprehensive introduction. For regression models involving a functional predictor, the dimension of the predictor is much higher than the number of subjects. For example, in the EEG data the number of numerical values in each signal depends on time length and sample rate, which is typically huge, while the number of subjects is often limited. To extract EEG features and quantify the correlation between neural dynamics and cognitive performance, FDA takes into account the smoothness of the signals underlying the discrete observations. Any signal can be approximated by a finite linear combination of basis functions, which are linearly independent. The basis can be predetermined (e.g., Fourier series, wavelets, or B-splines), or data driven. In effect, basis expansion represents the infinite-dimensional functions within the finite-dimensional space. Owing to the dimension of the expansion, a functional dataset reduces to a vector space instead of a functional space with ideally appropriate basis functions. Once basis functions are well estimated from the observed signals, a linear approximation is derived and certain characteristics of the signals are then included in the coefficients, which become appropriate descriptive variables of the signals. The number of the basis functions is regarded as a tuning parameter defined according to the characteristics of the data.

At each of the $J = 32$ scalp locations and for each of the $n = 142$ subjects, EEG signals were truncated from 5 min to the first 3 s at a 1000-Hz sample rate with subjects’ eyes closed due to the signal periodicity, resulting in $N = 3000$ recorded signal values. For convenience of representation, the signal time of 3 s is rescaled to unit interval $[0, 1]$, so the EEG signal value $S_{l,t,j}$ for the $l$th location of the $t$th subject at the $j$th millisecond is considered at time $j/N$ instead, $1 \leq l \leq L, 1 \leq t \leq n, 1 \leq j \leq N$. Meanwhile, a transformation is applied to the N-back score $S_{l}$, $i = 1, \ldots, n$, by letting $Y_i = \log \left( \left( S_{i+1} / (17 - S_i) \right) \right)$, a scalar response associated with EEG signals $S_{l,t,j}$, $1 \leq i \leq L, 1 \leq i \leq n, 1 \leq j \leq N$. If one were to naively build a regression model of $Y_i$ in terms of $S_{l,t,j}$ the number of covariates would be $N L = 96000$ relative to a much smaller data size of 142.

Let $X$ be the latent variable from the EEG signal. The proposed method is then formed of a pre-processing step, a feature-extraction step, an electrode-selection step, and, finally, a prediction step (Fig. 2).

3.2. B-spline estimation

A non-parametric smoothing method, such as the spline regression proposed in (Cao et al., 2012), is designed to recover smooth signals from discretely recorded measurements contaminated with errors in the pre-processing step. Denote by $\{t_j\}_{J+1}$ a sequence of equally spaced points, $t_j = J \cap \{J, J+1\}, 1 \leq J \leq J, 0 < t_1 < \cdots < t_J < 1$, called interior knots, which divide the interval $[0, 1]$ into $(J+1)$ equal subintervals $I_0 = [0, t_1], I_j = (t_j, t_{j+1}), J = 1, \ldots, J-1, I_J = [t_J, 1]$. For any positive integer $p$, let $H^{[p+1]} = H^{[p]}(0, 1)$ be the space of a polynomial spline of order $p$ on $[0, 1]$, which consists of all $(p + 2)$ times continuously differentiable functions on $[0, 1]$ that are polynomials of degree $(p - 1)$ on sub-intervals $I_j, J = 0, \ldots, J$. Following the notation in De Boor (2001), we denote by
For each 1 ≤ i ≤ N, the estimator in (4) to the EEG signals are viewed as noisy realizations of some smooth random functions via the following equations:

\[ S_{l,i,j} = X_{l,j}(j/N) + \sigma_i(j/N) \varepsilon_{l,i,j}, \]

where the coe\(^\text{cients} \) \( \varepsilon_{l,i,j} \) are latent random functions representing the noiseless “true” EEG signal trajectories for the \( i \)th subject at the \( j \)th location, \( \varepsilon_{l,i,j} \) random noises, or measurement errors. It is then natural to view the N-back scores \( Y_i \) as being driven by the true signals \( X_{l,i}(t) \). A widely used statistical model associating a scalar response \( Y \) and a functional predictor \( X(t) \) is the functional linear model (FLM) in Cardot et al. (1999, 2003); Fan and Zhang (2000), namely

\[ Y = \int_0^1 \beta(t) X(t) dt + \epsilon, \]

where \( X(t) \) is a random process on \([0, 1]\), \( \beta(t) \) a smooth regression slope function, and \( \epsilon \) a zero mean regression error. Extending (6) to the EEG experiment, a multiple functional linear model (MFLM) is

\[ Y_i = \sum_{l=1}^L \int_0^1 \beta_i(t) X_{l,i}(t) dt + \epsilon_i, i = 1, \ldots, n \]

where \( \beta_i(t) \) is the regression slope function for EEG trajectories \( X_{l,i}(t) \) at the \( i \)th location.

For each \( l = 1, \ldots, L \), random signals \( X_{l,i}(t) \), \( i = 1, \ldots, n \) are independent random processes with the same probability distribution, and thus the same mean and covariance functions \( m_i(t) \equiv E[X_{l,i}(t)], G_j(t, t') \equiv \text{Cov} \{ X_{l,i}(t), X_{l,j}(t') \} \). According to functional analysis,

\[ G_j(t, t') = \int_0^1 \int_0^1 \rho_{l,i,j}(t, t') \psi_{l,i,j}(t) \psi_{l,j}(t') dt \equiv \lambda_{l,i,j} \psi_{l,i,j}(t), \]

\[ \lambda_{l,i,j} \equiv \sum_{k=1}^\infty \rho_{l,i,j}(t, t') \psi_{l,i,j}(t) \psi_{l,j}(t') dt \]

As depicted in Figs. 3(a)-(c), the estimator in (3) converges better toward the underlying variable \( X_{l,i}(t) \) as the number of basis functions in the sum increases. Applying elementary algebra, one obtains

\[ \hat{X}_{l,i}(t) = B(t)^T \hat{B}^{-1} B^T S_{l,i}, \]

where the vector \( S_{l,i} \) consists of noisy signals from raw data \( \{ S_{l,i,j} \}_{1 \leq j \leq N} \), \( S_{l,i} = (S_{l,i,1}, \ldots, S_{l,i,L})^T \), and the design matrix \( B \) for spline regression is

\[ B = (B(1/N), \ldots, B(N/N))^T, \]

with \( B(.) = \{ B_{1,p}(.), \ldots, B_{N,p}(.) \}^T. \]
Prediction of working memory ability based on EEG

with rapidly decreasing eigenvalues \( \lambda_{l,1} \geq \lambda_{l,2} \geq \cdots \geq 0 \), nuclear norm \( \int_0^1 G_l(t, t) \, dt = \sum_{k=1}^{\infty} \lambda_{l,k} < \infty \), and eigenfunctions \( \{ \psi_{l,k}(\cdot) \}_{k=1}^{\infty} \) an orthonormal basis of the square integrable function space. The well-known Karhunen-Loève \( L^2 \) representation provides that

\[
X_{l,i}(t) = m_l(t) + \sum_{k=1}^{\infty} \tilde{\xi}_{l,i,k} \phi_{l,k}(t),
\]

where the random coefficients \( \tilde{\xi}_{l,i,k} \), called functional principal component (FPC) scores, have mean 0 and variance 1, and are uncorrelated. The rescaled eigenfunctions, \( \phi_{l,k} \), called FPC, satisfy that

\[
\phi_{l,k} = \sqrt{\lambda_{l,k}} \psi_{l,k}, \quad f_0^1 \{ X_{l,i}(t) - m_l(t) \} \phi_{l,k}(t) \, dt = \lambda_{l,k} \tilde{\xi}_{l,i,k},
\]

for \( k \geq 1 \). For all these functional principal component analysis (FPCA), see Bosq (2000).

The above FPCA allows one to write

\[
S_{l,i,j} = m_l(j/N) + \sum_{k=1}^{\infty} \tilde{\xi}_{l,i,k} \phi_{l,k}(j/N),
\]

\[
+ \sigma_l(j/N) \epsilon_{l,i,j}, 1 \leq l \leq L, 1 \leq i \leq n, 1 \leq j \leq N,
\]

where \( \epsilon_{l,i,j} \) are independent random errors with mean 0 and variance 1, \( \sigma_l(j/N) \) the variances of \( S_{l,i,j} \), at time \( j/N \), and \( \epsilon_{l,i,j} \)'s are independent of the \( \tilde{\xi}_{l,i,k} \)'s. Model (7) is then transformed to a linear form of the FPCA scores \( \{ \tilde{\xi}_{l,i,k}, k = 1, 2, \ldots \} \), which is

\[
Y_i = \beta_0 + \sum_{l=1}^{L} \sum_{k=1}^{\infty} \tilde{\xi}_{l,i,k} \beta_{l,k} + \epsilon_i, 1 \leq i \leq n,
\]

where \( \beta_0 = \sum_{l=1}^{L} \int_0^1 \beta_l(t) \, m_l(t) \, dt \) and \( \beta_{l,k} = \int_0^1 \beta_l(t) \phi_{l,k}(t) \, dt \). Although (9) resembles a standard linear regression, one is reminded that the FPC scores \( \{ \tilde{\xi}_{l,i,k} \}^\infty_{k=1} \) are all unobservable; thus, they cannot be treated as predictor variables in linear regression. Likewise, the sequences of eigenvalues \( \{ \lambda_{l,k} \}^\infty_{k=1} \) and FPCs \( \{ \phi_{l,k}(\cdot) \}^\infty_{k=1} \) are all unknown, and together with the FPC scores \( \{ \tilde{\xi}_{l,i,k} \}^\infty_{k=1} \) must be first estimated from the raw data \( \{ S_{l,i,j} \}_{1 \leq l \leq L, 1 \leq i \leq n, 1 \leq j \leq N} \).

3.5. Algorithm

As is discussed in Section 3, the FPCA theory enables isomorphic transformation of EEG signals to their FPC scores, which brings tremendous convenience to model fitting and in the functional linear model. To establish a framework in (7), we consider regressing the N-back scores (scalar responses) \( \{ Y_i \} \) directly on the sequences of FPC scores \( \{ \tilde{\xi}_{l,i,k} \}^\infty_{k=1} \) of \( \{ X_{l,i}(\cdot) \} \). However, the FPC scores \( \{ \tilde{\xi}_{l,i,k} \} \) serving as predictor variables in expression (9) cannot be observed. Therefore, we must estimate the FPC scores first, before variable selection, due to the high-dimensional linear regression, e.g., the least absolute shrinkage and selection operator-penalized regressions (LASSO). The algorithm is then given as follows.

Step 1. For scalp location \( l \) and subject \( i \), compute the spline estimator \( \hat{X}_{l,i}(t) \) in Section 3.2 with the number of knots \( N_j \) in Section 3.3;

Step 2. Compute the bivariate spline estimator \( \hat{G}_l(t,t') \) of covariance function \( G_l(t,t') \) in the tensor product spline space in Section 3.6;

Step 3. Implement FPCA to estimate the FPC scores \( \{ \tilde{\xi}_{l,i,k} \} \) of \( X_{l,i} \), and then the eigenvalues and eigenfunctions \( \lambda_{l,k}, \phi_{l,k}(t) \) up to \( k = N_j \), and then choose the largest \( k_j \) to account for 95% of the nuclear norm according to (8).

Step 4. Apply LASSO in (Bühlmann and Van, 2011) to the linear regression (13) to arrive at a model with a much smaller number of parameters than \( 1 + \sum_{l=1}^{L} k_j \) and use the corresponding prediction model (14) as the final model.

3.6. Covariance-function estimation

When constructing the regression model in (7), for scalp location \( l \), the covariance function \( G_l(\cdot, \cdot) \), and its eigenfunctions \( \phi_{l,k}(\cdot) \) and eigenvalues \( \lambda_{l,k} \), are estimated from the FPCA analysis. Define the tensor product spline space, which represents bivariate spline functions defined on the unit square \([0, 1]^2 \), \( H^{(p-2)} \otimes H^{(p-2)} = \{ \sum_{j=1}^{N_j} \sum_{j'=1}^{N_j} b_{l,j,j'} \, B_{J_{l,j,j'}}(t) B_{J_{l,j,j'}}(t') \mid b_{l,j,j'} \in \mathbb{R}, t, t' \in [0, 1] \} \). Making use of the spline representation (3) of the trajectories, the pilot bivariate spline estimator \( \hat{G}_l(t, t') \) of covariance function \( G_l(t, t') \) in the tensor product spline space is

\[
\hat{G}_l(t, t') = n^{-1} \sum_{i=1}^{n} \hat{X}_{l,i}(t) \hat{X}_{l,i}(t')
\]

\[
= \sum_{j=1}^{N_j} \sum_{j'=1}^{N_j} \hat{a}_{l,j,j'} B_{J_{l,j,j'}}(t) B_{J_{l,j,j'}}(t'),
\]

where \( \hat{a}_{l,j,j'} = n^{-1} \sum_{i=1}^{n} \hat{b}_{l,i,j,j'} \hat{b}_{l,i,j,j'} \). The estimates of eigenfunctions and eigenvalues correspond to \( \psi_{l,k} \) and \( \lambda_{l,k} \) can be obtained by solving the eigenequations

\[
\int_0^1 \hat{G}_l(t, t') \psi_{l,k}(t') \, dt' = \lambda_{l,k} \psi_{l,k}(t), \quad k = 1, \ldots, N_j.
\]

The above integral equations are easily solved with the aid of discretizing the integral and by spline approximation for \( \psi_{l,k}(\cdot) \): \( \hat{\psi}_{l,k}(t') = \sum_{j=1}^{N_j} \hat{\psi}_{l,k,j} B_{J_{l,k,j}}(t') \). The B-spline estimator coefficients \( \hat{\psi}_{l,k,j} \)'s are subject to \( \hat{\psi}_{l,k}^T B^T \hat{\psi}_{l,k} = N \) with \( \hat{\psi}_{l,k} = \left( \hat{\psi}_{l,k,1}, \ldots, \hat{\psi}_{l,k,N_j} \right)^T \). According to (10), solving (11) is equivalent to solving the following: \( B^T(t) \hat{a}_{l,k} B^T(t) = N \hat{\lambda}_{l,k} B^T(t) \hat{\psi}_{l,k} \), where \( \hat{a}_{l,k} = \left( \hat{a}_{l,j,j'} \right)_{j,j'}^{N_j} \). These equations are equivalent to \( \hat{a}_{l,k} B^T \hat{\psi}_{l,k} = N \hat{\lambda}_{l,k} \hat{\psi}_{l,k} \) by simple algebra and Lemma 3.1 in Wang and Yang (2009).

Making use of a Cholesky decomposition: \( B^T B / N = D_B^T D_B \), solving (11) is equivalent to solving \( \hat{\lambda}_{l,k} D_B^T \hat{\psi}_{l,k} = D_B^T \hat{a}_{l,k} D_B \hat{\psi}_{l,k} \); that is, \( \hat{\lambda}_{l,k} \) and \( D_B^T \hat{\psi}_{l,k} \) are the eigenvalues and unit eigenvectors of \( D_B^T \hat{\psi}_{l,k} D_B \). Thus, \( \hat{\lambda}_{l,k} \) is obtained by multiplying \( (D_B^T)^{-1} \) by the unit eigenvectors of \( D_B^T \hat{\psi}_{l,k} D_B \).
and since $\mathbf{B}^\top \mathbf{B} / N$ and $\hat{d}_i$ are $N \times N$ matrices and so is $\mathbf{D}_B^\top \mathbf{D}_B$, a total of $N^2$ eigenvalues $\hat{\lambda}_{i,k}$ are computed. The orthonormal eigenfunction $\tilde{\psi}_{i,k}(\cdot)$ is obtained next, and the rescaled eigenfunction $\hat{\phi}_{i,k}(t') = \hat{\lambda}_{i,k}^{3/2} \tilde{\psi}_{i,k}(t')$. The $k$th FPC score of the $ith$ signal on scalp location $l$ is then estimated by a numerical integration,

$$
\hat{\phi}_{i,l,k}(t') = \frac{1}{N} \sum_{j=1}^{N} \hat{\lambda}_{i,k}^{3/2} \left\{ \frac{j}{N} - \bar{m}_t \left( \frac{j}{N} \right) \right\} \hat{\phi}_{i,k} \left( \frac{j}{N} \right).
$$

(12)

Next, one chooses the effective number $\kappa_i$ of eigenfunctions for EEG signals at location $l$ by a rule-of-thumb criterion, i.e.,

$$
\kappa_i = \arg\min_{\sum_{k=1}^{L} \hat{\lambda}_{i,k} / \sum_{k=1}^{N} \hat{\lambda}_{i,k} > 0.95} \{ \sum_{k=1}^{N} \hat{\lambda}_{i,k} \}^\kappa_i
$$

which results in $\hat{\kappa}_i$, $\hat{\lambda}_{i,k} / \sum_{k=1}^{N} \hat{\lambda}_{i,k}$ account for approximately 95% of the nuclear norm $\sum_{k=1}^{L} \hat{\lambda}_{i,k}$.

From (1)-(12), an approximate of (9) for the N-back score $\{Y_i\}$ is

$$
Y_i = \beta_0 + \sum_{l=1}^{L} \sum_{k=1}^{\hat{\kappa}_i} \hat{\phi}_{i,l,k} \hat{\beta}_{l,k} + \epsilon_i, 1 \leq i \leq n.
$$

(13)

The regression parameters are estimated by

$$
\{\hat{\beta}_0, \hat{\beta}_{l,k}\} = \arg\min \left\{ \sum_{i=1}^{n} \left( Y_i - \beta_0 - \sum_{l=1}^{L} \sum_{k=1}^{\hat{\kappa}_i} \hat{\beta}_{l,k} \hat{\phi}_{i,l,k} \right)^2 \right\},
$$

which result in $\hat{\beta}_0 = \sum_{l=1}^{L} \int \hat{\beta}_l(t) \, dt$, $\hat{\beta}_i(t) = \sum_{k=1}^{\hat{\kappa}_i} \hat{\beta}_{l,k} \hat{\phi}_{i,l,k}(t)$. This leads to a prediction model

$$
\hat{Y}_i = \hat{\beta}_0 + \sum_{l=1}^{L} \int \hat{\beta}_l(t) \hat{X}_{i,l}(t) \, dt + \epsilon_i, 1 \leq i \leq n.
$$

(14)

for the score $Y_i$.

4. Results and discussion

In this section, the prediction model (14) for the N-back score $Y_i$ is built with the functional EEG data. The 142 EEG signals were randomly split into a training set (with 122 samples, $i = 1, \ldots, 122$) and a testing set (with 20 samples, $i = 123, \ldots, 142$) with 50 replications. Prediction of the N-back score for the 20 samples in testing set is carried out using a model (14) built from the training data.

Of all of the 32 electrode positions, for the retained training dataset ($i = 1, \ldots, 122$), through first three steps in Section 3.5, a total of 94 FPC scores $\hat{\phi}_{i,l,k}, k = 1, \ldots, \kappa_i, l = 1, \ldots, 32$ are selected for the preliminary full model (13). To further reduce the regression model to a much lower dimension, LASSO is conducted with tenfold cross-validation to select FPC scores predictive of the full model based on the criteria giving a minimum mean cross-validated error. This fourth step was performed using glmnet (https://cran.r-project.org/web/packages/glmnet/index.html). Since the raw scores $S_i$ take only integer values from 0 to 16 and $Y_i = \log \{ (S_i + 1) / (17 - S_i) \}$, further refinement is achieved by rounding the predicted raw scores $\hat{S}_i = \lceil \{ (\exp \{ \hat{Y}_i \} - 1) / (1 + \exp \{ \hat{Y}_i \}) \} \rceil$ to whole numbers, $\hat{S}_i, \hat{Y}_i, \hat{Y}_i, \hat{Y}_i, \hat{Y}_i$ adj = round ($\hat{S}_i$), and recomputing the predicted values $\hat{Y}_i, \hat{Y}_i, \hat{Y}_i, \hat{Y}_i$, adj = log $\{ (\hat{S}_i, adj + 1) / (17 - \hat{S}_i, adj) \}$. The $R^2$ is computed for the testing set, which is

$$
R^2 = \frac{\sum_{i=123}^{142} (Y_i - \bar{Y}) \left( \hat{Y}_i, adj - \bar{\hat{Y}}_i \right)}{\sum_{i=123}^{142} (Y_i - \bar{Y})^2 \sum_{i=123}^{142} \left( \hat{Y}_i, adj - \bar{\hat{Y}}_i \right)^2}.
$$

Among the 50 testing sets, the median of $R^2$ is 0.68, while the highest equals 0.72; see Fig. 7. The boxplot of the results shows that the proposed method significantly outperforms the conventional signal processing related to WM.

In the case of $R^2 = 0.72$, approximately 72% of the total variation in WM is explained by the EEG signals of eight selected electrode positions in Fig. 1, indicating a rather strong functional regression relationship. For the training subjects, applying (14), spline estimators $\hat{X}_{i,l}(t)$ of the subject $i = 1$ from eight, selected channels are shown in Figs. 8(a)-(h) and
the corresponding estimated coefficient functions are shown in Figs. 9(a)-(h). Figure 5 illustrates the first six eigenfunctions estimated by using FPCA of the Fp2 scalp location (i.e., $\hat{\phi}_{1,k}(\cdot), k = 1, 2, 3, 4, 5, 6$). Finally, a parsimonious regression model is obtained using 16 FPC scores at eight EEG channels out of the 32 as follows:

$$
\begin{align*}
Fp2 &: \hat{\xi}_{1,1,1}, \hat{\xi}_{1,1,3}; \\
F3 &: \hat{\xi}_{2,1,6}, \hat{\xi}_{2,1,9}; \\
Fz &: \hat{\xi}_{3,1,10} - \hat{\xi}_{3,1,13}; \\
F4 &: \hat{\xi}_{4,1,5}, \hat{\xi}_{4,1,9}; \\
F8 &: \hat{\xi}_{5,1,1}, \hat{\xi}_{5,1,4}; \\
Fc5 &: \hat{\xi}_{6,1,4}; \\
Fc1 &: \hat{\xi}_{7,1,6} - \hat{\xi}_{7,1,11}; \\
Fc6 &: \hat{\xi}_{8,1,2} - \hat{\xi}_{8,1,11}.
\end{align*}
$$

The plot of the predicted scores $\{\hat{Y}_{i,adj}\}$ versus the observed scores $\{Y_i\}$ for testing data (i.e., $i = 123, \ldots, 142$) is given in Fig. 6.

To our knowledge, there is no literature regarding the establishment of a model for WM ability prediction. Most of the existing EEG studies related to WM and using other feature-extraction methods focus on the estimation of WM load. In those studies, energy or connectivity features were used for classification, with less than desirable accuracy (Roy et al., 2013; Charbonnier et al., 2016). Our study shows that a multiple functional linear-model-based method could be a more advanced approach to predict the WM from eyes-closed resting-state EEG signals and potentially has wider applications in cognitive neuroscience.

5. Conclusions

The goal of this research is to extract EEG features that can be used for prediction of the human cognitive ability. We have proposed a novel multiple functional linear model method that relates individual WM ability measured with the N-back task to EEG signals in a resting position with eyes closed.

As previous studies suggested, our critical electrode positions are concentrated in the frontal region of the brain, which indicates that WM ability is associated with frontal activity (Charbonnier et al., 2016). Further study is expected to illustrate how the functional data analysis method can be applied to other cognitive functions as well, such as attention or decision making, and used for rehabilitation application (Li et al., 2016, 2019). Extended study with a larger cohort of subjects and multiple cognitive function indicators will result in a much more powerful prediction model with further enhanced $R^2$.

Declaration of conflicts of interest

The authors declare that there are no conflicts of interest relevant to this article.

Acknowledgements

The research was supported in part by the National Natural Science Foundation of China (Grant Nos. 20161301325 and 11771240). The authors thank three anonymous reviewers and Editor-in-Chief Dr. Giuseppe Di Giovanni for helpful comments.
Prediction of working memory ability based on EEG

Figure 9: Plots of estimated coefficient functions: (a) $\hat{\beta}_1(t)$ of Fp2, (b) $\hat{\beta}_2(t)$ of F3, (c) $\hat{\beta}_1(t)$ of Fz, (d) $\hat{\beta}_1(t)$ of F4, (e) $\hat{\beta}_1(t)$ of F8, (f) $\hat{\beta}_1(t)$ of Fc5, (g) $\hat{\beta}_1(t)$ of Fc1, (h) $\hat{\beta}_1(t)$ of Fc6.

References


