12-30-2018

Nonparametric Collective Spectral Density Estimation with an Application to Clustering the Brain Signals

Mehdi Maadooliat
Marquette University, mehdi.maadooliat@marquette.edu

Ying Sun
King Abdullah University of Science and Technology

Tianbo Chen
King Abdullah University of Science and Technology

Statistics in Medicine, Vol. 37, No. 30 (December 30, 2018): 4789-4806. DOI.

Nonparametric Collective Spectral Density Estimation with An Application to Clustering the Brain Signals

Mehdi Maadooliat
Department of Mathematics, Statistics and Computer Science, Marquette University, Milwaukee, Wisconsin
Center for Precision Medicine Research, Marshfield Clinic Research Institute, Marshfield, Wisconsin
Ying Sun
CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia
Tianbo Chen
CEMSE Division, King Abdullah University of Science and Technology, Thuwal, Saudi Arabia

Abstract

In this paper, we develop a method for the simultaneous estimation of spectral density functions (SDFs) for a collection of stationary time series that share some common features. Due to the

similarities among the SDFs, the log-SDF can be represented using a common set of basis functions. The basis shared by the collection of the log-SDFs is estimated as a low-dimensional manifold of a large space spanned by a prespecified rich basis. A collective estimation approach pools information and borrows strength across the SDFs to achieve better estimation efficiency. Moreover, each estimated spectral density has a concise representation using the coefficients of the basis expansion, and these coefficients can be used for visualization, clustering, and classification purposes. The Whittle pseudo-maximum likelihood approach is used to fit the model and an alternating blockwise Newton-type algorithm is developed for the computation. A web-based shiny App found at "**https://ncsde.shinyapps.io/NCSDE**" is developed for visualization, training, and learning the SDFs collectively using the proposed technique. Finally, we apply our method to cluster similar brain signals recorded by the for identifying synchronized brain regions according to their spectral densities.

1 INTRODUCTION

Nonparametric techniques for estimating functional structures have been developed in a variety of settings including regression, density estimation, and survival analysis. In time series analysis, the spectral density function plays an important role in characterizing the frequency content of a signal. Mainly, the estimated spectral density can be used to detect the periodicities of the signals in the frequency domain.

In practice, it is common to utilize a discrete Fourier transform (DFT) of the input signal and provide a mathematical approximation of the full integral solution of the Fourier transformation. The squared-magnitude of a DFT of the data is called periodogram. However, the raw periodogram is not a consistent estimator for the spectral density of a stationary random process. One classical method to obtain a consistent estimator is to smooth the periodogram across frequencies. Yuen[1] analyzed the performance of three methods of periodogram smoothing for spectrum estimation. Wahba[2] developed an objective optimum smoothing procedure for estimating the log-spectral density using the spline to smooth the log-periodogram. A discrete spectral average estimator and lag window estimators were introduced in the work of Brockwell and Davis.[3] Both of the two methods are consistent. Brillinger[4] introduced periodogram kernel smoothing. One critical issue in periodogram smoothing is span selection. Lee[5] proposed a span selector based on unbiased risk estimation. Ombao[6] proposed using the Gamma-deviance generalized cross-validation (Gamma GCV), and Lee[7] proposed a bandwidth selection based on coupling of the so-called plug-in and the unbiased risk estimation ideas. Iannaccone and Coles[8] introduced a semiparametric random effect model for spectral analysis in a Bayesian framework using the Markov Chain Monte Carlo (MCMC) methodology. Freyermuth et al[9] presented a functional mixed effects model based on the methodology of tree-structured wavelets to rigorously treat the random effects for estimating the spectrum of a stationary process. Krafty et al[10] proposed a more general mixed-effects model based on spline smoothing of the empirical log-spectra using the first two moments of the log-spectra. Chau and van Sachs[11] extended the last two approaches[9, 10] by allowing different time series replicates within a subject to be correlated based on known covariates.

Another popular method for spectral density estimation is based on likelihoods. For example, Capon[12] used the maximum likelihood method to estimate the spectral density of signals with noise. In the work of Güler et al,[13] the spectral density of brain signal data was analyzed using the maximum likelihood. The well-known Whittle likelihood was developed for time series analysis in the work of

Whittle.[14-16] Pawitan and O'Sullivan[17] proposed using a penalized Whittle likelihood to estimate spectral density function (SDF). Chow and Grenander[18] developed a penalized likelihood-type method for the nonparametric estimation of the spectral density of Gaussian processes. Fan and Kreutzberger[19] used local polynomial techniques to fit the Whittle likelihood for spectral density estimation.

It is challenging to estimate spectral densities from multiple time series. Pawitan[20] developed a penalized Whittle likelihood estimator for bivariate time series only and not suitable for high-dimensional time series. Dai and Guo[21] proposed to smooth the Cholesky decomposition of a raw estimate of a multivariate spectral density, allowing different degrees of smoothness for different elements. Rosen and Stoffer[22] proposed a Bayesian approach that applies MCMC techniques to fit smoothing splines to each component of the Cholesky decomposition of the periodograms, and then the spectral estimator is obtained by reconstructing the spectral estimator from these components. Krafty and Collinge[23] introduced a penalized likelihood approach for multivariate spectral density estimation through the minimization of a penalized Whittle negative log-likelihood, but it is computationally expensive for large dimensionality.

In this paper, we propose collectively estimating multiple spectral densities that share common features. The collective estimation was first proposed by Maadooliat et al[24] for the estimation of probability density functions in protein structure analysis. For multiple stationary time series that may share similar SDF, we propose a nonparametric spectral density estimation approach that minimizes a penalized Whittle negative log-likelihood. The dimension is reduced by representing the raw periodograms on a set of common basis functions while allowing each one to differ by introducing random effects into the model. The collectively estimated spectral densities can then be used for clustering purposes. Collective estimation is a novel approach for estimating multiple spectral densities that share common features in a more statistically efficient way.

The rest of the paper is organized as follows. Section **2** presents the core of the proposed method. Section **2.3** provides the blockwise Newton-Raphson algorithm, and Sections **2.4-2.8**-**2.4-2.8** provide the implementation details. Section **3** reports the simulation results to illustrate the proposed collective estimation approach and to compare it with competitive noncollective estimation approaches. The application to EEG data is given in Section **4**. Section **5** concludes the paper.

2 METHODOLOGY

2.1 Definitions and background

We suppose that X_t represents a zero-mean, weakly stationary time series with the autocovariance function $\gamma(h)$ defined as $\gamma(h) = E(X_t X_{t+h})$, $h = 0, \pm 1, \pm 2, \ldots$ If the autocovariance function is absolutely summable, ie, $\sum_{h=-\infty}^{\infty} |\gamma(h)| < \infty$, then the autocovariance sequence $\gamma(h)$ has the spectral representation

$$\gamma(h) = \int_{-1/2}^{1/2} f(\omega) e^{2\pi i \omega h} d\omega,$$

where $f\ (\omega)$ is the spectral density of X_t, which has the inverse Fourier representation

$$f(\omega) = \sum_{h=-\infty}^{\infty} \gamma(h) e^{-2\pi i \omega h}, -1/2 \leq \omega \leq 1/2.$$

Given the time series $\{x_t,\ t\ =\ 1, \ldots, n\}$, when $\gamma(h)$ is replaced by the sample covariance $\hat{\gamma}(h)$, the periodogram is defined as

$$I_n(\omega_j)\ = \sum_{h=-(n-1)}^{n+1} \hat{\gamma}(h) e^{-2\pi i \omega_j h}, j = 0, 1, \ldots, n-1,$$

which can be calculated as $I_n(\omega_j)\ =\ |d(\omega_j)|^2$, where $d(\omega_j) = n^{-1/2} \sum_{t=1}^{n} x_t\, e^{-2\pi i \omega_j h}$ is the DFT of x_t at the fundamental frequencies $\omega_j\ =\ j/n$.

Now, suppose $\mathbf{X}$ is a time series of length n from a mean-zero stationary Gaussian process with a parametric covariance function $\gamma(h; \boldsymbol{\theta})$, where $\boldsymbol{\theta}$ is the unknown parameter. If we let $\Gamma n, \boldsymbol{\theta}$ be the covariance matrix of the random vector $\mathbf{X}$, then the likelihood function is

$$L(\theta) = \frac{1}{(2\pi)^{n/2} \left|\Gamma_{n,\theta}\right|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{X}^{\top} \Gamma_{n,\theta}^{-1} \mathbf{X}\right),$$

and $\ell(\theta) = -2 \times \log L(\theta)$ is the negative log-likelihood

$$\ell(\theta) \propto \log\left(\left|\Gamma_{n,\theta}\right|\right) + \mathrm{tr}\left(\mathrm{X}\mathrm{X}^{\top} \Gamma_{n,\theta}^{-1}\right).$$

Whittle[25] proposed an approximation of the aforementioned log-likelihood function, known as the Whittle likelihood approximation,

$$\ell w(\theta) = n \int_{-\frac{1}{2}}^{\frac{1}{2}} \left\{\log\left(f_\theta(\omega)\right) + I_n(\omega) f_\theta(\omega)^{-1}\right\} d\omega,$$

(1)

where $f_\theta(\omega)$ is the spectral density function. For a discrete frequency range, the Whittle approximation 1 can be written as

$$\ell_w(\theta) \doteq n \sum_{-1/2 < \omega_j < 1/2} \left\{\log\left(f_\theta\left(\omega_j\right)\right) + I_n\left(\omega_j\right) f_\theta\left(\omega_j\right)^{-1}\right\}.$$

Using the fast Fourier transform, $\ell_w(\theta)$ can be obtained efficiently with only $O(n \log_2 n)$ operations.

2.2 Collective modeling of SDF

We consider a collection of stationary time series, ie, $\mathbf{X}_i$s, where $i = 1, \dots, m$ and $\mathbf{X}_i^\top = (X_{i1}, X_{i2}, \dots, X_{in})$ with the associated spectral density f_i. We want to estimate the SDF together. The rationale of this collective spectral density estimation approach is to improve the efficiency of the estimates by using a shared basis to obtain the SDFs.

We assume that each log-spectral density function can be represented by a linear combination of a common set of basis functions $\{\phi_k(\omega), k = 1, \dots, K\}$ and that each has its own set of coefficients. More specifically, we assume that $\log\{f_i(\omega)\} = u_i(\omega)$, with

$$u_i(\omega) = \sum_{k=1}^{K} \phi_k(\omega)\alpha_{ik}, \qquad i = 1, \dots, m.$$

(2)

Equivalently, the SDF can be written as

$$f_i(\omega) \exp u_i(\omega) = \exp\left\{\sum_{k=1}^{K} \phi_k(\omega)\alpha_{ik}\right\}, \qquad i = 1, \dots, m.$$

(3)

For identifiability, we require that $1, \phi_k, k = 1, \dots, K$, to be linearly independent. We would like K to be a small number so that the number of parameters to be estimated remains manageable even when we estimate a large number of spectral densities (ie, when m is large).

In our setting, the basis functions are not prespecified and need to be determined from the data. To this end, we suppose that these basis functions fall in a low-dimensional subspace of a function space spanned by a rich family of fixed basis functions, ie, $\{b_\ell(\omega), \ell = 1, \dots, L\}$ $(L \gg K)$, such that

$$\phi_k(\omega) = \sum_{\ell=1}^{L} b_\ell(\omega)\theta_{\ell k}, \qquad k = 1 \dots, K.$$

(4)

For identifiability, we require that $1, b_\ell, \ell = 1, \dots, L$, to be linearly independent. A large enough L ensures the necessary flexibility to represent the unknown spectral densities. For univariate cases, the fixed basis can be the monomials, the B-splines, or the Fourier basis. Bivariate splines can be used as the fixed basis functions for bivariate spectral densities. In this work, without loss of generality, we initialize the algorithm with a cubic B-spline basis functions with 40 degrees of freedom.

To simplify the presentation, we now introduce some vectors and matrices to denote the quantities of interest, ie, $\boldsymbol{\phi}(\omega) = (\phi_1(\omega), \dots, \phi_K(\omega))^\top$, $\boldsymbol{\alpha}_i = (\alpha_{i1}, \dots, \alpha_{iK})^\top$, $\mathbf{b}(\omega) = (b_1(\omega), \dots, b_L(\omega))^\top$, $\boldsymbol{\theta}_k =$

$(\theta_{1k}, \ldots, \theta_{Lk})^\top$, and $\boldsymbol{\Theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_K)$. Then, from 2 and 4, we can rewrite $u_i(\omega)$ in the vector-matrix form as

$$u_i(\omega) = \boldsymbol{\phi}(\omega)^\top \boldsymbol{\alpha}_i = \mathbf{b}(\omega)^\top \boldsymbol{\Theta} \boldsymbol{\alpha}_i, \qquad i = 1, \ldots, m.$$

(5)

One may combine the equations previously for $i = 1, \ldots, m$ into a matrix form by evaluating the log-SDFs over the discrete frequencies $\omega = (\omega_1, \ldots, \omega_{\tilde{n}})^\top$. The m equations given in 5 can be written as $\mathbf{U} = \mathbf{B}\boldsymbol{\Theta}\mathbf{A}^\top$, where $\boldsymbol{U} = (u_1(\boldsymbol{\omega}), \ldots, u_m(\boldsymbol{\omega}))$ is an $\tilde{n} \times m$ matrix that represents the log-SDFs, $\mathbf{B} = (\mathbf{b}(\omega_1), \ldots, \mathbf{b}(\omega_{\tilde{n}}))^\top$ is an $\tilde{n} \times L$ matrix that represents the rich basis functions at the discrete frequencies $\boldsymbol{\omega}$, and $\boldsymbol{A} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_m)^\top$. The unknown parameters can then be collectively written as the pair $(\boldsymbol{\Theta}, \mathbf{A})$. There is an identifiability issue caused by the nonuniqueness of the parametrization of $(\boldsymbol{\Theta}, \mathbf{A})$. This issue can be resolved by introducing some restrictions on the parameterization (see Section 2.8).

We could have used the fixed basis $\{b_\ell(\omega), \ell = 1, \ldots, L\}$ in 2 and 3; however, that would be either too restrictive (if L is small) or produce a large number of parameters (if L is large). Alternatively, if we were to model the individual density functions separately using the fixed basis $\{b_\ell(\omega), \ell = 1, \ldots, L\}$, then we would write

$$u_i(\omega) = \mathbf{b}(\omega)^\top \boldsymbol{\psi}_i, \qquad i = 1, \ldots, m.$$

(6)

We let $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_m)$ be the $L \times m$ matrix of coefficients from the basis expansions given in 6. By comparing 5 and 6, we obtain $\boldsymbol{\Psi} = \boldsymbol{\Theta}\mathbf{A}^\top$, which is a rank K matrix. Thus, the collective modeling approach introduces a low-rank structure to the coefficient matrix in the basis expansion of the log-spectral densities. This dimensionality reduction allows us to significantly reduce the number of parameters to be estimated and, thus, gain estimation efficiency.

2.3 Estimation using penalized Whittle likelihood

The periodogram $I_{i,n}$ is a rough estimate of the spectral density, ie, f_i, associated with the time series $\mathbf{X}_i$ observed over n time points

$$I_{i,n}(\omega) = \left| \frac{1}{n} \sum_{t=1}^{n} X_{it} e^{-2\pi i t \omega} \right|^2.$$

We consider the periodogram at the Fourier frequencies $\boldsymbol{\omega} = 2\pi \boldsymbol{J}/n$, for $\boldsymbol{J} = \{-\lfloor (n-1)/2 \rfloor, \ldots, n - \lfloor n/2 \rfloor\}$.

The Whittle likelihood for estimating *m* spectral densities has the following form:

$$\ell_w(\boldsymbol{\Theta}, \mathbf{A}) = \sum_{i=1}^{m} \sum_{j \in J} \{u_i(\omega_j) + I_{i,n}(\omega_j) \exp[-u_i(\omega_j)]\},$$

(7)

where $u_i(\omega)$ are defined in 5. It is concave in $\boldsymbol{\alpha}_i$ when other parameters are fixed and also concave in $\boldsymbol{\theta}_k$ when other parameters are fixed. Applying the roughness penalty approach of function estimation,[26] we estimate the model parameters by minimizing the penalized likelihood criterion

$$-2\ell_w(\boldsymbol{\Theta}, \mathbf{A}) + \lambda \sum_{k=1}^{K} \mathrm{PEN}(\phi_k),$$

(8)

where $\mathrm{PEN}(\phi_k)$ is a roughness penalty function that regularizes the estimated basis function ϕ_k to ensure that it is a smooth function and $\lambda > 0$ is a penalty parameter. The penalty function can be written in a quadratic form as

$$\sum_{k=1}^{K} \mathrm{PEN}(\phi_k) = \sum_{i=k}^{K} \boldsymbol{\theta}_k^{\top} \boldsymbol{R} \theta_k = \mathrm{tr}\{\boldsymbol{\Theta}^{\top} \mathbf{R} \boldsymbol{\Theta}\}.$$

(9)

Two choices for the penalty matrices ($\mathbf{R}_1$ and $\mathbf{R}_2$), based on the second derivative of the basis functions and the difference operator, are given in Section **2.5**.

We use an alternating blockwise Newton-Raphson algorithm to minimize the penalized Whittle likelihood approximation. Our algorithm cycles through updating $\boldsymbol{\alpha}_i$, for $i = 1, \ldots, m$ and $\boldsymbol{\theta}_k$ for $k = 1, \ldots, K$ until convergence. Following the usual step-halving strategy for the Newton-Raphson iteration, the updating formulas are

$$\begin{aligned}
\boldsymbol{\alpha}_i^{new} &= \boldsymbol{\alpha}_i^{old} - \tau \left[\frac{\partial^2}{\partial \boldsymbol{\alpha}_i \partial \boldsymbol{\alpha}_i^{\top}} \{\ell_w(\boldsymbol{\Theta}, \mathbf{A})\} \right]^{-1} \left[\frac{\partial}{\partial \boldsymbol{\alpha}_i} \{\ell_w(\boldsymbol{\Theta}, \mathbf{A})\} \right] \Big|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{old}, \mathbf{A}=\mathbf{A}^{old}} \\
&= \boldsymbol{\alpha}_i^{old} - \tau \left[\boldsymbol{\Theta}^{\top} \sum_j \left\{ \mathbf{b}(\omega_j) \boldsymbol{I}_{i,n}(\omega_j) \exp\left[-u_i(\omega_j)\right] \mathbf{b}(\omega_j)^{\top} \right\} \boldsymbol{\Theta} \right]^{-1} \times \\
&\quad \left[\boldsymbol{\Theta}^{\top} \sum_j \left\{ \mathbf{b}(\omega_j) - \mathbf{b}(\omega_j) \boldsymbol{I}_{i,n}(\omega_j) \exp\left[-u_i(\omega_j)\right] \right\} \right] \Big|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{old}, \mathbf{A}=\mathbf{A}^{old}}
\end{aligned}$$

(10)

and

$$\theta_k^{new} = \theta_k^{old} - \tau\left[\frac{\partial^2}{\partial\boldsymbol{\theta}_k\partial\boldsymbol{\theta}_k^{\top}}\{\ell_w(\boldsymbol{\Theta},\mathbf{A})\} - \lambda\mathbf{R}\right]^{-1} \times \left[\frac{\partial}{\partial\theta_k}\{\ell_w(\boldsymbol{\Theta},\mathbf{A})\} - \lambda\mathbf{R}\theta_k\right]\Big|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{old},\mathbf{A}=\mathbf{A}^{old}}$$

$$= \boldsymbol{\theta}_k^{old} - \tau\left[\sum_{\mathbf{1=1}}^{\boldsymbol{m}} \alpha_{ik}^2 \sum_{\boldsymbol{j}} \left\{\mathbf{b}\left(\omega_j\right)\boldsymbol{I}_{i,n}\left(\omega_j\right)\exp\left[-u_i\left(\omega_j\right)\right]\mathbf{b}\left(\omega_j\right)^{\top}\right\} - \lambda\boldsymbol{R}\right]^{-1} \times \left[\sum_{i=1}^{m} \alpha_{ik} \sum_{\boldsymbol{j}} \left\{\mathbf{b}\left(\omega_j\right) - \mathbf{b}\left(\omega_j\right)\boldsymbol{I}_{i,n}\left(\omega_j\right)\exp\left[-u_i\left(\omega_j\right)\right]\right\} - \lambda\boldsymbol{R}\theta_k\right]^{-1}\Big|_{\boldsymbol{\Theta}=\boldsymbol{\Theta}^{old},\mathbf{A}=\mathbf{A}^{old}}$$

(11)

where τ is the first result from the sequence $\{(1/2)^{\delta}, \delta = 0,1,\ldots\}$ such that the objective function in 8 is reduced. The initial values of the Newton-Raphson iteration can be obtained by projecting the raw spectral density estimates (eg, periodograms) to the model space of 3.

2.4 Selecting the tuning parameter

We may select the penalty parameter by minimizing the AIC[27]

$$\mathrm{AIC}(\lambda) =* 2\ell_W(\widehat{\boldsymbol{\Theta}},\widehat{\mathbf{A}}) + 2\mathrm{df}(\lambda),$$

(12)

where $\ell_W(\widehat{\boldsymbol{\Theta}},\widehat{\mathbf{A}})$ is the log-likelihood defined in 7, and the degrees of freedom $\mathrm{df}(\lambda)$ is defined as

$$\mathrm{df}(\lambda) = \sum_{k=1}^{K} \mathrm{trace}\left\{\begin{array}{l}\left[\sum_{\mathbf{1=1}}^{\boldsymbol{m}} \alpha_{ik}^2 \sum_{\boldsymbol{j}} \left\{\mathbf{b}(\omega_j)\boldsymbol{I}_{i,n}(\omega_j)\exp\left[-u_i(\omega_j)\right]\mathbf{b}(\omega_j)^{\top}\right\} - \lambda\boldsymbol{R}\right]^{-1} \\ \times\left[\sum_{i=1}^{m} \alpha_{ik}^2 \sum_{\boldsymbol{j}} \left\{\mathbf{b}(\omega_j)\boldsymbol{I}_{i,n}(\omega_j)\exp\left[-u_i(\omega_j)\right]\mathbf{b}(\omega_j)^{\top}\right\}\right].\end{array}\right\}$$

(13)

The parameters in these formulas are replaced by their estimated values. The AIC can be derived as an approximation of the leave-one-out cross-validation.[28, 29]

Selecting the tuning parameter that minimizes the AIC requires training the model for different values of λs and then picking the one that minimizes the criterion function, which can be very expensive in time. Instead, we present an alternative procedure that updates the value of the tuning parameter within the Newton-Raphson iterations. This idea has been used in a generalized mixture model to iteratively update the smoothing parameter.[30] Schellhase and Kauermann[31] and Najibi et al[32] extended

this approach for density estimation. We borrow their formulation and use the parameter estimates in the i^{th} step to update the tuning parameter, ie, $\hat{\lambda}_{i+1}$, through

$$\hat{\lambda}_{i+1}^{-1} = \frac{\text{trace}\left(\widehat{\boldsymbol{\Theta}}_{\text{i}}^{\top}\mathbf{R}\widehat{\boldsymbol{\Theta}}_{i}\right)}{\text{df}(\lambda) - (a-1)},$$

(14)

where a is the order of the differences (derivative) used in the penalty matrix $\mathbf{R}$ (see Section 2.5). From what we have seen in the implementation of the new procedure, updating the tuning parameter within the Newton-Raphson iterations, on average, does not increase the number of iterations required for convergence. Therefore, the new procedure obtains the final result p times faster than the old procedure, where p is the number of λs used in the grid search to minimize the AIC (see the work of Najibi et al[32] for details).

2.5 Choices of penalties

We consider two different choices of penalties.

- **Second derivative of the basis**. For the univariate spectral density estimation, if we use the usual squared-second-derivative penalty $\text{PEN}(\phi_k) = \int\{\phi_k''(\omega)\}^2\, dx$ and $\phi_k(\omega) = \mathbf{b}(\omega)^{\top}\boldsymbol{\theta}_k$, then $\mathbf{R}_1 = \int \ddot{\mathbf{b}}(\boldsymbol{\omega})\ddot{\mathbf{b}}(\omega)^{\top}\, d\omega$ with $\ddot{\mathbf{b}}(\boldsymbol{\omega}) = \left(b_1''(\omega), \ldots, b_L''(\omega)\right)$.
- **Difference operator**. We can also control the roughness of the estimated functions by using the difference penalty[33] to achieve the appropriate level of smoothness. The variability is controlled through a difference function of order a, ie, Δ_a, where $\Delta_1\boldsymbol{\theta}_k := \boldsymbol{\theta}_k - \boldsymbol{\theta}_{k-1}$, and Δ_a is obtained recursively. For example, the second-order difference function, ie, Δ_2, has the following form:

$$\Delta_2\theta_k := \Delta_1\Delta_1\theta_k = \theta_k - 2\theta_{k-1} + \theta_{k-2}.$$

We can write the difference functions Δ_a into a matrix form, ie, $\boldsymbol{L}_a$. For example, when $a = 1$, we have

$$\boldsymbol{L}_1 = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 0 & 1 & -1 & \ddots & 0 \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & 1 & -1 \end{bmatrix}_{(M-1\times M)} \quad 0$$

The positive definite penalty matrix used to control the smoothness is denoted as $\mathbf{R}_2$, and it has the quadratic form $\mathbf{R}_2 = \boldsymbol{L}_{\boldsymbol{a}}^{\top}\boldsymbol{L}_{\boldsymbol{a}}$.

In the web-application, where we implemented the proposed nonparametric collective spectral density estimation (NCSDE), we incorporated both penalties and either one can be used by the user.

2.6 Clustering

The NCSDE result can be utilized as an input to different clustering algorithms (eg, k-means, hierarchical, …). We used the *Euclidean* distance between (*i*) the estimated SDFs and (*ii*) the basis expansion coefficients (score matrix **A**), associated to six competitive approaches given in Section **3.2**, as inputs to the hierarchical clustering algorithm. In our implementation, we used the hclust function

with the *Euclidean* distances as input and option {method="ward.D2"} in the R package stats to obtain the clustering results.[34]

2.7 Number of basis functions (clusters)

A critical step in dealing with real data is identifying the number of common basis functions (choice of *K* in the context of NCSDE), which is directly related to the number of clusters, ie, $\tilde{k}$. We identify the number of basis functions (clusters) using the *elbow method* as typically used in clustering analysis and it can be traced back to the work of Thorndike.[35]

In the elbow method, we use hierarchical clustering for partitioning the data and obtain the total within-cluster sum of squares (WSS) based on different numbers of clusters $\tilde{k}$ (eg, by varying $\tilde{k}$ from 1 to 10). We note that WSS measures the concentration of the clusters and it is desired to be as small as possible. Therefore, the optimal number of clusters can be obtained by plotting WSS against the number of clusters $\tilde{k}$. The location of an elbow (turning point) in this plot is generally considered an indicator of the appropriate number of clusters (see Figures **1**A, **6**A, and **7**A for illustration purposes).

In the first step of the NCSDE algorithm, we obtain the logarithm of the smoothed periodograms, $\log(\mathrm{S.Ps}) = \mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top \log(\mathbf{I})$, which is then used as an input to the elbow method, implemented in the R package factoextra,[36] to determine the number of clusters and basis functions. From what we have seen in the implementation of NCSDE, a clear-cut elbow is achievable in longer time series ($n \geq 400$).

2.8 Identifiability of $(\boldsymbol{\Theta}, \mathbf{A})$

The nonuniqueness of the parametrization of $(\boldsymbol{\Theta}, \mathbf{A})$ causes an identifiability issue. Specifically, if $\mathbf{U}$ is a $K \times K$ orthogonal matrix, then $\boldsymbol{\Theta}\boldsymbol{\alpha}_i = (\boldsymbol{\Theta}\mathbf{U})(\mathbf{U}^\top\boldsymbol{\alpha}_i)$. Thus, $\widetilde{\boldsymbol{\Theta}} = \boldsymbol{\Theta}\mathbf{U}$ and $\boldsymbol{\alpha}_i = \mathbf{U}^\top\boldsymbol{\alpha}_i$ give the same representation of **5**. To gain identifiability, we require that (i) $\boldsymbol{\Theta}^\top\boldsymbol{\Theta} = \mathbf{I}$, (ii) $\mathbf{A}^\top\mathbf{A} = \mathbf{D}^2$ is a diagonal matrix, (*iii*) the columns of $\mathbf{A}$ are ordered such that the diagonal elements of $\mathbf{D}^2$ are in strictly descending order, and (*iv*) the first nonzero element of each column of $\boldsymbol{\Theta}$ is positive. With such $\boldsymbol{\Theta}$ and $\mathbf{A}$, if the diagonal elements of $\mathbf{D}$ are all different and we set $\overline{\mathbf{A}} = \mathbf{A}\mathbf{D}^{-1}$ so that $\overline{\mathbf{A}}^\top\overline{\mathbf{A}} = \mathbf{I}$ $\overline{\mathbf{A}}^\top\overline{\mathbf{A}} = \mathbf{I}$, then we have $\boldsymbol{\Theta}\mathbf{A}^\top = \boldsymbol{\Theta}\mathbf{D}\overline{\mathbf{A}}^\top$, which is a uniquely defined singular value decomposition (SVD). The desired identifiability of $(\boldsymbol{\Theta}, \mathbf{A})$ then follows from the uniqueness of the SVD.

3 SIMULATION STUDY

We conducted a simulation study to evaluate the proposed NCSDE method and compare it with noncollective spectral density estimation approaches. First, we describe the simulation setup and, in the next section, we list the competitive noncollective alternatives.

3.1 Simulation setup

In the simulation study, to mimic the real data structure of the EEG signals, we consider the single-channel EEG data described in Section **4.1**. We first select three sets of data from healthy volunteers (A), the epilepsy patients in seizure-free status (D), and the patients with seizure activities (E), respectively. There are 300 EEG signals in total with 100 from each set. We then fit $AR(p)$ models to each of the 300 time series in R using the function spec.ar with a sufficiently large order of *p* fixed at $p = 36$ and obtained the Yule-Walker estimates of the *AR* coefficients.

The *AR*(*p*) model is given by $X_t = \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$, with the associated SDF

$$f(\omega) = \sigma_\varepsilon^2 \left| 1 - \sum_{k=1}^{p} \varphi_k e^{-2\pi i k \omega} \right|^{-2},$$

(15)

where $\sigma_\varepsilon^2 = \mathrm{Var}\,(\varepsilon_t)$ (see the work of von Storch and Zwiers[37] for details).

We denoted the averaged estimates of the 100 sets of the *AR* coefficients for each class by $\hat{\varphi}_A$, $\hat{\varphi}_D$, and $\hat{\varphi}_e$, respectively. In each simulation run, we generated an $n \times m$ matrix $\mathbf{X} = (\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_\mathrm{m})$, where $\mathbf{X}_i^\top = (\mathbf{X}_{i1}, \mathbf{X}_{i2}, \dots, \mathbf{X}_{i\mathrm{n}})$, n is the length of the time series and m is the number of the times series. Each $\mathbf{X}_i$ is generated from one of the following models:

$$\begin{cases} Model\ I & AR(p): \boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}_A, \text{with prob. } p_1, \\ Model\ II & AR(p): \boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}_D, \text{with prob. } p_2, \\ Model\ III & AR(p): \boldsymbol{\varphi} = \hat{\boldsymbol{\varphi}}_E, \text{with prob. } p_3, \end{cases}$$

We set $n = (100,200,400)$, $m = (6,15,30)$, and $p_k = \frac{1}{3}$, for $k = 1,2,3$. Therefore, we had nine different pairs of (n, m) and, for each pair, we ran the simulation $N = 100$ times.

In reality, similar spectral densities within a cluster might exhibit small variability. To incorporate this variability into account, for each *AR* model, we obtained the Cholesky decomposition of the *n* × *n* autocovariance matrix $\boldsymbol{\Gamma} = \boldsymbol{\Lambda}\boldsymbol{\Lambda}^\top$. Then, we reconstructed a perturbed version of the autocovariance matrix, ie, $\tilde{\boldsymbol{\Gamma}} = \tilde{\boldsymbol{\Lambda}}\tilde{\boldsymbol{\Lambda}}^\top$, by adding an $n \times n$ matrix of random errors, ie, $\boldsymbol{E}$, to the Cholesky factor, ie, $\boldsymbol{\Lambda}$, to derive $\tilde{\boldsymbol{\Lambda}}$ (ie,) $\tilde{\boldsymbol{\Lambda}} = \boldsymbol{\Lambda} + \boldsymbol{E}$. We generated ϵ_{i_j} 's, the entries of $\boldsymbol{E}$, independently from normal distribution with zero-mean and $\sigma_e = 0.05$. In each simulation run, for each *AR* model, we used a perturbed autocovariance matrix, ie, $\tilde{\boldsymbol{\Gamma}}$, to simulate the $AR(p)$ signals.

3.2 Competitive approaches

In each simulation run, we used the $n \times m$ data matrix $\mathbf{X}$ as an input to obtain estimates of *m* SDFs ($\hat{\mathbf{f}}$: $\tilde{n} \times m$ matrix) from the following six methods.

- Periodograms (Ps).

  We used the periodogram $\mathbf{I}$ as a rough estimate. For each run, the *i*th column of the $\tilde{n} \times m$ matrix $\mathbf{I}$ was a vector of size $\tilde{n}$ obtained via

  $$I_{i,n}(\omega) = \left| \frac{1}{n} \sum_{t=1}^{n} X_{it} e^{-2\pi i t \omega} \right|^2,$$

  evaluated at the discrete frequencies $\omega = (\omega_1, \dots, \omega_{\tilde{n}})$

- Smoothed periodograms (S.Ps).

The next estimate was obtained by smoothing **I** using the rich set of basis functions, $\mathrm{S.Ps} = \exp[\mathbf{B}(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top \log(\mathbf{I})]$.

- tSVD periodograms (tSVD.Ps).

  We used the truncated SVD to obtain the rank *K* approximation of the smoothed periodograms $(\mathbf{B}^\top\mathbf{B})^{-1}\mathbf{B}^\top \log(\mathbf{I})$, called $\boldsymbol{\Theta_1}, \mathbf{A_1^\top}$. This was referred to in the work of Stewart[38] as the approximation theorem (a.k.a. the Eckart-Young theorem). Therefore, the third estimate, ie, "tSVD Periodograms", is $\mathrm{tSVD.Ps} = \exp[\mathbf{B\Theta_1}, \mathbf{A_1^\top}]$.

- Separate estimations (NSDE).

  We avoided collective estimation and obtained the nonparametric spectral density estimates (NSDEs) by maximizing the Whittle likelihood using the rich family of basis functions **B** *separately*. From **6**, $u_i(\boldsymbol{\omega}) = \boldsymbol{B\Psi}_i$. Therefore, our fourth estimate is $\mathrm{NSDE} = \exp[\mathbf{B\Psi}]$, where $\boldsymbol{\Psi} = (\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_m)$.

- tSVD separate estimations (tSVD.NSDE).

  Having obtained $\boldsymbol{\Psi}$ by maximizing the Whittle likelihood separately, we used the truncated SVD to obtain the rank *K* approximation of $\boldsymbol{\Psi}$ called $\boldsymbol{\Theta_s}\mathbf{A_s^\top}\boldsymbol{\Theta_s}, \mathbf{A_s^\top}$. The fifth estimate is $\mathrm{tSVD.NSDE} = \exp[\mathbf{B\Theta_s}, \mathbf{A_s^\top}]$.

- Collective estimations (NCSDE).

Finally, we used the proposed method from Section **2** to obtain $\mathrm{NCSDE} = \exp[\mathbf{B}\widehat{\boldsymbol{\Theta}}\widehat{\mathbf{A}}^\top]$$\mathrm{NCSDE} = \exp\left[\mathbf{B}\widehat{\boldsymbol{\Theta}}\,\widehat{\mathbf{A}}^\top\right]$..

For comparison purposes, we employed the *Euclidean* distance between the estimated SDFs of Ps, S.Ps, NSDE, and NCSDE, as well as the basis expansion coefficients (score matrix $\mathbf{A}$) of tSVD.Ps($\mathbf{A}_1$), tSVD.NSDE($\mathbf{A}_s$), and NCSDE($\widehat{\mathbf{A}}$) as inputs to the hierarchical clustering algorithm, as described in Section **2.6**.

3.3 Measures of quality

To evaluate the performance of the proposed method in discovering the correct labels (gold standard), in a more systematic framework, we used two popular external measures, ie, adjusted Rand index (ARI) and SIM index in the "TSClust" package, which are commonly used in the clustering evaluation literature.

3.3.1 Adjusted Rand similarity coefficient

The adjusted Rand similarity coefficient (adjusted Rand index) was used to compare the (dis)similarity between two clustering results. It is defined as follows[39]:

$$\mathrm{ARI} = \frac{\sum_{i=0}^{1}\ \sum_{j=0}^{1}\binom{n_{ij}}{2} - \left[\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}\right]/\binom{m}{2}}{\frac{1}{2}\left[\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}\right] - \left[\sum_i \binom{n_{i\cdot}}{2} + \sum_j \binom{n_{\cdot j}}{2}\right]/\binom{m}{2}}.$$

To calculate the adjusted Rand index, we computed the 2×2 contingency table, which consists of the following four cells:

- n_{11}: the number of observation pairs where both observations are comembers in both clusterings;
- n_{10}: the number of observation pairs where the observations are comembers in the first clustering but not the second;
- n_{01}: the number of observation pairs where the observations are comembers in the second clustering but not the first;
- n_{00}: the number of observation pairs where no pairs are comembers in either clustering result.

Furthermore, $n_{i\cdot}$ and $n_{\cdot j}$ are defined as $n_{i0} + n_{i1}$ and $n_{0j} + n_{1j}$, respectively.

3.3.2 SIM index in "TSClust" package

We also used the SIM index, provided in "TSClust" package,[40] for further comparison. The SIM index measures the amount of agreement between the true cluster partition $\mathcal{G} = \{G_1, \cdots . G_k\}$ (the "ground-truth"), assumed to be known, and the experimental cluster solution $\mathcal{A} = \{A_1, \cdots . A_k\}$ obtained by a clustering method. It is defined as follows:

$$\mathrm{SIM}(\mathcal{G}, \mathcal{A}) = \frac{1}{k} \sum_{i=1}^{k} \max_{1 \le j \le k} \mathrm{Sim}(G_i, A_j),$$

where $\mathrm{Sim}(G_i, A_j) = \frac{|G_i \cap A_j|}{|G_i| + |A_j|}$, with $|\cdot|$ denoting the cardinality of the elements in the set.

Note that the ARI and SIM range from 0 to 1, with 0 indicating that the two clusters do not agree on any pairs and 1 indicating that the clusters are exactly the same. In Section **3.4**, we present and compare the clustering results from the approaches given in Section **3.2**, considering the original labels from the simulation as the gold standard. We did not use the class labels when applying the clustering algorithms; we only use the class labels to evaluate the clustering results.

3.4 Simulation results

In each simulation run, we used the approaches given in Section **3.2** for nine different pairs of (*m*,*n*) to obtain matrices of size $\tilde{n} \times m$ that represent the associated estimates of *m* SDFs for each approach. Figure **1**A illustrates how to use the elbow method to obtain the number of clusters ($\tilde{k} = 3$), and the remaining subfigures (Figure **1**B-F) provide the results of the five approaches (S.Ps, tSVD.Ps, NSDE, tSVD.NSDE, and NCSDE) when estimating the SDFs for a randomly selected simulation run with $m = 30$ and $n = 400$. NCSDE obtained the results that are the smoothest and the closest to the true SDFs.

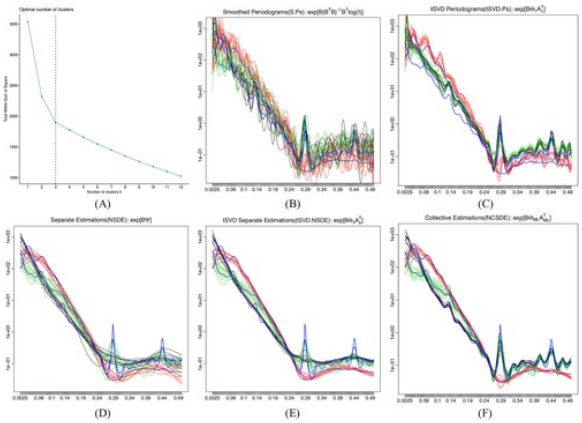

Figure 1 For a randomly selected simulation run with *m* = 30 and *n* = 400. A, The elbow method suggests picking $\tilde{k} = 3$ clusters; B-F, Comparison of results from five different estimates of spectral density function (SDFs) (S.Ps, tSVD.Ps, NSDE, tSVD.NSDE, and NCSDE) vs. true SDFs (blue)

To compare the performance of different techniques in hierarchical clustering, Figure **2** provides the dendrograms of four of the approaches, ie, Ps, tSVD.Ps($\mathbf{A}_1$), NSDE, and NCSDE($\widehat{\mathbf{A}}$), on a randomly selected simulation run with $m = 30$ and $n = 400$. Since we knew the number of clusters and the cluster labels (gold standard) in advance, we cut the dendrograms into the original number of clusters.

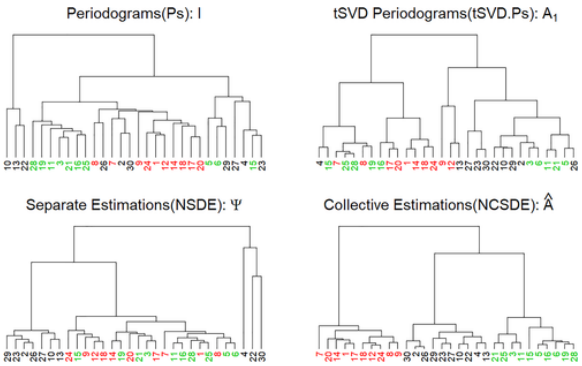

Figure 2 Dendrograms from hierarchical clustering. Comparing four different estimates (Ps, tSVD.Ps($\mathbf{A}_1$), NSDE, and NCSDE($\widehat{\mathbf{A}}$)) for a randomly selected simulation run with $m = 30$ and $n = 400$

A comprehensive comparison on the performance of the clustering results are given in Table **1**. NCSDE($\widehat{\mathbf{A}}$) clearly outperformed the competing approaches in this clustering task, based on both ARI and SIM measures.

However, another advantage of the proposed method is that the coefficients of the basis expansion for the fitted densities ($\mathbf{A}$ matrix) provide a low-dimensional representation that can be directly used to visualize and cluster the spectral densities. Figure **3** shows the scatter plot of the first three coefficients ($\mathbf{A}_{.1}$, $\mathbf{A}_{.2}$, and $\mathbf{A}_{.3}$) for the last two approaches in Section **3.2** (tSVD.NSDE and NCSDE). We observe a clear segregation into three classes using NCSDE, indicating that these coefficients are useful for clustering purposes.

Table 1. Comparison based on 100 simulation runs. The mean and standard errors (in parentheses) of the ARI and SIM are reported

Measure	**ts-*n*-*m***	**Ps**	**tSVD.Ps($\mathbf{A_1}$)**	**NSDE**	**tSVD.NSDE($\mathbf{A_s}$)**	**NCSDE**	**NCSDE($\widehat{\mathbf{A}}$)**
	ts-100-6	0.665(0.013)	0.643(0.012)	0.659(0.012)	0.649(0.012)	0.653(0.012)	**0.743**(0.013)
	ts-100-15	0.586(0.010)	0.562(0.005)	0.599(0.012)	0.481(0.008)	0.511(0.009)	**0.726**(0.009)
	ts-100-30	0.628(0.006)	0.554(0.003)	0.645(0.007)	0.441(0.005)	0.545(0.005)	**0.725**(0.005)
	ts-200-6	0.669(0.012)	0.630(0.010)	0.785(0.013)	0.777(0.013)	0.782(0.013)	**0.795**(0.012)
ARI	ts-200-15	0.644(0.009)	0.562(0.005)	0.736(0.008)	0.623(0.016)	0.641(0.015)	**0.794**(0.008)
	ts-200-30	0.658(0.006)	0.551(0.003)	0.747(0.005)	0.580(0.018)	0.657(0.011)	**0.798**(0.006)
	ts-400-6	0.709(0.012)	0.663(0.012)	0.797(0.011)	0.799(0.011)	0.785(0.011)	**0.867**(0.013)
	ts-400-15	0.670(0.009)	0.576(0.005)	0.788(0.006)	0.763(0.008)	0.772(0.008)	**0.904**(0.009)
	ts-400-30	0.691(0.008)	0.555(0.002)	0.786(0.004)	0.767(0.006)	0.769(0.006)	**0.925**(0.006)
	ts-100-6	0.695(0.012)	0.645(0.013)	0.708(0.010)	0.706(0.009)	0.702(0.009)	**0.755**(0.013)
	ts-100-15	0.544(0.011)	0.493(0.007)	0.556(0.012)	0.456(0.007)	0.476(0.009)	**0.693**(0.011)
	ts-100-30	0.574(0.008)	0.453(0.005)	0.581(0.008)	0.358(0.005)	0.454(0.007)	**0.690**(0.007)
	ts-200-6	0.706(0.011)	0.639(0.010)	0.797(0.013)	0.794(0.012)	0.798(0.012)	**0.801**(0.013)
SIM	ts-200-15	0.606(0.010)	0.501(0.007)	0.676(0.009)	0.578(0.013)	0.597(0.013)	**0.769**(0.011)
	ts-200-30	0.601(0.008)	0.446(0.005)	0.664(0.009)	0.492(0.017)	0.572(0.011)	**0.778**(0.009)
	ts-400-6	0.732(0.011)	0.682(0.012)	0.805(0.011)	0.803(0.012)	0.791(0.012)	**0.872**(0.013)
	ts-400-15	0.643(0.010)	0.521(0.008)	0.746(0.010)	0.717(0.010)	0.726(0.010)	**0.906**(0.009)
	ts-400-30	0.643(0.010)	0.460(0.005)	0.725(0.010)	0.689(0.009)	0.695(0.009)	**0.934**(0.006)

Abbreviations: NCSDE, nonparametric collective spectral density estimation; NSDE, nonparametric spectral density estimate.

It is common to proceed with a two-step procedure to visualize the level of similarity among the SDFs.

i.Obtain the $\tilde{n} \times m$ matrix ($\hat{\mathbf{f}}$) from any of the approaches given; then

ii.Use a multidimensional scaling technique.[41]

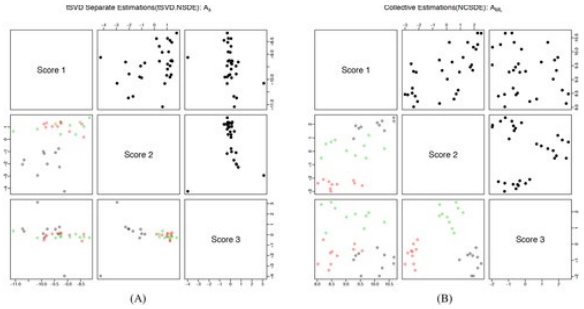

Figure 3A matrix plot that assesses the relationships among several pairs of scores, ie, $\mathbf{A}$, in tSVD.NSDE ($\mathbf{A}_s$) and NCSDE ($\widehat{\mathbf{A}}$), when clustering a randomly selected simulation run with $m = 30$ and $n = 400$

4 EEG DATA APPLICATIONS

We apply our method to two neuroscience studies in clustering of brain signals with electroencephalogram (EEG) data. The EEG is a method of monitoring spontaneous electrical activity in the brain over a period of time. EEGs are typically recorded from single or multiple electrodes, referred as EEG channels, placed on the scalp. The clustered EEG signals are useful for understanding how different diseases may affect the brain activity or how different brain regions communicate with each other.

4.1 Application 1: epilepsy - EEG data

For the first application, the publicly available EEG data in the work of Andrzejak et al[42] is used. The complete data includes five sets (denoted as $\boldsymbol{A}$, $\boldsymbol{B}$, $\boldsymbol{C}$, $\boldsymbol{D}$, and E), each one containing 100 single-channel EEG segments of 23.6-second duration. Each set was selected from EEG records after removing artifacts caused by strong eye and muscle movements. The first two sets ($\boldsymbol{A}$, $\boldsymbol{B}$) consisted of segments from surface EEG recordings taken from five healthy volunteers with eyes open (A) and eyes closed ($\boldsymbol{B}$). The last three sets ($\boldsymbol{C}$, $\boldsymbol{D}$, and $\boldsymbol{E}$) were intracranially taken from five epilepsy patients. Sets $\boldsymbol{C}$ and $\boldsymbol{D}$ are measured in seizure-free intervals from epileptic hemisphere and the opposite hemisphere of the brain, respectively. Finally, the last set ($\boldsymbol{E}$) captured the seizure EEG activity. Orhan et al[43] used this data set to introduce a deep neural network for classification as a diagnostic decision support mechanism in the epilepsy treatment. In this study, we used the proposed NCSDE method to obtain the SDF for different groups of unit variance EEG signals and evaluate the performance of the proposed technique in clustering those signals based on the estimated spectral densities. We should note that we did not use the set labels ($\boldsymbol{A}, \dots, \boldsymbol{E}$) when we applied the clustering algorithms. We only used the set labels for the evaluation of the clustering results. We conducted four different experiments, by varying the level of complexity, to evaluate the performance of the NCSDE in clustering different mixtures of healthy segments, epileptic seizure free segments, and epileptic seizure segments. The details of each experiment is provided here.

1. Easy Task ($\boldsymbol{A} - \boldsymbol{E}$): The aim of this experiment is to cluster the epileptic seizure patients ($\boldsymbol{E}$) and the healthy volunteers ($\boldsymbol{A}$). Therefore, the NCSDEs of 200 EEG signals in sets $\boldsymbol{A}$ and $\boldsymbol{E}$ are obtained.
2. Somewhat-Hard Task ($\boldsymbol{D} - \boldsymbol{E}$): The aim of this experiment is to cluster the epileptic seizure patients in seizure-free intervals ($\boldsymbol{D}$) and the patients with seizure activity ($\boldsymbol{E}$). Here, the NCSDEs of 200 EEG signals in sets $\boldsymbol{D}$ and $\boldsymbol{E}$ is obtained.

3. Hard Task ($\boldsymbol{A}-\boldsymbol{D}$): The aim of this experiment is to cluster the healthy volunteers ($\boldsymbol{A}$) and the epileptic seizure patients in seizure-free intervals ($\boldsymbol{D}$). Here, the NCSDEs of 200 EEG signals in sets $\boldsymbol{A}$ and $\boldsymbol{D}$ is obtained.
4. Challenging Task ($\boldsymbol{A}-\boldsymbol{D}-\boldsymbol{E}$): The aim of this experiment is to cluster the EEG signals of the healthy volunteers ($\boldsymbol{A}$), the epilepsy patients in seizure-free status ($\boldsymbol{D}$), and the patients with seizure activity ($\boldsymbol{E}$).

In each experiment, we used the *Euclidean* distance between the estimates SDFs of NCSDE procedure as well as the score matrix, ie, $\widehat{\mathbf{A}}$, as inputs to the hierarchical clustering algorithm to obtain the associated clusters. The results from the challenging task are presented in Figure 4. For comparison, we used some novel clustering methods implemented in the "TSClust" package.[40] Montero and Vilar provided more than 30 different methods (dissimilarity measures) for clustering the time series in this package. We picked the following four dissimilarity measures that are based on nonparametric spectral estimators.

- LLR.DLS : General spectral dissimilarity measure using the local lineal smoothers of the periodograms, obtained via generalized least squares.
- LLR.LK : General spectral dissimilarity measure using the local-linear estimation of the log-spectra. Here, the estimation is obtained via the maximum local likelihood criterion.
- ISD : Computes the dissimilarity between two time series in terms of the integrated squared difference between nonparametric estimators of their log-spectra.
- GLK : The dissimilarity between two time series is computed by using an adaptation of the generalized likelihood ratio test to check the equality of two log-spectra.

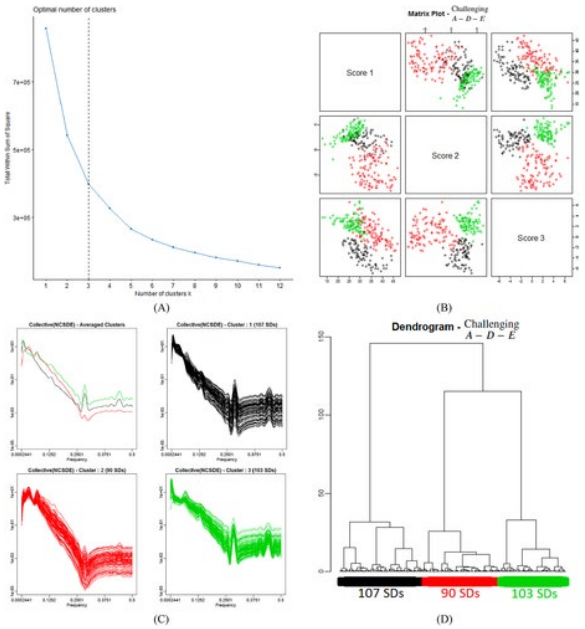

Figure 4 The clustering results from the epilepsy data, the challenging experiment ($\boldsymbol{A}-\boldsymbol{D}-\boldsymbol{E}$). The top left plot ($\boldsymbol{A}$) is the result of the elbow method based on the $\mathbf{log}(\mathbf{S}.\mathbf{Ps})$ and the top right plot ($\boldsymbol{B}$) is the result of matrix plot of the scores ($\widehat{\mathbf{A}}$ matrix). The bottom left plots ($\boldsymbol{C}$) are the estimated spectral density functions (SDFs) using

the nonparametric collective spectral density estimation (NCSDE) method, and the last plot ($\boldsymbol{D}$) is the result of the hierarchical clustering based on the $\widehat{\mathbf{A}}$ matrix

The results of the aforementioned dissimilarity measures are used as inputs to the hierarchical clustering algorithm. For more details about these four techniques, we refer the interested readers to the Montero and Vilar package.[40]

For evaluation purpose, we used the ARI and SIM that are introduced in Section **3.3** as two indices to evaluate the clustering results.

The associated results that compares our proposed approach with the competitive ones are given in Table **2**. In this table, we also presented the running time (TIME) of each method on an Intel(R) Xeon(R) with E5-1660 processor @ 3.3 GHz speed with 64 GB of memory to compare the computational complexity of different approaches.

Table 2. Comparison of four experimental tasks between our proposed method nonparametric collective spectral density estimation (NCSDE) and LLR.DLS, LLR.LK, ISD, GLK based on ARI, SIM, and TIME

Experiment	**Measure**	**NCSDE**	**NCSDE ($\widehat{\mathbf{A}}$)**	**LLR.DLS**	**LLR.LK**	**ISD**	**GLK**
Easy	ARI	0.98	0.98	0.98	0.88	0.98	0.36
A-E	SIM	0.99	0.99	0.99	0.97	0.99	0.80
	TIME(s)	52	52	427	30757	30805	15321
Somewhat hard	ARI	0.49	**0.88**	0.10	0.11	0.11	0.28
A-D	SIM	0.85	**0.97**	0.62	0.63	0.63	0.75
	TIME(s)	28	28	423	31235	31257	15437
Hard	ARI	0.49	**0.85**	0.05	0.38	0.38	0.43
D-E	SIM	0.85	**0.96**	0.59	0.81	0.81	0.83
	TIME(s)	34	34	495	27783	27744	15863
Challenging	ARI	0.37	**0.85**	0.29	0.36	0.36	0.25
A-D-E	SIM	0.68	**0.95**	0.60	0.64	0.63	0.66
	TIME(s)	99	99	732	42814	42842	23585

4.2 Application 2: resting state - EEG data

The EEG data was collected during the resting state from a single subject, a male student randomly selected from the 17 subjects in the experiment, which is described in detail in the work of Wu et al.[44] During the experiment, the EEG signals were recorded from 256 channels on the scalp surface, with a millisecond resolution (1000 recordings per second). From the 256 channels, 62 channels were eliminated due to the high muscle artifact content in EEG electrodes recording from cheek and neck areas.[44] The locations of the channels are shown in Figure **5**. In this application, we cluster the remaining 194 channels using the time series data from the first and the last minutes. The total length of the time series is 60 000. However, we only consider the low-frequency bands (ie, we truncate the periodograms at length 3000).

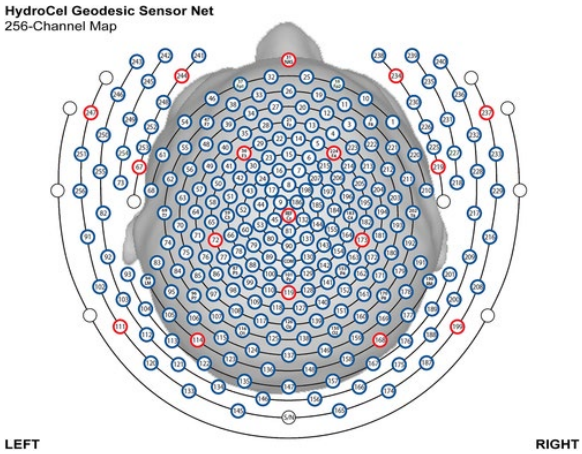

Figure 5 The locations of the 256 channels on the scalp surface

The elbow method (Figures **6**A and **7**A) suggested the presence of four different clusters in both cases, the first and the last minutes. Therefore, we set $\hat{k} = 4$ for both cases and proceeded with the NCSDE iterative procedure to obtain the Whittle maximum likelihood estimator.

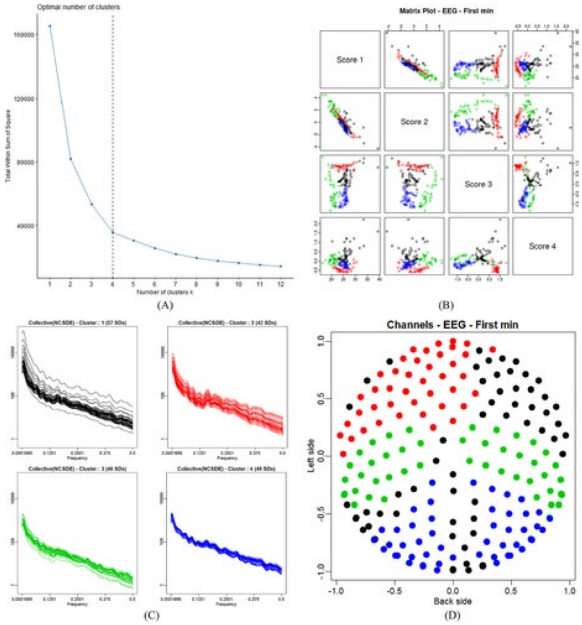

Figure 6 The clustering results from the first minute. The top left plot ($\boldsymbol{A}$) is the result of the elbow method based on the $\mathbf{log}(\mathbf{S}.\mathbf{Ps})$ and the top right plot ($\boldsymbol{B}$) is the result of the hierarchical clustering based on the $\widehat{\mathbf{A}}$ matrix. The bottom left plots ($\boldsymbol{C}$) are the estimated spectral density functions (SDFs) using the nonparametric collective spectral density estimation (NCSDE) method, and the last plot ($\boldsymbol{D}$) is the 2D brain map that illustrates the clustering results from the first minute of data

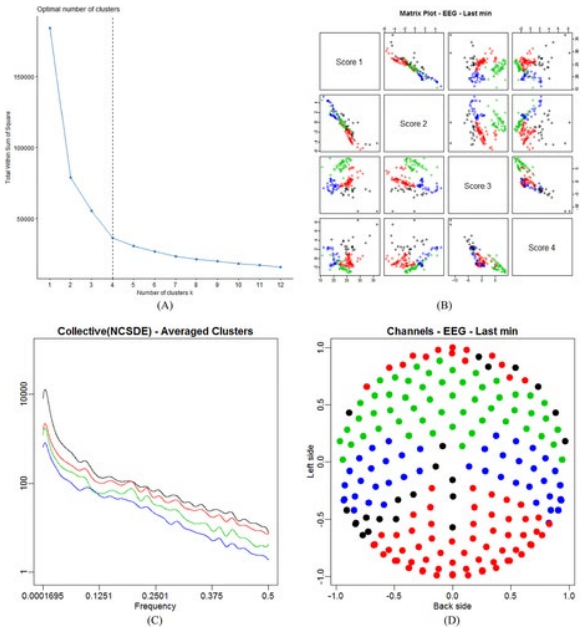

Figure 7 The clustering result from the last minute. The top left plot ($\boldsymbol{A}$) is the result of the elbow method based on the $\mathbf{log}(\mathbf{S}.\mathbf{Ps})$ and the top right plot ($\boldsymbol{B}$) is the result of matrix plot of the scores ($\widehat{\mathbf{A}}$ matrix). The bottom left plot ($\boldsymbol{C}$) is the functional mean of the estimated spectral density functions (SDFs) using the nonparametric collective spectral density estimation (NCSDE) method for each cluster, and the last plot ($\boldsymbol{D}$) is the 2D brain map that illustrates the clustering results from the last minute of data

For the data from the first minute, the algorithm converged after 33 iterations. Figure **6**B provides the result of the hierarchical clustering algorithm based on the estimated coefficients of the basis expansion, ie, the $\widehat{\mathbf{A}}$ matrix, where the associated tree has been cut into $\tilde{k} = 4$ clusters. Out of the 194 channels, 57,42,46, and 49 SDFs were assigned respectively to the four clusters. Figure **6**C shows the 194 estimated SDFs associated with the four different clusters. Figure **6**D presents the clustering results in the 2D brain map.

The results from the last minute of data are presented in Figure **7**. The algorithm converged after 170 iterations, and Figure **7**B provides the hierarchical clustering result based on the estimated $\widehat{\mathbf{A}}$ matrix for the last minute of data. By cutting the tree into $\tilde{k} = 4$ clusters, we determined that 22,75,54, and 43 SDFs should be assigned to the first through fourth clusters, respectively. Figure **7**C shows the functional means of the estimated SDFs in each cluster to illustrate the difference between the four groups. Figure **7**D presents the clustering results in the 2D brain map.

As shown in Figure 6 and Figure 7, we have the following findings.

1. For both datasets (first and last minutes), the elbow method suggested using four clusters, and both of the brain maps provide relatively well-separated, symmetric regions from the clustering results.
2. All of the brain maps give three regions, ie, the front, middle, and back regions. From the functional mean of each cluster, we find that the channels in the middle region of the brain have a lower density in the low-frequency band than the front and back regions.
3. Both data are associated with the resting state; when we compare the results from the first and last minutes, we find that there are no significant differences between them, except in the front

region where we obtain two clusters in first minute (Figure 6D), but just one cluster in the last minute (Figure 7D).

4. The most left-front channels in Figure 7D have a different pattern from the other channels.

5 DISCUSSION

A novel approach for collectively estimating multiple SDFs was developed in this paper. By pooling data from different time series in the frequency domain and using a shared basis to represent the SDFs, the collective estimation approach is statistically more efficient than noncollective estimation approaches. The proposed method uses the penalized Whittle likelihood approximation to yield a flexible family of spectral densities. As an output of the new method, each estimated log-spectral density is expressed in a basis expansion, where the basis is estimated from the data, assuming that the SDFs lie in a low-dimensional manifold of the large space spanned by a prespecified rich basis. The collective spectral density estimation approach is widely applicable when there is a need to estimate multiple SDFs from different populations. Moreover, the coefficients of the basis expansion for the fitted spectral densities provide a concise, low-dimensional representation that could be useful for visualization and clustering. Another advantage of the new procedure is that it speeds up the process by updating the smoothness parameter within the Newton-Raphson iterations and avoids a grid search over the space of the smoothing parameter, ie, λ, which could be very time consuming.

The proposed model could suffer from some limitations. For example, the Whittle estimates from this model are likely to be inconsistent for a collection of nonstationary time series. One may consider using adequate data tapers (eg, Whittle tapered estimate)[45] as an extension of NCSDE and a remedy to deal with nonstationary time series. Another direction of extension is to accommodate the correlation in complex studies, where subunits are nested within units. Potentially, a general framework of functional mixed effects model could be utilized to model the correlation within the subunit functions[46] in the NCSDE. It is also interesting to investigate the impact of using different fixed basis functions (eg, wavelet) and further study the effect of varying number of common basis functions ($\boldsymbol{K}$) on the estimation performance.

A web application that can be used by the research community to reproduce the results in this paper or to estimate SDFs collectively based on NCSDE for any other related applications is available at "**https://ncsde.shinyapps.io/NCSDE**."

ACKNOWLEDGEMENTS

We would like to thank two anonymous referees for their constructive and thoughtful comments, which helped us tremendously in revising the manuscript. The research reported in this publication was supported by funding from King Abdullah University of Science and Technology (KAUST) to Ying Sun and Tianbo Chen.

REFERENCES

1 Yuen CK. On the smoothed periodogram method for spectrum estimation. *Signal Process*. 1979; **1**(1): 83- 86.

2 Wahba G. Automatic smoothing of the log periodogram. *J Am Stat Assoc*. 1980; **75**(369): 122- 132.

3 Brockwell PJ, Davis RA. *Time Series: Theory and Methods*. New York, NY: Springer Science & Business Media; 2013.
4 Brillinger DR. *Time Series: Data Analysis and Theory*. Philadelphia, PA: SIAM; 2001.
5 Lee TCM. A simple span selector for periodogram smoothing. *Biometrika*. 1997; **84**(4): 965- 969.
6 Ombao H, Raz JA, Strawderman RL, von Sachs R. A simple generalised crossvalidation method of span selection for periodogram smoothing. *Biometrika*. 2001; **88**(4): 1186- 1192.
7 Lee TCM. A stabilized bandwidth selection method for kernel smoothing of the periodogram. *Signal Process*. 2001; **81**(2): 419- 430.
8 Iannaccone R, Coles S. Semiparametric models and inference for biomedical time series with extra-variation. *Biostatistics*. 2001; **2**(3): 261- 276.
9 Freyermuth J-M, Ombao H, von Sachs R. Tree-structured wavelet estimation in a mixed effects model for spectra of replicated time series. *J Am Stat Assoc*. 2010; **105**(490): 634- 646.
10 Krafty RT, Hall M, Guo W. Functional mixed effects spectral analysis. *Biometrika*. 2011; **98**(3): 583- 598.
11 Chau J, von Sachs R. Functional mixed effects wavelet estimation for spectra of replicated time series. *Electron J Statist*. 2016; **10**(2): 2461- 2510.
12 Capon J. Maximum-likelihood spectral estimation. In: *Nonlinear Methods of Spectral Analysis*. Berlin, Germany: Springer; 1983: 155- 179.
13 Güler I, Kiymik MK, Akin M, Alkan A. AR spectral analysis of EEG signals by using maximum likelihood estimation. *Comput Biol Med*. 2001; **31**(6): 441- 450.
14 Whittle P. Estimation and information in stationary time series. *Arkiv för Matematik*. 1953; **2**(5): 423- 434.
15 Whittle P. Some recent contributions to the theory of stationary processes. In: *A Study in the Analysis of Stationary Time Series*. vol. 2. Uppsala, Sweden: Almqvist & Wiksells Boktrycheri; 1954: 196- 228.
16 Whittle P. Gaussian estimation in stationary time series. *Bull Int Stat Inst*. 1962; **39**: 105- 129.
17Pawitan Y, O'Sullivan F. Nonparametric spectral density estimation using penalized Whittle likelihood. *J Am Stat Assoc*. 1994; **89**(426): 600- 610.
18 Chow Y-S, Grenander U. A sieve method for the spectral density. *Ann Stat*. 1985; **13**(3): 998- 1010.
19 Fan J, Kreutzberger E. Automatic local smoothing for spectral density estimation. *Scand J Stat*. 1998; **25**(2): 359- 369.
20 Pawitan Y. Automatic estimation of the cross-spectrum of a bivariate time series. *Biometrika*. 1996; **83**(2): 419- 432.
21 Dai M, Guo W. Multivariate spectral analysis using Cholesky decomposition. *Biometrika*. 2004; **91**(3): 629- 643.
22 Rosen O, Stoffer DS. Automatic estimation of multivariate spectra via smoothing splines. *Biometrika*. 2007; **94**(2): 335- 345.
23 Krafty RT, Collinge WO. Penalized multivariate Whittle likelihood for power spectrum estimation. *Biometrika*. 2013; **100**(2): 447- 458.
24 Maadooliat M, Zhou L, Najibi SM, Gao X, Huang JZ. Collective estimation of multiple bivariate density functions with application to angular-sampling-based protein loop modeling. *J Am Stat Assoc*. 2016; **111**(513): 43- 56.
25 Whittle P. On stationary processes in the plane. *Biometrika*. 1954; **41**(3/4): 434- 449.
26 Green PJ, Silverman BW. *Nonparametric Regression and Generalized Linear Models: A Roughness Penalty Approach*. Boca Raton, FL: Chapman & Hall/CRC; 1994.

27 Akaike H. A new look at the statistical model identification. *IEEE Trans Autom Control*. 1974; **19**(6): 716- 723.
28 O'Sullivan F. Fast computation of fully automated log-density and log-hazard estimators. *SIAM J Sci Stat Comput*. 1988; **9**(2): 363- 379.
29 Gu C. *Smoothing Spline ANOVA Models*: Springer; 2002. Springer Series in Statistics.
30 Schall R. Estimation in generalized linear models with random effects. *Biometrika*. 1991; **78**(4): 719- 727.
31 Schellhase C, Kauermann G. Density estimation and comparison with a penalized mixture approach. *Comput Stat*. 2012; **27**(4): 757- 777.
32 Najibi SM, Maadooliat M, Zhou L, Huang JZ, Huang J. Protein structure classification and loop modeling using multiple Ramachandran distributions. *Comput Struct Biotechnol J*. 2017; **15**: 243- 254.
33 Eilers PHC, Marx BD. Flexible smoothing with B-splines and penalties. *Stat Sci*. 1996; **11**(2): 89- 102.
34 RC Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2016.
35 Thorndike RL. Who belongs in the family? *Psychometrika*. 1953; **18**(4): 267- 276.
36 Kassambara A, Mundt F. Package 'factoextra': Extract and Visualize the Results of Multivariate Data Analyses. 2016. http://www.sthda.com/english/rpkgs/factoextra/
37 vonStorch H, Zwiers FW. *Statistical Analysis in Climate Research*. Cambridge, Uk: Cambridge University Press; 2001.
38 Stewart GW. On the early history of the singular value decomposition. *SIAM Rev*. 1993; **35**(4): 551- 566.
39 Vinh NX, Epps J, Bailey J. Information theoretic measures for clusterings comparison: Is a correction for chance necessary? In: Proceedings of the 26th Annual International Conference on Machine Learning (ICML'09); 2009; New York, NY.
40 Montero P, Vilar JA. TSclust: an R package for time series clustering. *J Stat Softw*. 2014; **62**(1): 1- 43.
41 Borg I, Groenen PJF. *Modern Multidimensional Scaling: Theory and Applications*. 2nd ed. New York, NY: Springer; 2005. Springer Series in Statistics.
42 Andrzejak RG, Lehnertz K, Mormann F, Rieke C, David P, Elger CE. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: dependence on recording region and brain state. *Phys Rev E*. 2001;061907: 64.
43 Orhan U, Hekim M, Ozer M. EEG signals classification using the K-means clustering and a multilayer perceptron neural network model. *Expert Syst Appl*. 2011; **38**(10): 13475- 13481.
44 Wu J, Srinivasan R, Kaur A, Cramer SC. Resting-state cortical connectivity predicts motor skill acquisition. *Neuroimage*. 2014; **91**: 84- 90.
45 Velasco C, Robinson PM. Whittle pseudo-maximum likelihood estimation for nonstationary time series. *J Am Stat Assoc*. 2000; **95**(452): 1229- 1243.
46 Zhou L, Huang JZ, Martinez JG, Maity A, Baladandayuthapani V, Carroll RJ. Reduced rank mixed effects models for spatially correlated hierarchical functional data. *J Am Stat Assoc*. 2010; **105**(489): 390- 400.