



Modeling and analyzing neural signals with phase variability using Fisher-Rao registration

Weilong Zhao^a, Zishen Xu^a, Wen Li^b, Wei Wu^{a,*}

^a Department of Statistics, Florida State University, 117 N Woodward Ave., Tallahassee, FL 32306-4330, USA

^b Department of Psychology, Florida State University, 1107 W. Call St., Tallahassee, FL 32306-4301, USA

ARTICLE INFO

Keywords:

Fisher-Rao registration
Time warping
Phase variability
Dynamic time warping
fMRI
EEG

ABSTRACT

Background: The dynamic time warping (DTW) has recently been introduced to analyze neural signals such as EEG and fMRI where phase variability plays an important role in the data.

New method: In this study, we propose to adopt a more powerful method, referred to as the Fisher-Rao Registration (FRR), to study the phase variability.

Comparison with existing methods: We systematically compare FRR with DTW in three aspects: (1) basic framework, (2) mathematical properties, and (3) computational efficiency.

Results: We show that FRR has superior performance in all these aspects and the advantages are well illustrated with simulation examples.

Conclusions: We then apply the FRR method to two real experimental recordings – one fMRI and one EEG data set. It is found the FRR method properly removes the phase variability in each set. Finally, we use the FRR framework to examine brain networks in these two data sets and the result demonstrates the effectiveness of the new method.

1. Introduction

Temporal phase lags have been commonly observed in neural signal recordings (Vinck et al., 2011; Williams et al., 2020). These lags may be introduced by the dynamic switching of brain states and exhibit linear or non-linear patterns across the time domain (Allen et al., 2014; Chen et al., 2015). To make meaningful inference on the recordings, it is often a prerequisite to measure and remove those phase lags. For example, the study of functional connectivity focuses on identifying statistical interdependencies between time series recording from different brain areas (Friston et al., 1997; Tononi et al., 1998). One major goal is usually to quantify the strength of phase synchronization, which constitutes a significant physiological mechanism for functional integration (Singer and Gray, 1995; Fries, 2005; Singer, 1999; Varela et al., 2001). Various mathematical measures, such as phase-locking statistics (Lachaux et al., 1999), mutual information (Hurtado, 2004), and partial directed coherence (Kamiński et al., 2001), have been applied to quantify the synchronization. However, all these methods have clear limitations. For example, the phase-locking statistics is used to detect stationary, non-zero phase lags between two signals, whereas phase differences between different brain areas may vary dynamically, even with a single

strong source. Similar limitation happens for the partial directed coherence method which is only appropriate for stationary and linear models.

In general, neurophysiological recordings often suffer from delay-induced bias. This includes those with high temporal resolutions such as electroencephalogram (EEG) and magnetoencephalography (MEG), and those with high spatial resolutions such as functional magnetic resonance imaging (fMRI). Indeed, small hemodynamic response delays or slice-timing differences may cause mismatches of data. To estimate the temporal latency, conventional approaches assume the delay-induced bias is stationary and shift signal curves within time scales of seconds to minutes. The estimated temporal delay is the time shift when strongest statistical interdependencies are obtained (Mørup et al., 2008). An increasing number of studies on fMRI (Chang and Glover, 2010; Sakoğlu et al., 2010; Kiviniemi et al., 2011; Handwerker et al., 2012; Jones et al., 2012; Smith, 2012; Allen et al., 2014), near-infrared spectroscopic and MEG (de Pasquale et al., 2010; Brookes et al., 2011; Keilholz, 2014; Li et al., 2015) indicate that functional connectivity reveals dynamic changes in human cognition process.

To capture a high-quality dynamic temporal-lag structure, various conventional approaches such as sliding-window analysis, wavelet

* Corresponding author.

transform coherence, and spontaneous co-activation patterns analysis, have been introduced to deal with this problem (Liu and Duyn, 2013; Chen et al., 2015). Traditional statistical methods such as linear and piecewise-linear models are also exploited (Williams et al., 2020). In particular, the dynamic time warping (DTW), a classical algorithm for measuring similarity between two temporal sequences (Sakoe and Chiba, 1978), has been a preferred method in recent studies of alignment on neural signals such as EEG (Huang and Jansen, 1985; Gupta et al., 1996; Karamzadeh et al., 2013), fMRI (Dinov et al., 2016; Meszlényi et al., 2017), and neuronal spike trains (Lawlor et al., 2018). Such temporal warping method offers a novel and enchanting insight to deal with common temporal-lag problems. The key idea in DTW is to perform a dynamic time warping on discrete sequence data to correct the non-stationary time lags. This method is immune to the common noise components in the data and can provide accurate estimate of the underlying templates such as the hemodynamic response functions (Meszlényi et al., 2017). Indeed, it has become one of the most commonly used tools for handling phase synchronization problems.

Nonetheless, the DTW method has apparent disadvantages: (1) It is a matching method between two sequences of temporal points. The matching process on points are not one-to-one, and therefore the sequence lengths change after the alignment (cannot be pre-determined beforehand). (2) The matching does not preserve feature in the given sequences and often leads to the pinching effect problem (Srivastava et al., 2011). (3) DTW is not a metric-based method, and can only do pairwise comparison. There is no principled way to conduct further statistical analysis (e.g. no notion of mean or covariance). (4) There is only pairwise comparison in the DTW method. It is very inefficient to use the method in practice. (5) Last but not least, DTW measures the distance after the phase-lag is removed, and therefore this distance can be referred to as “amplitude distance”. However, the method does not provide a measurement on how much warping is done. That is, it does not provide a measurement on the “phase distance” between two signals. The detailed discussion of these disadvantages is given in Section 2.3.

To overcome these problems in DTW, we propose to adopt a new framework, referred to as Fisher-Rao registration (FRR), to deal with the dynamic latency between signals. Treating signals as functions on a given time domain, FRR (Srivastava et al., 2011) offers a unified, comprehensive solution to conduct alignment (or registration) between two functions. Similar to DTW, the alignment in FRR is done with the notion of time warping, albeit as a continuous operation within the given domain. Any strictly increasing, nonlinear warping transformation is taken into account in this fixed domain. More importantly, FRR is a metric-based method, and can naturally introduce the notion of mean and covariance on the given signals and produce more powerful statistical analysis. It can also provide both “amplitude distance” and “phase distance” between two signals. We will provide a systematical comparison between FRR and DTW and demonstrate the superiority of FRR in this paper. We expect the FRR framework will turn out to be a useful method for neural signal processing where phase variability is a significant factor. The computational analysis in this study was implemented using Matlab (The Mathworks, Inc.), and scripts on examples can be accessed on GitHub.¹

The rest of this article is organized as follows. In Section 2, we will provide details of the FRR method and compare it with the DTW method in terms of basic framework, mathematical properties, and computational efficiency. In Section 3, we will illustrate the use of FRR method in real experimental data. We show that this method can effectively remove phase variability in given functional signals and such noise removal can be used to build more appropriate brain networks. Section 4 summaries the results and offers conclusions. Computational details on DTW and FRR are given in Appendices.

2. Methods

2.1. Review of the dynamic time warping method

The dynamic time warping (DTW) (Sakoe and Chiba, 1978) is a finite, discrete, algorithm-based method that can match two curves, in the form of two finite-length sequences of sampled points, to correct the phase difference. The main idea is to minimize the distance between the two sequences through copying and matching the points. To achieve this, one needs to select one sequence as the base sequence, while the other one as the operating sequence, and define a sliding window. For each point in the operating sequence, we can go over all the points in the sliding window on the base sequence to compute the cost value. It is computed as the distance between the starting point to the current point on the operating sequence and the starting point to the matched point on the base sequence. In this way, after all cost values for the last point on the operating sequence are obtained, the minimum distance between the two sequences is defined as the cost of matching the last point on the operating sequence to the last point on the base sequence. The implementation of DTW follows a shortest-path-finding idea. Suppose $\{u(t)\}_{t=1}^{l_u}$ and $\{v(t)\}_{t=1}^{l_v}$ are two sequences with integer time index t and the sliding window has window size to be w , the cost C_{ij} of matching $u(i)$ to $v(j)$, where $i = 1, 2, \dots, l_u$ and $j = \max\{1, i - w\}, \max\{1, i - w\} + 1, \dots, \min\{l_v, i + w\}$, is defined as:

$$C_{ij} = \begin{cases} d[u(1), v(1)] & i = 1, j = 1 \\ d[u(1), v(j)] + C_{1,j-1} & i = 1, j > 1 \\ d[u(i), v(1)] + C_{i-1,1} & i > 1, j = 1 \\ d[u(i), v(j)] + \min\{C_{ij}, C_{i-1,j}, C_{i,j-1}, C_{i-1,j-1}\} & i > 1, j > 1. \end{cases}$$

The $d(x, y)$ in the formula is used to measure the squared distance between two points x and y . For example, when x and y are real numbers, a commonly used d will be the squared Euclidean distance: $d(x, y) = (x - y)^2$. To compute the cost for matching $u(i)$ to $v(j)$, one needs to compute three adjacent “past” cost values $C_{i-1,j}$, $C_{i,j-1}$ and $C_{i-1,j-1}$ first. In this way, the last cost $C_{i,j}$ is obtained by an iterative way (loop over i , while for each i , loop over j). In each iteration, cost is computed and the three “past” costs given this cost are stored. Therefore, one can trace back from $C_{i,j}$ to get all previous matched costs one by one and figure out the shortest-path. The algorithm of the DTW method is given in detail in Appendix A.

A simple illustration of the DTW method is provided in Fig. 1. Suppose there are two sequences of points which are sampled from two curves. Sequence 1 has 9 sampled points and Sequence 2 has 11 sampled points (see Fig. 1(a)). The corresponding horizontal axis values are set to be just the index integers. The metric used is $d(x, y) = (x - y)^2$. For simplicity, no sliding window is applied, i.e. $w = 11$. We need to note that the number of sampled points is different for the two curves and the aligned curves have more points than the original. The optimal path is shown in the matrix in Fig. 1(b).

We note that in the framework of DTW method, the point matching is not one-to-one (see Fig. 1(c)), which may corrupt global features in the original curves. The one-to-many matching also results in duplicates of some points and stretches the original curves by introducing more sampled points, shown in Fig. 1(d). Moreover, since the DTW method is algorithm-based, it lacks solid mathematical background. One can only use finite, discrete numerical approximation to match two curves under the DTW method.

2.2. The Fisher-Rao registration method

The Fisher-Rao Registration (FRR) (Srivastava et al., 2011) is a nonparametric approach to conduct registration (or alignment) between two functions f and g on a time interval such as $[0, 1]$. The FRR method examines functions in the space $\mathcal{F} = \{f : [0, 1] \rightarrow \mathbb{R} \mid f \text{ is absolutely continuous}\}$. The alignment is represented by a time

¹ <https://github.com/WeilongZ/FRR-vs-DTW.git>

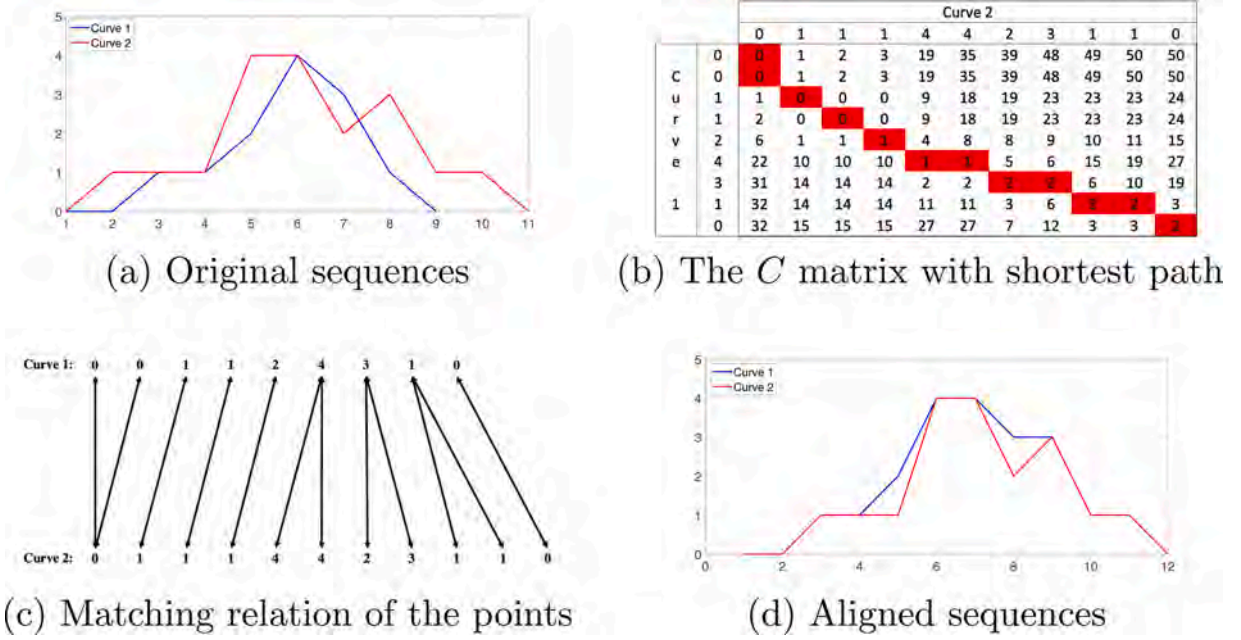


Fig. 1. An illustration of the DTW method to match two curves in the form of two finite-length sequences of sampled points. (a) The original sequences. (b) The cost matrix C where the shortest path has been marked in red. (c) The matching relation based on the shortest path. (d) The aligned sequences. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

warping function in the set $\Gamma = \{\gamma : [0, 1] \rightarrow [0, 1] | \gamma(0) = 0, \gamma(1) = 1, \dot{\gamma} > 0\}$, where $\dot{\gamma}$ denotes the derivative function of γ . The set Γ is a group with identity $\gamma_{id}(t) = t, t \in [0, 1]$. The main idea of FRR is to find the optimal time warping function such that the extended Fisher-Rao distance between the two curves after warping is minimized, i.e.

$$\min_{\gamma \in \Gamma} d_{FR}(f \circ \gamma, g).$$

To achieve this, the concept of squared root velocity function (SRVF) is introduced:

Definition 1. For a function $f \in \mathcal{F}$, its SRVF is $q_f(t) = \text{sign}[\dot{f}(t)] \sqrt{|\dot{f}(t)|}$.

Given the SRVF function $q(t)$ and function value at time 0, one can reconstruct the original function f by: $f(t) = f(0) + \int_0^t |q_f(x)| q_f(x) dx$. SRVF satisfies two important properties which build the registration framework:

Property 1. The Fisher-Rao distance between two functions is the same as the \mathbb{L}^2 distance between their SRVFs, i.e. $d_{FR}(f, g) = \|q_f - q_g\|$;

Property 2. For any function f and a time warping function γ , the SRVF of $f \circ \gamma$ is:

$$q_{f \circ \gamma} = (q_f \circ \gamma) \sqrt{\dot{\gamma}}.$$

Based on the above two properties, the optimal time warping function between two functions f and g can be computed as:

$$\hat{\gamma} = \arg \min_{\gamma \in \Gamma} d_{FR}(f \circ \gamma, g) = \arg \min_{\gamma \in \Gamma} \| (q_f \circ \gamma) \sqrt{\dot{\gamma}} - q_g \| . \quad (1)$$

Computation of $\hat{\gamma}$ in Eq. (1) can be efficiently done using a dynamic programming method (Srivastava et al., 2011). Once the optimal warping function $\hat{\gamma}$ is known, there is no phase difference between $f \circ \hat{\gamma}$ and g . The amplitude distance between f and g is given as:

$$d_{AMP}(f, g) = d_{FR}(f \circ \hat{\gamma}, g) = \| (q_f \circ \hat{\gamma}) \sqrt{\dot{\hat{\gamma}}} - q_g \| .$$

It is easy to verify that this amplitude distance is a semi-metric. That is, it satisfies nonnegativity, symmetry, and triangle inequality. In addition to using $\hat{\gamma}$ to measure the amplitude distance between f and g , we can use it to estimate the phase distance between them. Mathematically, we can

measure the difference between $\hat{\gamma}$ and γ_{id} . Based on the Fisher-Rao metric, the phase distance is computed as:

$$d_{PHA}(f, g) = \cos^{-1} \int_0^1 \sqrt{\dot{\hat{\gamma}}(t)} dt.$$

Note that the phase distance is in fact the arc length of γ_{id} and $\hat{\gamma}$ in their SRVF representations in the Hilbert unit sphere \mathbb{S}^∞ . The distance is in the range $[0, \pi/2]$.

In addition to the pairwise alignment for two functions, the FRR framework can also align a collection of functions. This is based on the fact that the amplitude distance is a semi-metric. In this case, we can define the notion of ‘‘mean’’ (or template) of the collection of functions and then align each function to the template. We briefly write down the estimation algorithm for the template of a collection of functions in Appendix B. The detailed procedure is given in Srivastava et al. (2011).

2.3. Advantages of FRR over DTW

In this subsection, we will systematically compare the FRR and DTW methods and demonstrate advantages of FRR. The comparison focuses on basic framework, mathematical properties, and computational efficiency.

2.3.1. Basic framework

As mentioned in Section 2.1, the DTW method is a point-to-point, finite, discrete method. In contrast, the FRR method is a function-based, continuous method. This difference makes the FRR method superior to the DTW method in three aspects: domain for alignment, unit invariance and feature preservation.

- Domain for alignment: The DTW method allows one-to-many and many-to-one mappings in order to obtain the minimal overall cost. Moreover, this method only focuses on the number of the sampled points, but not the exact time locations of these points. These properties often make the output of DTW method difficult to use, because the number of sample points will be changed in a non-systematical way after each alignment (the final number of points is case-by-

case in each alignment and it cannot be known until after the alignment). In contrast, by considering the points as discrete observations on a curve, the FRR method naturally takes interpolation methods to make the number of points invariant during the alignment process. This difference is illustrated using the following example.

As shown in Fig. 2(a), two curves: $f_1(t) = \sin(\pi t)$ and $f_2(t) = 3 \sin(\pi t)$ on $[0, 1]$ are sampled at $t = 0, \frac{1}{9}, \frac{2}{9}, \dots, 1$ (10 points overall). It is clear that these two curves are already well aligned, so the ideal warping function between them should be the identity $\gamma_{id}(t) = t$. This is what we obtain when the FRR alignment method is used and the result is shown in Fig. 2(b). In contrast, when the DTW method is applied on the 10 discrete points of the two curves, duplicate points are generated to minimize the cost. This is shown in Fig. 2(c). The number of sampled points increases from 10 to 14 after the alignment and the result is apparently not satisfactory. More importantly, there is no clear time locations for these 14 points in the warped curves.

- **Unit invariance:** Standardization and normalization are two common tools in data preprocessing. Both of them are done by adding constants and multiplying constants on all the data. They both aim at removing the impact of *units* or magnitudes (e.g. temperature in Celsius and Fahrenheit). For curve alignments, a desirable property will be that the matching relation should be invariant with respect to the unit. In other words, translation (adding constants to all curve points) or scaling (multiplying constants to all curve points) should not influence the alignment outcomes. For the DTW method, this property does not hold. The DTW alignment only considers point-to-point distances, so the shortest path will be different after translation or scaling. In contrast, it has been proven in Srivastava et al. (2011) that the optimal time warping between two functions is invariant with respect to the translation and scale. We will use the following examples in Figs. 3 and 4 to illustrate this result.

Basically, the original curves are $f_1(x) = 2 \sin(2\pi(x - 1))$ and $f_2(x) = 1.5 \sin(2\pi\sqrt{x} + 0.5)$ for $x \in [0, 1]$ with 100 uniform sampling points, which are shown in Fig. 3(a). We at first align the two functions using the DTW method. The result and shortest path are shown in Fig. 3(b) and (c), respectively. Similarly, we use the FRR method to match f_1 to f_2 . The alignment result and optimal warping function are shown in Fig. 3(d) and (e), respectively. For comparative purposes, we standardize each original curve by subtracting the mean and dividing the standard deviation (by treating 100 discrete points as a sample). Same as in the original curves, we apply both DTW and FRR methods and the results are shown in Fig. 4. According to Panels (b) and (c) in the two figures, we can see that the shortest paths in the DTW method are very different for the original and the standardized curves. The alignment on the original curves is not as good as the alignment on the standardized curves in terms of the matching on peaks and valleys. This shows that for the DTW method, translation and scaling will have impact on the result. In contrast, From Panels (d) and (e) in the two figures, we can see that the

optimal warping functions in the FRR method are identical and the alignment results are consistent. This well illustrates the invariance of the FRR alignment with respect to translation and scaling.

- **Feature preservation:** In the DTW algorithm, the goal is to find the minimal sum of squared distance on the discrete points. The method may distort the given curves' shapes for such purpose. This problem is well known as the *pinching effect* (Srivastava et al., 2011). This effect can corrupt the feature of the original curves by an improper alignment. On the contrary, the FRR method is based on a proper metric in the function space and can preserve features in the given functions. This comparison is illustrated with the following example.

In Fig. 5(a), we generate two curves $f_1(t) = 5 \sin(4\pi t^3)$ and $f_2(t) = \sin(4\pi t)$ on $[0, 1]$. The sampling points for f_1 and f_2 are both at $t = 0, \frac{1}{500}, \frac{2}{500}, \dots, 1$ (501 points overall). It is observed that both curves have two peaks and two valleys, so an ideal mapping is supposed to result in peak-to-peak and valley-to-valley alignment. The alignment result using the DTW method is shown in Fig. 5(b). We can see that DTW aligns the peak and valley, but destroys the original shapes entirely. On the contrary, the matching for FRR method not only aligns the peaks and valleys, but also keeps as much the original shapes as possible. This result is well demonstrated in Fig. 5(c) and (d).

2.3.2. Mathematical properties

As we have pointed out, DTW is a procedure-based method on discrete sequences. The cost function does not present any metric distance between two curves. In contrast, FRR follows a rigorous mathematical framework. The optimal distance between two curves naturally defines an amplitude metric and a phase metric. Based on these properties, we can introduce the notion of mean and covariance, and produce more powerful statistical analysis tools such as functional principal component analysis.

- **Metric property:** As mentioned in the previous sections, the FRR method is derived from the Fisher-Rao distance. As a result, its framework is metric-based and the alignment is a matching of two curves. In contrast, the DTW method aims at minimizing the cumulative point-to-point distance and is not metric-based. One advantage of the metric-based method is that the matching relation can be well captured using the warping functions. We can learn how one curve is warped to another in the given time domain and measure the degree of warping in the alignment process (using the phase distance d_{PHA}). On the contrary, the non-metric-based DTW can only obtain the point-to-point relation, and the mapping will change both of the curves to achieve the points' alignment. We will use Fig. 5 again to illustrate this difference.

Basically, Fig. 5(a) shows the original curves f_1 and f_2 , where the peaks and valleys present at different time locations. By the DTW algorithm, mapping from f_1 to f_2 has the same result as that from f_1 to f_2 . This alignment result is shown in Fig. 5(b). In this case, both sequences change length after the alignment (increasing from 501 to

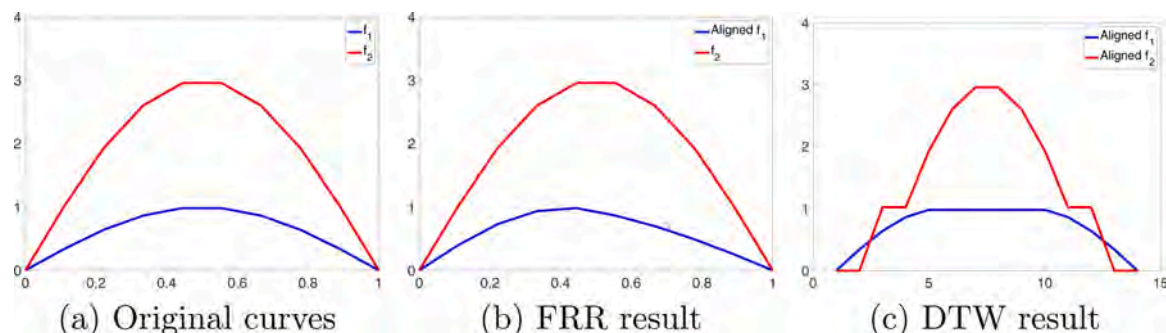


Fig. 2. Illustration on the properness of the alignment. (a) Given two curves. (b) Alignment result using the FRR method. (c) Alignment result using the DTW method.

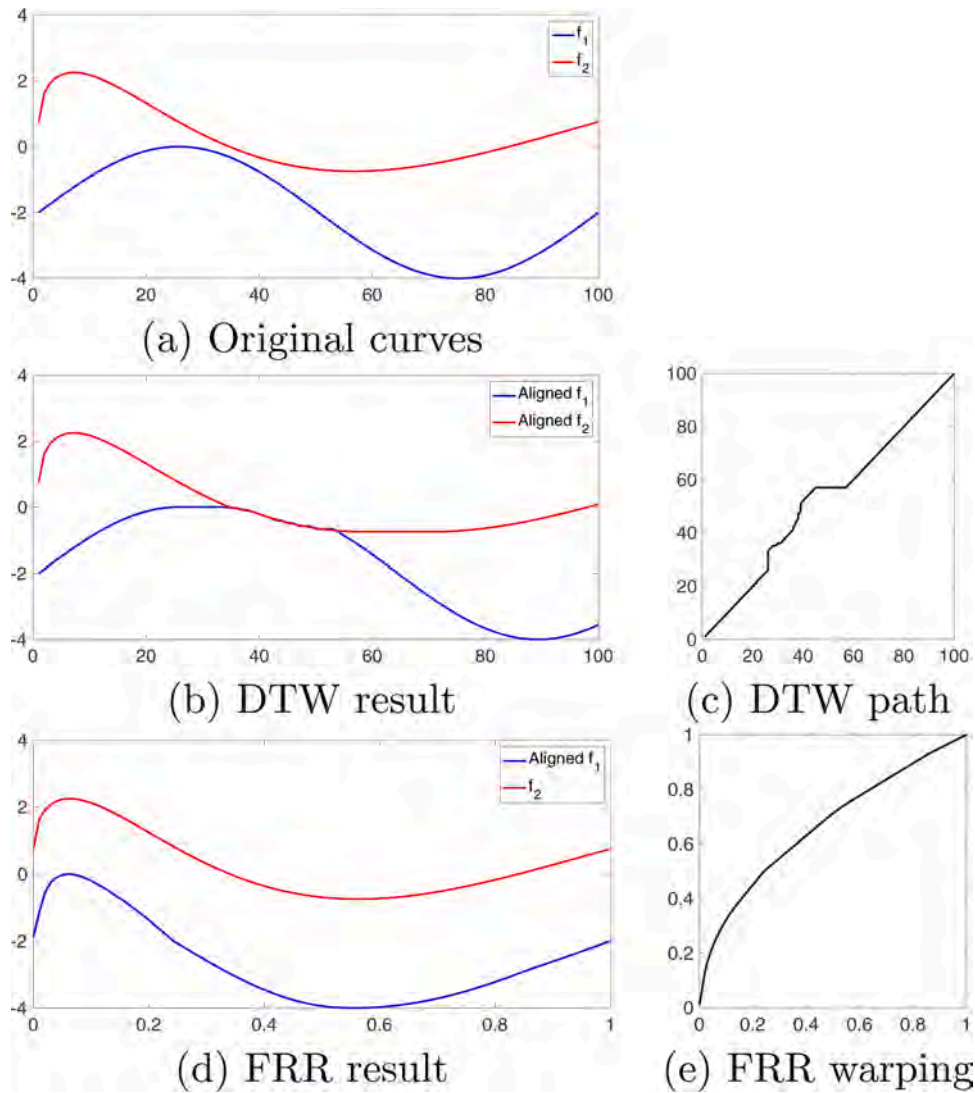


Fig. 3. Illustration of the unit invariance property: alignment on original curves. (a) The given original curves. (b) The alignment result by the DTW method. (c) The shortest path in the DTW method. (d) The alignment of f_1 to match f_2 by the FRR method. (e) The optimal warping function in the FRR method.

900). Using Eq. (1) in the FRR method, we can estimate the optimal warping function from one function to an objective function where the objective one remains unchanged. This is shown in Fig. 5(c) and (d). We can see that matching f_1 to f_2 and vice versa are two opposite transformations.

- Statistical analysis on multiple observations: When dealing with multiple-curve alignment, the metric-based FRR method can result in more powerful analysis tools. Based on the metric distance, we can naturally define the mean of a set of functions $\{f_i\}_{i=1}^n$ as

$$\bar{f} = \operatorname{argmin}_f \sum_{i=1}^n d_{\text{AMP}}^2(f_i, f). \quad (2)$$

As the objective function can remain invariant, we can align every function to the mean (or the mapping template), and then all functions can be aligned altogether. Higher order-statistics such as covariance can also be computed using the aligned curves. Moreover, using the phase distance, we can measure the degree of phase change of each curve from the template. Further analysis tools can be adopted on the amplitude and phase result from the FRR method. For example, one can do functional principle component analysis (fPCA) to learn the detailed structure of the given curves (Tucker et al.,

2013). One can also do functional analysis of variance (fANOVA, Zhang, 2013) to compare different groups of curves. All these analytical methods are not available to the DTW method because it is not a metric-based method, where the pairwise alignment is the only result one can get.

We here use one example to illustrate the FRR-based functional PCA. Let $f_i(t) = A_i \sin(2\pi t^{\alpha_i})$, $i = 1, 2, \dots, 10$, $t \in [0, 1]$, where A_i and α_i are independently and uniformly distributed with $A_i \sim U(1, 5)$ and $\alpha_i \sim U(1, 3)$. All the ten curves are shown in Fig. 6(a). It is apparent that A_i determines the amplitude variation and α_i determines the phase variation for each curve. We at first examine functional PCA on the given 10 curves and the result is shown in Fig. 6(d). We show the mean curve, $\text{mean} \pm \sqrt{\text{first eigenvalue}} \times (\text{first eigenfunction})$ and $\text{mean} \pm 2\sqrt{\text{first eigenvalue}} \times (\text{first eigenfunction})$. It is found that the first principle component obtained by applying functional PCA directly only explains around 65% of the total variance. That is, the first principal direction can only partially capture the variability in the given data.

As a comparison, we conduct phase-amplitude separation for the original curves under the FRR framework, and then do functional PCA on amplitude components and phase components separately. The alignment results on aligned curves and time warping functions are shown in Fig. 6(b) and (c), respectively. Here, the aligned curves

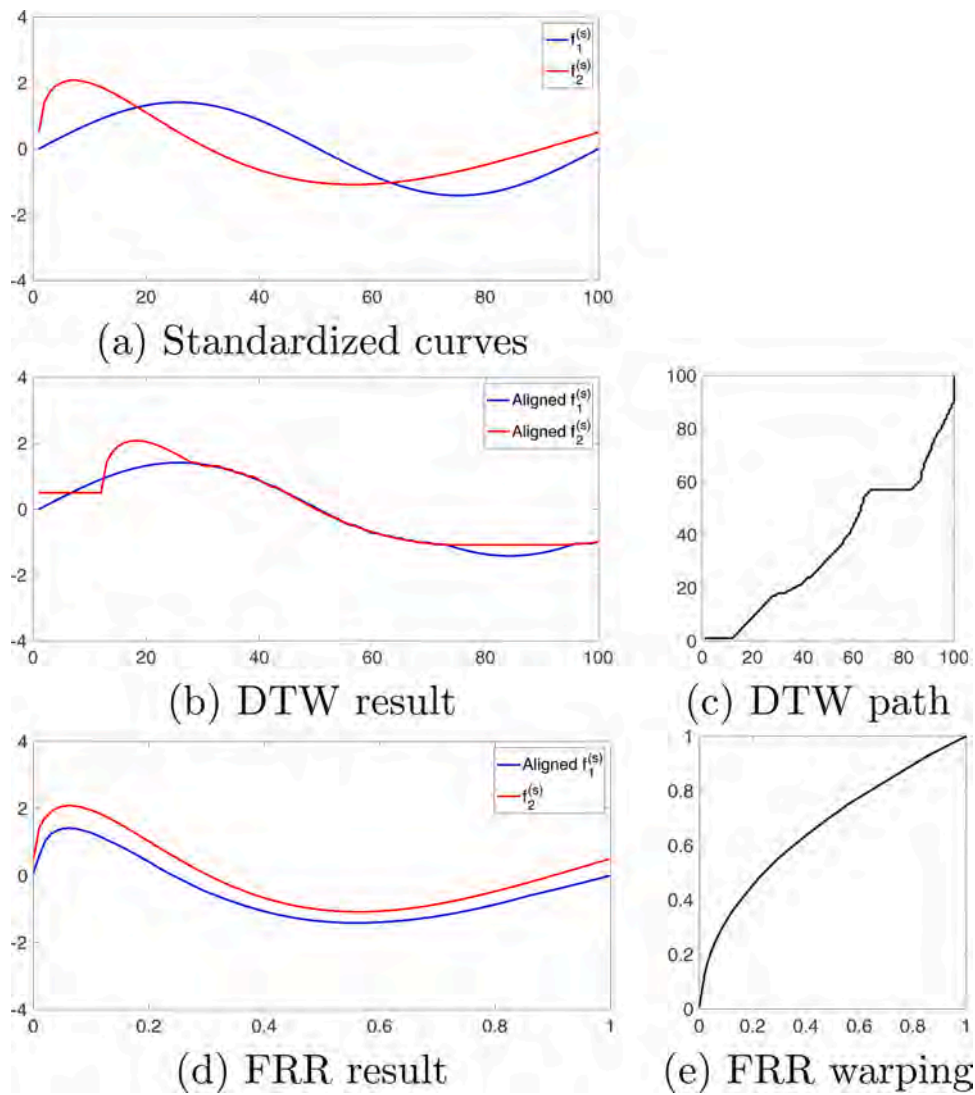


Fig. 4. Illustration of the unit invariance property: alignment on standardized curves. (a) The given curves after standardization (zero mean, unit variance). (b) The alignment result by the DTW method. (c) The shortest path in the DTW method. (d) The alignment of f_1 to match f_2 by the FRR method. (e) The optimal warping function in the FRR method.

represent the amplitude components and the warping functions represent the phase components. Functional PCA on aligned curves and warping functions are shown in Fig. 6(e) and (f), respectively. It is found that the first principle component for aligned curves explains about 99% of total amplitude variance and the first principle component for warping functions explains about 97% of total phase variance. That is, the variability in amplitude and phase can nearly entirely captured by the first principal direction, respectively. These results show that for functional observations with phase variability the FRR method on amplitude and phase component provides a better structure representation than that on the original curves. In contrast, the DTW method is not able to do any modeling analysis.

2.3.3. Computational cost

As mentioned in Section 2.2, $\hat{\gamma} = \operatorname{argmin}_{\gamma \in \Gamma} \| (q_f \circ \gamma) \sqrt{\hat{\gamma}} - q_g \|$ can be numerically solved using a dynamic programming procedure. This is the same framework as DTW, and therefore for pairwise alignment FRR and DTW have the same computational efficiency. However, as FRR can naturally introduce summary statistics, it will have more efficiency for group classification in large data sets.

For example, assume we have two groups of curves $\{\alpha_i(t)\}_{i=1}^n$ and $\{\beta_j(t)\}_{j=1}^n$, where the main difference between these two groups is the

amplitude (the phase variability is treated as noise). We then observe unlabeled new curves $\{f_k(t)\}_{k=1}^n$ and our goal is to classify these new curves to one of the two classes using the nearest neighborhood (NN) method.

For the DTW method, the minimum cost between each pair of function is naturally taken as the “distance” in the NN method. For each new curve, we need to calculate such DTW distance between it and every curve in $\{\alpha_i(t)\}_{i=1}^n$ and $\{\beta_j(t)\}_{j=1}^n$. These pairwise computations are in the order $\mathcal{O}(n)$. Therefore, the total cost for all testing curves are in the order $\mathcal{O}(n^2)$.

In contrast, using the FRR method, the *mean* curve of each group can be efficiently estimated in the computational order $\mathcal{O}(n)$ (Srivastava et al., 2011). Assume the means are proper representatives in their groups, respectively. For each new observation, we only need to compute the FRR distance to the two means, and then assign its label based on the smaller distance. Therefore, for all n test curves, the total computational cost is in the order of $\mathcal{O}(n)$ only. This is apparently a significant advantage over the DTW method.

Finally, Table 1 provides a summary of all above comparisons of DTW and FRR.

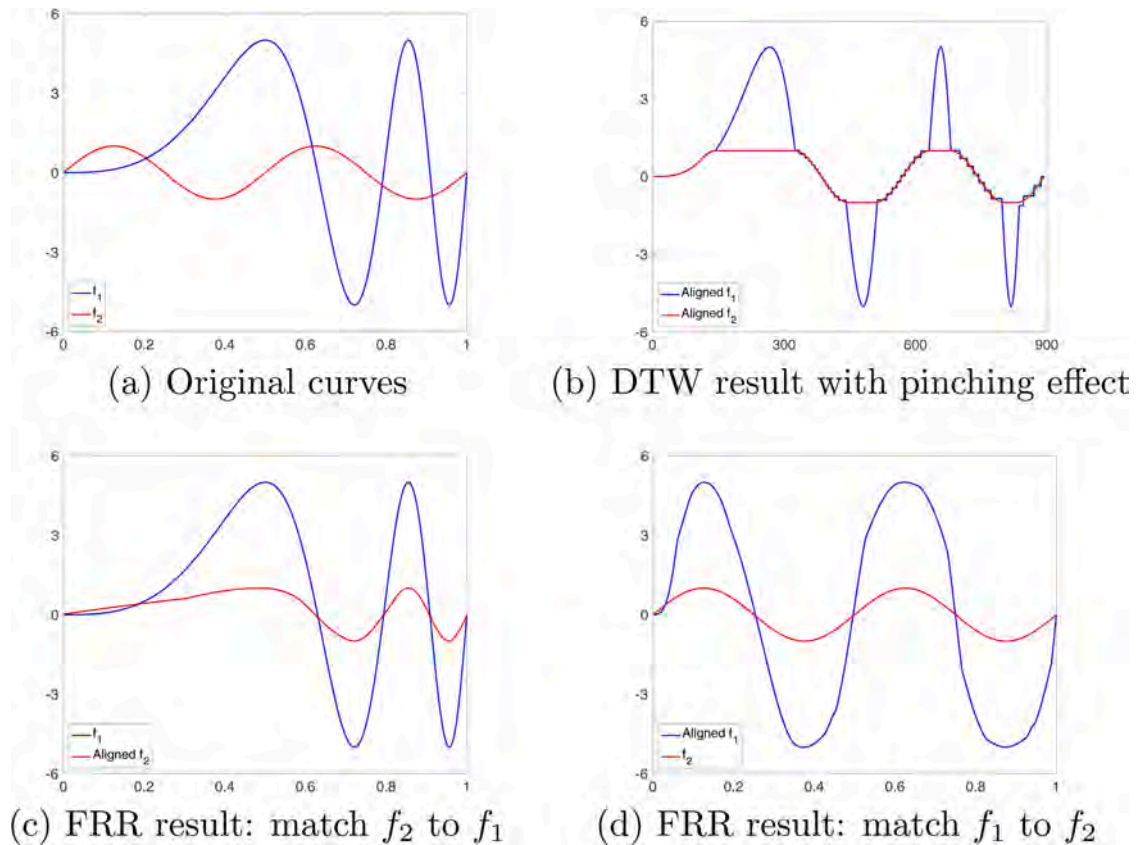


Fig. 5. Illustration on the feature preservation. (a) Given two curves f_1 and f_2 . (b) Alignment of f_1 and f_2 using the DTW method with a clear pinching effect. (c) Alignment of f_2 to f_1 using the FRR method. (d) Alignment of f_1 to f_2 using the FRR method.

3. Application in neural signal analysis

In this section, we apply the FRR method to neural signal analysis on an fMRI and an EEG data set. We will illustrate the FRR method using sample signals, and then use the FRR method to examine brain networks.

3.1. fMRI and EEG data

We will at first describe the two datasets we use in this study.

3.1.1. Data description

The fMRI data set was taken from Addiction Connectome Pre-processed Initiative.² In particular, we use a subset of rest-fMRI data in the project of Multimodal Treatment of Attention Deficit Hyperactivity Disorder (ADHD), which is pre-processed by the National Institute on Drug Abuse. The ADHD subset we use consists of 60 subjects, with 30 non-marijuana users and 30 marijuana users. The 6-min resting-state fMRI were band-pass filtered (0.01–0.1 Hz) and 116 brain regions-of-interest (ROIs) are drawn using the anatomical labeling atlas (Cradock et al., 2012), which includes precentral gyrus, olfactory cortex, hippocampus and other areas. The pre-processing procedure results in 180 total time points at 0.5 Hz temporal resolution for each subject. Fig. 7 shows preprocessed resting-state fMRI functional time series of two example ROIs in the marijuana-user group.

The EEG data were recorded at the Florida State University as a part of a prospective study (Clancy et al., 2018), where participants undertook four days of transcranial alternating current stimulation (tACS) at

the alpha frequency (8–12 Hz), and resting-state EEG (2 min long) was recorded immediately before (baseline), immediately after, and 30 min after tACS on Day 1 and Day 4. The raw data was collected from 32 scalp electrodes of 33 subjects' brains at 250 Hz temporal resolution. The scalp map for this experiment is shown in Fig. 8(a). The baseline EEG recording before tACS on Day 1 and Day 4 are taken in this study. The EEG signals were then bandpass (8–12 Hz) filtered to contain alpha oscillations. EEG data signals in two example electrodes (AF3 and AF7) on Day 1 are shown in Fig. 8(b) and (c).

3.1.2. Illustration of alignment

For illustrative purposes, we compare the alignment results by the FRR and DTW methods using samples from the fMRI and EEG data sets described above. In particular, we select two typical time series from all 30 marijuana-users from the same ROI (left precentral) in the fMRI set. Similarly, we select two typical time series from Electrode 1 (AF7) of the Day 1 group in the EEG set.

The alignment result using the FRR method is shown in Fig. 9. In Panel (a), we show two original fMRI signals with time length 3 min, denoted as f_1 to f_2 . These two signals have apparent nonlinear phase lag between them. We then adopt the FRR framework to conduct alignment. The aligned result is shown in Panel (b), where we align f_2 to match f_1 . We can see that all peaks and valleys are well matched after the alignment process. The optimal warping function is shown in Panel (c), which is close to the identical time warping function $\gamma_{id}(t) = t$. Similar alignment on a pair of EEG signals with a length of 1 s is shown in Panels (d), (e), and (f). We can see that the FRR method can also align peaks and valleys in the signals.

In contrast, we evaluate the performance of the DTW method on the same fMRI and EEG signals and the result is shown in Fig. 10. Note that the DTW method is constructed via aggregating point-wise deviations

² http://fcon_1000.projects.nitrc.org/indi/ACPI/html/index.html

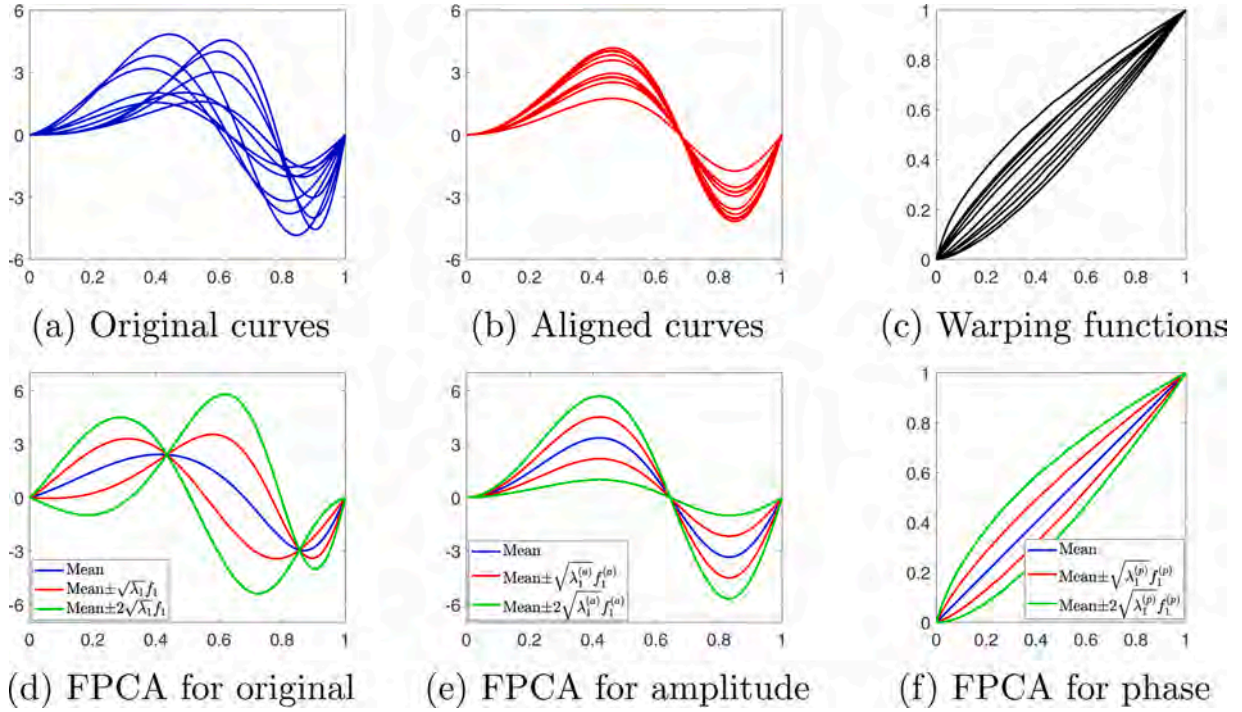


Fig. 6. Illustration on fPCA with FRR framework. (a) 10 simulated curves; (b) aligned curves obtained via the FRR method; (c) time warping functions obtained via the FRR method. (d) FPCA result on the original curves: blue line represents the mean curve, red curves represent mean $\pm \sqrt{\lambda_1}f_1$, and red curves represent mean $\pm 2\sqrt{\lambda_1}f_1$, where λ_1 denotes the first eigenvalue and f_1 denotes the first eigenfunction in the original data; (e) same as (d) except for the aligned curves, where $\lambda_1^{(a)}$ denotes the first eigenvalue and $f_1^{(a)}$ denotes the first eigenfunction in the aligned curves. (f) Same as (d) except for the warping functions, where $\lambda_1^{(p)}$ denotes the first eigenvalue and $f_1^{(p)}$ denotes the first eigenfunction in the warping functions. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 1
Summary of comparisons between DTW and FRR.

Method		DTW	FRR
Basic framework	Domain	Discrete sequence, length varied after alignment	Continuous curve, length-invariant
	Unit	Sensitive to translation and scaling	Invariant to translation and scaling
	Features	Features distorted (pinching effect)	Features well preserved
Mathematical properties	Metric property	Not metric-based	Metric-based
	In-depth analysis	Pairwise distance only, no statistical analysis	Summary statistics, fPCA, fANOVA, etc.
Computational cost	Pairwise alignment	$\mathcal{O}(m^2)$ (m : # of discrete points)	$\mathcal{O}(m^2)$
	Group classification	$\mathcal{O}(n^2)$ (n : sample size)	$\mathcal{O}(n)$

for discrete data, and the length of discrete observations are not consistent before and after DTW “alignment”. In Panel (a), we show the observed sequences with 90 discrete sampling points. The aligned result is shown in Panel (b), where pinching effect is clearly observed and the number of points increase to 120. Instead of aligning peaks and valleys, a lot of flat segments are generated which corrupt the basic features in the original data. This pinching effect is supported by the shortest path in Panel (c). We see several horizontal and vertical segments which indicate one-to-many and many-to-one matchings. Similar alignment on

a pair of EEG signals with 250 sampling points is shown in Panels (d), (e), and (f). We can also see the pinching effect in the alignment result.

3.2. Brain network via the FRR method

Recent study on brain network has been focusing on functional methods, where neural signals are treated as real-valued functions over a continuous time domain. A graph-based method, called functional additive semi-graphoid (FASG) model, was developed to build network using fMRI data from different brain areas (Li and Solea, 2018). This functional graphical model is based on the notion of Additive Conditional Independence (ACI) (Li et al., 2014), where random functions are assumed to reside in two level Hilbert spaces – $\{\mathcal{H}_i\}_{i=1,\dots,p}$ as the first level and a set of $\{\mathcal{M}_i\}_{i=1,\dots,p}$ RKHS as the second Hilbert spaces, where p denotes the number of vertices in the graph. These two level spaces are connected by a set of positive definite mapping $\{\kappa_i(\cdot, \cdot)\}_{i=1,\dots,p}$ from $\mathcal{H}_i \times \mathcal{H}_i \rightarrow \mathbb{R}$ as kernel functions.

$$X^i \perp\!\!\!\perp_A X^j | X^{-(i,j)}, \quad \forall (i,j) \notin E,$$

where E is edge set, and $\perp\!\!\!\perp_A$ represents the ACI relationship. X is said to follow an FASG if the above relationship holds. One of the key benefits of this method is its nonlinear, non-Gaussian property. This is more general and powerful than conventional Gaussian-based methods (Meinshausen et al., 2006; Qiao et al., 2018). In practice, the mapping kernel is often taken in the following form:

$$\kappa_i(f, g) = \exp\{-\gamma_i \|f - g\|_{L^2}^2\},$$

where $\gamma_i > 0$ is a scale parameter and the norm $\|\cdot\|_{L^2}^2$ denotes the Euclidean L^2 distance. In this study, we introduce FRR as a pre-processing procedure to remove the phase variability among one area

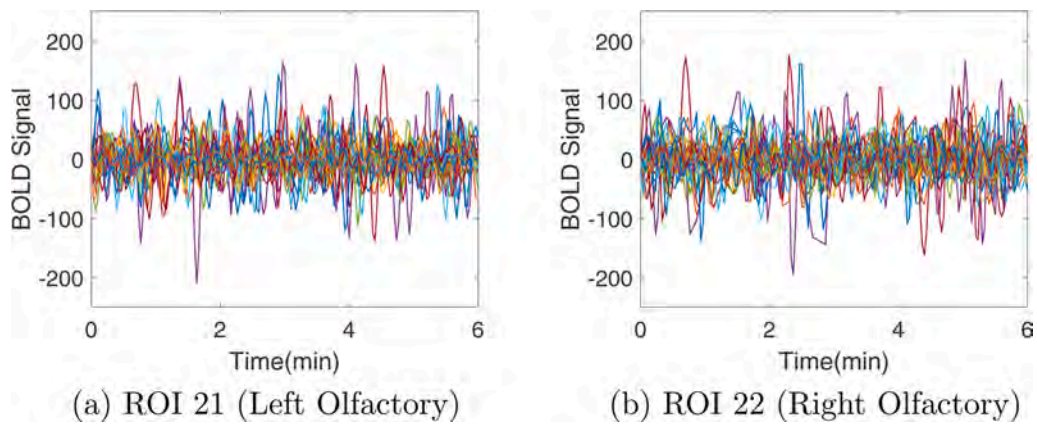


Fig. 7. fMRI data in two example brain regions of 30 marijuana users. Every user is represented by a curve where different users have different curve colors. (a) Left olfactory. (b) Right olfactory. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

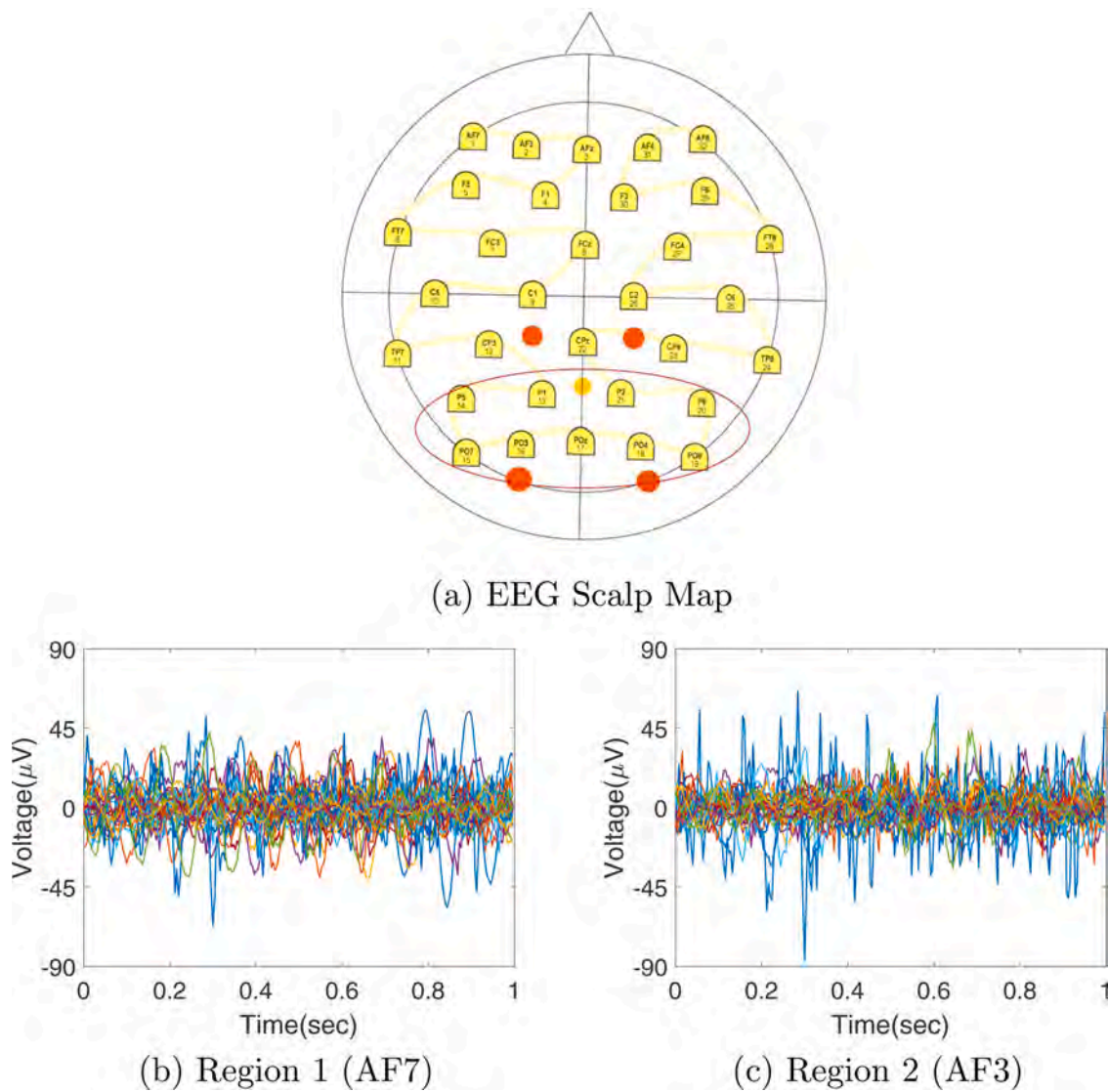


Fig. 8. EEG data. (a) EEG Scalp Map (Clancy et al., 2018). (b) Data in the brain region AF7 of 33 subjects. Every user is represented by a curve where different users have different curve colors. (c) Same as (b) except for region AF3. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

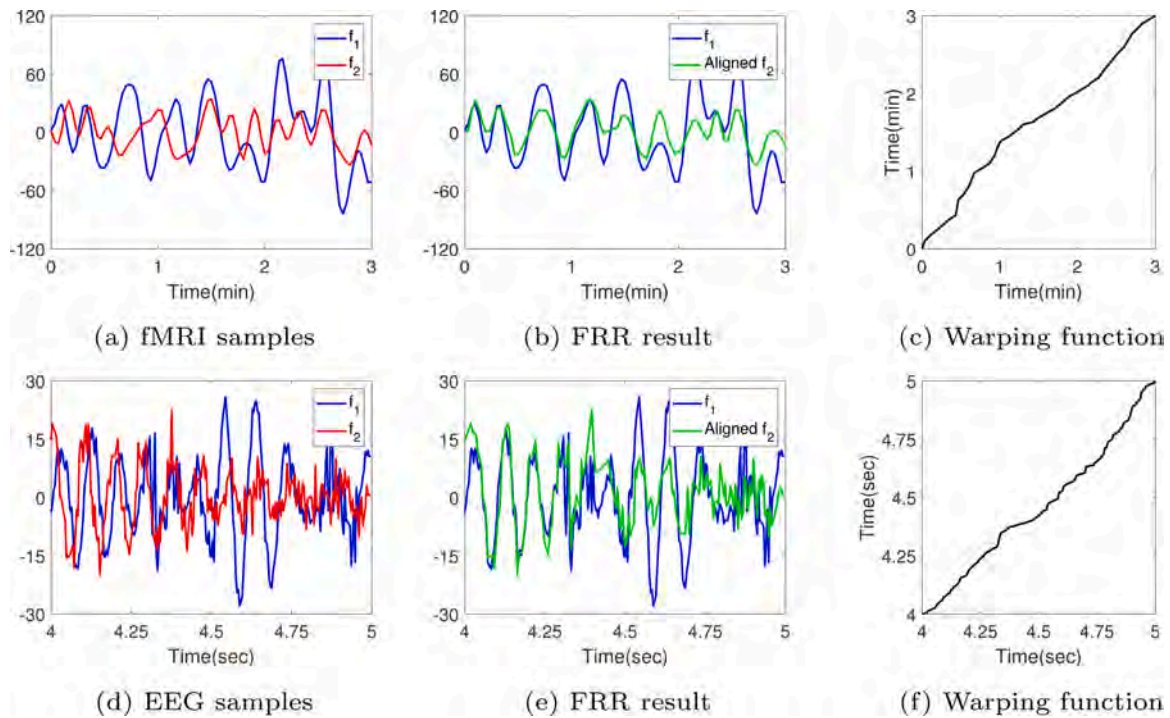


Fig. 9. Illustration of the FRR method on real data. (a) One pair of fMRI signals f_1 and f_2 ; (b) f_1 and the aligned $f_2 \circ \gamma$; (c) optimal time warping function; (d), (e), (f) Same as (a), (b) and (c) except for a pair of EEG signals.

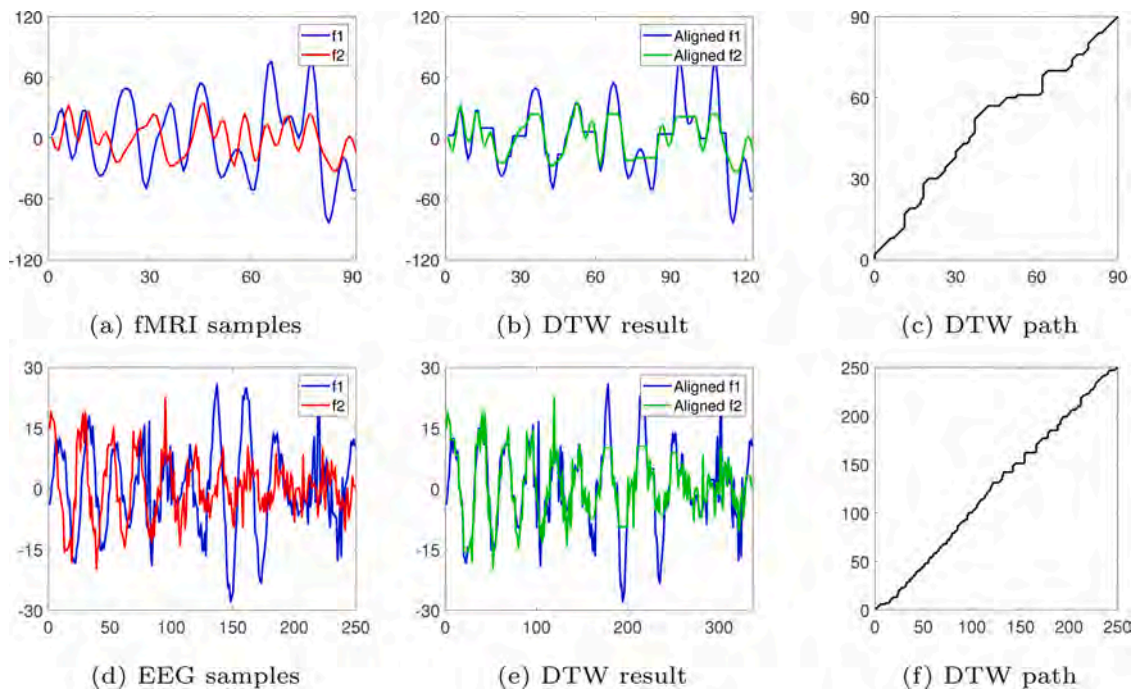


Fig. 10. Illustration of the DTW method on real data. (a) One pair of fMRI signals f_1 and f_2 , represented with 90 discrete sampling points. These signals are identical to those in Fig. 9(a); (b) aligned signals via DTW; (c) shortest path for the optimal DTW match; (d) one pair of EEG signals f_1 and f_2 , represented with 250 discrete sampling points. These signals are identical to those in Fig. 9(d); (e), (f) same as (b) and (c) except for the EEG signals.

and then construct the graph with the aforementioned setting. We will compare the network results of three methods: (1) Euclidean distance on the original data, (2) the DTW distance, and 3) Euclidean distance on the aligned data via FRR.

3.2.1. Network analysis – fMRI data

We now apply the FASG model under three settings to the fMRI data

set, described in Section 3.1.1. With 116 ROIs, there are 116 nodes in the network. The number of edges in a fully connected network is $116 \times 115/2 = 6670$. In this study, we construct brain networks by connecting edges with top 5% strongest additive dependence in the FASG method. That is, we will have $6670 \times 0.05 = 334$ edges in our estimated network.

In Fig. 11, the constructed networks are shown for two groups by

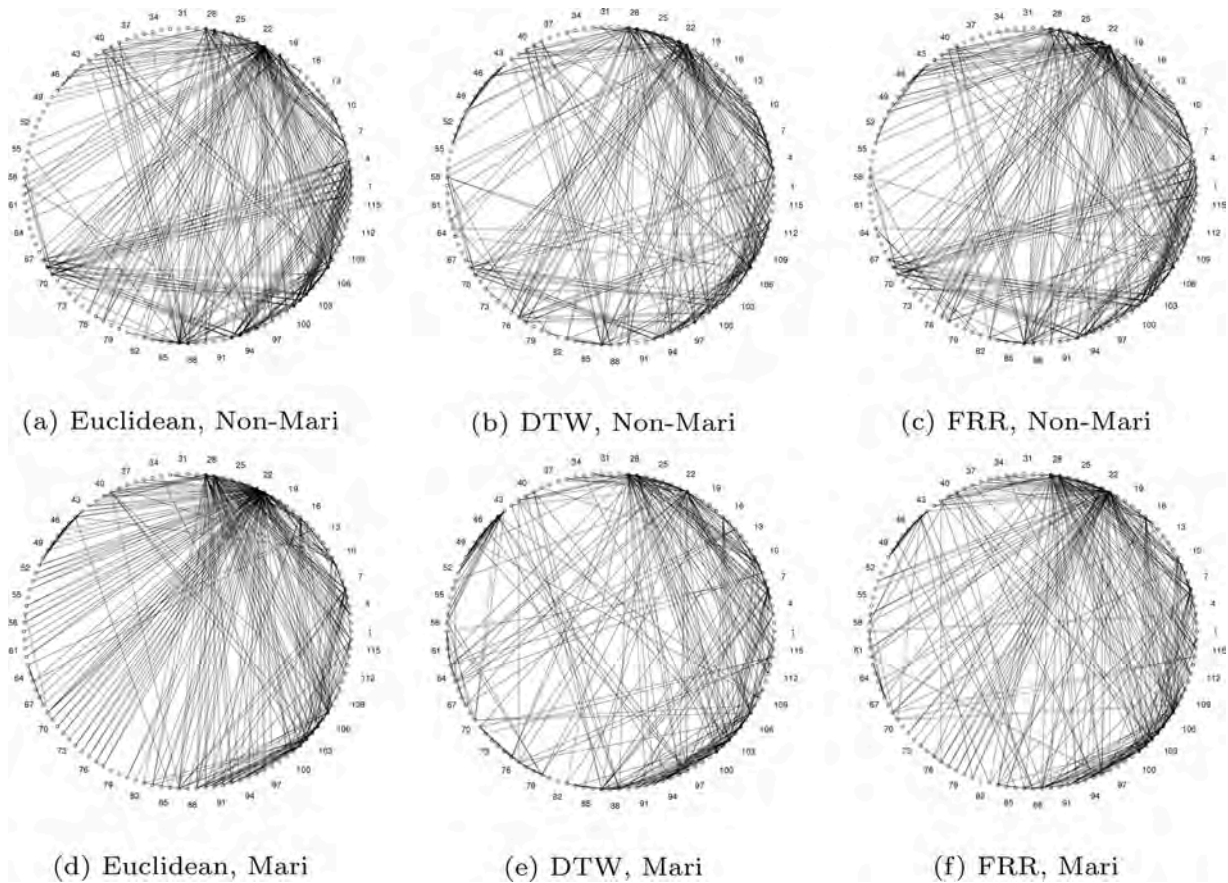


Fig. 11. Brain networks constructed by the FASG model on the fMRI data. (a) Network obtained by Euclidean distance for the non-marijuana group; (b) network obtained by DTW distance for the non-marijuana group; (c) network obtained by Euclidean distance after FRR alignment for the non-marijuana group; (d), (e), (f) same as (a), (b), (c) except for the marijuana group.

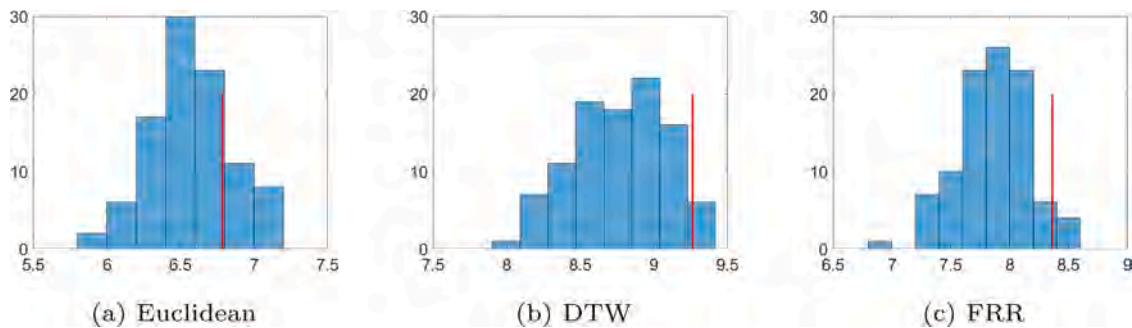


Fig. 12. Histogram for covariance operator difference $R_s = \|\hat{\Sigma}_{XX}^{(1)}(s) - \hat{\Sigma}_{XX}^{(2)}(s)\|$, for $s = 1, \dots, 100$ randomly split samples and the position of R_0 for true sample under different settings, where the red line indicates the position of R_0 : (a) Euclidean – R_0 is larger than 79% of $\{R_s\}_{s=1}^{100}$; (b) DTW – 96%; (c) FRR – 97%.

Table 2
Vertices with more than 20 edges.

	Euclidean	DTW	FRR
Non-marijuana	21, 22, 87	21, 22, 27, 28	21, 22, 27, 28, 87
Marijuana	21, 22, 27, 28	21, 27, 28, 103	21, 22, 27, 28, 105, 107

Boldface indicates the three identified cerebellum regions.

three settings, namely, {marijuana users, non-marijuana users} × {Euclidean distance, DTW distance, Euclidean distance after FRR alignment}. Visually speaking, the three networks in the upper row (for non-marijuana users) look very similar. At a more detailed level, we can see high similarity between networks obtained by FRR and DTW, which

are somewhat different from that by Euclidean distance. In contrast, the three networks in the lower row (for marijuana users) look very much different. These results indicate that the three methods can yield substantial differences in the marijuana user group.

For each method, we can also observe some difference between the non-marijuana-user group and marijuana-user group. We can assess whether the difference is statistically significant by a random permutation approach (Li and Solea, 2018). That is, we can check if the difference between the two groups is due to group variation or random variation among individuals. This is quantified by the difference of the covariance operators of the two groups. Here we use $\hat{\Sigma}_{XX}^{(1)}$ and $\hat{\Sigma}_{XX}^{(2)}$ to denote the covariance operators of the non-marijuana-user and marijuana-user groups, respectively. We compute the difference $R_0 = \|\hat{\Sigma}_{XX}^{(1)} - \hat{\Sigma}_{XX}^{(2)}\|$

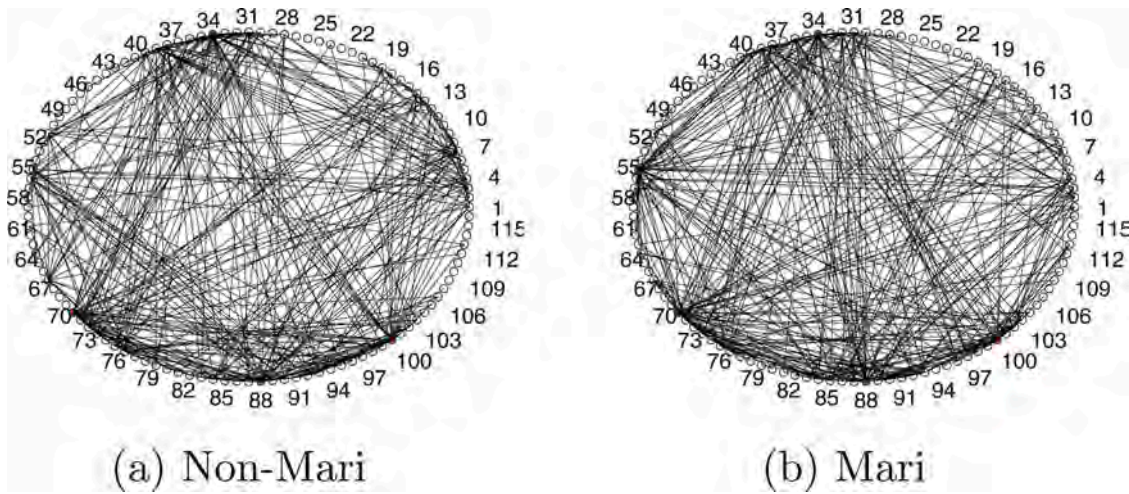


Fig. 13. Brain networks constructed by the partial correlation method on the fMRI data. (a) Network obtained after the FRR alignment for the non-marijuana group; (b) network obtained after the FRR alignment for the marijuana group.

$\widehat{\Sigma}_{XX}^{(1)}(0) - \widehat{\Sigma}_{XX}^{(2)}(0)$ for the original group assignment, and $R_s = \|\widehat{\Sigma}_{XX}^{(1)}(s) - \widehat{\Sigma}_{XX}^{(2)}(s)\|$, $s = 1, \dots, 100$, for 100 rounds of random permutation. In each round, we randomly assign 30 out of the total 60 as marijuana users, and the other 30 as non-marijuana users. Fig. 12 shows the histograms of all R_s and the location of R_0 . We can see that under the DTW and FRR settings (after removing pairwise phase-lags), R_0 for the true sample locates near the right end of all R_s produced by random permutations. This indicates that, for DTW and FRR, the difference between the two groups is statistically significant. In contrast, the R_0 obtained by Euclidean distance is near the center of the distribution of R_s , which implies we cannot tell if the detected difference obtained in Fig. 11 is due to group variation or individual variation.

To further investigate the difference between groups, we list the

regions with more than 20 edges in Table 2. From this table, we can see that in each of the three methods, the number of high-connectivity regions is always slightly higher (by 1) in the marijuana-user group. In particular, the DTW and FRR models both identify high-connectivity regions in the cerebellum (103, 105 and 107) in the marijuana user. The extant literature indicates that long-term daily cannabis users showed an increase in cerebellar blood volume (Sneider et al., 2006). This is well captured by the DTW and FRR methods when the phase variability is removed in the given fMRI signals. In summary, we see more high-connectivity regions in the marijuana group compared to the non-marijuana group, and the increase in cerebellar blood volume is identified by removing phase variability using the DTW and FRR methods.

Finally, we briefly compare the FASG-based network method with

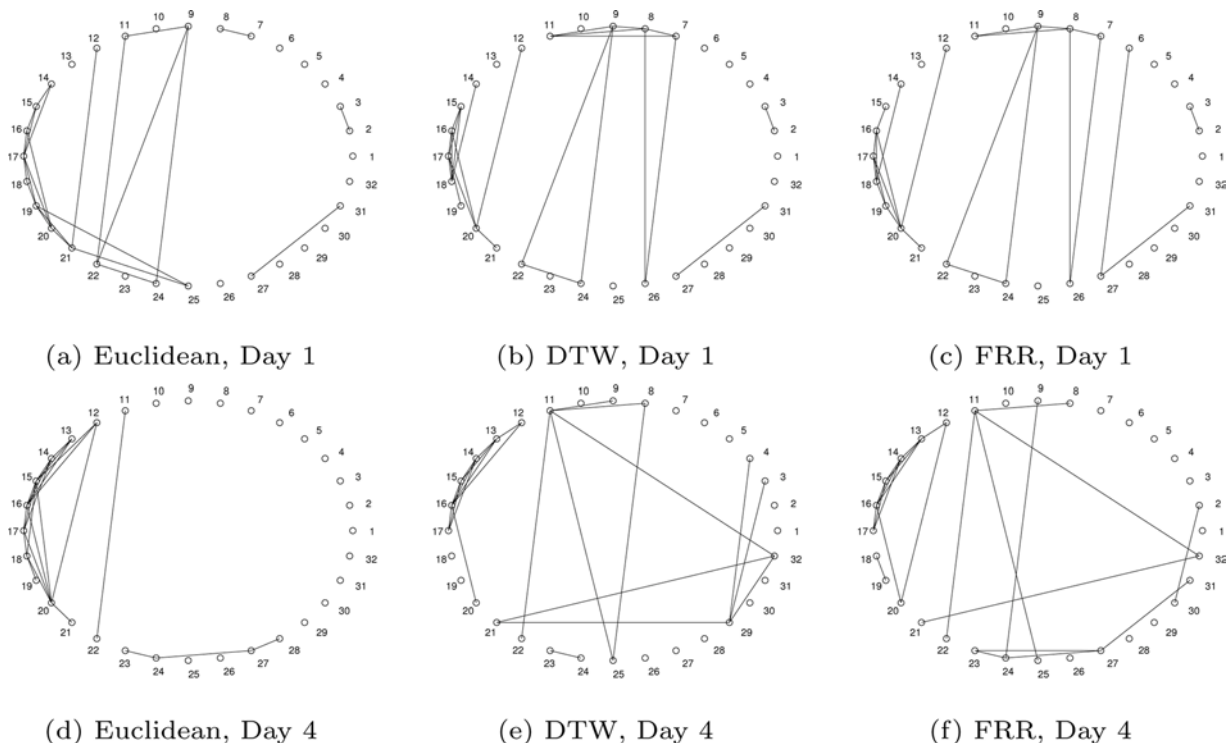


Fig. 14. Brain networks constructed by the FASG model on the EEG data. (a) Network obtained by Euclidean distance for Day 1; (b) network obtained by DTW distance for Day 1; (c) network obtained by Euclidean distance after FRR for Day 1; (d), (e), (f) same as (a), (b), (c) except for Day 4.

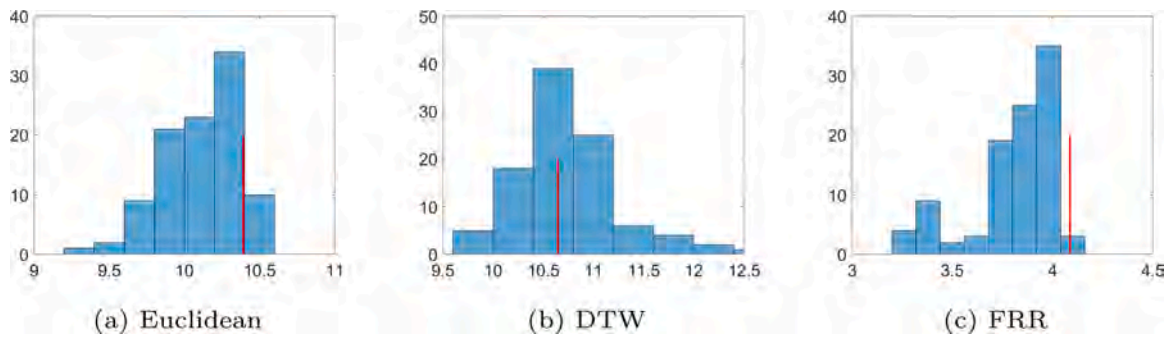


Fig. 15. Histogram for covariance operator difference $R_s = \|\hat{\Sigma}_{XX}^{(1)} - \hat{\Sigma}_{XX}^{(2)}\|$, for $s = 1, \dots, 100$ randomly split samples and the position of R_0 for true sample under different settings, and the red line indicates the position of R_0 : (a) Euclidean – R_0 is larger than 89% of $\{R_s\}_{s=1}^{100}$; (b) DTW – 51%; (c) FRR – 97%. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the traditional Gaussian-based partial correlation method. To utilize the FRR framework, the network is constructed using the aligned fMRI signals. The result (on the non-marijuana-user group and marijuana-user group) is shown in Fig. 13. Similar to that in the FASG method, we at first calculate difference of the two precision matrices in the two groups and then conduct permutation test to measure the significance. It is found that the difference between the two groups is larger than 83% of the differences after permutations. Basically, this corresponds to a p -value of 0.17, which implies the networks of the two groups are not statistically different.

3.2.2. Network analysis – EEG data

We also apply the FASG model under three settings to the EEG data set. With 32 electrodes, there are 32 nodes in the network. The number of edges in a fully connected network is $32 \times 31/2 = 496$. Same as the fMRI case, we construct brain networks by connecting edges with top 5% strongest additive dependence. That is, we will have $496 \times 0.05 = 25$ edges in each network.

The estimated networks using three distances methods on Day 1 and Day 4 are shown in Fig. 14. Basically, there is no much difference between three networks for Day 1 (shown in the upper panels). All of them show many connections in the posterior area (17, 18, 19, 20), some connections in the central region (7, 8, 9, 11, 22, 24), and very few common edges ($\{(2, 3), (27, 31)\}$) in the frontal area. However, for Day 4, the network under Euclidean setting is very different from the other two. That is, high connectivity is only detected in the posterior area using the Euclidean distance, whereas the networks obtained by DTW and FRR are constructed alike – they both show high connections between the frontal area and the posterior area. It was found that tACS induces long-term increases in posterior-to-frontal alpha connectivity (Clancy et al., 2018). The networks indicate that this long-term enhancement effect is confirmed by the FASG graph model under the DTW and FRR settings (pairwise phase-lags are removed).

Similar to the result on the fMRI data, we also observe difference between Day 1 and Day 4 using each distance. To evaluate if the difference is significant, we again conduct the permutation-based resampling. Results are shown in Fig. 15. It is found that only R_0 for FRR locates at 97% (above the 95% significance threshold), whereas R_0 for Euclidean and DTW is below the significance threshold. That is, the difference between two different days, reflecting the effect of tACS, can be effectively identified using the FRR method, whereas the Euclidean

and DTW fail to detect the difference.

4. Summary and conclusion

Phase variability is commonly observed in various neural signals such as EEG and fMRI. Modeling and analyzing those signals has been a major challenge, especially when the variability is nonstationary and nonlinear. In this article, we propose to adopt the Fisher-Rao registration (FRR) framework to deal with this problem. We systematically compare our method to the well-known DTW method in three aspects – basic framework, mathematical properties, and computational efficiency. The comparative result clearly demonstrates the superior performance of the FRR method. We then apply the FRR method to brain network problems using an fMRI and an EEG dataset. It is found that the FRR method can successfully identify difference between states/groups in both datasets, whereas DTW fails the task in the EEG dataset.

The FRR method shows its potentials in analyzing neural signals with phase variability. However, more thorough study is needed to understand the results we obtained in the real data analysis. We will also study if pre-processing procedures (e.g. normalization) could lead to any change in results. So far we have only tested the method with one state-of-the-art nonparametric graph model in the brain network analysis. We will explore other network models and test the method using more extensive data. We will also investigate our framework to study the problem of phase synchrony between brain regions (Lachaux et al., 1999). Finally, the current study focuses on continuous signals such as fMRI and EEG. We will explore possibilities to examine latency in discrete spike trains between neuronal units.

Author contributions

W.Z. and Z.X. contributed equally to this paper. They both designed the proposed framework, conducted computational analysis, and wrote the paper. W.L. provided experimental data, interpreted analysis result, and wrote the paper. W.W. designed the proposed framework, supervised data analysis, and wrote the paper.

Declarations of interest

None.

Appendix A. The DTW algorithm

Algorithm 1

Dynamic time warping **Input:** $\{u(t)\}_{t=1}^{l_u}$ and $\{v(t)\}_{t=1}^{l_v}$ are two curves with observed points; w is the warping window size; d is the metric used to compare two points' distance;

Initialize $C \in \mathbb{R}^{l_u \times l_v}$ to be an empty matrix;

for each $i = 1, 2 \dots l_u$ **do**

for each $j = \max\{1, i - w\}, \max\{1, i - w\} + 1 \dots \min\{l_v, i + w\}$ **do**

if $i = 1$ and $j = 1$ **then**

$C(i, j) = d[u(1), v(1)];$

else if $i = 1$ and $j > 1$ **then**

$C(i, j) = d[u(1), v(j)] + C(1, j - 1);$

else if $i > 1$ and $j = 1$ **then**

$C(i, j) = d[u(1), v(j)] + C(i - 1, 1);$

else

$C(i, j) = d[u(i), v(j)] + \min\{C(i, j), C(i - 1, j), C(i, j - 1), C(i - 1, j - 1)\};$

end if

end for

end for

$\text{dist} = C(l_u, l_v); \text{id1} = l_u; \text{id2} = l_v; \text{path} = \{(id1, id2)\};$

while $\text{id1} \neq 1$ or $\text{id2} \neq 1$ **do**

if $\text{id1} > 1$ and $\text{id2} > 1$ **then**

$\text{temp1} = [\text{id1} - 1, \text{id1}, \text{id1} - 1]; \text{temp2} = [\text{id2}, \text{id2} - 1, \text{id2} - 1];$

$\text{temp3} = \text{argmin}\{C(\text{id1} - 1, \text{id2}), C(\text{id1}, \text{id2} - 1), C(\text{id1} - 1, \text{id2} - 1)\};$

$\text{id1} = \text{temp1}(\text{temp3}); \text{id2} = \text{temp2}(\text{temp3});$

else if $\text{id1} = 1$ and $\text{id2} > 1$ **then**

$\text{id2} = \text{id2} - 1;$

else if $i > 1$ and $j = 1$ **then**

$\text{id1} = \text{id1} - 1;$

end if

Insert $\{(id1, id2)\}$ to path on top;

end while

Output: dist is the DTW distance between u and v ; path is the point-matching path

Appendix B. The template estimation algorithm

The optimal template can be computed under the FRR method by:

1. Compute the SRVFs for the given curves: $q_1, q_2 \dots q_n$;
2. Iteratively: find the mean of the SRVFs μ ; match each of the SRVFs to μ and replace the original ones; do until convergence to obtain $\hat{\mu}$, the estimated mean in amplitude (the Karcher mean);
3. Match each of the SRVFs to $\hat{\mu}$ to obtain the estimated phase $\hat{\gamma}_1, \hat{\gamma}_2 \dots \hat{\gamma}_n$;
4. Since the SRVF of a time warping function is in \mathbb{S}^∞ , the mean of $\hat{\gamma}_1, \hat{\gamma}_2 \dots \hat{\gamma}_n$ can be computed by the mean of their SRVFs (the Karcher mean), $\hat{\psi}_1, \hat{\psi}_2 \dots \hat{\psi}_n$, on the infinite-dimension ball, using the shooting vectors' average and exponential mapping. In this way, the estimated mean in phase, $\hat{\omega}$, can be obtained;
5. Combining the average in amplitude and phase, one can obtain the optimal template:

$$\hat{q}_{\text{temp}} = (\hat{\mu} \circ \hat{\omega}) * \sqrt{\hat{\omega}}$$

References

- Allen, E.A., Damaraju, E., Plis, S.M., Erhardt, E.B., Eichele, T., Calhoun, V.D., 2014. Tracking whole-brain connectivity dynamics in the resting state. *Cereb. Cortex* 24, 663–676.
- Brookes, M.J., Hale, J.R., Zumer, J.M., et al., 2011. Measuring functional connectivity using MEG: methodology and comparison with fMRI. *NeuroImage* 56, 1082–1104.
- Chang, C., Glover, G.H., 2010. Time-frequency dynamics of resting-state brain connectivity measured with fMRI. *NeuroImage* 50, 81–98.
- Chen, J.E., Chang, C., Greicius, M.D., Glover, G.H., 2015. Introducing co-activation pattern metrics to quantify spontaneous brain network dynamics. *NeuroImage* 111, 476–488.
- Clancy, K.J., Baisley, S.K., Albizu, A., Kartvelishvili, N., Ding, M., Li, W., 2018. Lasting connectivity increase and anxiety reduction via transcranial alternating current stimulation. *Soc. Cogn. Affect. Neurosci.* 13, 1305–1316.
- Craddock, R.C., James, G.A., Holtzheimer III, P.E., Hu, X.P., Mayberg, H.S., 2012. A whole brain fMRI atlas generated via spatially constrained spectral clustering. *Hum. Brain Mapp.* 33, 1914–1928.
- Dinov, M., Lorenz, R., Scott, G., Sharp, D.J., Fagerholm, E.D., Leech, R., 2016. Novel modeling of task vs. rest brain state predictability using a dynamic time warping spectrum: comparisons and contrasts with other standard measures of brain dynamics. *Front. Comput. Neurosci.* 10, 46.
- Fries, P., 2005. A mechanism for cognitive dynamics: neuronal communication through neuronal coherence. *Trends Cogn. Sci.* 9, P474–480.
- Friston, K.J., Buechel, C., Fink, G.R., et al., 1997. Psychophysiological and modulatory interactions in neuroimaging. *NeuroImage* 6, 218–229.
- Gupta, L., Molfese, D.L., Tammana, R., Simos, P.G., 1996. Nonlinear alignment and averaging for estimating the evoked potential. *IEEE Trans. Biomed. Eng.* 43, 348–356.
- Handwerker, D.A., Gonzalez-Castillo, J., D'Esposito, M., Bandettini, P.A., 2012. The continuing challenge of understanding and modeling hemodynamic variation in fMRI. *NeuroImage* 62, 1017–1023.
- Huang, H.C., Jansen, B., 1985. EEG waveform analysis by means of dynamic time-warping. *Int. J. Bio-med. Comput.* 17, 135–144.
- Hurtado, J.E., 2004. An examination of methods for approximating implicit limit state functions from the viewpoint of statistical learning theory. *Struct. Saf.* 26, 271–293.
- Jones, D.T., Vemuri, P., Murphy, M.C., et al., 2012. Non-stationarity in the “resting brain’s” modular architecture. *PLOS ONE* 380, 899–907.
- Kamiński, M., Ding, M., Truccolo, W.A., Bressler, S.L., 2001. Evaluating causal relations in neural systems: Granger causality, directed transfer function and statistical assessment of significance. *Biol. Cybern.* 85, 145–157.
- Karamzadeh, N., Medvedev, A., Azari, A., et al., 2013. Capturing dynamic patterns of task-based functional connectivity with EEG. *NeuroImage* 66, 311–317.
- Keilholz, S.D., 2014. The neural basis of time-varying resting-state functional connectivity. *Brain Connect.* 4, 769–779.
- Kiviniemi, V., Vire, T., Remes, J., et al., 2011. A sliding time-window ica reveals spatial variability of the default mode network in time. *Brain Connect.* 1, 339–347.
- Lachaux, J., Rodriguez, E., Martinerie, J., Varela, F.J., 1999. Measuring phase synchrony in brain signals. *Hum. Brain Mapp.* 8, 194–208.

- Lawlor, P.N., Perich, M.G., Miller, L.E., Kording, K.P., 2018. Linear-nonlinear-time-warp-poisson models of neural activity. *J. Comput. Neurosci.* 45, 173–191.
- Li, B., Chun, H., Zhao, H., 2014. On an additive semigraphoid model for statistical networks with application to pathway analysis. *J. Am. Stat. Assoc.* 109, 1188–1204.
- Li, B., Solea, E., 2018. A nonparametric graphical model for functional data with application to brain networks based on fMRI. *J. Am. Stat. Assoc.* 113, 1637–1655.
- Li, Z., Liu, H., Liao, X., 2015. Dynamic functional connectivity revealed by resting-state functional near-infrared spectroscopy. *Biomed. Opt. Express* 6, 2337–2352.
- Liu, X., Duyn, J.H., 2013. Time-varying functional network information extracted from brief instances of spontaneous brain activity. *Proc. Natl. Acad. Sci. U. S. A.* 110, 4392–4397.
- Meinshausen, N., Bühlmann, P., et al., 2006. High-dimensional graphs and variable selection with the lasso. *Ann. Stat.* 34, 1436–1462.
- Meszlényi, R.J., Hermann, P., Buza, K., et al., 2017. Resting state fMRI functional connectivity analysis using dynamic time warping. *Front. Neurosci.* 11, 75.
- Mørup, M., Hansen, L.K., Arnfred, S.M., 2008. Algorithms for sparse nonnegative Tucker decompositions. *Neural Comput.* 20, 2112–2131.
- de Pasquale, F., Penna, S.D., Snyder, A.Z., et al., 2010. Temporal dynamics of spontaneous MEG activity in brain networks. *Proc. Natl. Acad. Sci. U. S. A.* 107, 6040–6045.
- Qiao, X., Guo, S., James, G.M., 2018. Functional graphical models. *J. Am. Stat. Assoc.* 1–12.
- Sakoe, H., Chiba, S., 1978. Dynamic programming algorithm optimization for spoken word recognition. *IEEE Trans. Acoust. Speech Signal Process.* 26, 43–49.
- Sakoglu, Ü., Pearson, G.D., Kiehl, K.A., et al., 2010. A method for evaluating dynamic functional network connectivity and task-modulation: application to schizophrenia. *Magn. Reson. Mater. Phys. Biol. Med.* 23, P351–P366.
- Singer, W., 1999. Neuronal synchrony: a versatile code for the definition of relations. *Neuron* 24, 49–65.
- Singer, W., Gray, C.M., 1995. Visual feature integration and the temporal correlation hypothesis. *Annu. Rev. Neurosci.* 18, 555–586.
- Smith, S.M., 2012. The future of fMRI connectivity. *NeuroImage* 62, 1257–1266.
- Sneider, J.T., Pope Jr., H.G., Silveri, M.M., Simpson, N.S., Gruber, S.A., Yurgelun-Todd, D.A., 2006. Altered regional blood volume in chronic cannabis smokers. *Exp. Clin. Psychopharmacol.* 14, 422.
- Srivastava, A., Wu, W., Kurtek, S., Klassen, E., Marron, J., 2011. Registration of Functional Data Using Fisher-Rao Metric arXiv preprint arXiv:1103.3817.
- Tononi, G., Edelman, G.M., Sporns, O., 1998. Complexity and coherency: integrating information in the brain. *Trends Cogn. Sci.* 2, 474–484.
- Tucker, J.D., Wu, W., Srivastava, A., 2013. Generative models for functional data using phase and amplitude separation. *Comput. Stat. Data Anal.* 61, 50–66.
- Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: phase synchronization and large-scale integration. *Nat. Rev. Neurosci.* 2, 229–239.
- Vinck, M., Oostenveld, R., van Wingerden, M., et al., 2011. An improved index of phase synchronization for electrophysiological data in the presence of volume-conduction, noise and sample-size bias. *NeuroImage* 55, 1548–1565.
- Williams, A.H., Poole, B., Maheswaranathan, N., Dhawale, A.K., Fisher, T., Wilson, C.D., Brann, D.H., Trautmann, E.M., Ryu, S., Shusterman, R., et al., 2020. Discovering precise temporal patterns in large-scale neural recordings through robust and interpretable time warping. *Neuron* 105, 246–259.
- Zhang, J.T., 2013. *Analysis of Variance for Functional Data*. CRC Press.