

## TOPOLOGICAL DATA ANALYSIS OF SINGLE-TRIAL ELECTROENCEPHALOGRAPHIC SIGNALS

BY YUAN WANG<sup>\*,1</sup>, HERNANDO OMBAO<sup>†,‡,2</sup> AND MOO K. CHUNG<sup>\*,1</sup>

*University of Wisconsin–Madison<sup>\*</sup>, University of California, Irvine<sup>†</sup> and  
King Abdullah University of Science and Technology<sup>‡</sup>*

Epilepsy is a neurological disorder marked by sudden recurrent episodes of sensory disturbance, loss of consciousness, or convulsions, associated with abnormal electrical activity in the brain. Statistical analysis of neurophysiological recordings, such as electroencephalography (EEG), facilitates the understanding of epileptic seizures. Standard statistical methods typically analyze amplitude and frequency information in EEG signals. In the current study, we propose a topological data analysis (TDA) framework to analyze single-trial EEG signals. The framework denoises signals with a weighted Fourier series (WFS), and tests for differences between the topological features—persistence landscapes (PLs) of denoised signals through resampling in the frequency domain. Simulation studies show that the test is robust for topologically similar signals while bearing sensitivity to topological tearing in signals. In an application to single-trial epileptic EEG signals, EEG signals in the diagnosed seizure origin and its symmetric site are found to have similar PLs before and during a seizure attack, in contrast to signals at other sites showing significant statistical difference in the PLs of the two phases.

**1. Introduction.** Epilepsy is a neurological disorder marked by sudden recurrent episodes of sensory disturbance, loss of consciousness, or convulsions, associated with abnormal electrical activity in the brain. During an epileptic seizure attack, the patient may experience idiosyncratic symptoms ranging from visual hallucinations to a sense of disassociation [Bancaud et al. (1994), Fried (1997)]. Findings by the World Health Organization (WHO) indicate that approximately nine in 1000 people around the world suffered from epilepsy in 1998 [WHO (2005)]. The Centers for Disease Control and Prevention (CDC) have recently reported that an estimated 1% of adults in the United States currently suffer from active epilepsy [Kobau et al. (2012)]. Neuroscientific researchers are pursuing all possible avenues to gain better understanding and management of the disease. One key area of research is the study of the epileptogenic zone, that is, the set of brain sites involved

---

Received September 2016; revised June 2017.

<sup>1</sup>Supported in part by NIH Brain Initiative Grant EB022856.

<sup>2</sup>Supported in part by NSF DMS, NSF SES and the KAUST Baseline Research Fund.

*Key words and phrases.* Persistence landscape, persistent homology, weighted Fourier series, electroencephalogram, epilepsy.

in the generation of seizures, particularly for the purpose of epilepsy surgery. However, high failure rate in epilepsy surgery suggests that epileptogenicity remains elusive [Bartolomei, Chauvel and Wendling (2008)].

In our current study, we aim to gain fresh insight on the epileptogenic zone through exploratory analysis of electroencephalography (EEG) from an epilepsy patient. EEG is an important electrophysiological modality for understanding the function and dysfunction of the brain. It is popular in studying epileptic seizures because of its noninvasive procedure and high temporal resolution. EEG signals are synchronous discharges from cerebral neurons detected by electrodes placed on the scalp or intracranially implanted in the patient's brain. Epileptic seizures are associated with brief and episodic neuronal synchronous discharges with dramatically changing amplitude and frequency in EEG signals. State-of-the-art statistical methods have been developed to study EEG signals between and during epileptic seizures [Donoho, Mallat and von Sachs (1998), Mitra and Pesaran (1999), Ombao et al. (2001), Ombao, von Sachs and Guo (2005)]. In the medical literature, more accessible methods have been developed for the purpose of understanding epileptogenicity with EEG signals observed between and during seizure attacks [Martinerie et al. (1998), McSharry, Smith and Tarassenko (2003), Bartolomei, Chauvel and Wendling (2008)]. These methods tend to utilize transformed information, such as time-frequency and phase space information, of the epileptic EEG signals. There are also methods that utilize temporal information of the EEG signals. The local variance method, where variances of signal amplitude are computed in a moving window across time, often serve as a baseline. The simplistic approach has shown to be more effective in some cases than methods using information from a transformed domain [Martinerie et al. (1998), McSharry, Smith and Tarassenko (2003), Mohseni, Maghsoudi and Shamsollahi (2006)].

The epileptogenic zone has been shown to generate EEG pattern distinct from other brain sites at the onset of a seizure [Lange et al. (1983), Van Quyen et al. (2001), Burns et al. (2014)]. To gain fresh insight on epileptic EEG pattern through topology, we need to account for the fact that amplitude and frequency of EEG signals in the whole brain undergo drastic changes before and during a seizure attack. We need the topological method to be reliable in detecting true topological changes before and during seizure.

A promising exploratory approach is topological data analysis (TDA), an umbrella term for various topological techniques to analyze scientific data [Carlsson (2009)]. A key TDA technique is persistent homology (PH) developed by [Edelsbrunner, Letscher and Zomorodian (2002)]. PH tracks changes in topological features of EEG data based on the content of oscillations across multiple resolutions and dimensions. PH on an EEG signal summarizes the evolution of the connected components (based on the Fourier coefficients of the signal). PH descriptors, such as the barcode and persistence diagram (PD), keep track of the birth and death times of connected components as they appear and disappear in

the sublevel set when the sublevel threshold  $\lambda$  increases—the parameter  $\lambda$  is analogous to but more general than the thresholding level in multi-resolution wavelet analysis [Chung et al. (2014)]. It is a challenge to perform inference on barcode and PD directly due to theoretical and practical reasons, as detailed in Section 3. Bubenik (2015) proposed an alternative PH descriptor persistence landscape (PL) that builds landscape-like features with a rigorous statistical framework.

PH has been applied to a wide range of data: Sousbie, Pichon and Kawahara (2011) used PH for scale-free and parameter-free identification of the voids, walls, filaments, clusters and their configuration within the cosmic web; Ahmed, Fasy and Wenk (2014) proposed a localized version of PH to compare the intrinsic structures of two road networks; Reininghaus et al. (2015) and Zhu et al. (2016) proposed stable multiresolution PH kernels to capture global topology (and also local topology in the latter work) for shape, motion and eye disease classification; existing PH applications on medical imaging data have been focused on static multivariate random samples, typically from positron emission tomography (PET), magnetic resonance imaging (MRI) and functional MRI (fMRI) studies [Gamble and Heo (2010), Lee et al. (2011), Heo, Gamble and Kim (2012), Chung et al. (2015)]. A popular approach in neuroimaging applications is to model the data as a network or graph based on sites on the brain and join two sites when the distance between them exceeds a certain threshold  $\lambda$  [Lee et al. (2011), Chung et al. (2015)]. This approach has also been applied to EEG functional networks in a mouse model of depression [Khalid et al. (2014)].

In this paper, we propose an inference framework for comparing the PLs of two phases of a single-trial EEG signal. Our key contributions are as follows:

(a) This is the first PL application to study the topology in the sublevel set of a univariate EEG signal. Related research is mainly based on Euler characteristics [Worsley (1995), Adler et al. (2010), Turner, Mukherjee and Boyer (2014)].

(b) The proposed inference framework incorporates a frequency resampling method for comparing two phases of a single-trial EEG signal by shuffling their Fourier coefficients. Simulation studies show that the method is robust to signal similarity in the time domain.

(c) We set topological performance criteria for the inference framework. We want TDA to provide new insight on EEG data by controlling the “topological false positive” rate. In other words, we want to control the rate of detecting false difference between PLs. For instance, if what we detect is only a scaling difference between signals, then we are not getting new insight by TDA. Our simulation studies show that the proposed TDA framework is robust to topology-preserving transformations such as translation, amplitude and frequency scaling. It is also sensitive to topology-destroying transformations.

(d) We treat statistically insignificant difference between PLs before and during a seizure as a “signal” in the test results. Using the proposed TDA framework, we are able to identify PL indifference/difference before and during seizure of signals

in/outside the seizure zone, which was not previously discovered by other methods that used the same data, for example, SLEX methods in [Ombao et al. (2001), Ombao, von Sachs and Guo (2005)]. The results are promising and provide new insight on epileptogenicity through single-trial multichannel EEG signals.

In Section 3, we present our procedure for fitting and resampling thresholded WFS of single-trial multichannel EEG signals, and for inference on the PLs of the resampled thresholded WFS before and during a seizure attack. Sections 4 and 5 are designated for simulation studies and data application.

**2. Preliminary on persistent homology.** In Section 3.1, we will see how an EEG signal becomes a one-dimensional Morse function after denoising. The homological information of a one-dimensional Morse function  $g : \mathcal{X} \subset \mathbb{R} \rightarrow \mathbb{R}$  is the connected components in its sublevel set

$$S(\lambda) = \{t \in \mathcal{X} : g(t) \leq \lambda\},$$

for some  $\lambda \in \mathbb{R}$ . Increasing  $\lambda$ , the structure of connected components changes in a process called Morse filtration. The birth and death of connected components in the Morse filtration is characterized by the pairing of local minimums and maximums of the Morse function [Milnor (1963), Chung, Bubenik and Kim (2009), Bubenik et al. (2010)]. To see this, let  $\#S(\lambda)$  be the number of connected components in the sublevel set  $S(\lambda)$ . We denote the local minimums as  $g_1, \dots, g_m$  and the local maximums as  $h_1, \dots, h_n$ . Since the critical values of the Morse function are unique, we can order the local minimums from the smallest to the largest:

$$g(1) < g(2) < \dots < g(m),$$

and order the local maximums:

$$h(1) < h(2) < \dots < h(n).$$

We also collect all the local minimums and maximums:

$$z_1 = g_1, \quad \dots, \quad z_m = g_m, \quad z_{m+1} = h_1, \quad \dots, \quad z_{m+n} = h_n,$$

and order them as

$$z(1) < z(2) < \dots < z(m+n).$$

At each local minimum  $g_i$ , a new component is born and

$$\#S(g_i) = \#S(g_i - \varepsilon) + 1$$

for sufficiently small  $\varepsilon$ . The new component is thus identified with the local minimum  $g_i$ . Similarly, at each local maximum  $h_i$ , a component merges with an old component and

$$\#S(h_i) = \#S(h_i - \varepsilon) - 1.$$

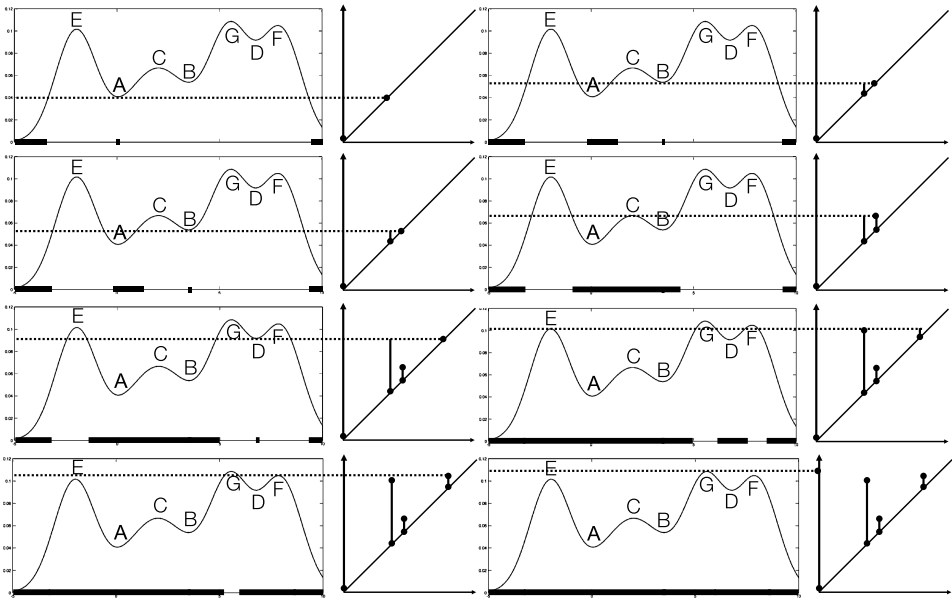


FIG. 1. The evolution of connected components in the sublevel set of a Morse function, and the construction of a barcode encoding the birth and death times of the connected components.

The number of connected components changes only when we pass through a local minimum or maximum  $z_j$ :

$$\#S(z_{(i+1)}) = \#S(z_{(i)}) \pm 1,$$

where the sign depends on whether  $z_{(i+1)}$  is a local maximum ( $-1$ ) or a local minimum ( $+1$ ).

In Figure 1, we illustrate how local minimums and maximums characterize the birth and death of connected components in the sublevel set of a Morse function. A horizontal line of reference indicating the threshold  $\lambda$  moves from the minimum value of the function and up. Before the line hits the point A, the sublevel set of the function is empty except for the boundaries. After the line touches A, the sublevel set becomes a line segment on the  $x$ -axis underneath A that grows as the reference line keeps moving upward. Another line segment under B joins the sublevel set when the reference line hits point B. The two line segments join up when the reference line reaches point C. According to the *Elder Rule* [Edelsbrunner and Harer (2010), Chung (2014)] that says *older features live on at merging junctions*, we pair point B with point C and leave point A with the oldest line segment in the sublevel set to be paired with a later local maximum. As we reach point D, a new line segment emerges in the sublevel set and is annihilated. The next point reached is E, where the left boundary merges with the big component joined by line segments A and B. When the reference line reaches the point F, the line segment created at D

is merged with the right boundary. So we pair D with F. Lastly, the two big components are joined as one at the point G and we pair the global minimum value of the function with G, using an idea from extended persistence [Cohen-Steiner and Edelsbrunner (2009)]. The pairing of birth and death times of the connected components in the sublevel set are encoded in the barcode. Throughout the paper, the PH of sublevel sets of any Morse function is calculated by a pairing algorithm based on the Elder Rule [Chung, Bubenik and Kim (2009), Chung (2014)]. The critical points used for the pairing algorithm were found numerically by checking sign changes in the differences  $g(t_{i+1}) - g(t_i)$  of neighboring observations in an EEG signal.

*PH features.* The collection of bars in the lower right plot of Figure 1 when rotated 90° clockwise is called a barcode (Figure 2). The barcode or its equivalent persistence diagram (PD) (collection of the upper end points of the bars in the lower right plot of Figure 1) are the originally proposed PH features. PD has been shown to possess desirable properties such as Lipschitz stability with respect to the bottleneck distance [Cohen-Steiner, Edelsbrunner and Harer (2007), Cohen-Steiner and Edelsbrunner (2009)]. It was also shown in [Mileyko, Mukherjee and Harer (2011)] that the PDs under the Wasserstein distance form a Polish space—a complete and separable metric space. The mean and variance appropriate for the space are the so-called Fréchet mean and variance. The Fréchet mean, however, is not unique [Turner et al. (2014)], rendering it a challenging statistical issue to perform inference on PDs directly. Notable PD applications in medical imaging studies also show that inference on PD is by no means straightforward in practice [Chung, Bubenik and Kim (2009), Gamble and Heo (2010), Heo, Gamble and Kim (2012)].

An alternative PH feature—the persistence landscape (PL)—was proposed by Bubenik (2015) for the purpose of statistical analysis. Given a bar  $(a, b)$  in a barcode with  $a \leq b$ , we can define the piecewise linear bump function  $h_{(a,b)} : \mathbb{R} \rightarrow \mathbb{R}$  by

$$(1) \quad h_{(a,b)}(\lambda) = \max(\min(\lambda - a, b - \lambda), 0).$$

The geometric representation of the bump function (1) is a right-angled isosceles triangle with height equal to half of the base of the corresponding interval in the

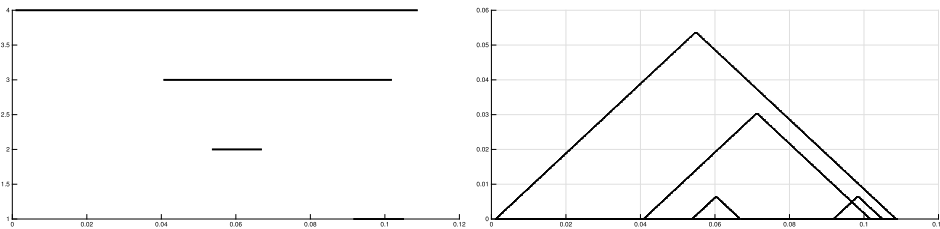


FIG. 2. Barcode and persistence landscape (PL) based on the example shown in Figure 1.

barcode. The PL  $\nu$  of the barcode  $\{(a_i, b_i)\}_{i=1}^N$  is a multi-valued function defined by

$$(2) \quad \nu_\ell(\lambda) = \begin{cases} \ell\text{th largest value of } \{h_{(a_i, b_i)}(\lambda)\}_{i=1}^N, & \ell = 1, \dots, N, \\ 0, & \ell > N, \end{cases}$$

where  $\nu_\ell$  will be referred to as the  $\ell$ th layer of the PL  $\nu$ . Figures 2 illustrate the PL of the barcode obtained in Figure 1. The main technical advantage of PL is that, as a function on a separable Banach space, the theory (and results) of random variables can be applied [Bubenik (2015), Chazal et al. (2014)].

To compare the PLs of two groups of signals

$$\{v^{11}, \dots, v^{1n}\} \quad \text{and} \quad \{v^{21}, \dots, v^{2n}\},$$

a parametric approach is to apply a real-valued functional to the PLs and then conduct a two-sample  $t$ -test on the means of the functionals [Bubenik (2015)]. However, in the case of single-trial epileptic EEGs, the dataset does not contain multiple signals; we observe only two PLs

$$\{v^1\} \quad \text{and} \quad \{v^2\}$$

of the respective WFS (13) and (14) for Phases 1 and 2 of a seizure. Resampling the signals requires additional work on the covariance estimation of the corresponding PLs. In Section 3, we propose a permutation test on the PLs of single-trial EEGs resampled in the frequency domain.

**3. Methods.** We propose a permutation test on the PLs of single-trial EEG signals by resampling in the frequency domain. Since TDA of raw EEG signals may pick up more artifacts than features, particularly in the analysis of connected component evolution [Bendich et al. (2016)], we first denoise an EEG signal to stabilize the subsequent topological permutation test.

3.1. *Signal denoising by weighted Fourier series.* An EEG signal is recorded at regular time intervals  $-T = t_1 < t_1 < \dots < t_N = T$ . We denoise a signal by estimating  $\mu$  in the model:

$$(3) \quad f(t) = \mu(t) + \epsilon(t), \quad -T \leq t \leq T,$$

where we assume the time to be continuous. The additive model (3) is the most fundamental and flexible scientific model for a stochastic process. We estimate the underlying signal  $\mu(t)$  by a weighted Fourier series (WFS) since an EEG signal is often considered as superpositions of sine and cosine waveforms with varying amplitudes, and the weighting governs the relative weights of the high-frequency components to the low-frequency ones. On a deeper level, the weighted Fourier approach is motivated by the connection between kernel estimation and heat diffusion equation. The popular Gaussian kernel estimator is noted to be equivalent

to the solution of a Fourier heat equation [Chaudhuri and Marron (2000)], and a more general heat diffusion equation generates a large class of kernel estimators with desirable statistical properties [Botev, Grotowski and Kroese (2010)].

Now consider the series solution to the linear diffusion equation

$$(4) \quad \frac{\partial}{\partial \sigma} g(t, \sigma) = \frac{\partial^2}{\partial x^2} g(t, \sigma), \quad \sigma \geq 0, t \in [-T, T].$$

By treating the observed signal  $f$  as the initial condition of the diffusion equation:

$$g(t, 0) = f(t),$$

we are able to obtain a closed-form WFS estimate [Chung et al. (2007), Chung et al. (2014)] for the signal  $f$ :

$$(5) \quad \hat{\mu}(t) = \sum_{j=0}^{\infty} e^{-\gamma_j \sigma} a_j \phi_{j1}(t) + \sum_{j=1}^{\infty} e^{-\gamma_j \sigma} b_j \phi_{j2}(t),$$

with the eigenvalues  $\gamma_j = (j\pi/T)^2$  for  $j \geq 1$ , the Fourier coefficients

$$\begin{aligned} a_0 &= \frac{1}{2T} \int_{-T}^T f(t) dt, \\ a_j &= \frac{1}{T} \int_{-T}^T f(t) \cos(j\pi t/T) dt, \\ b_j &= \frac{1}{T} \int_{-T}^T f(t) \sin(j\pi t/T) dt, \end{aligned}$$

and the basis functions

$$(6) \quad \phi_{j1}(t) = \cos(j\pi t/T), \quad j \geq 0,$$

$$(7) \quad \phi_{j2}(t) = \sin(j\pi t/T), \quad j \geq 1.$$

The degree- $k$  representation of (5) is

$$(8) \quad \hat{\mu}^k(t) = a_0 + \sum_{j=1}^k e^{-(j\pi/T)^2 \sigma} [a_j \cos(j\pi t/T) + b_j \sin(j\pi t/T)],$$

where the degree  $k$  decides the highest frequency  $[k/T]$  to be included in the representation (e.g., 100 Hz for  $k = 499$  and  $T = 5$ ), and the relative weights of the high frequency components to the low frequency components is governed by the parameter  $\sigma$ .

WFS also effectively reduces the Gibbs phenomenon in the FS estimation of data at discontinuities [Chung et al. (2014)]. Figure 3 shows FS and WFS in a simple example. The underlying function takes step values 1 and  $-1$  on the intervals  $[-\pi, 0)$  and  $[0, \pi]$  respectively. All series estimation is based on the first 50 terms



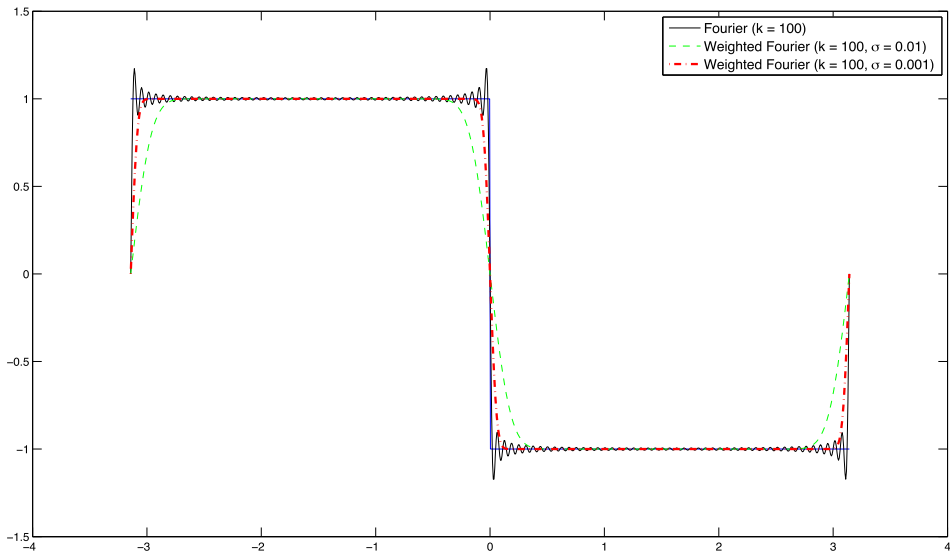


FIG. 3. *Gibbs phenomenon (ringing artifacts) is visible in the Fourier series expansion of a step function, whereas the weighted Fourier series approximation shows less visible artifacts.*

of finite approximation. The discontinuity at  $\pi$  and the two end-points cause the FS to overshoot, whereas the WFS is not affected in the same way.

In practice, some low-frequency components may not matter in the degree- $k$  WFS estimator (8). So we include in the representation only those sine or cosine waveform (even those at high frequencies) whose amplitudes exceed some threshold. We borrow strength from wavelet thresholding [Donoho and Johnstone (1994), Donoho and Johnstone (1995), Abramovich and Benjamini (1996)] to delete frequency components that are of lesser importance:

$$(9) \quad \hat{\mu}_{T_u}^k(t) = \sum_{j \in \mathcal{I}_1} e^{-(j\pi/T)^2\sigma} a_j \cos(j\pi t/T) + \sum_{j \in \mathcal{I}_2} e^{-(j\pi/T)^2\sigma} b_j \sin(j\pi t/T),$$

with

$$(10) \quad \mathcal{I}_1 = \{j = 1, \dots, k : |a_j| > T_u\}, \quad \mathcal{I}_2 = \{j = 1, \dots, k : |b_j| > T_u\},$$

where  $T_u$  is the universal threshold

$$(11) \quad T_u = s\sqrt{2\log n},$$

where  $n$  is the number of data points in each phase and  $s$  is the median of the absolute deviation (MAD) of the Fourier coefficients:

$$(12) \quad s = \text{median}\{j = 1, \dots, k : |a_j - \text{median}\{i = 1, \dots, k : |a_i|\}|, \\ |b_j - \text{median}\{i = 1, \dots, k : |b_i|\}|\}.$$

In summary, the signal  $\mu(t)$  is estimated as superpositions of sinusoidal functions corresponding to high-frequency oscillations, and relatively lower contribution to the total variation in the signal are removed.

We fit the thresholded degree- $k$  WFS estimator to the signals  $f_1(t)$  and  $f_2(t)$  before and during seizure (Phases 1 and 2):

$$\begin{aligned}
 \hat{\mu}_{T_u^1}^k(t) &= \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} A_{1j} \cos(j\pi t/T) \\
 &+ \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} B_{1j} \sin(j\pi t/T),
 \end{aligned}
 \tag{13}$$

$$\begin{aligned}
 \hat{\mu}_{T_u^2}^k(t) &= \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} A_{2j} \cos(j\pi t/T) \\
 &+ \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} B_{2j} \sin(j\pi t/T),
 \end{aligned}
 \tag{14}$$

where

$$\begin{aligned}
 A_{1j} &= a_{1j} \mathbb{I}_{j \in \mathcal{I}_{11}}, & B_{1j} &= b_{1j} \mathbb{I}_{j \in \mathcal{I}_{12}}, \\
 A_{2j} &= a_{2j} \mathbb{I}_{j \in \mathcal{I}_{21}}, & B_{2j} &= b_{2j} \mathbb{I}_{j \in \mathcal{I}_{22}},
 \end{aligned}$$

with  $a_{ij}$  and  $b_{ij}$  being the Fourier coefficients for Phase  $i$ ,  $i = 1, 2$ :

$$\begin{aligned}
 a_{i0} &= \frac{1}{2T} \int_{-T}^T f_i(t) dt, \\
 a_{ij} &= \frac{1}{T} \int_{-T}^T f_i(t) \cos(j\pi t/T) dt, \\
 b_{ij} &= \frac{1}{T} \int_{-T}^T f_i(t) \sin(j\pi t/T) dt,
 \end{aligned}
 \tag{15}$$

and the indicator functions  $\mathbb{I}_{j \in \mathcal{I}_{i1}}$  and  $\mathbb{I}_{j \in \mathcal{I}_{i2}}$  setting the universal threshold on  $a_{ij}$  and  $b_{ij}$ ,  $j = 0, \dots, k$ , by

$$\mathcal{I}_{i1} = \{j = 0, \dots, k : |a_{ij}| > T_u^i\}, \quad \mathcal{I}_{i2} = \{j = 0, \dots, k : |b_{ij}| > T_u^i\},
 \tag{16}$$

where  $T_u^i = s_i \sqrt{2 \log n}$  is the universal threshold with

$$\begin{aligned}
 s_i &= \text{median}\{j = 1, \dots, k : |a_{ij} - \text{median}\{l = 1, \dots, k : |a_{il}|\}|\}, \\
 &|b_{ij} - \text{median}\{l = 1, \dots, k : |b_{il}|\}|\}.
 \end{aligned}$$

3.2. *Statistical inference on PLs of denoised EEGs.* By denoising an EEG signal with a finite sum of weighted sinusoidal functions, we make it a Morse function. In Section 2 we describe how the evolution of connected components in the sublevel sets of a Morse function is summarized by PLs. We now want to compare the PLs of denoised signals  $\hat{\mu}_1(t)$  and  $\hat{\mu}_2(t)$  before and during seizure.

The before- and during-seizure phases of an EEG signal are modeled by

$$(17) \quad f_1(t) = \mu_1(t) + \varepsilon_1(t) \quad \text{and} \quad f_2(t) = \mu_2(t) + \varepsilon_2(t), \quad -T \leq t \leq T.$$

With the hypothesis

$$(18) \quad H_0: \quad \mu_1(t) = \mu_2(t) \quad \forall -T \leq t \leq T,$$

we test the pointwise difference between the two phases of the signal in the time domain. As emphasized in Section 2, only the local minimums and maximums of WFS-denoised EEG amplitudes characterize the evolution of connected components in the sublevel sets of the signal. So a statistical test on the pointwise difference in the amplitudes contains redundant information and can yield a false conclusion about the difference between the PH features of two phases of an EEG signal. Instead, we conduct a test on the PLs  $v^1$  and  $v^2$  of the two phases:

$$(19) \quad H_0: \quad v^1 = v^2.$$

As we have only one epoch in each phase of the EEG signal, we require an inference framework on PLs incorporating a natural resampling approach. Since we estimate  $\mu_1(t)$  and  $\mu_2(t)$  by the degree- $k$  WFS  $\hat{\mu}_{T_u^1}^k(t)$  and  $\hat{\mu}_{T_u^2}^k(t)$  in (13) and (14), we resample in the frequency domain by randomly exchanging the frequency components in the two WFS. In other words, we randomly exchange the labels (Phase 1 and Phase 2) of the thresholded Fourier coefficients in (13) and (14) at each frequency:

$$\begin{array}{cccccc} A_{10} & \dots & A_{1k} & B_{10} & \dots & B_{1k} \\ \Downarrow & \dots & \Downarrow & \Downarrow & \dots & \Downarrow \\ A_{20} & \dots & A_{2k} & B_{20} & \dots & B_{2k}, \end{array}$$

and reconstruct two WFS in the time domain:

$$(20) \quad \begin{aligned} \hat{\mu}_{T_u^1}^{k'}(t) &= \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} A'_{1j} \cos(j\pi t/T) \\ &\quad + \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} B'_{1j} \sin(j\pi t/T), \end{aligned}$$

$$(21) \quad \begin{aligned} \hat{\mu}_{T_u^2}^{k'}(t) &= \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} A'_{2j} \cos(j\pi t/T) \\ &\quad + \sum_{j=0}^k e^{-(j\pi/T)^2\sigma} B'_{2j} \sin(j\pi t/T), \end{aligned}$$

where the  $A'_{1j}$ ,  $A'_{2j}$ ,  $B'_{1j}$ ,  $B'_{2j}$  are the resampled thresholded Fourier coefficients.

We measure the difference between the PLs  $v^1$  and  $v^2$  of the WFS  $\hat{\mu}_{T_u^1}^k(t)$  and  $\hat{\mu}_{T_u^2}^k(t)$  by their  $L_2$  distance:

$$(22) \quad L_2(v^1, v^2) = \left( \int \sum_{\ell=1}^N |v_\ell^1(\lambda) - v_\ell^2(\lambda)|^2 d\lambda \right)^{1/2},$$

where  $v_\ell^i$  is the  $\ell$ th layer of  $v^i, i = 1, 2$ , and  $N$  is the larger of the numbers of layers of the two PLs. For each permutation, we calculate the  $L_2(v^{1'}, v^{2'})$  between the PLs  $v^{1'}$  and  $v^{2'}$  of the two reconstructed WFS  $\hat{\mu}_{T_u^{1'}}^k(t)$  and  $\hat{\mu}_{T_u^{2'}}^k(t)$ . We then measure the statistical significance of  $L_2(v^1, v^2)$  between the PLs by comparing it with the distribution of the  $L_2(v^{1'}, v^{2'})$ . Two phases of a signal are said to be *topologically invariant* in the statistical sense if the difference between their PLs is not statistically significant.

In Section 4, we test out the performance of the proposed framework with respect to simulated ground truth. In Section 5, we use the proposed permutation test to study how statistical topological invariance manifests in and out of the epileptogenic zone in a single-trial epileptic EEG dataset.

3.3. *Baseline comparison.* In Sections 4.2 and 5, we compare the proposed inference framework with the following baseline statistical tests on standard EEG features.

*Paired t-test on local variance.* Local variance is a simplistic statistical approach that has given surprisingly accurate results in epileptogenic analysis [McSharry, Smith and Tarassenko (2003), Mohseni, Maghsoudi and Shamsollahi (2006)]. It is defined as

$$\sigma^2 = \langle \mathbf{x}^2 \rangle - \langle \mathbf{x} \rangle^2,$$

where  $\mathbf{x}$  is a signal at one channel and  $\langle \cdot \rangle$  is the average taken over a time window of a certain size. For inference, we split up the range into before and during seizure phases. Then we perform paired  $t$ -test on the local variances in the two phases of each channel.

*Paired t-test on spectrogram.* We first performed the discrete short-time Fourier transform (DSTFT) to obtain the windowed Fourier coefficients

$$X_{j,\ell} = (a_{j,\ell}, b_{j,\ell})$$

before and during seizure:

$$(23) \quad a_{j,\ell} = \sum_{n=0}^{L-1} w(n) f(x_{\ell+n}) \cos(2\pi jn/L),$$

$$(24) \quad b_{j,\ell} = \sum_{n=0}^{L-1} w(n) f(x_{\ell+n}) \sin(2\pi jn/L),$$

where  $j = 1, \dots, k$  ( $k$  is the degree of the discrete Fourier transform for a windowed segment),  $\ell$  is the number of time points by which the initial time points of consecutive segments differ and  $L$  is the pre-specified window size. The window functions  $w(n)$  are meant to smooth out discontinuities at boundaries. The most popular windows in practice are Hamming, Hanning, Kaiser and Gaussian [Oppenheim and Schaffer (1989)]. Here we used the Hamming window of pre-specified length  $L$ :

$$w(n) = 0.54 - 0.46 \cos(2\pi n/L), \quad 0 \leq n \leq L - 1.$$

A signal is driven through the window in individual segments; each segment is multiplied pointwise to the window function. The power spectral density (PSD) estimates are given by

$$(25) \quad |X_{j,\ell}|^2 = a_{j,\ell}^2 + b_{j,\ell}^2.$$

**4. Simulations.** The proposed inference framework tests whether the two phases of a single-trial EEG signal have different PLs. We conduct a permutation test on the PLs of replicates of the EEG signal in the time domain obtained through resampling in the frequency domain (by shuffling the respective Fourier coefficients of the signal in two phases, the procedure reconstructs two new signals in the time domain out of the permuted Fourier coefficients). The complexity of the framework calls for validation in two aspects: (1) reliability of this frequency resampling step at the presence of noise; and (2) the robustness and sensitivity of the topological permutation test under certain transformations of the underlying signal. In this section, we use two independent sets of simulation studies to assess the performance of the frequency resampling method and topological permutation test.

4.1. *Performance of frequency resampling method.* This set of simulations is meant to demonstrate that the frequency resampling approach preserves signal similarity and difference in the time domain. We simulate two signals  $f_1(t_i)$  and  $f_2(t_i)$  with  $-5 \leq t_1, \dots, t_{500} \leq 5$  by

$$(26) \quad f_1(t_i) = \sum_{j=0}^k e^{-(j\pi/5)^2\sigma} (c_{1j} \cos(j\pi t_i/5) + d_{1j} \sin(j\pi t_i/5)),$$

$$(27) \quad f_2(t_i) = \sum_{j=0}^k e^{-(j\pi/5)^2\sigma} (c_{2j} \cos(j\pi t_i/5) + d_{2j} \sin(j\pi t_i/5)),$$

where in each simulation

$$(28) \quad c_{ij} = a_{ij} + N(0, \sigma_c^2),$$

$$(29) \quad d_{ij} = b_{ij} + N(0, \sigma_d^2),$$

with fixed  $a_{ij} \sim N(0, \sigma_a^2)$  and  $b_{ij} \sim N(0, \sigma_b^2)$  for all simulations.

We model the simulated signals  $f_1$  and  $f_2$  with

$$f_1(t) = \mu_1(t) + \varepsilon_1(t) \quad \text{and} \quad f_2(t) = \mu_2(t) + \varepsilon_2(t), \quad -5 \leq t \leq 5,$$

and then test the null hypothesis

$$H_0: \quad \mu_1(t) = \mu_2(t) \quad \forall -5 \leq t \leq 5.$$

The first step is to estimate  $\mu_1$  and  $\mu_2$  with the WFS  $\hat{\mu}_{T_u^k}^k(t)$  and  $\hat{\mu}_{T_u^k}^k(t)$  in (13) and (14) ( $k = 499$ ,  $\sigma = 0.001$ ). We then resample in the frequency domain by randomly exchanging the thresholded Fourier coefficients in  $\hat{\mu}_{T_u^k}^k(t)$  and  $\hat{\mu}_{T_u^k}^k(t)$ . We compute the difference between the WFS  $\hat{\mu}_{T_u^k}^k(t)$  and  $\hat{\mu}_{T_u^k}^k(t)$  using the  $L_2$  distance

$$(30) \quad L_2(\hat{\mu}_{T_u^k}^k, \hat{\mu}_{T_u^k}^k) = \left( \int_{-5}^5 |\hat{\mu}_{T_u^k}^k(t) - \hat{\mu}_{T_u^k}^k(t)|^2 dt \right)^{1/2}.$$

With each random exchange of Fourier coefficients, we compute the  $L_2$  distance  $L_2(\hat{\mu}_{T_u^k}^{k'}, \hat{\mu}_{T_u^k}^{k'})$  between the reconstructed WFS  $\hat{\mu}_{T_u^k}^{k'}(t)$  and  $\hat{\mu}_{T_u^k}^{k'}(t)$  in (20) and (21). The  $p$ -value for each simulation is calculated as the proportion of the  $L_2(\hat{\mu}_{T_u^k}^{k'}, \hat{\mu}_{T_u^k}^{k'})$  that exceed  $L_2(\hat{\mu}_{T_u^k}^k, \hat{\mu}_{T_u^k}^k)$ . Note that the WFS estimation and frequency resampling in the testing procedure are identical to the test on PLs described in Section 3.2.

As a performance measure, we collect the  $p$ -values in 1000 simulations and compute the percentages of those below 0.05. Due to different magnitudes of variance, we expect the percentages in Study 1 and Study 2 to be small and large respectively.

*Study 1.* In this setting, the noise components are generated with

$$\sigma_a = \sigma_b = 1, \quad \sigma_c = 0.00001, \quad \sigma_d = 0.00001$$

for signals constructed at  $k = 99$  and  $\sigma = 0$ . The percentage of  $p$ -values below 0.05 is 1%.

*Study 2.* In this setting, the noise components are generated with

$$\sigma_a = 1, \quad \sigma_b = 10, \quad \sigma_c = 0.00001, \quad \sigma_d = 0.00001$$

for signals constructed at  $k = 99$  and  $\sigma = 0$ . The percentages of  $p$ -values below 0.05 is 100%.

*Discussion of results.* The results demonstrate that the frequency resampling approach preserves signal similarity and difference in the time domain under reasonable noise contamination.

*4.2. Performance of topological permutation test.* The TDA inference framework proposed in Section 3.2 tests the difference between the PLs of two signals.

We want to set a topological criterion to evaluate the performance of the test in simulation studies. A standard performance measure for a statistical test is the false positive rate. Here, we want to control the rate of “topological false positives”; in other words, we require the proposed test on PLs to stay robust when the underlying signals have identical PH. We know from algebraic topology that the homology of a topological space is preserved under continuous transformation (without tearing or glueing) of the space [Hatcher (2002), Turner, Mukherjee and Boyer (2014)]. Yet, continuity alone in signal transformation does not guarantee preservation of PH in the sublevel set of the signal. To preserve PH, signal transformations also need to respect the pairing of birth and death times of connected components in the sublevel sets—we call these *topology-preserving transformations* of a signal. In Study 1.1, 1.2 and 1.3, we check the false positive rates of the proposed test under the most basic topology-preserving transformations of signals: amplitude translation and scaling, and frequency scaling. Defining general topology-preserving transformations is part of our ongoing research. On the other hand, there are indefinitely many ways of defining transformations that destroy the PH in the sublevel set of a signal. For the purpose of performance evaluation, we define *tearing of a signal* as a transformation that pointwise multiplies the signal by a step function with all piecewise time intervals containing at least one but not all critical points of the signal. Note that tearing a signal results in discontinuities, but all signals undergo WFS-denoising in the proposed framework, which ensures the smoothness of the denoised signal with reduced Gibbs phenomenon. In Study 2, we test the power of the proposed test on tearing-transformed signals.

In each study, we also compare the proposed TDA framework with two baseline methods on the signals.

*Study 1. Topological invariance.* In each study, we first simulate a signal  $g(t)$  with unique critical points. The second signal is simulated in three settings with respect to three basic types of topology-preserving transformations of  $g(t)$ :

1. *amplitude translation:*

$$h(t) = g(t) + a, \quad a \in \mathbb{R};$$

2. *amplitude scaling:*

$$h(t) = cg(t);$$

for  $c > 1$  ( $c < 1$ ), the function is stretched (squeezed) in the amplitude.

3. *frequency scaling:*

$$h(t) = g(\omega t);$$

for  $\omega < 1$  ( $\omega > 1$ ), the function is stretched (squeezed) in the direction of  $t$ .

For each parameter setting of a study, 1000 datasets are simulated. In each dataset, two blocks of signals are generated according to the topologically invari-

ant transformations. Independent Gaussian noises are added to each signal at the time points where the signal is generated. We test the null hypothesis of equality between PLs

$$H_0 : \nu^1 = \nu^2.$$

The testing procedure is as described in Section 3.2. Percentages out of the 1000 simulations are then computed for  $p$ -values below 0.05. Percentages below 5% are considered as good performance in robustness.

*Study 1.1. Amplitude translation.* We simulate four pairs of signals  $y_1(t_i)$  and  $y_2(t_i)$ , at regular time intervals  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$(31) \quad \begin{aligned} y_1(t_i) &= t_i \cos(\omega t_i), \\ y_2(t_i) &= t_i \cos(\omega t_i) + 10, \end{aligned}$$

where  $\omega$  takes on four values: (1)  $\omega = 1$ ; (2)  $\omega = 2$ ; (3)  $\omega = 5$ ; (4)  $\omega = 10$ . In each simulation, independent Gaussian noises  $N(0, 2^2)$  are added to the signals at  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$\begin{aligned} y_1(t_i) &= t_i \cos(\omega t_i) + \epsilon_{1i}, \\ y_2(t_i) &= t_i \cos(\omega t_i) + 10 + \epsilon_{2i}, \end{aligned}$$

where  $\epsilon_{1i}, \epsilon_{2i} \sim N(0, 2^2)$ . Figure 4 shows examples of noisy signals in each simulation.

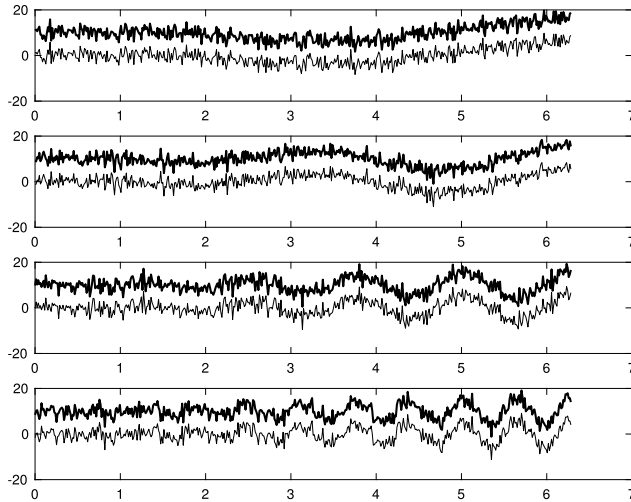


FIG. 4. Study 1.1—amplitude translation—examples of noisy translated signals generated from (31). Tests are performed between each pair of curves with independent Gaussian noises in each simulation.



TABLE 1

Study 1.1—amplitude translation—percentages of  $p$ -values  $< 0.05$  in 1000 simulated datasets by baseline methods and the topological permutation test

Percentages of $p$ -value $< 0.05$	$\omega = 1$	$\omega = 2$	$\omega = 5$	$\omega = 10$
Paired $t$ -test on local variance	27.2%	25%	19.8%	16.7%
Paired $t$ -test on PSD estimates	100%	100%	100%	100%
Topological permutation test	0.7%	0.3%	0.2%	0.3%

Results of percentages of  $p$ -values below 0.05 by baseline methods and the proposed topological permutation test are summarized in Table 1. The results show that the paired  $t$ -tests on local variance and PSD estimates are sensitive to translation, whereas the topological permutation test is robust to the topology-preserving transformation.

Study 1.2. *Amplitude scaling.* We simulate four pairs of signals  $y_1(t_i)$  and  $y_2(t_i)$ , at regular time intervals  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$(32) \quad \begin{aligned} y_1(t_i) &= t_i \cos(\omega t_i), \\ y_2(t_i) &= 2t_i \cos(\omega t_i), \end{aligned}$$

where  $\omega$  takes on four values: (1)  $\omega = 1$ ; (2)  $\omega = 2$ ; (3)  $\omega = 5$ ; (4)  $\omega = 10$ . In each simulation, independent Gaussian noises  $N(0, 2^2)$  are added to the signals at  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$\begin{aligned} y_1(t_i) &= t_i \cos(\omega t_i) + \epsilon_{1i}, \\ y_2(t_i) &= 2t_i \cos(\omega t_i) + \epsilon_{2i}, \end{aligned}$$

where  $\epsilon_{1i}, \epsilon_{2i} \sim N(0, 2^2)$ . Figure 5 shows examples of noisy signals. Results of percentages of  $p$ -values below 0.05 by baseline methods and the topological permutation test are summarized in Table 2. The results show that the paired  $t$ -tests on local variance and PSD estimates are sensitive to amplitude scaling, whereas the topological permutation test is robust to the topology-preserving transformation.

Study 1.3. *Frequency scaling.* We simulate four signals  $y(t_i)$  at regular time intervals  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$(33) \quad y(t_i) = t_i \cos(\omega t_i),$$

where  $\omega$  takes on four values: (1)  $\omega = 1$ ; (2)  $\omega = 2$ ; (3)  $\omega = 5$ ; (4)  $\omega = 10$ . In each simulation, independent Gaussian noise  $N(0, 2^2)$  are added to the signals at  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$y(t_i) = t_i \cos(\omega t_i) + \epsilon_i,$$

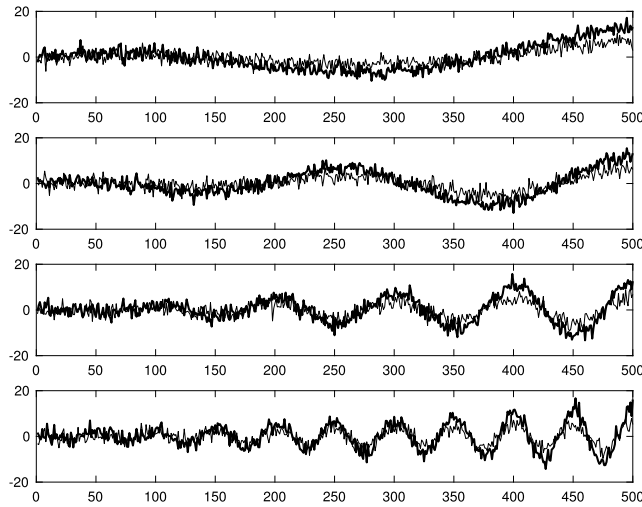


FIG. 5. Study 1.2—amplitude scaling—examples of noisy scaled signals generated from (32). Tests are performed between each pair of curves with independent Gaussian noises in each simulation.

TABLE 2  
Study 1.2—amplitude scaling—percentages of  $p$ -values  $< 0.05$  in 1000 simulated datasets by baseline methods and the topological permutation test

Percentages of $p$ -value $< 0.05$	$\omega = 1$	$\omega = 2$	$\omega = 5$	$\omega = 10$
Paired $t$ -test on local variance	33.7%	63.1%	100%	100%
Paired $t$ -test on PSD estimates	100%	100%	100%	100%
Topological permutation test	0%	0%	0%	0%

where  $\epsilon_i \sim N(0, 2^2)$ . Figure 6 shows an example of the four signals without and with noise in each simulation.

Results of percentages of  $p$ -values below 0.05 are summarized in Table 3. The results show that the paired  $t$ -test on local variance is sensitive to frequency scaling, whereas the paired  $t$ -test on PSD estimates and the topological permutation test are robust to the topology-preserving transformation.

Study 2. Topological difference. We simulate four pairs of signals  $y_1(t_i)$  and  $y_2(t_i)$ , at regular time intervals  $0 \leq t_1, \dots, t_{500} \leq 2\pi$ :

$$\begin{aligned}
 & y_1(t_i) = t \cos(\omega t_i), \\
 (34) \quad & y_2(t_i) = \begin{cases} y_1(t_i), & 0 \leq t_i \leq 0.4\pi, 1.6\pi < t_i \leq 2\pi; \\ y_1(t_i) - 200, & 0.4\pi < t_i \leq 0.96\pi, 1.04\pi < t_i \leq 1.6\pi, \\ y_1(t_i) + 200, & 0.96\pi < t_i \leq 1.04\pi, \end{cases}
 \end{aligned}$$

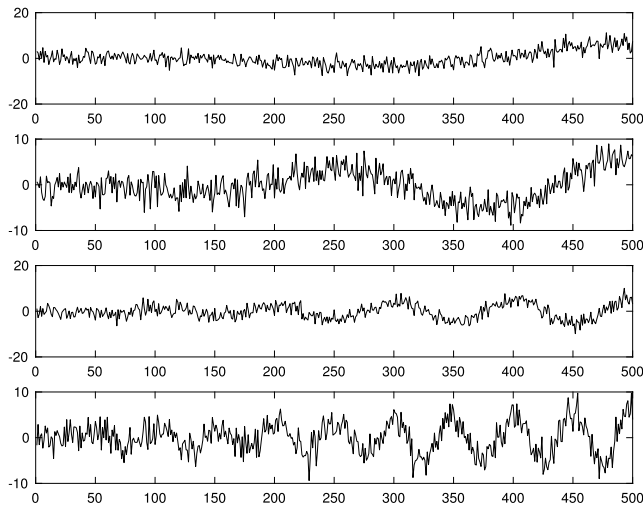


FIG. 6. Study 1.3—frequency scaling—examples of noisy signals generated from (33). Tests are performed between the four curves with independent Gaussian noises in each simulation.

TABLE 3  
Study 1.3—frequency scaling—percentages of  $p$ -values  $< 0.05$  in 1000 simulated datasets by baseline methods and the topological permutation test

Percentages of $p$ -value $< 0.05$	$\omega = 1$ vs $\omega = 2$	$\omega = 1$ vs $\omega = 5$	$\omega = 1$ vs $\omega = 10$
Paired $t$ -test on local variance	14.7%	79.8%	99.9%
Paired $t$ -test on PSD estimates	0%	0%	1.8%
Topological permutation test	0%	0%	0.1%

where  $\omega$  takes on one of four values: (1)  $\omega = 1$ ; (2)  $\omega = 2$ ; (3)  $\omega = 5$ ; (4)  $\omega = 10$ . In each simulation, independent Gaussian noise  $N(0, 50^2)$  are added to the signals at  $0 \leq t_1, \dots, t_{500} \leq 2\pi$  (Figure 7). We test whether  $y_1$  and  $y_2$  have different PLs.

Percentages of  $p$ -values below 0.05 by different tests are summarized in Table 4. The results show that all three tests are fairly sensitive to the topological difference in the signals at a threshold of 95%.

*Discussion.* In all four simulation studies, the performance of the topological permutation test shows low rates of detecting false topological difference and high power in detecting true topological difference. The results are also far more stable than those of the two baseline methods.

**5. Application to epileptic EEGs.** In this section, we use the proposed permutation test to study how topological invariance manifests in and out of the epileptogenic zone in a single-trial epileptic EEG dataset.

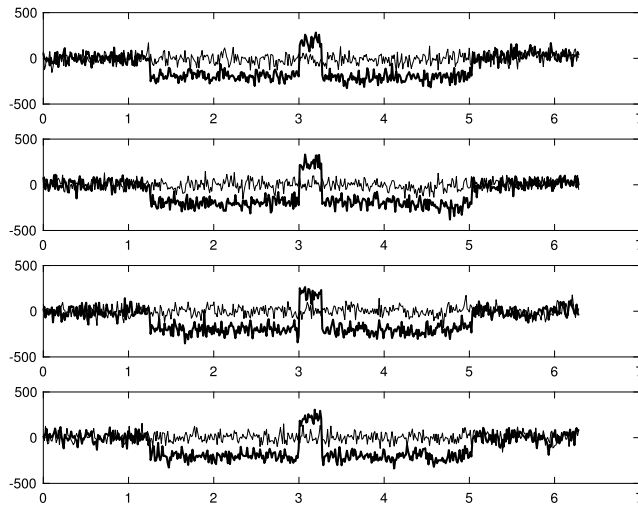


FIG. 7. Examples of the four signals given by (34) with noise. Tests are performed on each pair of curves with independent Gaussian noises in each simulation.

TABLE 4

Study 2—percentages of  $p$ -values  $< 0.05$  in 1000 simulated datasets by baseline methods and the topological permutation test

Percentages of $p$ -value $< 0.05$	$\omega = 1$	$\omega = 2$	$\omega = 5$	$\omega = 10$
Paired $t$ -test on local variance	91.7%	91.5%	91.2%	90.8%
Paired $t$ -test on PSD estimates	100%	100%	100%	100%
Topological permutation test	90%	100%	93%	100%

5.1. Description of the EEG data. EEG signals are sampled at the rate of 100 Hz (1 sample per 10 millisecond) from channels C3,C4,Cz,P3,P4,T3,T4,T5 in the central, parietal and temporal brain regions (Figure 8) of a female patient of Dr. Malow, an attending neurologist at the University of Michigan. Due to a lesion located on the cortical surface of the patient below the T3 channel, epileptic seizures are more likely to initiate from the left temporal lobe and abnormal electrical fluctuations are expected in EEG signals from channels around T3. During the entire EEG recording of 32,680 time points, a seizure initiates approximately halfway at the left temporal site (T3 channel). Visual inspection of Figure 9 (left) shows the signals before seizure to be more stable (more stationary, lower variation, smaller magnitude of the waveforms) than the later period. Highly volatile oscillations in the seizure period also seem to be concentrated in channels located near the T3 channel.

In subsequent analysis, the first and second phases of 16,340 time points in the EEG recording are designated as “before” and “during” seizure.

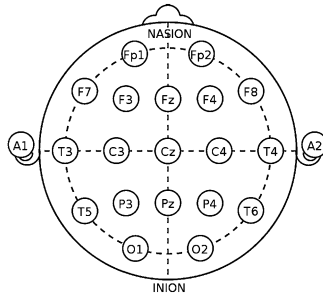


FIG. 8. EEG channels in accordance with the international 10–20 system;  $T$  = temporal,  $P$  = parietal,  $C$  = central; odd (even) numbers indicate the left (right) hemisphere. Image source: [https://en.wikipedia.org/wiki/10-20\\_system\\_\(EEG\)](https://en.wikipedia.org/wiki/10-20_system_(EEG)).

**5.2. Topological data analysis.** We first analyze the eight EEG signals by the proposed topological permutation test. As detailed in Section 3, each signal in the dataset undergoes denoising by a thresholded WFS (Figure 9 (right)). Sine or cosine waveforms are included in the representation if the corresponding amplitudes exceed some threshold. PLs are then constructed on each of the denoised signals (Figure 10). In the topological permutation test, the  $p$ -values are computed by counting the proportion of frequency-resampled PLs having  $L_2$  distances exceeding that of the observed PLs before and during seizure. The results obtained through 10,000 resamples with respect to multiple combinations of WFS degrees and bandwidths are summarized in Table 5. Our approach here was to be very conservative and thus we imposed a Bonferroni corrected 5% significance level for the eight simultaneous topological permutation tests.

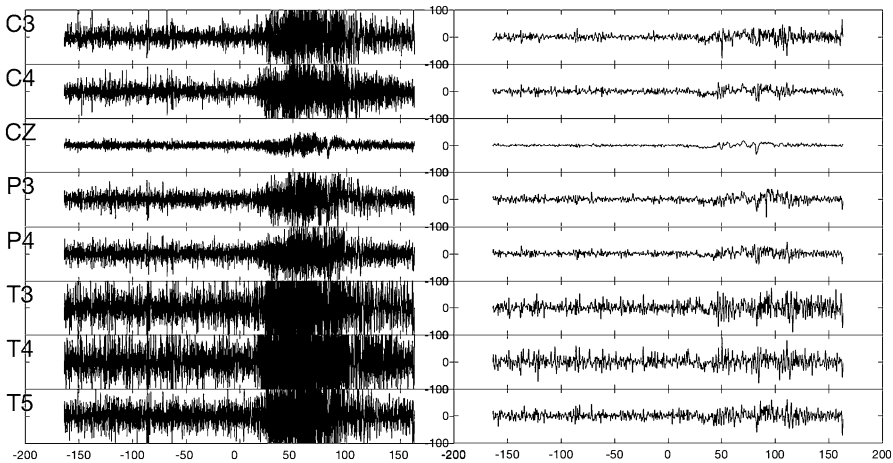


FIG. 9. Left: EEG recording with a sampling rate of 100 Hz over 163.4 s; time 0 indicates the start of the seizure attack. Right: thresholded WFS denoising with degree  $k = 499$  and bandwidth  $\sigma = 0.0001$ .

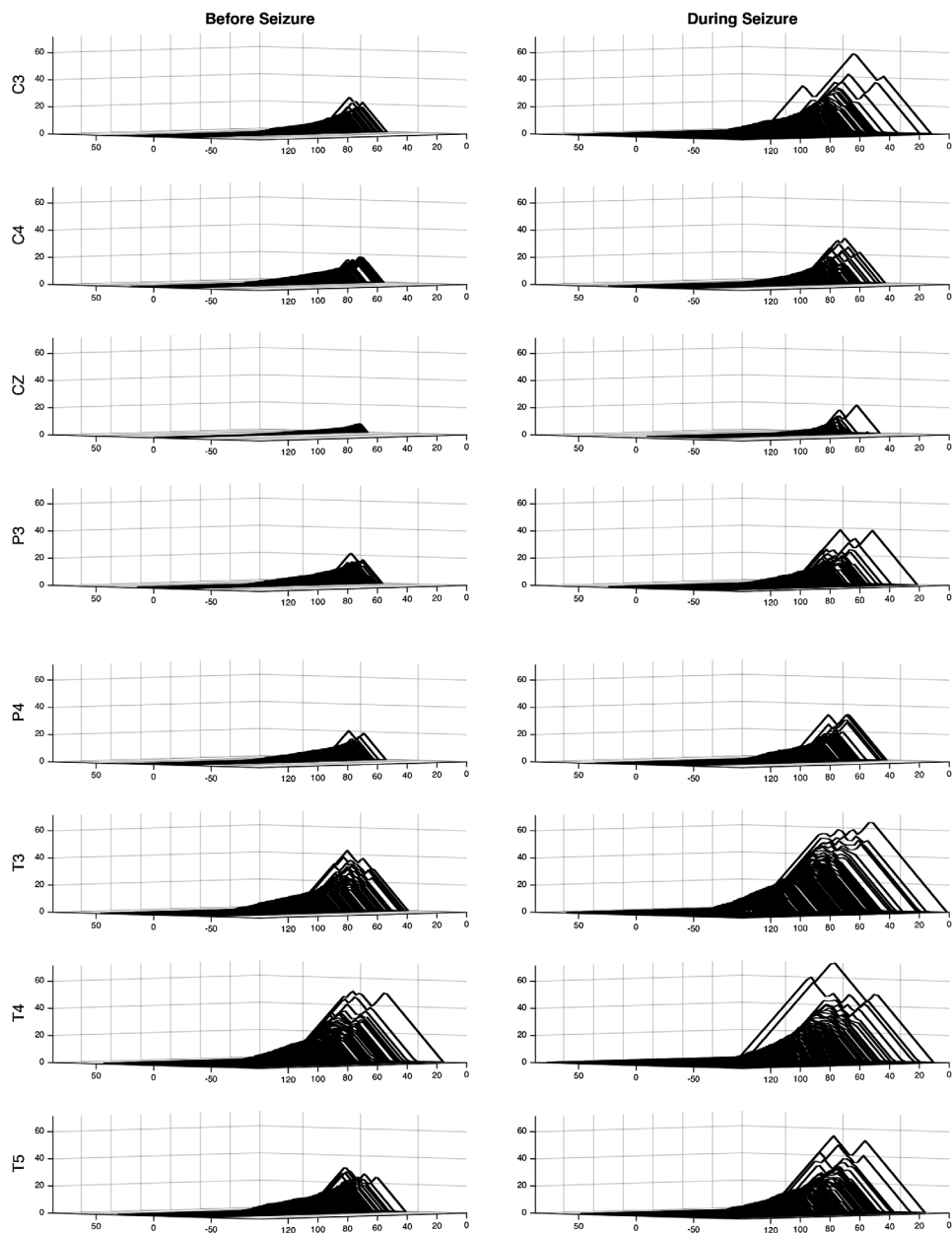


FIG. 10. Persistence landscapes at all channels before and during seizure for signals denoised with WFS of degree  $k = 499$  and bandwidth  $\sigma = 0.0001$ .

TABLE 5

Summary of  $p$ -values computed with the proposed topological method with respect to different combination of thresholded WFS degrees and bandwidths. The  $p$ -values above the Bonferroni threshold of  $0.05/8 = 0.0063$  are shaded in gray. We observe that T3, T4 and T5 consistently show up as topologically invariant before and during seizure according to the Bonferroni threshold

		$k = 99$	$k = 499$	$k = 999$
$\sigma = 0.0005$	C3	0.0001	0.0001	0.0001
	C4	0.0009	0.0014	0.0026
	Cz	0.0001	0.0001	0.0001
	P3	0.0072	0.0020	0.0022
	P4	0.0003	0.0001	0.0003
	T3	0.1975	0.0960	0.0934
	T4	0.0255	0.0228	0.0464
	T5	0.0435	0.1522	0.1723
$\sigma = 0.0001$	C3	0.0057	0.0010	0.0006
	C4	0.0054	0.0639	0.0445
	Cz	0.0001	0.0001	0.0001
	P3	0.0029	0.0001	0.0001
	P4	0.0001	0.3238	0.2848
	T3	0.0661	0.0406	0.0682
	T4	0.0162	0.1286	0.1397
	T5	0.1044	0.3065	0.2608

An EEG channel is said to have *topological invariance before and during seizure* if the  $p$ -value of the proposed permutation test at the channel is insignificant at the Bonferroni corrected 5% significance level. We observe that T3, T4 and T5 consistently show topological invariance (i.e., the topology before seizure is highly similar to that during seizure) after the Bonferroni correction. In particular, T3 and T5 have the least significant difference between PLs compared with T4. The observation coincides with the diagnosis of the left temporal lobe being the epileptogenic zone. It suggests that EEG signals in the epileptogenic zone *even before* a seizure attack already have similar topological patterns as those during the seizure. There is a strong likelihood that the seizure originates from a neighborhood around the left temporal and left central regions (as suggested by the method) but due to volume conduction in EEGs, one cannot precisely pinpoint the location of the seizure. The fact that T4, which is symmetric to T3, shows up as vaguely topologically invariant suggests the seizure is spreading from the left temporal region. Neuroscientific research has shown evidence that the path followed by spontaneous seizure activity in spreading to other structures is a feature peculiar to each individual [Brazier (1972)]. It is interesting that the proposed method was able to capture these features of the EEG signals which were overlooked by other methods that previously analyzed this same dataset.

To investigate the robustness of our method, additional analysis was also performed on the middle 24,000 and 30,000 instead of the full 32,680 time points of each of the eight EEG signals denoised with WFS of degree  $k = 499$  and bandwidths  $\sigma = 0.0005$  and  $\sigma = 0.0001$ . Channels with  $p$ -values above the Bonferroni threshold  $0.05/8 = 0.0063$  are again considered as topologically invariant. For the bandwidth  $\sigma = 0.0001$ , the spatial pattern of channel topological invariance remains the same for the two datasets of shorter lengths; for  $\sigma = 0.0005$ , the pattern varies while T5 remains topologically invariant.

5.3. *Baseline comparison.* As described in Section 3.3, we compare the results of the proposed inference procedure with two baseline statistical tests of standard EEG features.

*Paired t-test on local variance.* Table 6 summarizes the results of  $p$ -values of the paired  $t$ -tests on local variances of raw signals. The results are not informative as all the  $p$ -values are uniformly too small.

*Paired t-test on spectrogram.* Table 7 summarizes the results of  $p$ -values of the paired  $t$ -tests on PSD estimates of raw signals. The results are not informative as all the  $p$ -values are uniformly too small.

*Results on denoised signals.* The same methods were applied to signals denoised with WFS at degree 99 and bandwidth 0.001. Table 8 summarizes the results of  $p$ -values of the paired  $t$ -tests on local variances and PSD estimates of signals denoised with WFS. Although the overall conclusions remain the same, we can see that denoising raises the  $p$ -values dramatically.

TABLE 6  
 Summary of  $p$ -values of the paired  $t$ -tests  
 on local variances of raw signals

Channel	$p$ -values (in $10^{-45}$ )
C3	0.0001
C4	0.0001
Cz	0.0001
P3	0.0001
P4	0.0001
T3	0.6721
T4	0.0001
T5	0.0001



TABLE 7  
*Summary of  $p$ -values of the paired  $t$ -tests  
on PSD estimates of raw signals*

Channel	$p$ -values (in $10^{-18}$ )
C3	0.0001
C4	0.1112
Cz	0.0001
P3	0.0001
P4	0.0001
T3	0.0001
T4	0.0001
T5	0.0001

**6. Discussion.** This work explores the topological information in epileptic EEGs through the evolution of connected components in the signals. The novelty and contribution of the proposed procedure are unique in several aspects. Denoising univariate EEG signals by a thresholded WFS helps stabilize the subsequent analysis of connected components in the signals. The frequency resampling procedure based on the WFS serves the purpose of preserving waveforms characteristic of epileptic discharges.

The data application shows topological invariance (topological difference) before and during seizure in (out of) the epileptogenic zone—sites close to the brain lesion causing the epileptic seizures. The analysis adds new insight on epileptogenicity as existing neurophysiological research shows only that the seizure onset zone has isolated electric patterns from other brain sites at the onset of a seizure [Lange et al. (1983), Van Quyen et al. (2001), Burns et al. (2014)], and previous analysis on the same epileptic EEG data shows only the evolution of the spectral power and coherence during the seizure episode but does not identify seizure

TABLE 8  
*Summary of  $p$ -values of the paired  $t$ -tests on local variances of signals  
denoised with WFS of degree 99 and bandwidth 0.001*

Channel	Local variance (in $10^{-7}$ )	PSD estimates (in $10^{-7}$ )
C3	0.0001	0.0012
C4	0.0001	0.2013
Cz	0.0001	0.0472
P3	0.0001	0.0192
P4	0.0001	0.0245
T3	0.0001	0.0004
T4	0.4691	0.2884
T5	0.0001	0.2701

location [Ombao et al. (2001), Ombao, von Sachs and Guo (2005)]. More importantly, our method is shown to be reliable in detecting true topological changes with noise. As shown by simulation studies in Section 4.2, results from a baseline statistical test on local variance and spectrogram is sensitive to amplitude and frequency changes even when they do not correspond to topological changes, such as translation and scaling of signals. In contrast, the proposed inference framework stays robust to nontopological changes and sensitive to true topological changes.

In the current application, we demonstrate the robustness of the proposed procedure by testing it out for various combinations of large degree  $k$  and small bandwidth  $\sigma$ . A data-driven automatic selection of these parameters is beyond the scope of this paper. We recommend practitioners base their applications with respect to a reasonable range of  $k$  and  $\sigma$ . For future methodological extension, a direction worth pursuing is to weigh PLs across a range of  $k$  and  $\sigma$  to obtain a topological overview over the parameter space.

It is important to note that the purpose of the current application is to demonstrate the potential utility of the proposed TDA framework in epileptogenic studies per individual patients. Neurologists may use this tool to guide them to investigate further the regions in a patient's brain that show topological invariance before and during the seizure episode. However, conclusions on the clinical population of epilepsy patients should be avoided [Maris (2012)]. As part of our ongoing work, we are testing the proposed framework on large samples of multi-trial EEG recordings.

**Acknowledgments.** The authors would like to thank Robert Nowak and Xiaojin Zhu at University of Wisconsin–Madison and Steve Marron at University of North Carolina–Chapel Hill for their valuable comments and discussion on the methods in this paper, and neurologists Erwin Montgomery at McMaster University and Nigel Pedersen at Emory University for their expert opinion on epilepsy. We would also like to thank the four anonymous reviewers, and the Area and Associate Editors for highly constructive comments for us to substantially improve the paper.

## REFERENCES

- ABRAMOVICH, F. and BENJAMINI, Y. (1996). Adaptive thresholding of wavelet coefficients. *Comput. Statist. Data Anal.* **22** 351–361. [MR1411575](#)
- ADLER, R. J., BOBROWSKI, O., BORMAN, M. S., SUBAG, E. and WEINBERGER, S. (2010). Persistent homology for random fields and complexes. In *Borrowing Strength: Theory Powering Applications—a Festschrift for Lawrence D. Brown. Inst. Math. Stat. (IMS) Collect.* **6** 124–143. IMS, Beachwood, OH. Available at [arXiv:1003.1001](#). [MR2798515](#)
- AHMED, M., FASY, B. T. and WENK, C. (2014). Local persistent homology based distance between maps. In *Proceedings of the 22nd ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems* 43–52.
- BANCAUD, J., BRUNET-BOURGIN, F., CHAUVEL, P. and HALGREN, E. (1994). Anatomical origin of déjà vu and vivid ‘memories’ in human temporal lobe epilepsy. *Brain* **117** 71–90.

- BARTOLOMEI, F., CHAUVEL, P. and WENDLING, F. (2008). Epileptogenicity of brain structures in human temporal lobe epilepsy: A quantified study from intracerebral EEG. *Brain* **131** 1818–1830.
- BENDICH, P., MARRON, J. S., MILLER, E., PIELOCH, A. and SKWERER, S. (2016). Persistent homology analysis of brain artery trees. *Ann. Appl. Stat.* **10** 198–218. [MR3480493](#)
- BOTEV, Z. I., GROTHOWSKI, J. F. and KROESE, D. P. (2010). Kernel density estimation via diffusion. *Ann. Statist.* **38** 2916–2957. [MR2722460](#)
- BRAZIER, M. A. B. (1972). Spread of seizure discharges in epilepsy: Anatomical and electrophysiological considerations. *Exp. Neurol.* **36** 263–272.
- BUBENIK, P. (2015). Statistical topological data analysis using persistence landscapes. *J. Mach. Learn. Res.* **16** 77–102. [MR3317230](#)
- BUBENIK, P., CARLSON, G., KIM, P. T. and LUO, Z.-M. (2010). Statistical topology via Morse theory persistence and nonparametric estimation. In *Algebraic Methods in Statistics and Probability II. Contemporary Mathematics*. **516** 75–92.
- BURNS, S. P., SANTANIELLO, S., YAFFE, R. B., JOUNY, C. C., CRONE, N. E., BERGEY, G. K., ANDERSON, W. S. and SARMA, S. V. (2014). Network dynamics of the brain and influence of the epileptic seizure onset zone. *Proc. Natl. Acad. Sci. USA* **111** E5321–E5330.
- CARLSSON, G. (2009). Topology and data. *Bull. Amer. Math. Soc. (N.S.)* **46** 255–308. [MR2476414](#)
- CHAUDHURI, P. and MARRON, J. S. (2000). Scale space view of curve estimation. *Ann. Statist.* **28** 408–428. [MR1790003](#)
- CHAZAL, F., FASY, B. T., LECCI, F., MICHEL, B., RINALDO, A. and WASSERMAN, L. (2014). Subsampling methods for persistent homology. Available at [arXiv:1406.1901](#).
- CHUNG, M. K. (2014). *Statistical and Computational Methods in Brain Image Analysis*. Chapman & Hall/CRC, London.
- CHUNG, M. K., BUBENIK, P. and KIM, P. T. (2009). Persistence diagrams of cortical surface data. In *Proceedings of the 21st International Conference on Information Processing in Medical Imaging (IPMI)* 386–397.
- CHUNG, M., DALTON, K., SHEN, L., EVANS, A. C. and DAVIDSON, R. J. (2007). Weighted Fourier series representation and its application to quantifying the amount of gray matter. *IEEE Trans. Med. Imag.* **26** 566–581.
- CHUNG, M. K., SCHAEFER, S. M., VAN REEKUM, C. M., PESCHKE-SCHMITZ, L., SUTTERER, M. J. and DAVIDSON, R. J. (2014). A unified kernel regression for diffusion wavelets on manifolds detects aging-related changes in the amygdala and hippocampus. In *Proceedings of the 17th International Conference on Medical Image Computing and Computer Assisted Intervention (MICCAI)*. *LNCS* **8674** 791–798.
- CHUNG, M. K., HANSON, J. L., YE, J., DAVIDSON, R. J. and POLLAK, S. D. (2015). Persistent homology in sparse regression and its application to brain morphometry. *IEEE Trans. Med. Imag.* **34** 1928–1939.
- COHEN-STEINER, D. and EDELSBRUNNER, H. (2009). Lipschitz functions have  $L_p$ -stable persistence. *Found. Comput. Math.* **10** 127–139.
- COHEN-STEINER, D., EDELSBRUNNER, H. and HARER, J. (2007). Stability of persistence diagrams. *Discrete Comput. Geom.* **37** 103–120.
- DONOHO, D. L. and JOHNSTONE, I. M. (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika* **81** 425–455. [MR1311089](#)
- DONOHO, D. L. and JOHNSTONE, I. M. (1995). Adapting to unknown smoothness via wavelet shrinkage. *J. Amer. Statist. Assoc.* **90** 1200–1224.
- DONOHO, D. L., MALLAT, S. and VON SACHS, R. (1998). Estimating covariances of locally stationary processes: Rates of convergence of best basis methods. Technical report, Dept. Statistics, Stanford Univ. Stanford, CA.
- EDELSBRUNNER, H. and HARER, J. L. (2010). *Computational Topology: An Introduction*. Amer. Math. Soc., Providence, RI. [MR2572029](#)

- EDELSBRUNNER, H., LETSCHER, D. and ZOMORODIAN, A. (2002). Topological persistence and simplification. *Discrete Comput. Geom.* **28** 511–533. [MR1949898](#)
- FRIED, I. (1997). Auras and experiential responses arising in the temporal lobe. *J. Neuropsychiatry Clin. Neurosci.* **9** 420–428.
- GAMBLE, J. and HEO, G. (2010). Exploring uses of persistent homology for statistical analysis of landmark-based shape data. *J. Multivariate Anal.* **101** 2184–2199.
- HATCHER, A. (2002). *Algebraic Topology*. Cambridge Univ. Press, Cambridge.
- HEO, G., GAMBLE, J. and KIM, P. T. (2012). Topological analysis of variance and the maxillary complex. *J. Amer. Statist. Assoc.* **107** 477–492. [MR2980059](#)
- KHALID, A., KIM, B. S., CHUNG, M. K., YE, J. C. and JEON, D. (2014). Tracing the evolution of multi-scale functional networks in a mouse model of depression using persistent brain network homology. *NeuroImage* **101** 351–363.
- KOBAU, R., LUO, Y., ZACK, M., HELMERS, S. and THURMAN, D. (2012). Epilepsy in adults and access to care—United States, 2010. *Morb. Mort. Wkly. Rep.* **61** 910–913.
- LANGE, H., LIEB, J., ENGEL, J. J. and CRANDALL, P. (1983). Temporo-spatial patterns of pre-ictal spike activity in human temporal lobe epilepsy. *Electroencephalogr. Clin. Neurophysiol.* **1978** 543–555.
- LEE, H., CHUNG, M. K., KANG, H., KIM, B. N. and LEE, D. S. (2011). Computing the shape of brain networks using graph filtration and Gromov–Hausdorff metric. In *Proceedings of the 14th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*. *LNCS* **6892** 302–309.
- MARIS, E. (2012). Statistical testing in electrophysiological studies. *Psychophysiology* **49** 549–565.
- MARTINERIE, J., ADAM, C., LE VAN QUYEN, M., BAULAC, M., CLÉMENCEAU, S., RENAULT, B. and VARELA, F. (1998). Can epileptic seizure be anticipated by nonlinear analysis? *Nat. Med.* **4** 1173–1176.
- MC SHARRY, P. E., SMITH, L. A. and TARASSENKO, L. (2003). Prediction of epileptic seizures: Are nonlinear methods relevant? *Nat. Med.* **9** 241–242; author reply 242.
- MILEYKO, Y., MUKHERJEE, S. and HARER, J. (2011). Probability measures on the space of persistence diagrams. *Inverse Probl.* **27** 1–21.
- MILNOR, J. (1963). *Morse Theory*. Princeton Univ. Press, Princeton.
- MITRA, P. P. and PESARAN, B. (1999). Analysis of dynamic brain imaging data. *Biophys. J.* **76** 691–708.
- MOHSENI, H. R., MAGHSOUDI, A. and SHAMSOLLAHI, M. B. (2006). Seizure detection in EEG signals: A comparison of different approaches. In *Proceedings of the 28th Annual International Conference of the IEEE Engineering in Medicine and Biology Society* 6724–6727.
- OMBAO, H., VON SACHS, R. and GUO, W. (2005). SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.* **100** 519–531.
- OMBAO, H. C., RAZ, J. A., VON SACHS, R. and MALOW, B. A. (2001). Automatic statistical analysis of bivariate nonstationary time series. *J. Amer. Statist. Assoc.* **96** 543–560.
- OPPENHEIM, A. V. and SCHAFER, R. W. (1989). *Discrete-Time Signal Processing*. Prentice Hall, New York.
- REININGHAUS, J., HUBER, S., BAUER, U., TU, M. and KWITT, R. (2015). A stable multi-scale kernel for topological machine learning. Available at [arXiv:1412.6821](#).
- SOUSBIE, T., PICHON, C. and KAWAHARA, H. (2011). The persistent cosmic web and its filamentary structure. *Mon. Not. R. Astron. Soc.* **414** 384–403.
- TURNER, K., MUKHERJEE, S. and BOYER, D. M. (2014). Persistent homology transform for modeling shapes and surfaces. *Inf. Inference* **3** 310–344.
- TURNER, K., MILEYKO, Y., MUKHERJEE, S. and HARER, J. (2014). Frechet means for distributions of persistence diagrams. *Discrete Comput. Geom.* **52** 44–70.
- VAN QUYEN, M. L., MARTINERIE, J., NAVARRO, V., BAULAC, M. and VARELA, F. J. (2001). Characterizing neurodynamic changes before seizures. *J. Clin. Neurophysiol.* **18** 191–208.

- WHO (2005). Atlas: Epilepsy care in the world. Technical report.
- WORSLEY, K. J. (1995). Estimating the number of peaks in a random field using the Hadwiger characteristic of excursion sets, with applications to medical images. *Ann. Statist.* **23** 640–669. [MR1332586](#)
- ZHU, X., VARTANIAN, A., BANSAL, M., NGUYEN, D. and BRANDL, L. (2016). Stochastic multi-resolution persistent homology kernel. In *Proceedings of the 25th International Joint Conference on Artificial Intelligence (IJCAI)* 2449–2455.

Y. WANG  
M. K. CHUNG  
DEPARTMENT OF BIostatISTICS AND  
MEDICAL INFORMATICS  
UNIVERSITY OF WISCONSIN–MADISON  
MADISON, WISCONSIN 53705  
USA  
E-MAIL: [ywang259@wisc.edu](mailto:ywang259@wisc.edu)  
[mkchung@wisc.edu](mailto:mkchung@wisc.edu)

H. Ombao  
DEPARTMENT OF STATISTICS  
UNIVERSITY OF CALIFORNIA, IRVINE  
IRVINE, CALIFORNIA 92697  
USA  
E-MAIL: [hombao@uci.edu](mailto:hombao@uci.edu)