Manuscript submitted to
AIMS' Journals
Volume **X**, Number **0X**, XX **20XX**

doi:10.3934/xx.xxxxxxx

pp. **X–XX**

BAYESIAN TOPOLOGICAL SIGNAL PROCESSING

Christopher Oballe*

University of Notre Dame, Department of Aerospace and Mechanical Engineering
Fitzpatrick Hall of Engineering and Cushing Hall
112 N Notre Dame Ave, Notre Dame, IN 46556

Alan Cherne

University of Tennessee, Department of Mathematics
1403 Circle Drive.
Knoxville, TN 37996-1320

Dave Boothe and Scott Kerick

US Army Research Laboratory
7101 Mulberry Point Road, Bldg. 459
Aberdeen Proving Ground, MD 21005-5425

Piotr J. Franaszczuk

US Army Research Laboratory and Johns Hopkins University
7101 Mulberry Point Road, Bldg. 459
Aberdeen Proving Ground, MD 21005-5425

Vasileios Maroulas*

University of Tennessee, Department of Mathematics
1403 Circle Drive.
Knoxville, TN 37996-1320

Abstract. Topological data analysis encompasses a broad set of techniques that investigate the shape of data. One of the predominant tools in topological data analysis is persistent homology, which is used to create topological summaries of data called persistence diagrams. Persistent homology offers a novel method for signal analysis. Herein, we aid interpretation of the sublevel set persistence diagrams of signals by 1) showing the effect of frequency and instantaneous amplitude on the persistence diagrams for a family of deterministic signals, and 2) providing a general equation for the probability density of persistence diagrams of random signals via a pushforward measure. We also provide a topologically-motivated, efficiently computable statistical descriptor analogous to the power spectral density for signals based on a generalized Bayesian framework for persistence diagrams. This Bayesian descriptor is shown to be competitive with power spectral densities and continuous wavelet transforms at distinguishing signals with different dynamics in a classification problem with autoregressive signals.

Date: November 24, 2021.

2020 *Mathematics Subject Classification.* Primary: 55N31, 68T07.

Key words and phrases. Topological data analysis, Bayesian, signal processing, autoregressive, EEG, machine learning.

1. **Introduction.** Data from measurements sampled in time create time series, which are useful descriptors of natural phenomena and artificially-created signals alike. Time series analysis is performed to extract meaningful statistics and other characteristics from data **?**. Frequency-domain methods are particularly suited for data exhibiting oscillatory behavior. A commonly used tool for nonparametric spectral analysis is the discrete Fourier transform **?**. Alternatively, parametric methods are frequently used to estimate the power spectral density of stochastic, stationary signals **?**. For non-stationary signals, time-frequency methods, e.g. wavelets **?** or matching pursuits **??**, provide useful characterizations of different time series. One of the important applications of time series analysis is in neuroscience and neurology where measurements of electrical activity in the brain (electroencephalography, EEG) can provide information about brain function and serves as an important clinical diagnostic tool. In this paper, we present a novel approach to time series analysis using persistent homology within a Bayesian framework. We illustrate the method on data generated by an autoregressive model simulating frequency characteristics of EEG signals **?**.

Persistent homology is the workhorse of topological data analysis (TDA), a set of ideas and tools to make inferences about topological and geometric aspects of data **?**. Efficient methods to compute the persistent homology of data were unveiled in **?**. Later, persistence barcodes, topological summaries to describe the shape of data, were introduced **?** and their theoretical properties were investigated **??**. In particular, **?** showed that persistence barcodes were stable under perturbations of underlying data. A comprehensive survey of persistence barcodes may be found in **?**.

TDA, persistent homology in particular, has been used for signal processing in many applications, including detection of periodic behavior, change point monitoring in genetic regulatory systems, detection of bifurcation in stochastic delayed differential systems, acoustic signal identification, and prediction of financial crashes from time series of stock returns **??????**. The work **?** describes how one can leverage persistent homology to create error tolerant systems for signal processing. These applications are bolstered by numerous theoretical results concerning the homology of simplicial complexes built on point clouds obtained from delay embeddings of time series **???**. Homology of these simplicial complexes is summarized in a persistence diagram, an equivalent representation of the persistence barcode as a multiset of points in $\mathbb{R}^2$. Recently, the work in **?** generated a topological signal processing approach by viewing a signal as a hypergraph based on the simplicial complex. Another way to create persistence diagrams is with the sublevel sets of a continuous function **?**. This method is attractive for signal processing as it may be used directly on signals to summarize their shape without selection of a delay parameter; however, a framework for inference about qualities such as frequency and instantaneous amplitude of a signal from its sublevel set persistence diagram remains wanting. Herein, we examine links between frequency content and instantaneous amplitude of signals and their sublevel set persistence diagrams, then explore a Bayesian method for inference with these topological objects.

A large body of work in TDA is concerned with developing concrete inferential tools for persistent homology. The work **?** introduces persistence landscapes, a Hilbert space representation of persistence diagrams built with collections of tent functions; **?** introduces another Hilbert space representation, the persistence image, which is a pixellated version of a persistence diagram created by convolving its

points with a Gaussian kernel then binning. In a different direction, **?** provides a notion of confidence sets for persistent diagrams based on kernel density estimators constructed from bootstrapped samples of point cloud data along with stability results for widely-used persistence diagram metrics (e.g., the bottleneck distance). Probability densities of persistence diagrams themselves, viewed as point processes, were considered in **?** wherein the authors construct a kernel for nonparametric density estimation. To the best of our knowledge, a Bayesian viewpoint for persistence diagrams was first espoused in **?**. Other works followed that blended Bayesian ideas with persistent homology. **?** gives a framework for Bayesian optimization with persistence diagrams, and the work **?** discusses persistent-homology-inspired likelihoods after using tropical geometry to introduce a sufficient statistic for persistence diagrams. Finally, a generalized Bayesian paradigm to describe probability distributions of persistence diagrams was introduced in **?**.

The work **?** is based on a point process model for persistence diagrams and creates probability densities for the space in which they live. In this fashion, the Bayesian model in **?** abstracts away from the data, choosing instead to focus on the statistics of its persistence diagrams. Other authors have shown that persistence diagrams are sufficient summaries of signals to answer pertinent research questions **???** like signal classification and seizure detection when concrete stochastic models for signals of interest are difficult to specify. On a similar note, we investigate the capabilities of the Bayesian topological model for signals by testing it to classify autoregressive signals. We compare this method to other model-free approaches for signal classification, specifically those based on feature extraction and machine learning architectures in **?** and **?**.

Our major contributions in this work are:

1. The establishment of interpretable links between sublevel set persistence diagrams of signals and their frequency domain,
2. The exploration of a Bayesian framework for time series classification, and
3. Evidence that Bayesian topological features of time series are competitive with well-established features, specifically those from power spectral densities and continuous wavelets, as distinguishing characteristics.

The paper is organized as follows. Section 2.1 contains our signal processing background and an overview of the autoregressive model we use for data generation. Section 2.2 outlines persistent homology of signals, persistence diagrams, and the high-level stochastic model we use to describe them, namely point processes. Section 3 begins with Proposition 3.1 and Corollary 3.1, which provide concrete equations for the sublevel set persistence diagrams of a family of deterministic signals and general random signals. Proposition 3.1 motivates the use of a Bayesian framework for describing persistence diagram densities by showing the complexity of persistence diagram distributions. The latter part of Section 3 details this Bayesian framework. In Section 4.1, we establish heuristic links between parameters of power spectral densities of autoregressive time series and features of their persistence diagrams. Then, in Section 4.2 we compare Bayesian topological features to power spectral density and wavelet features for autoregressive time series classification. Section 5 concludes with a discussion of our results.

2. **Background.**

Throughout, $x(t)$ denotes a real-valued, continuous time signal. We suppose x is continuous to avoid complications when we introduce topological concepts. Unless otherwise noted, time t is given in seconds and units of frequency

are in Hz. In general, x may be a sample from a stochastic process, which we denote by X. For applications, one often uses a discretized approximation to x obtained by fixing a sampling rate $\Delta t > 0$ then measuring x at regular intervals. We denote elements of the discretized collection $\{x(n\Delta t)\}_{n=1}^{N}$ by x_n where N is the total number of grid points.

2.1. **Signal Processing Preliminaries.** In this section, we establish concepts and tools from signal processing employed in Section 4. During our analyses, we investigate autoregressive signals satisfying the $AR(p)$ model $x_n = \sum_{i=1}^{p} a_i x_{n-i} + w_n$ where p is the model order, a_i are real-valued coefficients, and w_n are independent, identically distributed samples from a Gaussian random variable $\mathcal{N}(0, \sigma)$ with standard deviation σ. This autoregressive model is used in various applications across many fields with an easily computable power spectral density, p_X, that is described by finitely many parameters β_k and f_k,

$$\log p_X(f) = 2\left(-\sum_{k=1}^{P} \log\left|1 - e^{-\beta_k + \frac{2\pi i}{\Delta t}(f_k - f)}\right| + \log\sigma\right), \tag{1}$$

where σ is the standard deviation of w_n in $AR(p)$ model. A derivation of Equation (1) may be found in **?**. Through inspection of Equation (1), one surmises that the f_k are locations of local maxima in the log PSD (and therefore also local maxima of the PSD) and the β_k are corresponding damping factors controlling the width of each peak. Namely, small and large β_k are associated with narrow and broad peaks, respectively. As a peak broadens, the presence of oscillations of its corresponding frequency diminish in the average signal. Therefore, we expect power spectral densities with narrow peaks to yield signals with stronger oscillations.

2.2. **Persistent Homology of Signals.** In this section, we briefly discuss how to employ the sublevel set filtrations of continuous functions in order to convert signals to persistence diagrams. Persistence diagrams usually contain fewer elements than their corresponding signals, so in a sense they compress information about a signal into a lower dimensional representation.

For a fixed real number V, the sublevel set C_V of a signal x is defined by $C_V := \left\{x^{-1}\Big((-\infty, V]\Big)\right\}$, where $x^{-1}\Big((-\infty, V]\Big) := \{t \in \mathbb{R} : x(t) \leq V\}$. $x^{-1}\Big((-\infty, V]\Big)$ is the set of times where $x(t)$ is less than or equal to V. We refer to the collection $C := \{C_V\}_{V \in \mathbb{R}}$ as the sublevel set filtration of x. From the sublevel filtration C, we obtain a persistence diagram for x by tracking the birth and death values of connected components (which are in this case disjoint continuous intervals) in C_V as V increases. Specifically, as V increases, connected components appear or merge in C_V. To create a persistence diagram, we record the values, b and d, respectively, at which a connected component appears and disappears by merging into one that appeared earlier (a convention known as the Elder rule **?**). For most continuous signals encountered in applications, it can be shown b is always a local minimum and d is always a local maximum **?**. In the end, we aggregate each connected component's birth-death pair (b, d) to create a collection, $\mathcal{D}$, known as a persistence diagram. For a visual summary of persistence diagram generation through sublevel set filtrations, see Figure 1, which shows a damped cosine $s(t; f, \beta) := e^{-\beta t}\cos(2\pi f t)$ with $f = 4$ and $\beta = 2$. Notice in general that $s(t; f, \beta)$ has f local minima (ignoring units of f) on the interval $[0, 1]$. As the number of points in a persistence diagram is equal to the number of its local minima, we immediately conclude that the diagram's

cardinality is f. To compute the coordinates of points in the persistence diagram, notice that $e^{-\beta t}$ is monotonically decreasing while each oscillation of $\cos(2\pi f t)$ starts at a local maximum and has a unique local minimum. The latter implies that each oscillation gives rise to a connected component during the sublevel set filtration of s and that this connected component merges into another when the filtration reaches the value of the local maximum where the oscillation begins. The former implies that connected components for oscillations occurring later in time disappear sooner than those born earlier due to the Elder Rule.

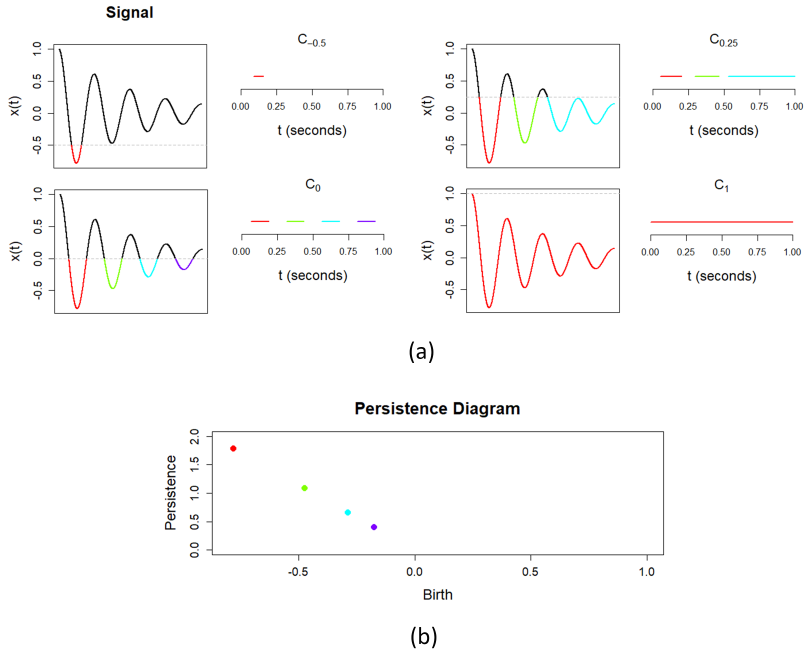

FIGURE 1. Shown above (a) are the sublevel sets $C_{-0.5}$, C_0, $C_{0.25}$, and C_1 for a damped cosine $e^{-2t}\cos(8\pi t)$. (b) shows the persistence diagram of the sublevel set filtration. The points in (b) are colored to match the connected components their birth coordinates correspond to. The transition from C_0 to $C_{0.25}$ depicts the Elder rule; notice that in C_0, there are light blue and purple connected components, which merge together in $C_{0.25}$. A similar merging happens in the transition from $C_{0.25}$ to $C_{0.5}$. Since the purple component has a later birth value, it disappears into the light blue component, which persists until it merges into the green component by the same line of reasoning.

2.3. **Point Processes.** In practice, we observe random signals. Uncertainty in signals arises independently from stochasticity in underlying data-generating phenomena and measurement noise due to limitations in data-collection methods. The latter source of uncertainty is pictured in Figure 2 where a signal is embedded in different levels of white Gaussian noise. The persistent diagrams created from stochastic signals (subsequently referred to as random persistence diagrams) inherit

randomness, necessitating a probablisitic description. Intuitively, a random persistence diagram is a random collection of points in a subset of the plane. Elements of a persistence diagram lack an intrinsic ordering, and moreover persistence diagrams sampled from the same random signal can have different cardinalities. The theory of random variables is ill-equipped to handle such objects directly since it concerns itself with random elements that take values in a Hilbert space. However, the theory of point processes rigorously treats random collections like random persistence diagrams and provides machinery to model them. A point process is a set-valued random element characterized by a random variable K that prescribes the cardinality distribution and a set of spatial distributions conditioned on K that dictate where to place points in samples once cardinality is known. Formally,

Definition 2.1. For a discrete random variable K, a (finite) point process $P := \{x_1, \ldots, x_K\}$ is a random collection of elements in a Polish Space $\mathbb{X}$ characterized by a collection of symmetric conditional probability distributions $P|K = k \sim \mathbb{P}_k$.

In general, an exact description for K and the spatial distributions of a random persistence diagram is computationally intractable, even in scenarios where a reliable stochastic model for X is known a priori. Thus, studying random persistence diagrams requires a well grounded framework for estimating their distributions. The method we choose is based on the Bayesian paradigm for persistence diagrams introduced in **?**. Central to this framework are Poisson point processes, which are defined below.

Definition 2.2. Let $\lambda : \mathbb{R}^2 \to \mathbb{R}$ be a non-negative function satisfying $0 < \Lambda < \infty$ where $\Lambda := \int_{\mathbb{R}^2} \lambda(\mathbf{x})\, d\mathbf{x}$. The function λ is called the intensity function or simply the intensity. A Poisson point process, $\mathcal{P}$ is a point process satisfying:

$$K \sim \text{Poisson}(\Lambda) \tag{2}$$

$$\mathcal{P}|K = k \sim \frac{1}{\Lambda^n} \prod_{i=1}^{n} \lambda(\mathbf{x}_i). \tag{3}$$

Definition 2.2 prescribes a natural recipe to sample from a Poisson point process. First, one determines the cardinality of $\mathcal{P}$ by drawing from Poisson(Λ). With this number in hand, say n, one then makes n independent draws from the probability density $\frac{\lambda}{\Lambda}$ to spatially distribute the points in $\mathbb{R}^2$. Poisson point processes are completely characterized by their intensities. This makes them a convenient tool for Bayesian inference on the space of persistence diagrams since they can be used to specify prior distributions using a single, real-valued function on $\mathbb{R}^2$. With prior distributions in hand, the last ingredient one needs for Bayesian inference is a likelihood, which is obtained via marked point processes (defined below).

Definition 2.3. Let $\ell(\mathbf{y}|\mathbf{x})$ be a probability density parameterized by $\mathbf{x}$. A marked point process is a collection (P, M) where

$$M|P = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\} \sim \prod_{i=1}^{k} \ell(\mathbf{y}_i|\mathbf{x}_i). \tag{4}$$

Elements of M in Definition 2.3 are determined by independent draws from a stochastic kernel $\ell(\cdot|\mathbf{p})$, which is a probability density parameterized by the elements of samples from P.

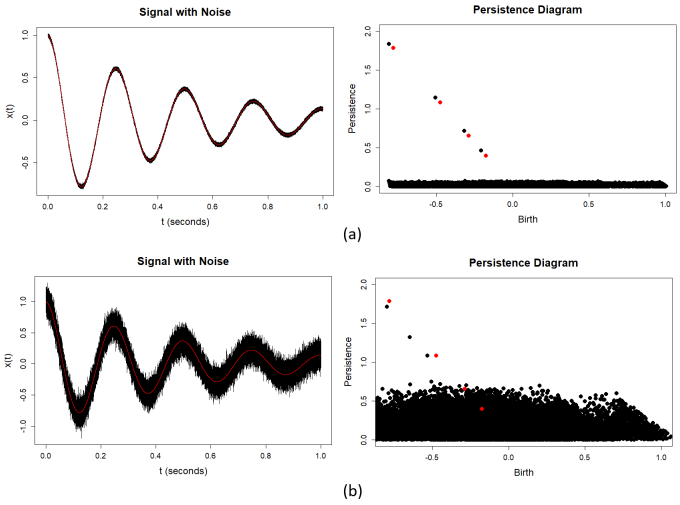

FIGURE 2. This figure illustrates sources of uncertainty in persistence diagrams. Shown above are signals with additive noise (a) $\mathcal{N}(0, 0.01)$, and (b) $\mathcal{N}(0, 0.1)$ along with their persistence diagrams. The persistence diagram for the true underlying signal is shown in red. Spurious features arise due to noise and additionally, true features also shift around.

3. **Methods: Uncertainty Quantification for Random Persistence Diagrams.** The persistence diagram of a signal encodes information about its frequency and peak-to-peak amplitude. This is exemplified in Proposition 3.1 and Corollary 3.1 below.

Proposition 3.1. Let $a_\beta(t)$ be a monotonically non-increasing function of time parameterized by a real number β and suppose that $\phi(2\pi ft)$ is a periodic signal parameterized by a positive frequency f such that $\phi(0)$ is a local maximum and each cycle of ϕ has a unique local minimum. The persistence diagram $\mathcal{D}_{f,\beta}$ of $a_\beta(t)\phi(2\pi ft)$ on the interval [0,1] is given by $\mathcal{D}_{f,\beta} = \left\{\left(a_\beta(t_i + t_m)\phi(2\pi f(t_i + t_m)), a_\beta(t_i)\phi(2\pi ft_i)\right)\right\}_{i=1}^{\lfloor f \rfloor}$ where, $\lfloor f \rfloor$ denotes the integer part of f, $t_i = \frac{i-1}{f}$ and t_m is the time of the first local minimum of $a_\beta(t)\phi(2\pi ft)$ in $[0, 1]$.

A proof of Proposition 3.1 may be found in the Appendix. Proposition 3.1 establishes direct links between frequencies and instantaneous amplitudes to the persistence diagrams for a large class of deterministic signals that naturally occur in applications. The parameter β in Proposition 3.1 can have various effects on the statistics of the persistence diagram of the signal. Figure 1 considers the signal $e^{-\beta t}\cos(8\pi t)$, and in that example increasing the value of β increases the variance in persistence coordinates of the diagram. When in the presence of noise or high-frequency oscillations, such as in Figures 2 and 3, the effect is inverted - increasing β leads to a decrease in variance.

Next, we turn our attention to relationships between stochastic signals and their persistence diagrams. Figure 2(a) shows the signal $s(t; 4, 2)$ with additive Gaussian white noise and its persistence diagram. Although the presence of noise introduces

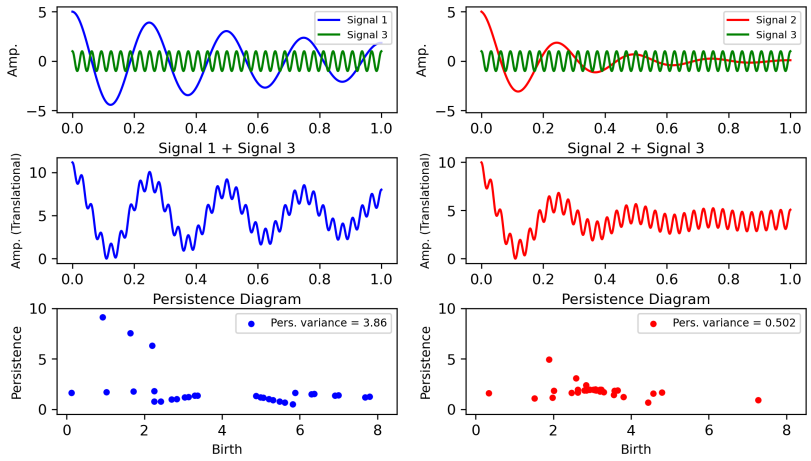

FIGURE 3. *Top:* We consider three signals. The blue signal (Signal 1) and the red signal (Signal 2) are modeled by $a_\beta(t)\cos(8\pi t)$ where $a_\beta(t) = 5e^{-\beta t}$ with $\beta = 1, 4$ in Signals 1 and 2 respectively. The green signal (Signal 3) is then added to each case and the amplitudes are translated to have global minima equal to zero. *Bottom:* The associated persistence diagrams are plotted using the method described in Section 2.2. We observe that as β increases, the high-frequency oscillations are less affected by the low-frequency signal and converge faster towards the uniform shape of the green signal. This leads to a decrease in the variance of the persistence coordinates in the red diagram.

spurious low persistence features in the persistence diagram, the relationship unveiled in Proposition 3.1 is evident in the higher persistence points. In this case, noticeable relationships still exist between the frequency and instantaneous amplitude of the deterministic signal embedded in noise and its persistence diagram.

The relatively simple closed form in Proposition 3.1 is a consequence of the min-max pairs in the signal, which determine points of its persistence diagram, having a clear ordering in time. In general, suppose x is sampled from a stochastic process and consider its discretization $\{x_n\}_{n=0}^N$, which we momentarily refer to as x by a slight abuse of notation. Ideally, the probability distribution $\mathbb{P}(D^x)$ for the random persistence diagram of D^x, would be expressed as $\mathbb{P}(D^x) = \cup_{i,j\in\mathcal{I}}\{\mathbb{P}(x_i, x_j)\}$ for a fixed set of index pairs $\mathcal{I}$. In this situation, which is similar to the setting of Proposition 3.1, D^x is a fixed union of joint random variables. Unfortunately, the aforementioned decomposition of $\mathbb{P}(D^x)$ does not hold in general, specifically because distinct orderings of values of elements in x are associated to different minimum-maximum pairings. The next lemma establishes events $\mathcal{E}$ in which $\mathbb{P}(D^x|\mathcal{E}) = \cup_{i,j\in\mathcal{I}}\{\mathbb{P}(x_i, x_j)\}$ holds.

Proposition 3.2. Let x be a signal randomly sampled from a stochastic process X. Consider the time series $\{x_n\}_{n=1}^L$ created by measuring x at regular intervals on $[0, 1]$. Denote the persistence diagram for $\{x_n\}_{n=1}^L$ by $\mathcal{D}^x$. Let Π_L be the set of permutations of $\{1, 2, \ldots, L\}$ and suppose for a fixed element π of Π, say

$\pi = \{i_1, i_2, \ldots, i_L\}$, O_π is the event that $\{x_{i_1} < x_{i_2} < \cdots < x_{i_L}\}$. $D^x|O_\pi$ is a fixed union of joint random variables; we denote the probability distribution for this collection by $\mathbb{P}(D^x|O_\pi)$.

Proof. The event O_π corresponds to an exact ordering of the values of elements in x. The exact ordering preserves the minimum-maximum associations used to construct sublevel sets to create D^x. In particular, $D^x = \cup_{i,j\in\mathcal{I}_\pi}\{(x_i, x_j)\}$ for a fixed set of index pairs $\mathcal{I}_\pi$ given O_π. This establishes the claim. □

Corollary 3.1. Let x be a signal randomly sampled from a stochastic process X. Consider the time series $\{x_n\}_{n=1}^L$ created by measuring x at regular intervals on $[0, 1]$. Denote the persistence diagram for $\{x_n\}_{n=1}^L$ by $\mathcal{D}^x$. Let Π_L be the set of permutations of $\{1, 2, \ldots, L\}$ and suppose for a fixed element π of Π, say $\pi = \{i_1, i_2, \ldots, i_L\}$, O_π is the event that $\{x_{i_1} < x_{i_2} < \cdots < x_{i_L}\}$. The probability distribution for $\mathcal{D}^x$ is given by

$$\mathbb{P}(\mathcal{D}^x) = \sum_{\pi\in\Pi_L} \mathbb{P}(\mathcal{D}^x|O_\pi)\mathbb{1}_{O_\pi}(\mathcal{D}^x). \tag{5}$$

where $\mathbb{P}(\mathcal{D}^x|O_\pi)$ is the distribution for a random collection of points given the ordering O_π.

Let Π^k be the subset of Π such that each element of Π^k is an event with k local minima. Then, the cardinality distribution of $\mathcal{D}^x$ is given by

$$\mathbb{P}(|\mathcal{D}^x| = k) = \sum_{\pi\in\Pi^k} \mathbb{P}(\pi). \tag{6}$$

Proof. Equation (5) follows from Proposition 3.2 by the law of total probability and Equation (6) is established by a straightforward counting argument. □

Intuitively, the frequency content and distribution for instantaneous amplitudes of a random signal x influence the probability of events in Π in Corollary 3.1. In general, describing the persistence diagram in its entirety for a time series of length L arising from a random signal requires one to construct $L!$ different probability distributions. We circumvent this problem by using a flexible Bayesian framework to approximate distributions for persistence diagrams. We elaborate on this methodology for the remainder of Section 3. With this tool in hand, we can model the distributions of persistence diagrams in a computationally expedient manner.

For a random signal X, we refer to the random object D^X whose samples are drawn by creating persistence diagrams from samples of X as a random persistence diagram. Formally, D^X is a point process (Section 2.3). Corollary 3.1 demonstrates that an exact description for the spatial distributions of D^X and its cardinality N is computationally intractable in general, even in scenarios where a reliable stochastic model for X is known a priori. Thus, studying D^X requires a well grounded framework for estimating its distribution. The method we choose is based on a Bayesian paradigm for persistence diagrams. Let (D^X, D^Y) be a collection of point processes. D^X models the random persistence diagram for X while D^Y models observed persistence diagrams, which are samples from D^X obscured by measurement noise. We impose additional structure on D^X and D^Y to account for common experimental conditions and model dependencies between D^X and D^Y, namely:

(M1) $D^X = D^V \cup D^O$ where D^V and D^O are independent Poisson point processes with respective intensities $(1-\alpha)\lambda_{D^X}$ and $\alpha\lambda_{D^X}$, and

(M2) $D^Y = D^S \cup D^M$, where D^S is a Poisson point process with intensity λ_S and D^M forms a marked Poisson point process (D^O, D^M) with D^O whose stochastic kernel is denoted by ℓ.

D^V models the event that an element $\mathbf{x}$ from a sample of D^X does not appear in the observed, noisy diagram $\mathcal{D}^Y$, which happens with probability $1-\alpha$. On the other hand, an element from D^X may appear in $\mathcal{D}^Y$ as a mark with probability α. This is accounted for by D^O and D^M. D^S models the spurious points that arise in $\mathcal{D}^Y$ due to noise or other randomness. Theorem 3.1 gives a closed form for the posterior intensity of the model outlined in $(M1)$ and $(M2)$.

Theorem 3.1. Let $\mathcal{N}(\mathbf{x}; \mu, \sigma I)$ denote the two-dimensional normal density with mean μ and covariance matrix $\sigma^2 I$ and let $\mathcal{N}^*$ denote the restriction of $\mathcal{N}$ to $\mathbb{W} := \{(b,p) \in \mathbb{R}^2 : p > b > 0\}$. Suppose in (M1) and (M2) that $\lambda_{D^X}(\mathbf{x}) = \sum_{j=1}^K c_j^X \mathcal{N}^*(\mathbf{x}; \mu_j, \sigma_j I)$ for positive constants c_j^X, α is a constant, and $\ell(\mathbf{y}|\mathbf{x}) = \mathcal{N}^*(\mathbf{y}; \mathbf{x}, \sigma^Y I)$. Then, for m independent samples from D^Y,$\mathcal{D}^{Y_{1:m}}$, it follows that

$$\lambda_{D^X|\mathcal{D}^{Y_{1:m}}}(\mathbf{x}) \;=\; (1-\alpha)\lambda_{D^X}(\mathbf{x}) \;+\; \frac{\alpha}{m}\sum_{i=1}^{m}\sum_{\mathbf{y}\in\mathcal{D}^{Y_i}}\sum_{j=1}^{K} C_j^{\mathbf{y}}\mathcal{N}^*\left(\mathbf{x}; \mu_j^{\mathbf{y}}, \sigma_j^{\mathbf{y}} I\right) \quad (7)$$

where

$$C_j^{\mathbf{y}} := \frac{w_j^{\mathbf{y}}}{\lambda_S(\mathbf{y}) + \alpha\sum_{j=1}^K w_j^{\mathbf{y}} Q_j^{\mathbf{y}}}$$

for

$$w_j^{\mathbf{y}} := c_j \mathcal{N}\left(\mathbf{y}; \mu_j, (\sigma_j + \sigma^Y) I\right), Q_j^{\mathbf{y}} := \int_{\mathbb{W}} \mathcal{N}(\mathbf{u}; \mu_j^{\mathbf{y}}, \sigma_j^{\mathbf{y}} I)\, d\mathbf{u}$$

$$\mu_j^{\mathbf{y}} := \frac{\sigma_j \mathbf{y} + \sigma^Y \mu_j}{\sigma_j + \sigma^Y} \text{ , and } \sigma_j^{\mathbf{y}} := \frac{\sigma_j \sigma^Y}{\sigma_j + \sigma^Y}.$$

Theorem 3.1 depends on several parameters which need to be effectively chosen. As the distributions involved are Gaussian we can use bandwidth estimators such as Silverman's rule **?** to estimate covariances σ_j in the prior using a subset of the training data. In our example we relied on a grid search and cross-validation to find suitable σ_j^y. Notice in the equation for σ_j^y if we divide the numerator and denominator by σ^Y the denominator only depends on the ratio of σ_j to σ^Y, and therefore we can fix σ^Y in our search. In regards to the frequency and amplitude of signals, one consideration that must be made is the fineness of the partition for the sublevel sets. In general, increasing the fineness of a partition will increase the cardinality of the density as more local minima are captured - this in turn will yield a different intensity for the underlying point process. This effect is mitigated by the presence of the noise distribution λ_S in the formula for the intensity, as points with lower persistence are often related to noise. Figure 4 visually depicts the intensity in Equation (7) using a persistence diagram from a simulated signal. In particular, Figure 4 demonstrates how λ_S in Theorem 3.1 can be leveraged for regularization of posterior intensities to prevent overfitting to noise. Figure 5 illustrates more relationships between properties of signals and those of their posterior intensities computed according to Theorem 3.1.

Figures 5 (b) and (d) show the persistence diagrams for $5\cos(8\pi t) + \cos(64\pi t)$ and $\cos(8\pi t) + \cos(64\pi t)$ on the interval $[0,1]$, respectively. In both figures, the points near the Birth axis correspond to the small peaks that ride along the low

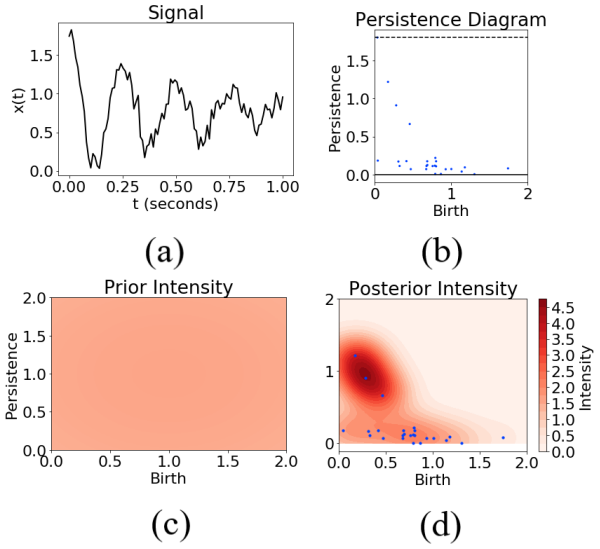

FIGURE 4. (a) The damped cosine $e^{-2t}\cos(8\pi t)$ with additive noise $\mathcal{N}(0, 0.01)$ and (b) its persistence diagram. (b) shows an uninformative prior intensity with a single component at $(1, 1)$ with covariance matrix $10I$. Using the model from Equation (7) with the prior in (c) and the observed diagram in (b) results in the posterior intensity shown in (d). To account for spurious points, which we suspected to be low persistence in this example, we placed components of λ_S at $(0.5, 0.1), (1, 0.1), (0.75, 0.1)$ and $(1.75, 0.1)$.

frequency wave while the group of points farthest from the Birth axis correspond to global maxima and minima. As can be seen in Figure 5 (b), more low frequency power than high frequency power causes a wider spread in the birth times of low persistence points. This causes greater variance in the birth dimension of the posterior intensity and also in the persistence dimension, similar to Figure 3. Moreover, the high persistence points in Figure 5 (b) have smaller birth coordinates and larger persistence coordinates than those in Figure 5 (d). As a result, the posterior intensity in Figure 5 (b) exhibits an isolated high-persistence mode, which is not as distinguished in Figure 5 (d).

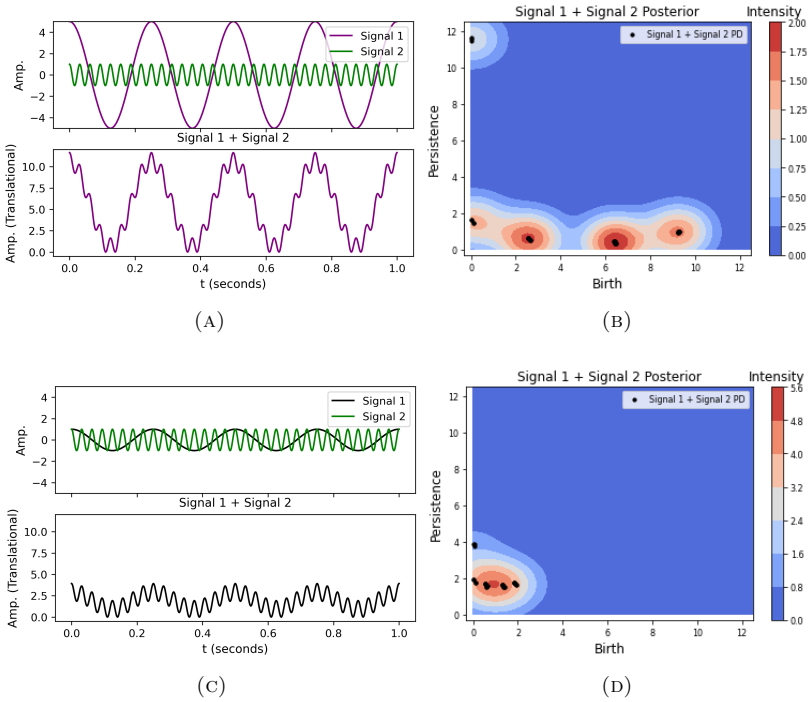

FIGURE 5. This figure demonstrates the effect of greater low frequency power on the persistence diagram of a signal. Figures (a) and (c) show two signals, respectively, which are the result of summing low-frequency and high-frequency oscillators. The power of the low-frequency signal is greater in (a) than in (c). To ensure that persistence diagrams in (b) and (d) lie in $\mathbb{W}$, the aggregate signals in both (a) and (c) have been translated so that their absolute minima are at zero. Notice in (b) that elements of the persistence diagram show greater spread along the Birth axis than in (d). This results in greater birth variance of the corresponding posterior intensity. Also notice the isolated high-persistence mode in (b), which is not present in (d). These phenomena arise because the low frequency signal scatters the higher frequency peaks along the Amplitude axis.

4. **Results.** Electroencephalography is a neuroimaging technique wherein electrodes are placed on a subject's head to measure local changes in voltage over time, which are reported as a collection of time series. In our experiments, we examine synthetic EEG signals generated according to Equation (1). We select EEG signals because they have well-studied power spectra. In particular, the log power spectral density for EEG is approximately inversely related to frequency, a phenomena we subsequently refer to as $1/f$ behavior. We can easily simulate this behavior with Equation (1) by including a frequency component f_1 equal to zero. Moreover, EEG signals associated with different brain states often exhibit a prominent peak in their power spectral density at a nonzero frequency, indicating the discernible presence

of oscillations at that frequency in the signals. For example EEG signals with peaks in the $4-7$ Hz, $8-12$ Hz, or $14-32$ Hz ranges, referred to as theta, alpha, and beta signals, respectively, may indicate heightened emotional states, eyes open, or drowsiness **??**. Oscillatory behavior can also be accounted for with Equation (1) by including a particular nonzero frequency component f_2.

Selection of the damping factor β_1 for the zero frequency component f_1, as well as ranges for f_2 and β_2 (the location of the oscillatory component in the PSD and its accompanying damping factor) was done by using the Burg method to fit autoregressive models to real EEG signals (model order was determined using the methodology from **?**). Two one second and two five second epochs were selected for their visually apparent oscillations. The exact location of spectral peaks along with their corresponding damping factors as determined after fitting autoregressive models are shown in Table 2 of the Appendix.

Informed by the parameters in Table 2, we created 29×29 sets of synthetic EEG signals $\mathcal{A}_f^\beta$ where $f \in \{4, 5, \ldots, 32\}$ and $\beta \in \{4, 5, \ldots, 32\}$. A fixed set $\mathcal{A}_f^\beta$ contained 30 signals simulated by our autoregressive filter with $f_1 = 0$, $f_2 = f$, $\beta_1 = 200$, and $\beta_2 = \beta$. The signals in $\mathcal{A}_f^\beta$ are draws from a stochastic process whose PSD has a peak at f with a damping factor of β, so we expect these signals to resemble an oscillator of frequency f; the strength of this resemblance diminishes as β increases. Since the PSD for $\mathcal{A}_f^\beta$ more closely resembles $1/f$ as β increases, we expect more low-power, high frequency signals riding on high-power, low frequency signals as β decreases. Moreover, the ratio of the peak at zero to that of the peak at f increases as β increases.

4.1. **Relationships between the Frequency and Diagram Domains.** In this section, we examine relationships between f and β and the persistence diagrams of $\mathcal{A}_f^\beta$. Specifically, we look at how f and β relate to the average cardinality and variance in birth values for persistence diagrams of signals in $\mathcal{A}_f^\beta$. As in the deterministic setting, we expect cardinality to show a strong association with the peak frequency f for $\mathcal{A}_f^\beta$. Our choice to inspect birth time variance is motivated by the discussion at the end of Section 3.

Fig. 6 shows the average cardinality of persistence diagrams in $\mathcal{A}_f^\beta$ against the location of peak frequencies f with colors showing the damping factors β. A strong monotonic trend is apparent for all damping factors. For smaller damping factors, the relationship more heavily resembles that in the deterministic setting (see Section 3) where PD cardinality from a one second epoch is in fact equal to the frequency of the signal. Notice the effect of β diminishes as f increases, which implies cardinality is more sensitive to higher frequencies.

Figure 7 shows the relationship between f, β, and the birth value variance for $\mathcal{A}_f^\beta$. Birth variance increased as the damping factors increased. This is consistent with the idea that higher low-frequency-power-to-high-frequency-power ratios result in more variation for low persistence birth times. Interestingly, birth variance decreased as frequency increased, suggesting this trend is less notable at higher frequencies.

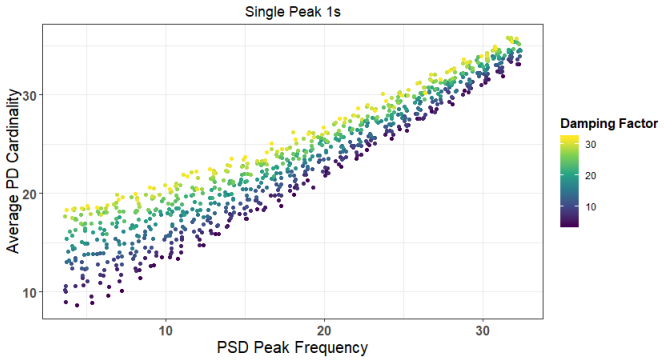

FIGURE 6. This plot depicts the relationship between the cardinality of persistence diagrams and the frequency of the dominant oscillation for one second autoregressive signals across various damping factors. For each included frequency and damping factor, we simulated thirty signals (each had a component fixed at zero to give the $1/f$ PSD commonly seen in EEG), computed their persistence diagrams, then recorded their average cardinality. We see a strong positive correlation between this average cardinality and the frequency of the dominant oscillation (i.e., PSD Peak Frequency) consistent with the idealized deterministic sinusoid case.

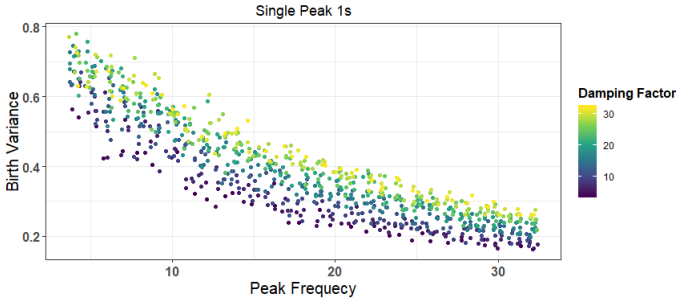

FIGURE 7. The peak frequency f for $\mathcal{A}_f^\beta$ plotted against the average birth variance for its persistence diagrams. Colors depict the damping factor β.

4.2. **Classification.** As a point of comparison to traditional signal processing techniques and to showcase the utility of topological methods, we used features derived from Bayesian persistent homology (Section 3) and discrete Fourier transforms to classify signals with different dynamics. Specifically, we considered five classes of signals in total, each with a distinct rhythm quantified by a nonzero f_k parameter in our autoregressive filter (additionally, all classes had a peak at zero with a fixed damping factor to simulate the $1/f$ behavior commonly present in the PSD of EEG). Four classes we considered were: alpha (a), high beta (hb), low beta (lb), and theta (t), which had (autoregressive) spectral peaks at 10, 21, 14, and 6 Hz, respectively. We also included a null (n) class that had no peaks aside from one at zero. These spectral peaks were chosen for their prevalence in EEG data. We

considered signals with damping factors of thirty-two to mimic weak oscillatory behavior to tackle a challenging problem. Each class of signals contained 30 examples. Visual descriptions of our data are provided in Figure 8 of the Appendix.

We used three different classifiers in our experiment. The first classifier, logistic regression (LR), is the most basic and models the probability that a signal belongs to class i as a function of a linear combination of input features. Another classifier we used was a support vector machine with a linear kernel (SVM-Lin) , which works by deciding optimal decision boundaries between the classes in feature space, then assigning new observations to classes based on these decision boundaries. A hyperparameter in SVM-Lin is a regularization term that penalizes mis-classification. To tune this hyperparameter, we relied on a grid search. The last classifier we considered was a multilayer perceptron (MLP), also known as a feedforward neural network. We chose a neural network architecture analogous to one used in **?**. Namely, we used 1 hidden layer of 4 neurons with a saturating linear activation function, followed by a softmax layer for classification.

We used three sets of features for each classifier. The first two sets were derived from standard signal processing methods. The first set was based upon PSDs estimated via discrete Fourier transforms. In particular, for each signal, we computed its DFT then recovered an estimate to the PSD. We then binned the total power in frequency bands of interest (0.5-4 Hz, 4-7Hz, 7-12Hz, 12-21 Hz, 21-32Hz, $>$ 32 Hz) to obtain a 6-dimensional feature vector. The second set of features were created from continuous wavelet transforms. Specifically, we used the Mexican hat mother wavelet to obtain time-frequency plots for each signal. These time-frequency estimates were then binned in the same frequency bands as the PSD estimates to obtain 6 dimensional vectorizations. The final set of features we considered were derived from the Bayesian method outlined in Section 3. For each signal, we computed its persistence diagram, then used Equation (7) to estimate a posterior intensity. We then took 6 features of the posterior intensity to obtain 6-dimensional vectorizations. As the estimated posterior intensities were sensitive to the prior used to fit them, we tuned parameters in our prior λ_D using a grid search and LR as LR had no hyperparameters to consider, unlike the other two classifiers (see Appendix). To compete with the 6-dimensional features above, the following features of the posterior for a diagram $\mathcal{D}$, $\lambda_{D|\mathcal{D}}$, were chosen:

1. The total intensity, $\int_{\mathbb{W}} \lambda_{D|\mathcal{D}}(\mathbf{u})\, d\mathbf{u}$, which gives the expected cardinality.
2. The birth coordinate of the center of mass of $\lambda_{D|\mathcal{D}}$, $\int_{\mathbb{W}} \mathbf{u}\lambda_{D|\mathcal{D}}(\mathbf{u})\, d\mathbf{u}$.
3. The persistence coordinate of the center of mass.
4. The variance in birth coordinates for the posterior means of $\lambda_{D|\mathcal{D}}$.
5. The variance in persistence coordinates for the posterior means of $\lambda_{D|\mathcal{D}}$.
6. The covariance of birth and persistence coordinates of posterior means in $\lambda_{D|\mathcal{D}}$.

The first feature is the most direct proxy to the frequency of the signal; the second and third account for defining features of the point process model. The fourth, fifth, and sixth features capture the variation in birth and persistence coordinates, as motivated by Figures 3 and 4. Based on the observations surrounding Figure 5, we expected different ratios of low-frequency to high-frequency power in the signals to yield posterior distributions with different properties.

For each classifier and vectorization, we used leave-one-out-cross-validation to assign a label for each signal. Results from our classification are shown in Table 1. Examining Table 1 shows competitive performance of Bayesian-derived persistence

diagram features for signal classification to that of well-established features for signal processing.

	Bayesian		**PSD**		**CWT**	
Classifier	Precision	Recall	Precision	Recall	Precision	Recall
LR	0.84 ± 0.06	0.85 ± 0.07	0.90 ± 0.04	0.90 ± 0.04	0.91 ± 0.03	0.90 ± 0.04
SVM - Lin.	0.92 ± 0.05	0.91 ± 0.04	0.91 ± 0.03	0.90 ± 0.05	0.91 ± 0.04	0.91 ± 0.03
MLP	0.89 ± 0.05	0.88 ± 0.04	0.90 ± 0.02	0.89 ± 0.02	0.92 ± 0.03	0.93 ± 0.02

TABLE 1. Precisions and recalls for each feature and classifier. Results are reported as mean ± standard error across each class.

5. **Discussion and Conclusions.** Herein, we give an interpretable framework for signal processing via sublevel persistent homology. Explicit representations for persistence diagrams of signals are provided in Propositions 3.1 and 3.1, which aid interpretability of persistence diagrams in a signal processing context. In future work, one may expand on Proposition 3.1 by finding closed forms for more general family of signals, and investigate families of stochastic signals for which Proposition 3.1 yields tractable probability densities. This would further enhance interpretation of persistence diagrams, making them more attractive objects for the scientific community.

We also provide a method to probabilistically describe persistence diagrams of signals by relying on a Bayesian paradigm for topological data analysis. This approach imposes no assumptions on the stochastic behavior of the signals. Our results in classification provide evidence that this probabilistic descriptor is competitive with well-established descriptors like the power spectral density and continuous wavelet transform for distinguishing time series dynamics. In the future, one may examine the effectiveness of higher resolution summaries of posterior intensities for use in advanced deep learning architectures. This is necessary for the incorporation of topological methods into state-of-the-art signal processing methods.

Acknowledgments. CO's research was sponsored by the Army Research Laboratory (ARL) and was accomplished under Cooperative Agreement Number W911NF-19-2-0302. VM's work was partially supported by the ARO W911NF- 17-1-0313, NSF DMS-1821241, and ARL and was accomplished under Cooperative Agreement Number W911NF-19-2-0328. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of the Army Research Laboratory or the U.S. Government. The U.S. Government is authorized to reproduce and distribute reprints for Government purposes notwithstanding any copyright notation herein.

REFERENCES.

Swaiman's Pediatric Neurology, Elsevier, 2017, URL `https://doi.org/10.1016/c2013-1-00079-0`.

H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta and L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, *The Journal of Machine Learning Research*, **18** (2017), 218–252.

M. Bandarabadi, A. Dourado, C. A. Teixeira, T. I. Netoff and K. K. Parhi, Seizure prediction with bipolar spectral power features using adaboost and svm classifiers, *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 6305–6308.

S. Barbarossa and S. Sardellitti, Topological signal processing over simplicial complexes, *IEEE Transactions on Signal Processing*.

R. J. Barry, A. R. Clarke, S. J. Johnstone, C. A. Magee and J. A. Rushby, EEG differences between eyes-closed and eyes-open resting conditions, *Clinical Neurophysiology*, **118** (2007), 2765–2773, URL `https://doi.org/10.1016/j.clinph.2007.07.028`.

P. Bromiley, Products and convolutions of gaussian probability density functions.

P. Bubenik, Statistical topological data analysis using persistence landscapes, *The Journal of Machine Learning Research*, **16** (2015), 77–102.

G. Carlsson, Topology and data, *Bulletin of the American Mathematical Society*, **46** (2009), 255–308.

G. Carlsson, A. Zomorodian, A. Collins and L. Guibas, Persistence Barcodes for Shapes, in *Symposium on Geometry Processing* (eds. R. Scopigno and D. Zorin), The Eurographics Association, 2004.

D. Cohen-Steiner, H. Edelsbrunner and J. Harer, Stability of persistence diagrams, *Discrete & Computational Geometry*, **37** (2006), 103–120, URL `https://doi.org/10.1007/s00454-006-1276-5`.

W. Crawley-Boevey, Decomposition of pointwise finite-dimensional persistence modules, *Journal of Algebra and Its Applications*, **14** (2015), 1550066, URL `http://dx.doi.org/10.1142/S0219498815500668`.

Edelsbrunner, Letscher and Zomorodian, Topological persistence and simplification, *Discrete & Computational Geometry*, **28** (2002), 511–533, URL `https://doi.org/10.1007/s00454-002-2885-2`.

H. Edelsbrunner and J. Harer, *Computational Topology*, American Mathematical Society, 2009, URL `https://doi.org/10.1090/mbk/069`.

B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, A. Singh et al., Confidence sets for persistence diagrams, *The Annals of Statistics*, **42** (2014), 2301–2339.

P. J. Franaszczuk and K. J. Blinowska, Linear model of brain electrical activity?EEG as a superposition of damped oscillatory modes, *Biological Cybernetics*, **53** (1985), 19–25, URL `https://doi.org/10.1007/bf00355687`.

P. J. Franaszczuk, G. K. Bergey, P. J. Durka and H. M. Eisenberg, Time–frequency analysis using the matching pursuit algorithm applied to seizures originating from the mesial temporal lobe, *Electroencephalography and Clinical Neurophysiology*, **106** (1998), 513–521.

S. Gholizadeh and W. Zadrozny, A short survey of topological data analysis in time series and systems analysis, 2018.

R. Ghrist, Barcodes: The persistent topology of data, 2007.

C. Ieracitano, N. Mammone, A. Bramanti, S. Marino, A. Hussain and F. C. Morabito, A time-frequency based machine learning system for brain states classification via eeg signal processing, in *2019 International Joint Conference on Neural Networks (IJCNN)*, 2019, 1–8.

M. G. Jesse Berwald, Critical transitions in a model of a genetic regulatory system, *Mathematical Biosciences and Engineering*, **11** (2014), 723.

F. Khasawneh and E. Munch, Exploring equilibria in stochastic delay differential equations using persistent homology, vol. 8, 2014.

J. F. C. Kingman, *Poisson processes*, vol. 3 of Oxford Studies in Probability, The Clarendon Press Oxford University Press, New York, 1993, Oxford Science Publications.

S. G. Mallat and Zhifeng Zhang, Matching pursuits with time-frequency dictionaries, *IEEE Transactions on Signal Processing*, **41** (1993), 3397–3415.

A. Marchese and V. Maroulas, Signal classification with a point process distance on the space of persistence diagrams, *Advances in Data Analysis and Classification*, **12** (2018), 657–682, URL `https://ideas.repec.org/a/spr/advdac/v12y2018i3d10.1007_s11634-017-0294-x.html`.

V. Maroulas, J. L. Mike and C. Oballe, Nonparametric estimation of probability density functions of random persistence diagrams, *Journal of Machine Learning Research*, **20** (2019), 1–49, URL `http://jmlr.org/papers/v20/18-618.html`.

V. Maroulas, F. Nasrin and C. Oballe, A bayesian framework for persistent homology, *SIAM Journal on Mathematics of Data Science*, **2** (2019), 48–74.

Y. Mileyko, S. Mukherjee and J. Harer, Probability measures on the space of persistence diagrams, *Inverse Problems*, **27** (2011), 124007, URL `https://doi.org/10.1088/2F0266-5611/2F27/2F12/2F124007`.

A. Monod, S. Kalisnik, J. A. Patino-Galindo and L. Crawford, Tropical sufficient statistics for persistent homology, *SIAM Journal on Applied Algebra and Geometry*, **3** (2019), 337–371, URL `https://doi.org/10.1137/17m1148037`.

F. Nasrin, C. Oballe, D. Boothe and V. Maroulas, Bayesian topological learning for brain state classification, in *2019 18th IEEE International Conference On Machine Learning And Applications (ICMLA)*, 2019, 1247–1252.

A. V. Oppenheim, J. R. Buck and R. W. Schafer, *Discrete-Time Signal Processing; 2nd ed.*, Prentice-Hall signal processing, Prentice-Hall, Upper Saddle River, NJ, 1999, URL `https://cds.cern.ch/record/389969`.

J. A. Perea and J. Harer, Sliding windows and persistence: An application of topological methods to signal analysis, *Found. Comput. Math.*, **15**.

R. Pintelon and J. Schoukens, Time series analysis in the frequency domain, *IEEE Transactions on Signal Processing*, **47** (1999), 206–210.

R. Pintelon and J. Schoukens, Time series analysis in the frequency domain, *IEEE Transactions on Signal Processing*, **47** (1999), 206–210.

M. Robinson, *Topological signal processing*, Springer, 2014.

M. D. Sacchi, T. J. Ulrych and C. J. Walker, Interpolation and extrapolation using a high-resolution discrete fourier transform, *IEEE Transactions on Signal Processing*, **46** (1998), 31–38.

N. Sanderson, E. Shugerman, S. Molnar, J. D. Meiss and E. Bradley, Computational topology techniques for characterizing time-series data, in *Advances in Intelligent Data Analysis XVI*, Springer International Publishing, 2017, 284–296, URL `https://doi.org/10.10072F978-3-319-68765-0_24`.

T. Shiraishi, T. Le, H. Kashima and M. Yamada, Topological bayesian optimization with persistence diagrams.

B. W. Silverman, *Density estimation for statistics and data analysis*, 1986.

P. Skraba, V. de Silva and M. Vejdemo-Johansson, Topological analysis of recurrent systems, in *NIPS 2012*, 2012.

Y. Umeda, Time series classification via topological data analysis, *Transactions of The Japanese Society for Artificial Intelligence*, **32**.

Y. Wang, H. Ombao and M. K. Chung, Topological data analysis of single-trial electroencephalographic signals, *Ann. Appl. Stat.*, **12** (2018), 1506–1534, URL `https://doi.org/10.1214/17-AOAS1119`.

Appendix. The following theorem is useful to the proof of Lemma 5.1. Its proof can be found in **?**.

Theorem 5.1 (The Superposition Theorem)**.** Let $\{\Pi_n\}_{n\in\mathbb{N}}$ be a collection of independent Poisson point processes with respective intensity measures Λ_n. Then, their superposition Π given by $\Pi := \bigcup_{n\in\mathbb{N}} \Pi_n$ is a Poisson PP with intensity measure $\Lambda = \sum_{n\in\mathbb{N}} \Lambda_n$.

The next lemma originally appeared in **?**. Here, we present a simplified proof without relying on probability generating functionals, as was done in **?**.

Lemma 5.1. Let D^V and D^O be independent Poisson point processes on $\mathbb{R}^2$ with intensities $(1-\alpha)\lambda(\mathbf{x})$ and $\alpha\lambda(\mathbf{x})$, respectively, for $\alpha \in [0,1]$. Suppose further that D^O is a marked Poisson point process with stochastic kernel $\ell(\mathbf{y}|\mathbf{x})$ and denote the marks of D^O by D^M. Let D^S be a Poisson point process whose intensity is λ_S, and suppose D^S is independent of D^O and D^V. Define $D^X := D^V \cup D^O$ and $D^Y := D^M \cup D^S$, and consider a sample $\mathcal{D}^Y$ from D^Y. Then, the posterior intensity, $\lambda_{D^X|\mathcal{D}^Y}$, for the point process $D_X|\mathcal{D}_Y$ is

$$\lambda_{D^X|\mathcal{D}^Y}(\mathbf{x}) = (1-\alpha)\lambda(\mathbf{x}) + \alpha \sum_{y\in\mathcal{D}^Y} \frac{\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})}{\lambda_S(\mathbf{y}) + \int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})} \tag{8}$$

Proof. By definition, $\lambda_{D^X|\mathcal{D}^\mathcal{Y}}(\mathbf{x})$ is uniquely characterized by the identity

$$\int_A \lambda_{D^X|\mathcal{D}^Y}(\mathbf{x})\, d\mathbf{x} = \mathbb{E}\left(|(D^X|\mathcal{D}^Y) \cap A|\right) \tag{9}$$

for any $A \subset \mathbb{R}^2$, where the expectation on the right hand side of Equation (9) is taken with respect to the probability measure of $D^X|\mathcal{D}^Y$. Hence, it suffices to show that Equation (8) satisfies Equation (9). Notice that

$$D^X|\mathcal{D}^Y = (D^V \cup D^O)|\mathcal{D}^Y \tag{10}$$

$$= D^V \cup (D^O|\mathcal{D}^Y) \tag{11}$$

since D^V is independent of D^Y. Thus, by Theorem 5.1,

$$\mathbb{E}\left(|(D^X|\mathcal{D}^Y) \cap A|\right) = \mathbb{E}\left(|D^V \cup (D^O|\mathcal{D}^Y) \cap A|\right) \tag{12}$$

$$= \mathbb{E}\left(|D^V \cap A|\right) + \mathbb{E}\left(|(D^O|\mathcal{D}^Y) \cap A|\right) \tag{13}$$

By assumption, the left hand term in Equation (13) is

$$\mathbb{E}\left(|D^V \cap A|\right) = \int_A (1-\alpha)\lambda(\mathbf{x})\, d\mathbf{x}. \tag{14}$$

It remains to compute the right hand term in Equation (13). To this end, let $\mathbb{P}(\mathbf{y} \in \mathcal{D}^M, \mathbf{x} \in A)$ be the joint probability that $\mathbf{y}$ is drawn from D^M and that its corresponding element in the draw from D^O, $\mathbf{x}$, lies in A. Notice,

$$\mathbb{E}\left(|(D^O|\mathcal{D}^Y) \cap A|\right) = \sum_{y\in\mathcal{D}^Y} \mathbb{P}(\mathbf{y} \in \mathcal{D}^M, \mathbf{x} \in A) \tag{15}$$

$$= \sum_{\mathbf{y}\in\mathcal{D}^Y} \mathbb{P}(\mathbf{y} \in \mathcal{D}^M)\mathbb{P}(\mathbf{x} \in A|\mathbf{y} \in \mathcal{D}^M) \tag{16}$$

since elements of $\mathcal{D}^Y$ are drawn from D^S or D^M; in the former case, the elements bear no association to those in draws from D^O while in the latter, each element corresponds to exactly one element in a draw from D^O. Moreover, each element in $\mathcal{D}^Y$ is an independent draw from the scaled intensity of D^Y since it can be shown D^Y is a Poisson process **?**. Now, since D^Y is an independent mixture of Poisson processes with intensities λ_M and λ_S, it follows

$$\mathbb{P}(\mathbf{y} \in \mathcal{D}^M) = \frac{\lambda_M(\mathbf{y})}{\lambda_M(\mathbf{y}) + \lambda_S(\mathbf{y})} \tag{17}$$

$$= \frac{\int \alpha\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}}{\lambda_S(\mathbf{y}) + \int \alpha\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}} \tag{18}$$

where the identity $\lambda_M(\mathbf{y}) = \int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}$ follows by the Marking Theorem. To compute the other component in Equation (16), we rely on Bayes rule to compute a posterior probability:

$$\mathbb{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{y}|\mathbf{x})}{\int \mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{y}|\mathbf{x})\, d\mathbf{x}}. \tag{19}$$

Using the fact that D^O is a marked Poisson point process, we substitute for each of the components in Equation (20):

$$\mathbb{P}(\mathbf{x}|\mathbf{y}) = \frac{\mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{y}|\mathbf{x})}{\int \mathbb{P}(\mathbf{x})\mathbb{P}(\mathbf{y}|\mathbf{x})\, d\mathbf{x}} \tag{20}$$

$$= \frac{\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})}{\int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}}. \tag{21}$$

Next, we integrate Equation (21) over A:

$$\mathbb{P}(\mathbf{x} \in A|\mathbf{y} \in \mathcal{D}^M) = \frac{1}{\int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{u})\, d\mathbf{u}} \int_A \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x} \tag{22}$$

Substituting Equations (18) and (22) in Equation (16) yields

$$\mathbb{E}\left(|(D^O|\mathcal{D}^Y) \cap A|\right) = \sum_{y \in \mathcal{D}^Y} \int_A \left(\frac{\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})}{\lambda_S(\mathbf{y}) + \int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}}\right) d\mathbf{x}. \tag{23}$$

Finally, substituting Equations (23) and (14) into Equation 9 reveals:

$$\mathbb{E}\left(|(D^X|\mathcal{D}^Y) \cap A|\right) = \int_A \Big((1-a)\lambda(\mathbf{x}) + \sum_{\mathbf{y} \in \mathcal{D}^Y} \frac{\alpha\lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})}{\lambda_S(\mathbf{y}) + \int \lambda(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\, d\mathbf{x}}\Big)\, d\mathbf{x}, \tag{24}$$

which upon inspection establishes the lemma. □

Lemma 5.2. Let $\mathcal{D}^{Y_{1:m}} = \{\mathcal{D}^{Y_1}, \mathcal{D}^{Y_2}, \ldots, \mathcal{D}^{Y_m}\}$ be independent samples from D^Y in Lemma 8. Then $\lambda_{D^X|\mathcal{D}^{Y_{1:m}}}(\mathbf{x}) = \frac{1}{m}\sum_{i=1}^m \lambda_{D^X|\mathcal{D}^{Y_i}}(\mathbf{x})$ is an unbiased estimator of the intensity $\lambda_{D^X|\mathcal{D}^Y}(\mathbf{x})$ in Lemma 5.1.

Proof. By Theorem 5.1, the intensity for the point process given by the superposition $\cup_{i=1}^m D^X|\mathcal{D}^{Y_i}$ is $m\lambda_{D^X|\mathcal{D}^Y}(\mathbf{x})$. The claim now follows. □

The last tool we need to prove Theorem 3.1 is a standard result that gives a closed form for the product of two Gaussian densities. Its proof may be found in **?**.

Lemma 5.3. Let $\mathcal{N}(\mathbf{x};\mu,P)$ denote the two-dimensional normal density with mean μ and covariance matrix P. For 2x2 matrices H and R, with H positive definite, $\mathcal{N}(\mathbf{y};H\mathbf{x},R)\mathcal{N}(\mathbf{x};\mu,P) = q(\mathbf{y})\mathcal{N}(\mathbf{x};\mu_*,P_*)$, where $q(\mathbf{y}) = \mathcal{N}(\mathbf{y};H\mu,R+HPH^T)$, $\mu_* = \mu + K(\mathbf{y}-H\mu)$, $P_* = (I-KH)P$, and $K = PH^T(HPH^T+R)^{-1}$.

With Lemma 5.1, Lemma 5.2, and Lemma 5.3 in hand, we are ready to prove Theorem 3.1.

Proof of Theorem 3.1. By Lemmas 5.1 and 5.2, we have:

$$\lambda_{D^X|\mathcal{D}^{Y_{1:m}}}(\mathbf{x}) = (1-\alpha)\lambda_{D^X}(\mathbf{x}) + \frac{\alpha}{m}\sum_{i=1}^{m}\sum_{\mathbf{y}\in\mathcal{D}^{Y_i}}\frac{\lambda_{D^X}(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})}{\lambda_P(\mathbf{y}) + \int_{\mathbb{W}}\lambda_{D^X}(\mathbf{x})\ell(\mathbf{y}|\mathbf{x})\,d\mathbf{x}}, \quad (25)$$

Theorem 3.1 follows by substituting the densities for λ_{D^X}, ℓ, λ_P, then appealing to the identities in Lemma 5.3. □

Next, we prove Proposition 3.1, which directly relates the persistence diagrams of a family of signals to their frequencies and instantaneuos amplitudes.

Proof. Without loss of generality, suppose f is an integer. The case for real-valued f is similar. By definition, $a_\beta(t)\phi(2\pi ft)$ has f cycles in $[0,1]$. Let M_i and m_i respectively denote the maximum and local minimum of the ith cycle of $a_\beta(t)\phi(2\pi ft)$ in $[0,1]$. Since each cycle has a unique local minimum, the only connected components that arise during the sublevel set filtration of $a_\beta(t)\phi(2\pi ft)$ are C_{m_i}, which are clearly born at values m_i of the sublevel set filtration. It remains to deduce the value at which each C_{m_i} disappears. From the assumption that $\phi(0)$ is a local maximum, each cycle begins at a local maximum and monotonicity of $a_\beta(t)$ ensures that M_i occurs at the beginning of the ith cycle and at the end of the $(i-1)$th cycle. Recall that connected components merge at values M_i during the sublevel set filtration. By monotonicity of a_β, the smallest of the M_i (and hence the first one encountered in the sublevel set filtration) is M_f. This merges connected components C_{m_f} and $C_{m_{f-1}}$. Once again by monotonicity of $a_\beta(t)$, $m_f \le m_{f-1}$, so by the Elder Rule, C_{m_f} disappears by merging into $C_{m_{f-1}}$. Thus, the first point in $D^{f,\beta}$ is (m_f, M_f). Using the monotonicity of $a_\beta(t)$ to continue in this fashion, we see that $\mathcal{D}^{f,\beta} = \{(m_i,M_i)\}_{i=1}^{f}$. The result now follows by explicitly computing m_i and M_i, which is done by exploiting the periodicity of ϕ. □

Signal Length	1 Second								5 Seconds							
	f_1	f_2	f_3	f_4	β_1	β_2	β_3	β_4	f_1	f_2	f_3	f_4	β_1	β_2	β_3	β_4
Signal 1	0	5.87	18.59	-	344.80	5.37	16.6	-	0	6.00	14.4	20.85	24.98	10.54	31.64	26.97
Signal 2	0	10.70	-	-	202.78	7.41	-	-	0	10.16	23.02	-	17.24	4.06	20.37	-

TABLE 2. Parameter values for autoregressive model determined by fitting to real EEG. Missing values indicate that the optimal AR model order did not include a corresponding frequency component.

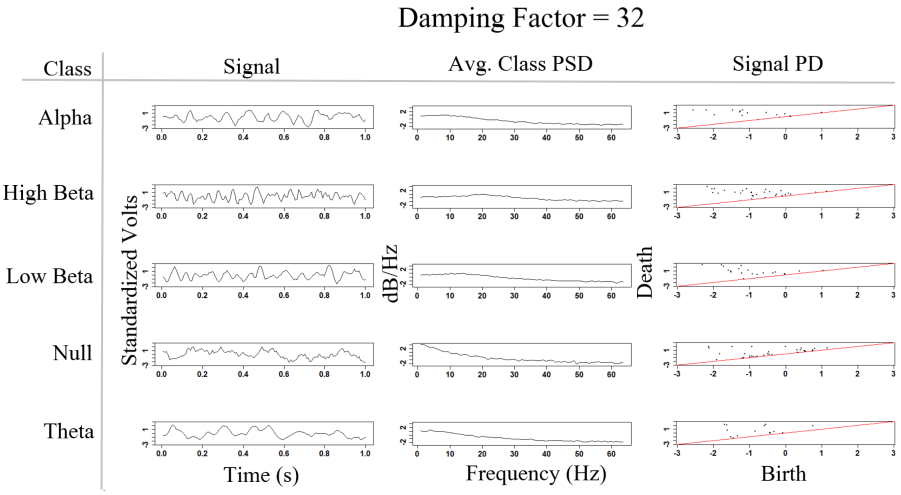

FIGURE 8. The average (log) power spectral densities along with examples of signals and persistence diagrams from each class for damping factors of top) 4, and bottom) 32.

Received xxxx 20xx; revised xxxx 20xx.

E-mail address: coballe@nd.edu
E-mail address: acherne@vols.utk.edu
E-mail address: david.l.boothe7.civ@mail.mil
E-mail address: scott.e.kerick.civ@nd.edu
E-mail address: piotr.j.franaszczuk.civ@mail.mil
E-mail address: vmaroula@utk.edu