A Bayesian Framework for Persistent Homology[∗]

Vasileios Maroulas[†], Farzana Nasrin[†], and Christopher Oballe[†]

Abstract. Persistence diagrams offer a way to summarize topological and geometric properties latent in datasets. While several methods have been developed that use persistence diagrams in statistical inference, a full Bayesian treatment remains absent. This paper, relying on the theory of point processes, presents a generalized Bayesian framework for inference with persistence diagrams relying on a substitution likelihood argument. In essence, we model persistence diagrams as Poisson point processes with prior intensities and compute posterior intensities by adopting techniques from the theory of marked point processes. We then propose a family of conjugate prior intensities via Gaussian mixtures to obtain a closed form of the posterior intensity. Finally, we demonstrate the utility of this generalized Bayesian framework with a classification problem in materials science using Bayes factors.

1. Introduction. A crucial first step in understanding patterns and properties of a crystalline material is determining its crystal structure. For highly disordered metallic alloys, such as high entropy alloys (HEAs), atom probe tomography (APT) gives a snapshot of the local atomic environment; see Figure 1. However, APT has two main drawbacks: experimental noise and an abundance of missing data. Approximately 65% of the atoms in a sample are not registered in a typical experiment [50], and the spatial coordinates of those identified atoms are corrupted by experimental noise [42]. Understanding the atomic pattern within HEAs using an APT image requires observation of atomic cubic unit neighborhood cells under a microscope. This is problematic, as APT may have a spatial resolution approximately the length of the unit cell under consideration [28, 42]. Hence, the process is unable to see the finer details of a material, rendering the determination of a lattice structure a challenging problem [55, 39]. Existing algorithms for detecting the crystal structure [14, 24, 25, 32, 43, 56] are not

[†]Department of Mathematics, University of Tennessee, Knoxville, TN 37996 (vmaroula@utk.edu, fnasrin@utk.edu, coballe@vols.utk.edu).

able to establish the crystal lattice of an APT dataset, as they rely on symmetry arguments based on identifying repeating parts of molecules. Consequently, the field of atom probe crystallography, i.e., determining the crystal structure from APT data, has emerged in recent years [22, 43]. Algorithms in this field rely on knowing the global lattice structure a priori and aim to determine local small-scale structures within a larger sample. For some materials, this information is readily known, while for others, such as HEAs, the global structure is unknown and must be inferred.
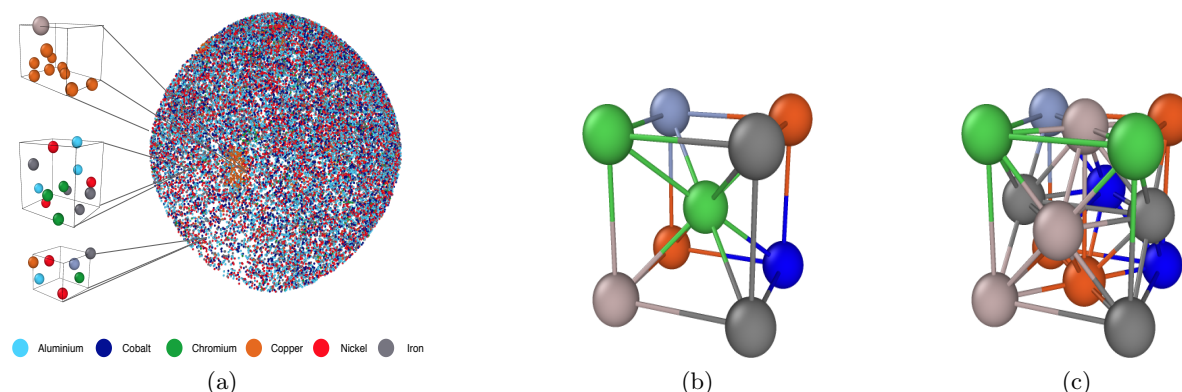

Figure 1. (a) *Image of APT data with atomic neighborhoods shown in detail on the left. Each pixel represents a different atom, the neighborhood of which is considered. Certain patterns with distinct crystal structures exist, e.g., the orange region is copper-rich (left), but overall no pattern is identified. Putting a single atomic cubic unit cell under a microscope, the true crystal structure of the material, which could be either body-centered cubic* (b) *or face-centered cubic* (c)*, is not revealed. This distinction is obscured due to further experimental noise. Notice there is an essential topological difference between the two structures in* (b) *and* (c)*: The BCC structure has one atom at its center, whereas the FCC is hollow in its center, but has one atom in the center of each of its faces.*

In this work, we specifically classify unit cells that are either body-centered cubic (BCC) or face-centered cubic (FCC). These lattice structures are the essential building blocks of HEAs [61] and have fundamental differences that set them apart in the case of noise-free, complete materials data. The BCC structure has a single atom in the center of the cube, while the FCC has a void in its center but has atoms on the center of the cubes' faces; see Figure 1(b)–(c). These two crystal structures are distinct when viewed through the lens of topology. Differentiating between the empty space and connectedness of these two lattice structures allows us to create an accurate classification rule. This fundamental distinction between BCC and FCC point clouds is captured well by topological methods and explains the high degree of accuracy in the classification scheme presented herein. Indeed, we offer a Bayesian classification framework for topological features, and although we focus on materials science data in this paper, the generalized Bayesian framework can handle any datasets.

Overall, topological data analysis (TDA) encompasses a broad set of techniques that explore topological structure in datasets [18, 23, 13, 59]. One of these techniques, persistent homology, associates shapes to data and summarizes salient features with persistence diagrams—multisets of points that represent homological features along with their appearance and disappearance scales [18]. Features of a persistence diagram that exhibit long persistence

describe global topological properties in the underlying dataset, while those with shorter persistence encode information about local geometry and/or noise. Hence, persistence diagrams can be considered multiscale summaries of data's shape. While there are several methods present in the literature to compute persistence diagrams, we adopt geometric complexes that are typically used for applications of persistent homology to data analysis in various settings, such as handwriting analysis [2], studying of brain arteries [5, 6], image analysis [8, 12, 11], neuroscience [15, 54, 4], sensor networks [17, 52], protein structure [21, 31], biology [51, 40, 45], dynamical systems [29], action recognition [58], signal analysis [35, 34, 47, 36], chemistry [60], genetics [26], object data [46], etc.

Researchers desire to use persistence diagrams in inference and classification problems. Several achieve this directly with persistence diagrams [37, 35, 7, 20, 41, 49, 10], while others elect to first map them into a Hilbert space [9, 48, 1, 57, 19]. The latter approach enables one to adopt traditional machine learning and statistical tools such as principal component analysis, random forests, support vector machines, and more general kernel-based learning schemes. Despite progress toward statistical inference, to the best of our knowledge, a full Bayesian treatment predicated upon creating posterior distributions of persistence diagrams is still absent in the literature. The first Bayesian considerations in a TDA context take place in [41], where the authors discuss a conditional probability setting on persistence diagrams where the likelihood for the observed point cloud has been substituted by the likelihood for its associated topological summary.

The homological features in persistence diagrams have no intrinsic order implying they are random sets as opposed to random vectors. To that end, we model random persistence diagrams as Poisson point processes. The defining feature of these point processes is that they are solely characterized by a single parameter known as the intensity. Utilizing the theory of marked point processes, we obtain a method for computing posterior intensities that does not require us to consider explicit maps between input diagrams and underlying parameters, alleviating the computational burden associated with deriving the posterior intensity from Bayes' rule alone.

In particular, for a given collection of observed persistence diagrams, we consider the underlying stochastic phenomena generating persistence diagrams to be Poisson point processes with prior uncertainty captured in presupposed intensities. In applications, one may select an informative prior by choosing an intensity based on expert opinion, or alternatively choose an uninformative prior intensity when information is not available. The likelihood functions in our model represent the level of belief that observed diagrams are representative of the entire population. We build this analogue using the theory of marked Poisson point processes [16]. A central idea of this paper is to use the topological summaries of point clouds in place of the actual point clouds. This provides a powerful tool with applications in wide ranging fields. The application considered in this paper is the classification of the crystal structure of materials, which allows scientists to predict the properties of a crystalline material. Our goal is to view point clouds through their topological descriptors, as this can reveal essential shape peculiarities latent in the point clouds. Our generalized Bayesian method adopts a substitution likelihood technique by Jeffreys in [27] instead of considering the full likelihood for the point cloud. A similar sort of discussion was considered in [41] for defining conditional probability on persistence diagrams.

Another key contribution of this paper is the derivation of a *closed form* of the posterior intensity, which relies on conjugate families of Gaussian mixtures. An advantage of this Gaussian mixture representation is that it allows us to perform Bayesian inference in an efficient and reliable manner. Indeed, this model can be viewed as an analogue of the ubiquitous example in standard Bayesian inference where a Gaussian prior and likelihood yield a Gaussian posterior. We present a detailed example of our closed form implementation to demonstrate computational tractability and showcase its applicability by using it to build a Bayes factor classification algorithm; we test the latter in a classification problem for materials science data.

The contributions of this work are

1. Theorem 3.1, which provides the generalized Bayesian framework for computing the posterior distribution of persistence diagrams;
2. Proposition 3.2, which yields a conjugate family of priors based on a Gaussian mixture for the proposed Bayesian framework; and
3. a classification scheme using Bayes factors considering the posteriors of persistence diagrams and its application to a materials science problem.

This paper is organized as follows. Section 2 provides a brief overview of persistence diagrams and general point processes. Our methods are presented in section 3. In particular, subsection 3.1 establishes the Bayesian framework for persistence diagrams, while subsection 3.2 contains the derivation of a closed form for a posterior distribution based on a Gaussian mixture model. A classification algorithm with Bayes factors is discussed in section 4. To assess the capability of our algorithm, we investigate its performance on materials data in subsection 4.1. Finally, we end with discussions and conclusions in section 5.

2. Background. We begin by discussing preliminary definitions essential for building our model. In subsection 2.1, we briefly review simplicial complexes and provide a formal definition for persistence diagrams (PDs). Pertinent definitions and theorems from point processes (PPs) are discussed in subsection 2.2.

2.1. Persistence diagrams. We start by discussing simplices and simplicial complexes, intermediary structures for constructing PDs.

Definition 2.1. *A d-dimensional collection of data $\{v_0, \ldots, v_n\} \subset \mathbb{R}^{\mathrm{d}} \setminus \{0\}$ is said to be geometrically independent if for any set $t_i \in \mathbb{R}$ with $\sum_{i=0}^{n} t_i = 0$, the equation $\sum_{i=0}^{n} t_i v_i = 0$ implies that $t_i = 0$ for all $i \in \{0, \ldots, n\}$.*

Definition 2.2. *A k-simplex is a collection of $k+1$ geometrically independent elements along with their convex hull: $[v_0, \ldots, v_k] = \left\{ \sum_{i=0}^{k} \alpha_i v_i : \sum_{i=0}^{k} \alpha_i = 1 \right\}$. We say that the vertices $v_0, \ldots, v_n$ span the k-dimensional simplex, $[v_0, \ldots, v_k]$. The faces of a k-simplex $[v_0, \ldots, v_k]$ are the $(k-1)$-simplices spanned by subsets of $\{v_0, \ldots, v_k\}$.*

Definition 2.3. *A simplicial complex S is a collection of simplices satisfying two conditions: (i) if $\xi \in S$, then all faces of ξ are also in S, and (ii) the intersection of two simplices in S is either empty or contained in S.*

Given a point cloud X, our goal is to construct a sequence of simplicial complexes that reasonably approximates the underlying shape of the data. We accomplish this by using the

Vietoris–Rips filtration.

Definition 2.4. *Let $X = \{x_i\}_{i=0}^{L}$ be a point cloud in $\mathbb{R}^{\mathrm{d}}$, and let $r > 0$. The Vietoris–Rips complex of X is defined to be the simplicial complex $\mathcal{V}_r(X)$ satisfying $[x_{i_1}, \ldots, x_{i_l}] \in \mathcal{V}_r(X)$ if and only if $\mathrm{diam}(x_{i_1}, \ldots, x_{i_l}) < r$. Given a nondecreasing sequence $\{r_n\} \in \mathbb{R}^+ \cup \{0\}$ with $r_0 = 0$, we denote its Vietoris–Rips filtration by $\{\mathcal{V}_{r_n}(X)\}_{n \in \mathbb{N}}$.*
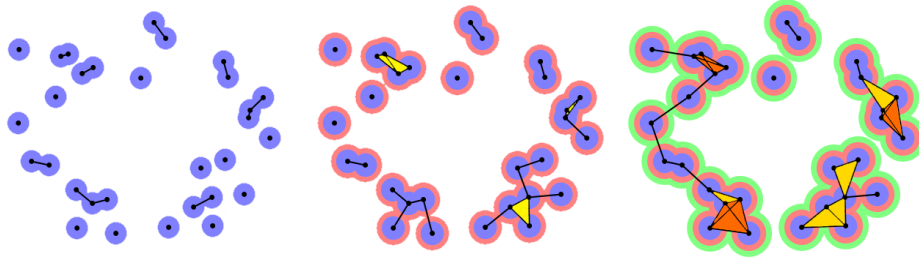

Figure 2. *Vietoris–Rips complexes.*

A PD $\mathcal{D}$ is a multiset of points in $\mathcal{W} := \mathbb{W} \times \{0, 1, \ldots, \mathrm{d} - 1\}$, where $\mathbb{W} := \{(b, d) \in \mathbb{R}^2 : d \geq b \geq 0\}$ and each element (b, d, k) represents a homological feature of dimension k that appears at scale b during a Vietoris–Rips filtration and disappears at scale d. Intuitively speaking, the feature (b, d, k) is a k-dimensional hole lasting for duration $d - b$. Namely, features with $k = 0$ correspond to connected components, $k = 1$ to loops, and $k = 2$ to voids. An illustration of Vietoris–Rips complexes is shown in Figure 2, and an example of a PD is shown in Figure 3.

2.2. Poisson point processes. This section contains basic definitions and fundamental theorems from PPs, primarily Poisson PPs. Detailed treatments of Poisson PPs can be found in [16] and references therein. For the remainder of this section, we take $\mathbb{X}$ and $\mathcal{X}$ to be a Polish space and its Borel σ-algebra, respectively.

Definition 2.5. *A finite PP $\mathcal{P}$ is a pair $(\{p_n\}, \{\mathbb{P}_n\})$ where $\sum_{n=0}^{\infty} p_n = 1$ and $\mathbb{P}_n$ is a symmetric probability measure on $\mathcal{X}^n$, where $\mathcal{X}^0$ is understood to be the trivial σ-algebra.*

The sequence $\{p_n\}$ defines a cardinality distribution, and the measures $\{\mathbb{P}_n\}$ give spatial distributions of vectors $(x_1, \ldots, x_n)$ for fixed n. Definition 2.5 naturally prescribes a method for sampling a finite PP: (i) determine the number of points n by drawing from $\{p_n\}$, and then (ii) spatially distribute $(x_1, \ldots, x_n)$ according to a draw from $\mathbb{P}_n$. As PPs model random collections of elements in $\{x_1, \ldots, x_n\} \subset \mathbb{X}$ whose order is irrelevant, any sensible construction relying on random vectors should assign equal weight to all permutations of $(x_1, \ldots, x_n)$. This is ensured by the symmetry requirement in Definition 2.5. We abuse notation and write $\mathcal{P}$ for samples from $\mathcal{P}$ as well as their set representations. It proves useful to describe finite PPs by a set of measures that synthesize p_n and $\mathbb{P}_n$ to simultaneously package cardinality and spatial distributions.

Definition 2.6. *Let $(\{p_n\}, \{\mathbb{P}_n\})$ be a finite PP. The Janossy measures $\{\mathbb{J}_n\}$ are defined as the set of measures satisfying $\mathbb{J}_n(A) = n! p_n \mathbb{P}_n(A)$ for all $n \in \mathbb{N}$ and $A \in \mathcal{X}^n$.*

Given a collection of disjoint rectangles $A_1, \ldots, A_n \subset \mathbb{X}$, the value $\mathbb{J}_n(A_1 \times \cdots \times A_n)$ is the probability of observing exactly one element in each of $A_1, \ldots, A_n$ and none in the complement of their union. For applications, we are primarily interested in Janossy measures $\mathbb{J}_n$ that admit densities j_n with respect to a reference measure on $\mathbb{X}$. We are now ready to describe the class of finite PPs that model PDs.

Definition 2.7. *Let Λ be a finite measure on $\mathbb{X}$, and let $\mu := \Lambda(\mathbb{X})$. The finite PP Π is Poisson if, for all $n \in \mathbb{N}$ and disjoint measurable rectangles $A_1 \times \cdots \times A_n \in \mathcal{X}^n$, $p_n = e^{-\mu}\frac{\mu^n}{n!}$, and $\mathbb{P}_n(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n}\left(\frac{\Lambda(A_i)}{\mu}\right)$. We call Λ an intensity measure.*

Equivalently, a Poisson PP is a finite PP with Janossy measures $\mathbb{J}_n(A_1 \times \cdots \times A_n) = e^{-\mu}\prod_{i=1}^{n}\Lambda(A_i)$. The intensity measure in Definition 2.7 admits a density, λ, with respect to some reference measure on $\mathbb{X}$. Notice that for all $A \in \mathcal{X}$,

$$\mathbb{E}(|\Pi \cap A|) = \sum_{n=0}^{\infty} p_n \mathbb{E}_{\mathbb{P}_n}\left(\sum_{k=0}^{n} k\binom{n}{k}\mathbb{1}_{A^k \times (A^c)^{n-k}}\right).$$

Elementary calculations then show that $\mathbb{E}(|\Pi \cap A|) = \Lambda(A)$. Thus, we interpret the intensity measure of a region A, $\Lambda(A)$ as the expected number of elements in Π that land in A. The intensity measure serves as an analogue to the first-order moment for a random variable (RV).

The next two definitions involve a joint PP wherein points from one space parameterize distributions for the points living in another. Consequently, we introduce another Polish space $\mathbb{M}$ along with its Borel σ-algebra $\mathcal{M}$ to serve as the mark space in a marked Poisson PP. These model scenarios in which points drawn from a Poisson PP provide a data likelihood model for Bayesian inference with PPs.

Definition 2.8. *Suppose $\ell : \mathbb{X} \times \mathbb{M} \to \mathbb{R}^+ \cup \{0\}$ is a function satisfying the following: (1) for all $x \in \mathbb{X}$, $\ell(x, \bullet)$ is a probability measure on $\mathbb{M}$, and (2) for all $B \in \mathcal{M}$, $\ell(\bullet, B)$ is a measurable function on $\mathbb{X}$. Then, ℓ is a stochastic kernel from $\mathbb{X}$ to $\mathbb{M}$.*

Definition 2.9. *A marked Poisson PP Π_M is a finite PP on $\mathbb{X} \times \mathbb{M}$ such that the following hold: (i) $(\{p_n\}, \{\mathbb{P}_n(\bullet \times \mathbb{M})\})$ is a Poisson PP on $\mathbb{X}$, and (ii) for all $(x_1, \ldots, x_n) \in \mathbb{X}^n$, measurable rectangles $B_1 \times \cdots \times B_n \in \mathcal{M}^n$ and $\mathbb{P}_n((x_1, \ldots, x_n) \times B_1 \times \cdots \times B_n) = \frac{1}{n!}\sum_{\pi \in \mathcal{S}_n}\prod_{i=1}^{n}\ell(x_{\pi(i)}, B_i)$, where $\mathcal{S}_n$ is the set of all permutations of $(1, \ldots, n)$ and ℓ is a stochastic kernel.*

Given a set of observed marks $M = \{y_1, \ldots, y_m\}$, after adopting Definition 2.6, it can be shown [53] that the Janossy densities for the PP induced by Π_M on $\mathbb{X}$ given M are

$$(2.1) \qquad j_{n|M}(x_1, \ldots, x_n) = \begin{cases} \mathbb{1}_\emptyset, & n = m = 0, \\ \sum_{\pi \in \mathcal{S}_n}\prod_{i=1}^{n} p(x_i|y_{\pi(i)}), & n = m > 0, \\ 0 & \text{otherwise}, \end{cases}$$

where p is the stochastic kernel for Π_M evaluated in $\mathbb{X}$ for a fixed value of $y \in \mathbb{M}$.

The following theorems allow us to construct new Poisson PPs from existing ones. Their proofs can be found in [30].

Theorem 2.10 (the superposition theorem). *Let $\{\Pi_n\}_{n\in\mathbb{N}}$ be a collection of independent Poisson PPs each having intensity measure Λ_n. Then their superposition Π given by $\Pi := \bigcup_{n\in\mathbb{N}}\Pi_n$ is a Poisson PP with intensity measure $\Lambda = \sum_{n\in\mathbb{N}}\Lambda_n$.*

Theorem 2.11 (the mapping theorem). *Let Π be a Poisson PP on $\mathbb{X}$ with σ-finite intensity measure Λ, and let $(\mathbb{T},\mathcal{T})$ be a measurable space. Suppose $f : \mathbb{X} \to \mathbb{T}$ is a measurable function. Write Λ^* for the induced measure on T given by $\Lambda^*(B) := \Lambda(f^{-1}(B))$ for all $B \in \mathcal{T}$. If Λ^* has no atoms, then $f \circ \Pi$ is a Poisson PP on T with intensity measure Λ^*.*

Theorem 2.12 (the marking theorem). *The marked Poisson PP in Definition 2.9 has the intensity measure given by $\Lambda_M(C) = \iint_C \Lambda(dx)\ell(x,dm)$ for all $C \in \mathcal{X}\times\mathcal{M}$, where Λ is the intensity measure for the Poisson PP that Π_M induces on $\mathbb{X}$, and ℓ is a stochastic kernel.*

The final tool we need is the probability generating functional, as it enables us to recover intensity measures using a notion of differentiation. The probability generating functional can be interpreted as the PP analogue of the probability generating function.

Definition 2.13. *Let $\mathcal{P}$ be a finite PP on a Polish space $\mathbb{X}$. Denote by $\mathcal{B}(\mathbb{C})$ the set of all functions $h : \mathbb{X} \to \mathbb{C}$ with $||h||_\infty < 1$. The probability generating functional of $\mathcal{P}$ denoted by $G : \mathcal{B}(\mathbb{C}) \to \mathbb{R}$ is given by*

$$(2.2) \qquad G(h) = J_0 + \sum_{n=1}^{\infty} \frac{1}{n!}\int_{\mathbb{X}^n}\left(\prod_{j=1}^{n}h(x_j)\right)\mathbb{J}_n(dx_1\ldots dx_n).$$

Definition 2.14. *Let G be the probability generating functional given in (2.2). The functional derivative of G in the direction of $\eta \in \mathcal{B}(C)$ evaluated at h, when it exists, is given by $G'(h;\eta) = \lim_{\epsilon\to 0}\frac{G(h+\epsilon\eta)-G(h)}{\epsilon}$.*

It can be shown that the functional derivative satisfies the familiar product rule [33]; namely, for $G = \prod_{i=1}^m G_i$,

$$(2.3) \qquad G'(h;\gamma) = \sum_{i=1}^{m}G_i'(h;\gamma)\prod_{j\neq 1}G_j(h).$$

As proved in [44], the intensity measure Λ of the Poisson PP in Definition 2.7 can be obtained by differentiating G, i.e.,

$$(2.4) \qquad \Lambda(A) = G'(1;\mathbb{1}_A),$$

where $\mathbb{1}_A$ is the indicator function for any $A \in \mathcal{X}$. Generally speaking, one obtains the intensity measure for a general PP through $\Lambda(A) = \lim_{h\to 1} G'(h;\mathbb{1}_A)$, but the preceding identity suffices for our purposes since we only consider PPs for which (2.2) is defined for all bounded h.

Corollary 2.15. *The intensity function for the PP whose Janossy densities are listed in (2.1) is $\sum_{i=1}^m p(x|y_i)$ if $m > 0$ and 0 otherwise.*

Proof. We substitute the Janossy measures $\mathbb{J}_n$ for the PP (see (2.1)) into (2.2). This yields $G(h) = \int_{\mathbb{X}^m}\frac{1}{m!}\sum_{\pi\in\mathcal{S}_n}\prod_{i=1}^n p(x_i|y_{\pi(i)})\,dx_i$ if $m > 0$, and $G(h) = \mathbb{1}_\emptyset$ otherwise. It is easy to
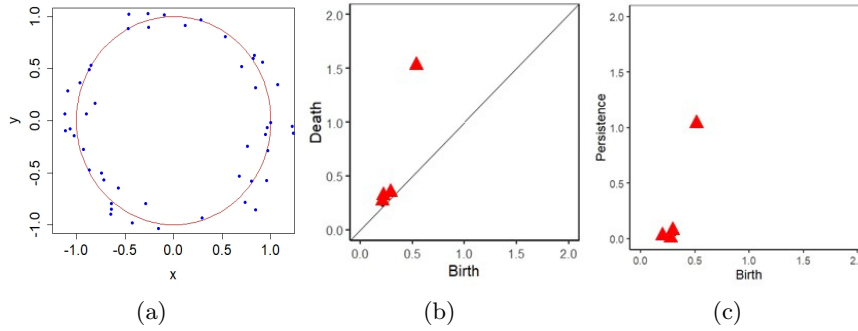
Figure 3. (a) *An example of a dataset;* (b) *its PD;* (c) *its tilted representation.*

see that the functional derivative (Definition 2.14) of a constant is zero, so our claim follows immediately in the latter case by appealing to (2.4). In the former case, we have

$$G(h) = \int_{\mathbb{X}^m} \frac{1}{m!} \sum_{\pi \in \mathcal{S}_m} \prod_{i=1}^{n} h(x_i) p(x_i|y_{\pi(i)}) \, dx_i$$

$$= \frac{1}{m!} \sum_{\pi \in \mathcal{S}_m} \int_{\mathbb{X}^m} \prod_{i=1}^{m} h(x_i) p(x_i|y_{\pi(i)}) \, dx_i \quad \text{(by linearity of the integral)}$$

$$= \frac{1}{m!} \sum_{\pi \in \mathcal{S}_m} \left(\prod_{i=1}^{m} \int_{\mathbb{X}} h(x) p(x|y_{\pi(i)}) \, dx \right) \quad \text{(by Fubini's theorem)}$$

$$= \prod_{i=1}^{m} \int_{\mathbb{X}} h(x) p(x|y_i) \, dx \quad \text{(by symmetry).}$$

Write $G_i(h)$ for the probability generating functional of the PP with $j_1 = p(x|y_i)$ and $j_n = 0$ for $n \neq 1$. We can summarize the preceding string of equalities by $G(h) = \prod_{i=1}^{m} G_i(h)$. We apply the product rule for functional derivatives (see (2.3)) and then substitute $h = 1$ and $\gamma = \mathbb{1}_{x'}$ (the indicator of a singleton) to obtain

$$G'(1; \mathbb{1}_{x'}) = \sum_{i=1}^{m} p(x'|y_i) \prod_{i \neq j} \int_{\mathbb{X}} p(x|y_j) \, dx = \sum_{i=1}^{m} p(x|y_i).$$

Appealing to (2.4) establishes the claim. ■

3. Bayesian inference. In this section, we construct a framework for Bayesian inference with PDs by modeling them as Poisson PPs. First, we derive a closed form for the posterior intensity given a PD drawn from a finite PP, and then we present a family of conjugate priors followed by an example.

3.1. Model. Given a PD $\mathcal{D}$, the map $T : \mathbb{W} \to T(\mathbb{W})$ given by $T(b, d) = (b, d - b)$ defines tilted representation of $\mathcal{D}$ as $T(\mathcal{D}) = \cup_{(b,d,k) \in \mathcal{D}}(T(b, d), k)$; see Figure 3. In what follows, we assume all PDs are given in their tilted representations and, unless otherwise noted, abuse

notation by writing $\mathbb{W}$ and $\mathcal{D}$ for $T(\mathbb{W})$ and $T(\mathcal{D})$, respectively. We also fix the homological dimension of features in a PD by defining $\mathcal{D}^k := \{(b,d) \in \mathbb{W} \; : \; (b,d,k) \in \mathcal{D}\}$.

According to Bayes' theorem, posterior density is proportional to the product of a likelihood function and a prior. To adopt the Bayesian framework to PDs, we need to define two models. In particular, our Bayesian framework views a random PD as a Poisson PP equipped with a prior intensity while observed PDs $\mathcal{D}_Y$ are considered to be marks from a marked Poisson PP. This enables modification of the prior intensity by incorporating observed PDs, yielding a posterior intensity based on data. Some parallels between our Bayesian framework and that for RVs are illustrated in Table 1.

Table 1
The parallels between the Bayesian framework for RVs and its counterpart for random PDs.

	Bayesian framework for RVs	Bayesian framework for random PDs
Prior	modeled by a prior density f	modeled by a Poisson PP with prior intensity λ
Likelihood	depends on observed data	stochastic kernel that depends on observed PDs
Posterior	compute the posterior density	a Poisson PP with posterior intensity

Let $(\mathcal{D}_X^k, \mathcal{D}_Y^k) \in \mathbb{W} \times \mathbb{W}$ be a finite PP, and consider the following:

(M1) For $k_1 \neq k_2$, $(\mathcal{D}_X^{k_1}, \mathcal{D}_Y^{k_1})$ and $(\mathcal{D}_X^{k_2}, \mathcal{D}_Y^{k_2})$ are independent.

(M2) For k fixed, $\mathcal{D}_X^k = \mathcal{D}_{X_O}^k \cup \mathcal{D}_{X_V}^k$, and some $\alpha : \mathbb{W} \to [0,1]$, $\mathcal{D}_{X_O}^k$ and $\mathcal{D}_{X_V}^k$ are independent Poisson PPs having intensity functions $\alpha(x)\lambda_{\mathcal{D}_X^k}(x)$ and $(1 - \alpha(x))\lambda_{\mathcal{D}_X^k}(x)$, respectively.

(M3) For k fixed, $\mathcal{D}_Y^k = \mathcal{D}_{Y_O}^k \cup \mathcal{D}_{Y_S}^k$, where the following hold:

    (i) $(\mathcal{D}_{X_O}^k, \mathcal{D}_{Y_O}^k)$ is a marked Poisson PP with a stochastic kernel density $\ell(y|x)$.

    (ii) $\mathcal{D}_{Y_O}^k$ and $\mathcal{D}_{Y_S}^k$ are independent finite Poisson PPs where $\mathcal{D}_{Y_S}^k$ has intensity function $\lambda_{\mathcal{D}_{Y_S}^k}$.

Hereafter we abuse notation by writing $\mathcal{D}_X$ for $\mathcal{D}_X^k$. The modeling assumption (M1) allows us to develop results independently for each homological dimension k and then combine them using independence. In (M2), the random PD $\mathcal{D}_X$ is modeled as a Poisson PP with prior intensity $\lambda_{\mathcal{D}_X}$. There are two cases we may encounter for any point x from the prior intensity due to the nature of PDs. We assign a probability function $\alpha(x)$ to accommodate these two possibilities. Depending upon the noise level in data, any feature x in $\mathcal{D}_X$ may not be represented in observations; this scenario happens with probability $1 - \alpha(x)$, and we denote this case as $\mathcal{D}_{X_V}$ in (M2). Otherwise, a point x is observed with a probability of $\alpha(x)$ and this scenario is presented as $\mathcal{D}_{X_O}$ in (M2). Consequently, the intensities of $\mathcal{D}_{X_O}$ and $\mathcal{D}_{X_V}$ are proportional to the intensity $\lambda_{\mathcal{D}_X}$ weighted by $\alpha(x)$ and $1 - \alpha(x)$, respectively, and the total prior intensity for $\mathcal{D}_X$ is given by their sum. (M3) considers observed PD $\mathcal{D}_Y$ and decomposes it into two independent PDs, $\mathcal{D}_{Y_O}$ and $\mathcal{D}_{Y_S}$. $\mathcal{D}_{Y_O}$ is linked to $\mathcal{D}_{X_O}$ via a marked PP with likelihood $\ell(y|x)$ defined in (2.1), whereas the component $\mathcal{D}_{Y_S}$ includes any point y that arises from noise or unanticipated geometry. See Figure 4 for a graphical representation of these ideas.

Theorem 3.1 (Bayesian theorem for PDs). *Let $\mathcal{D}_X$ be a PD modeled by a Poisson PP as in (M2). Suppose $\mathcal{D}_{X_O}$ and $\mathcal{D}_{X_V}$ have prior intensities $\alpha(x)\lambda_{\mathcal{D}_X}$ and $(1-\alpha(x))\lambda_{\mathcal{D}_X}$, respectively.*
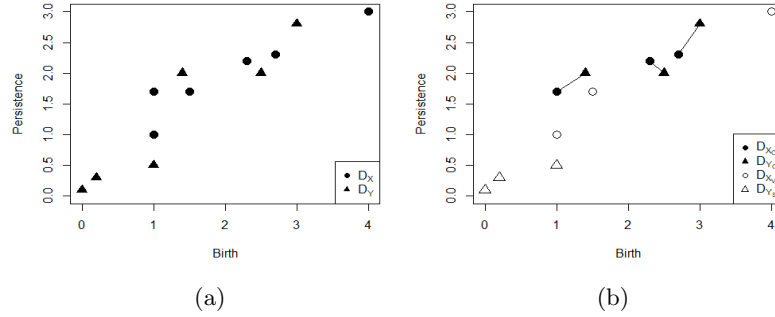
Figure 4. (a) *A sample from the prior PP $\mathcal{D}_X$ and an observed PD $\mathcal{D}_Y$. (b) Decomposition of $\mathcal{D}_X$ into $\mathcal{D}_{X_O}, \mathcal{D}_{X_V}$ and $\mathcal{D}_Y$ into $\mathcal{D}_{Y_O}, \mathcal{D}_{Y_S}$. The points in $\mathcal{D}_{X_V}$ have no relationship to those in $\mathcal{D}_Y$, while those in $\mathcal{D}_{X_O}$ only generate observed points in $\mathcal{D}_{Y_O}$. The remaining observed points in $\mathcal{D}_{Y_S}$ model unanticipated features that one may obtain due to uncertainty/noise.*

Consider $D_{Y^1}, \ldots, D_{Y^m}$ independent samples from the PP that characterizes the PD $\mathcal{D}_Y$ of (M3), and denote $D_{Y^{1:m}} := \cup_{i=1}^m D_{Y^i}$, where $D_{Y^i} = D_{Y_O^i} \cup D_{Y_S^i}$ for all $i = 1, \ldots, m$. Moreover, $\ell(y|x)$ is the likelihood associated for the stochastic kernel between $\mathcal{D}_{X_O}$ and $\mathcal{D}_{Y_O}$, and $\lambda_{\mathcal{D}_{Y_S}}$ is the intensity of $\mathcal{D}_{Y_S}$ as defined in (M3). Then, the posterior intensity of $\mathcal{D}_X$ given $D_{Y^{1:m}}$ is

$$(3.1) \quad \lambda_{\mathcal{D}_X|D_{Y^{1:m}}}(x) = (1 - \alpha(x))\, \lambda_{\mathcal{D}_X}(x)$$

$$+ \frac{1}{m}\alpha(x) \sum_{i=1}^m \sum_{y \in D_{Y^i}} \frac{\ell(y|x)\lambda_{\mathcal{D}_X}(x)}{\lambda_{\mathcal{D}_{Y_S}}(y) + \int_{\mathbb{W}} \ell(y|u)\alpha(u)\lambda_{\mathcal{D}_X}(u)du} \quad a.s.$$

The proof of Theorem 3.1 can be found in Appendix A. One important point about the above theorem is that, instead of relying on a likelihood function for the point cloud data, our Bayesian model considers the likelihood for the PD generated by the observed point cloud data at hand. This is analogous to the idea of substitution likelihood by Jeffreys in [27].

3.2. A conjugate family of prior intensities: Gaussian mixtures. This section focuses on constructing a family of conjugate prior intensities, i.e., a collection of priors that yield posterior intensities of the same form when used in (3.1). Exploiting Theorem 3.1 with Gaussian mixture prior intensities, we obtain Gaussian mixture posterior intensities. As PDs are stochastic PPs on the space $\mathbb{W}$, not $\mathbb{R}^2$, we consider a restricted Gaussian density restricted to $\mathbb{W}$. Namely, for a Gaussian density on $\mathbb{R}^2$, $\mathcal{N}(z; v, \sigma I)$, with mean v and covariance matrix σI, we restrict the Gaussian density on $\mathbb{W}$ as

$$(3.2) \qquad\qquad \mathcal{N}^*(z; v, \sigma I) := \mathcal{N}(z; v, \sigma I)\mathbb{1}_{\mathbb{W}}(z),$$

where $\mathbb{1}_{\mathbb{W}}$ is the indicator function of the wedge $\mathbb{W}$.

Consider a random PD $\mathcal{D}_X$ as in (M2) and a collection of observed PDs $\{D_{Y^1}, \ldots, D_{Y^m}\}$ that are independent samples from the Poisson PP characterizing the PD $\mathcal{D}_Y$ in (M3). We

denote $D_{Y^{1:m}} := \cup_{i=1}^{m} D_{Y^i}$. Below we specialize (M2) and (M3) so that applying Theorem 3.1 to a mixed Gaussian prior intensity yields a mixed Gaussian posterior:

(M2') $\mathcal{D}_X = \mathcal{D}_{X_O} \cup \mathcal{D}_{X_V}$, where $\mathcal{D}_{X_O}$ and $\mathcal{D}_{X_V}$ are independent Poisson PPs with intensities $\alpha \lambda_{\mathcal{D}_X}(x)$ and $(1 - \alpha) \lambda_{\mathcal{D}_X}(x)$, respectively, with

$$(3.3) \qquad \lambda_{\mathcal{D}_X}(x) = \sum_{j=1}^{N} c_j^{\mathcal{D}_X} \mathcal{N}^*(x; \mu_j^{\mathcal{D}_X}, \sigma_j^{\mathcal{D}_X} I),$$

where N is the number of mixture components.

(M3') $\mathcal{D}_Y = \mathcal{D}_{Y_O} \cup \mathcal{D}_{Y_S}$, where

    (i) the marked Poisson PP $(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})$ has density $\ell(y|x)$ given by

$$(3.4) \qquad \ell(y|x) = \mathcal{N}^*(y; x, \sigma^{\mathcal{D}_{Y_O}} I).$$

    (ii) $\mathcal{D}_{Y_O}$ and $\mathcal{D}_{Y_S}$ are independent finite Poisson PPs, and $\mathcal{D}_{Y_S}$ has the intensity function given below:

$$(3.5) \qquad \lambda_{\mathcal{D}_{Y_S}}(y) = \sum_{k=1}^{M} c_k^{\mathcal{D}_{Y_S}} \mathcal{N}^*(y; \mu_k^{\mathcal{D}_{Y_S}}, \sigma_k^{\mathcal{D}_{Y_S}} I),$$

where M is the number of mixture components.

Proposition 3.2. *Suppose that assumptions* (M1), (M2'), *and* (M3') *hold; then, the posterior intensity of* (3.1) *in Theorem* 3.1 *is a Gaussian mixture of the form*

$$(3.6) \qquad \lambda_{\mathcal{D}_X | D_{Y^{1:m}}}(x) = (1 - \alpha) \lambda_{\mathcal{D}_X}(x) + \frac{\alpha}{m} \sum_{i=1}^{m} \sum_{y \in \mathcal{D}_{Y^i}} \sum_{j=1}^{N} C_j^y \mathcal{N}^*(x; \mu_j^y, \sigma_j^y I),$$

$$where \quad C_j^y = \frac{w_j^y}{\lambda_{\mathcal{D}_{Y_S}}(y) + \alpha \sum_{j=1}^{N} w_j^y Q_j^y}; \; Q_j^y = \int_{\mathbb{W}} \mathcal{N}(u; \mu_j^y, \sigma_j^y I) du;$$

$$w_j^y = c_j^{\mathcal{D}_X} \mathcal{N}(y; \mu_j^{\mathcal{D}_X}, (\sigma^{\mathcal{D}_{Y_O}} + \sigma_j^{\mathcal{D}_X}) I);$$

$$and \quad \mu_j^y = \frac{\sigma_j^{\mathcal{D}_X} y + \sigma^{\mathcal{D}_{Y_O}} \mu_j^{\mathcal{D}_X}}{\sigma_j^{\mathcal{D}_X} + \sigma^{\mathcal{D}_{Y_O}}}; \sigma_j^y = \frac{\sigma^{\mathcal{D}_{Y_O}} \sigma_j^{\mathcal{D}_X}}{\sigma_j^{\mathcal{D}_X} + \sigma^{\mathcal{D}_{Y_O}}}.$$

The proof of Proposition 3.2 follows from well-known results about products of Gaussian densities given below; for more details, the reader may refer to [38] and references therein.

Lemma 3.3. *For $p \times p$ matrices H, R, P, with R and P positive definite, and a $p \times 1$ vector s, $\mathcal{N}(y; Hx, R) \mathcal{N}(x; s, P) = q(y) \mathcal{N}(x; \hat{s}, \hat{P})$, where $q(y) = \mathcal{N}(y; Hs, R + HPH^T)$, $\hat{s} = s + K(y - Hs)$, $\hat{P} = (I - KH)P$, and $K = PH^T(HPH^T + R)^{-1}$.*

Proof of Proposition 3.2. Using Lemma 3.3, we first derive $\ell(y|x) \lambda_{\mathcal{D}_X}(x)$ by seeing that, in our model, $H = I$, $R = \sigma^{\mathcal{D}_{Y_O}} I$, $s = \mu_j^{\mathcal{D}_X}$, and $P = \sigma_j^{\mathcal{D}_X} I$. By typical matrix operations

we obtain $K = \frac{\sigma_j^{\mathcal{D}_X}}{\sigma_j^{\mathcal{D}_X}+\sigma^{\mathcal{D}_{Y_O}}}$, $\hat{s} = \frac{\sigma_j^{\mathcal{D}_X} y+\sigma^{\mathcal{D}_{Y_O}}\mu_j^{\mathcal{D}_X}}{\sigma_j^X+\sigma^{\mathcal{D}_{Y_O}}}$, and $\hat{P} = \frac{\sigma^{\mathcal{D}_{Y_O}}\sigma_j^{\mathcal{D}_X}}{\sigma_j^{\mathcal{D}_X}+\sigma^{\mathcal{D}_{Y_O}}}$. The numerator and denominator of the second term in (3.1), $\sum_{j=1}^N c_j^{\mathcal{D}_X}\mathcal{N}(y;\mu_j^{\mathcal{D}_X},(\sigma^{\mathcal{D}_{Y_O}}+\sigma_j^{\mathcal{D}_X})I)\mathcal{N}^*(x;\mu_j^y,\sigma_j^y I)$, and $\lambda_{\mathcal{D}_{Y_S}}(y)+\alpha\sum_{j=1}^N c_j^{\mathcal{D}_X}\mathcal{N}(y;\mu_j^{\mathcal{D}_X},(\sigma^{\mathcal{D}_{Y_O}}+\sigma_j^{\mathcal{D}_X})I)\int_{\mathbb{W}}\mathcal{N}(u;mu_j^y,\sigma_j^y I)du$, respectively, yield

$$\sum_{j=1}^N \left[\frac{c_j^{\mathcal{D}_X}\mathcal{N}(y;\mu_j^{\mathcal{D}_X},(\sigma^{\mathcal{D}_{Y_O}}+\sigma_j^{\mathcal{D}_X})I)}{\lambda_{\mathcal{D}_{Y_S}}(y)+\alpha\sum_{j=1}^N c_j^{\mathcal{D}_X}\mathcal{N}(y;\mu_j^{\mathcal{D}_X},(\sigma^{\mathcal{D}_{Y_O}}+\sigma_j^{\mathcal{D}_X})I)\int_{\mathbb{W}}\mathcal{N}(u;\mu_j^{x|y},\sigma_j^{x|y}I)du}\right]\mathcal{N}^*(x;\mu_j^{x|y},\sigma_j^{x|y}I),$$

where the bracketed expression is the definition of C_j^y. ∎

3.2.1. Example. Here, we present a detailed example of computing the posterior intensity according to (3.6) for a range of parametric choices. Reproducing these results, the interested reader may download our R-package BayesTDA. We consider circular point clouds often associated with periodicity in signals [35] and focus on estimating homological features with $k = 1$ as they correspond to 1-dimensional holes, which describe the prominent topological feature of a circle. Precisely, our goals are to (i) illustrate posterior intensities and draw analogies to standard Bayesian inference; (ii) determine the relative contributions of the prior and observed data to the posterior; and (iii) perform sensitivity analysis. *Case*-I is ideal, while *Case*-II, *Case*-III, and *Case*-IV are purposefully disadvantageous.

Table 2

List of Gaussian mixture parameters of the prior intensities in (3.3). The means $\mu_i^{\mathcal{D}_X}$ are 2×1 vectors and the rest are scalars.

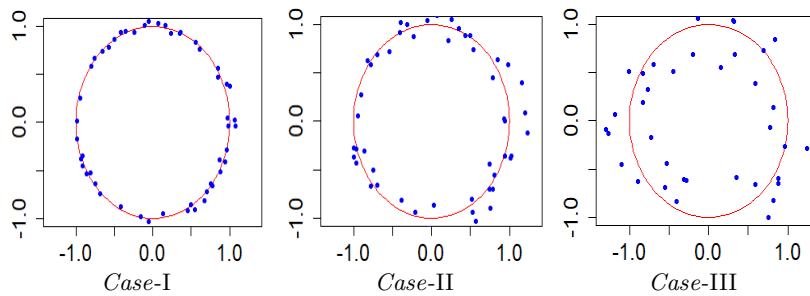|	$\mu_i^{\mathcal{D}_X}$	$\sigma_i^{\mathcal{D}_X}$	$c_i^{\mathcal{D}_X}$
informative prior	$(0.5, 1.2)$	0.01	1
weakly informative prior	$(0.5, 1.2)$	0.2	1
unimodal uninformative prior	$(1, 1)$	1	1
bimodal uninformative prior	$(0.5, 0.5)$	0.2	1
	$(1.5, 1.5)$	0.2	2

Case-I *Case*-II *Case*-III

Figure 5. *The observed datasets generated for Case-I, Case-II, and Case-III by sampling the unit circle and perturbing with Gaussian noise having variances $0.001I_2$, $0.01I_2$, and $0.1I_2$, respectively.*

We start by considering a Poisson PP with prior intensity $\lambda_{\mathcal{D}_X}$ that has the Gaussian mixture form given in (M2′). We take into account four types of prior intensities: (i) informative, (ii) weakly informative, (iii) unimodal uninformative, and (iv) bimodal uninformative; see Figures 6–8(a), (d), (g), (j), respectively. We use one Gaussian component in each of the first

three priors, as the underlying shape has a single 1-dimensional feature, and two for the last one to include a case where we have no information about the cardinality of the underlying true diagram. The parameters of the Gaussian mixture density in (3.3) used to compute these prior intensities are listed in Table 2. To present the intensity maps uniformly throughout this example while preserving their shapes, we divide the intensities by their corresponding maxima. This ensures all intensities are on a scale from 0 to 1, and we call it the scaled intensity. The observed PDs are generated from point clouds sampled uniformly from the unit circle and then perturbed by varying levels of Gaussian noise; see Figure 5, wherein we present three point clouds sampled with Gaussian noise having variances $0.001I_2$, $0.01I_2$, and $0.1I_2$, respectively. Consequently, these point clouds provide PDs D_{Y^i} for $i = 1, 2, 3$, which are considered as independent samples from Poisson PP $\mathcal{D}_Y$, exhibiting distinctive characteristics such as only one prominent feature with high persistence and no spurious features (*Case*-I), one prominent feature with high persistence and very few spurious features (*Case*-II), and one prominent feature with medium persistence and more spurious features (*Case*-III).

Table 3

Parameters for (M3′) *in* (3.4) *and* (3.5). *We set the weight and mean of the Gaussian component,* $c^{\mathcal{D}_{YS}} = 1$ *and* $\mu^{\mathcal{D}_{YS}} = (0.5, 0)$, *respectively, for all of the cases. The first row corresponds to parameters in the functions characterizing* $\mathcal{D}_Y$ *that are used in computing the posterior depicted in the first column of Figure* 9. *The second row corresponds to analogous parameters that are used in computing the posterior depicted in the second columns of Figures* 6–9. *Similarly, the third row corresponds to parameters in the functions characterizing* $\mathcal{D}_Y$ *used for computing the posterior presented in the third columns of Figures* 6–9.

Case-I	*Case*-II	*Case*-III	*Case*-IV
			$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 0.1$
$\sigma^{\mathcal{D}_{YO}} = 0.01$ $\sigma^{\mathcal{D}_{YS}} = 0.1$	$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 0.1$	$\sigma^{\mathcal{D}_{YO}} = 0.01$ $\sigma^{\mathcal{D}_{YS}} = 0.1$	$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 1$
$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 0.1$	$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 1$	$\sigma^{\mathcal{D}_{YO}} = 0.1$ $\sigma^{\mathcal{D}_{YS}} = 0.1$	$\sigma^{\mathcal{D}_{YO}} = 0.01$ $\sigma^{\mathcal{D}_{YS}} = 0.1$

For each observed PD, 1-dimensional persistence features are presented as green dots overlaid on their corresponding posterior intensity plots. For *Case*-I, *Case*-II, and *Case*-III, we set the probability α of the event that a feature in $\mathcal{D}_X$ appears in $\mathcal{D}_Y$ to 1, i.e., any feature in $\mathcal{D}_X$ is certainly observed through a mark in $\mathcal{D}_Y$, and later in *Case*-IV, we decrease α to 0.5 while keeping all other parameters the same for the sake of comparison. The choice of $\alpha = 0.5$ anticipates that any feature has equal probability to appear or disappear in the observation and in turn provides further intuition about the contribution of prior intensities to the estimated posteriors. We observe that in all cases, the posterior estimates the 1-dimensional hole but with different uncertainty each time. For example, for the cases where the data are trustworthy, expressed by a likelihood with tight variance, or in the case of an informative prior, the posterior accurately estimates the 1-dimensional hole. In contrast, when the data suffer from high uncertainty and the prior is uninformative, then the posterior offers a general idea that the true underlying shape is a circle, but the exact estimation of the 1-dimensional hole is not accurate. We examine the cases below.

Case-I. We consider informative, weakly informative, unimodal uninformative, and bimodal uninformative prior intensities as presented in Figure 6(a), (d), (g), and (j), respec-

tively, to compute corresponding posterior intensities. The prior intensities parameters are listed in Table 2. The observed PD is obtained from the point cloud in Figure 5 (left). The parameters associated to the observed PD are listed in Table 3. For the observed PD arising from data with very low noise, we observe that the posterior computed from any of the priors predicts the existence of a 1-dimensional hole accurately. First, with a low variability in observed PD ($\sigma^{\mathcal{D}_{Y_O}} = 0.01$ and $\sigma^{\mathcal{D}_{Y_S}} = 0.1$), the posterior intensities estimate the hole with high certainty (Figure 6(b), (e), (h), and (k), respectively). Next, to determine the effect of observed data on the posterior, we increase the variance of the observed PD component $\mathcal{D}_{Y_O}$, which consists of features in observed PDs that are associated to the underlying prior. Here, we observe that the posterior intensities still estimate the hole accurately due to the trustworthy data; this is evident in Figure 6(c), (f), (i), and (l). In Figure 6, the posteriors in (b), (e), (h), and (k) have lower variance around the 1-dimensional feature in comparison to (c), (f), (i), and (l), respectively.

Case-II. Here, we consider all four priors as in *Case*-I (see Figure 7(a), (d), (g), and (j)). The point cloud in Figure 5 (center) is more perturbed around the unit circle than that of *Case*-I (Gaussian noise with variance $0.01I_2$). Due to this, the associated PD exhibits spurious features. The parameters used for this case are listed in Table 3. We compute the posterior intensities for each type of prior. First, to illustrate the posterior intensity and check the capability of detecting the 1-dimensional feature, we use moderate noise for the observed PD ($\sigma^{\mathcal{D}_{Y_O}} = 0.1$ and $\sigma^{\mathcal{D}_{Y_S}} = 0.1$). The results are presented in Figure 7(b), (e), (h), and (k); overall, the posteriors estimate the prominent feature with different variances in their respective posteriors. Next, to illustrate the effect of observed data on the posterior, we increase the variance $\sigma^{\mathcal{D}_{Y_S}}$ of $\mathcal{D}_{Y_S}$. According to our Bayesian model, the PD component $\mathcal{D}_{Y_S}$ contains features that are not associated with $\mathcal{D}_X$, so increasing $\sigma^{\mathcal{D}_{Y_S}}$ yields that every observed point is linked to $\mathcal{D}_X$, and therefore one may expect to observe increased intensity skewed towards the spurious points that arise from noise. Indeed, posterior intensities with weakly informative, unimodal uninformative, and bimodal uninformative priors exhibit the skewness toward the spurious point in Figure 7(f), (i), and (l), respectively, but this is not the case when an informative prior is used. In (f), we observe increased intensity skewing towards the spurious points, and in (i) and (l) the intensity appears to be bimodal with two modes—one at the prominent point and the other at the spurious point. For the bimodal uninformative prior, since one mode is located close to the spurious point in the observed PD, we observe higher intensity for that mode in the posterior (Figure 7(l)) with another mode estimating the prominent feature.

Case-III. We again consider the four types of priors here. The observed PD is constructed from the point cloud in Figure 5 (right). The point cloud has Gaussian noise with variance $0.1I_2$, and due to the high noise level in sampling relative to the unit circle, the associated PD exhibits one prominent feature and several spurious features. We repeat the parameter choices as in *Case*-I for the variances of the observed PD. For the choice of $\sigma^{\mathcal{D}_{Y_O}} = 0.01$ and $\sigma^{\mathcal{D}_{Y_S}} = 0.1$, the posteriors computed from all of the four priors are able to detect the difference between the one prominent point and other spurious points. We increase the variance of $\mathcal{D}_{Y_O}$ to determine the effect of the observed PD on the posterior, and we observe that only the posterior intensity from the informative prior has evidence of the hole (Figure 8(c)). For the weakly informative and uninformative priors, while the posteriors in (f), (i), and (l) may not
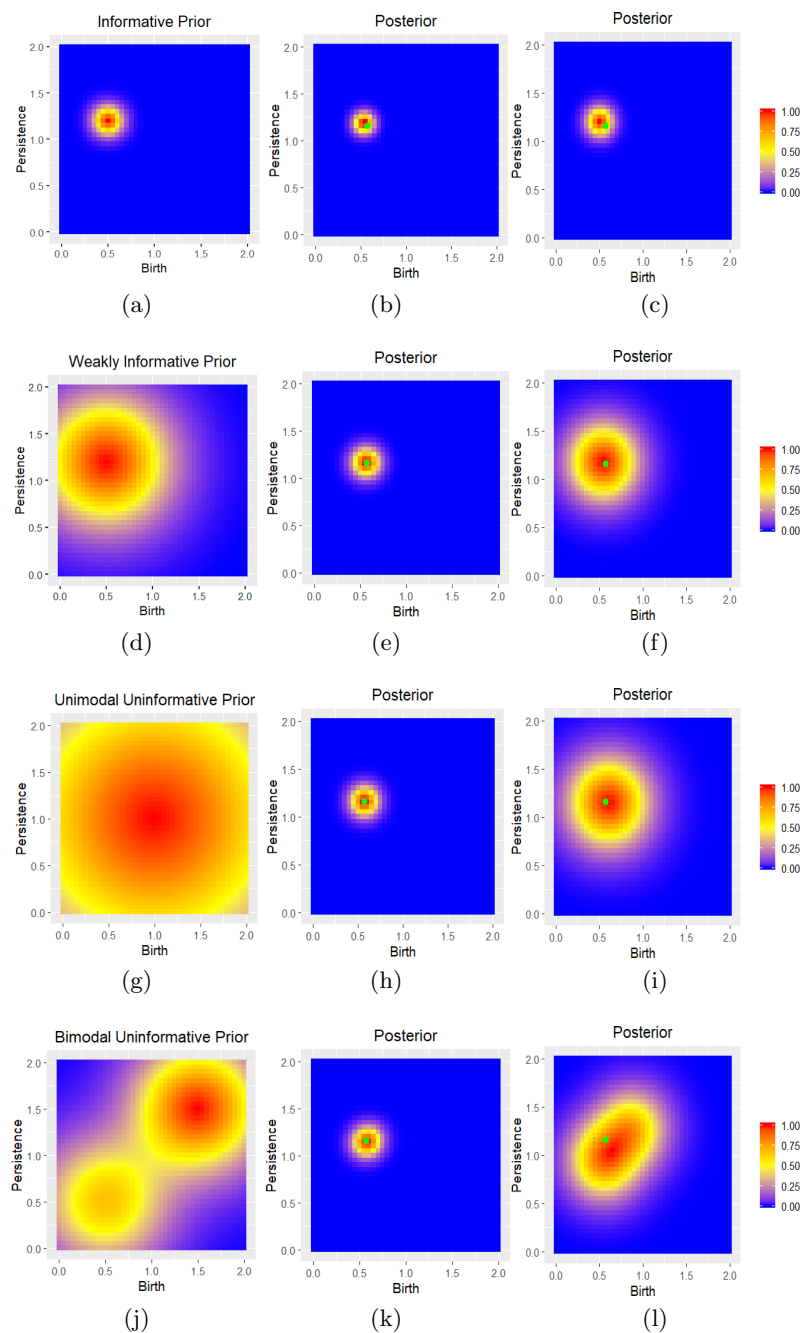
Figure 6. *Case-I: Posterior intensities obtained by using Proposition* 3.2. *We consider informative* (a), *weakly informative* (d), *unimodal uninformative* (g), *and bimodal uninformative* (j) *prior intensities. The color maps represent scaled intensities. The list of associated parameters of the observed PD used for this case is in Table* 3. *Posteriors computed from all of these priors estimate the 1-dimensional hole accurately for a choice of variances in the observed PD as* $\sigma^{\mathcal{D}_{Y_O}} = 0.01$ *and* $\sigma^{\mathcal{D}_{Y_S}} = 0.1$ *which are presented in* (b), (e), (h), *and* (k). *After increasing the variance to* $\sigma^{\mathcal{D}_{Y_O}} = 0.1$, *we observe that the posteriors can still estimate the hole with higher variance as presented in* (c), (f), (i), *and* (k).
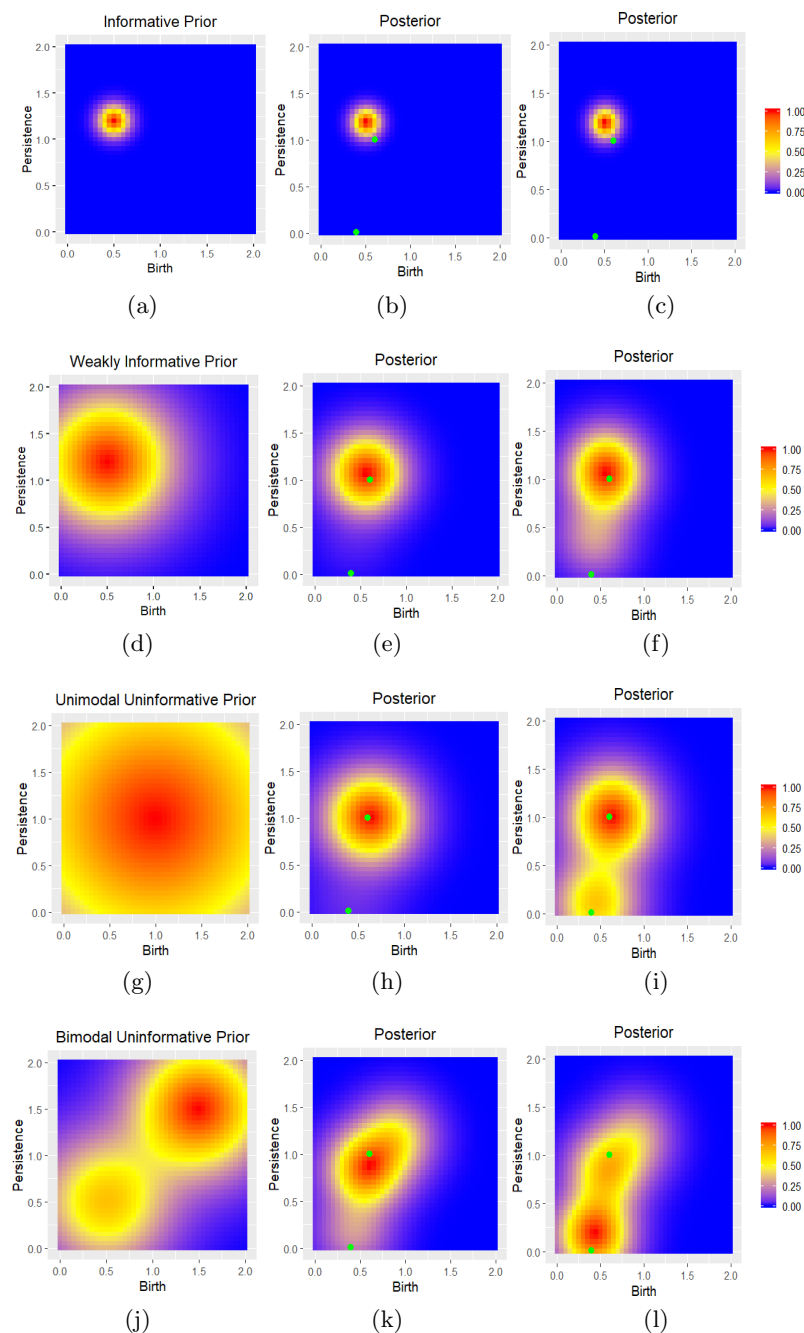
Figure 7. *Case*-II: *We consider informative* (a), *weakly informative* (d), *unimodal uninformative* (g), *and bimodal uninformative* (j) *prior intensities to estimate posterior intensities using Proposition 3.2. The color maps represent scaled intensities. The parameters we use for estimating the posterior intensity are listed in Table 3. The posterior intensities estimated from the informative prior in* (b) *and* (c) *estimate the 1-dimensional hole with high certainty. Also, the posterior intensities estimated from the weakly informative and uninformative priors in* (e), (h), *and* (k) *imply the existence of a hole with lower certainty.* (c), (f), (i), *and* (l) *represent the posterior with higher variance in observed PD component* $\mathcal{D}_{Y_S}$. *As this makes the assumption that every observed point is associated to* $\mathcal{D}_X$, *we observe increased intensity skewed towards the spurious point in* (f). *Furthermore, in* (i) *and* (l), *we observe bimodality in the posterior intensity.*
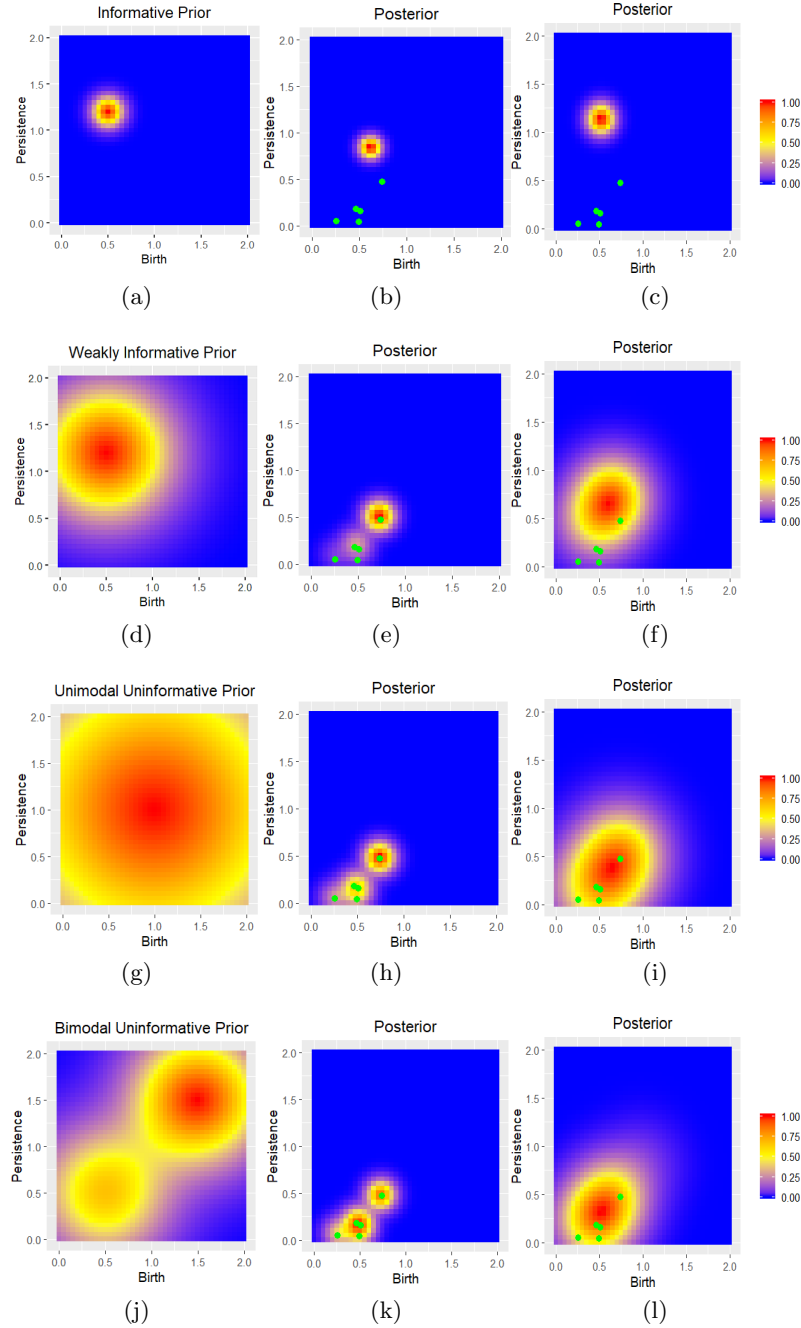
Figure 8. *Case-III: Posterior intensities obtained by using Proposition* 3.2. *We consider informative* (a), *weakly informative* (d), *unimodal uninformative* (g), *and bimodal uninformative* (j) *prior intensities. The color maps represent scaled intensities. Parameters of the observed PD used to estimate the posterior intensity are listed in Table* 3. *With a choice of* $\sigma^{\mathcal{D}_{Y_O}} = 0.01$ *and* $\sigma^{\mathcal{D}_{Y_s}} = 0.1$, *we observe that the posteriors can deduce the existence of the prominent feature as presented in* (b), (e), (h), *and* (k), *as we have more confidence in the component of observed data associated to the prior. Otherwise, with an increased variance* $\sigma^{\mathcal{D}_{Y_O}} = 0.1$, *only the posterior intensity from the informative prior is able to detect the hole with high certainty, as observed in* (c). *For the weakly informative and uninformative priors, the posteriors in* (f), (i), *and* (l) *may not detect the hole directly, but the mode of* (f) *with higher variance and the tail towards the prominent point in* (i) *and* (l) *imply the existence of a hole in the underlying PD.*
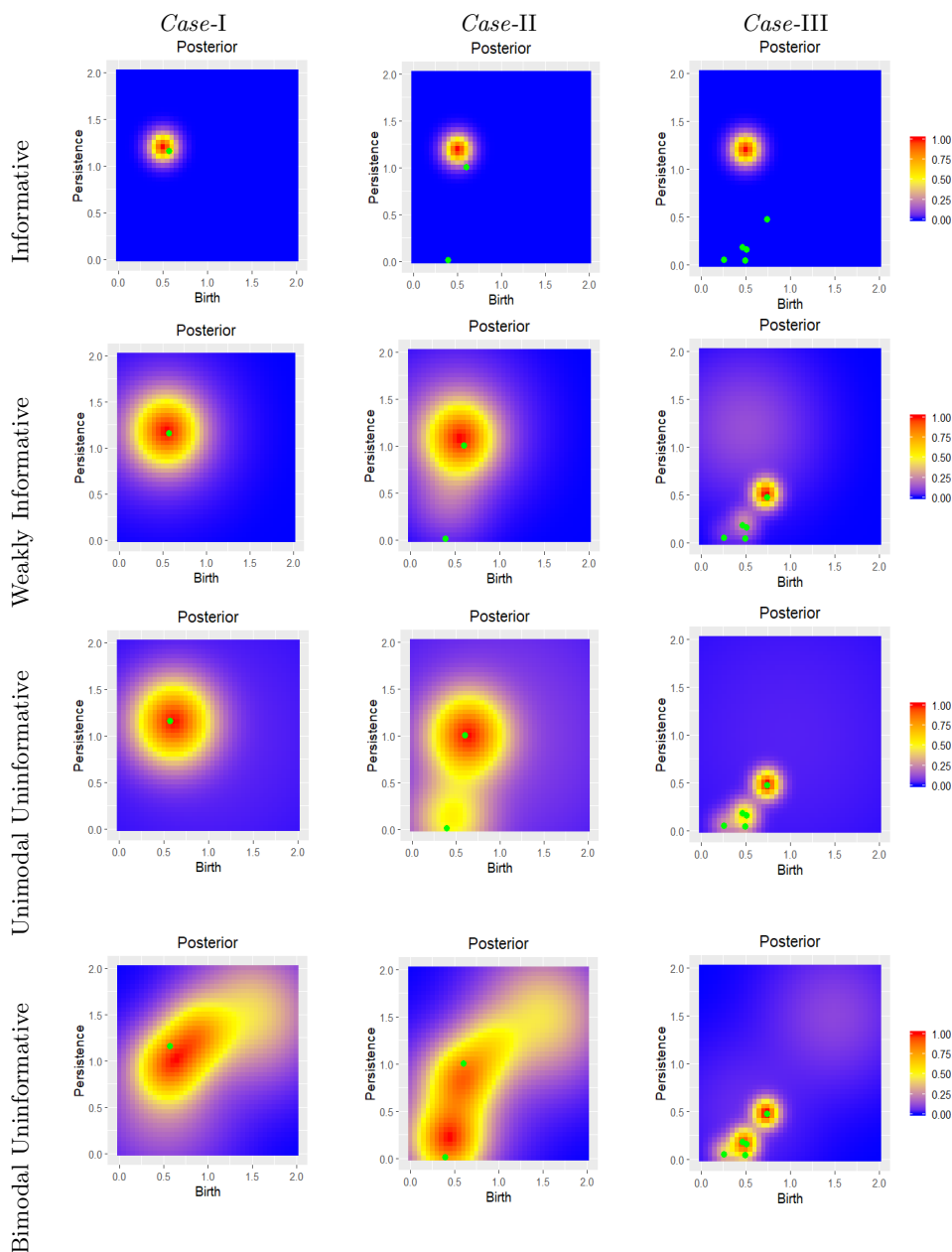
Figure 9. *Case*-IV: *The first, second, and third columns match the parameters of observed PD $\mathcal{D}_Y$ used in computing the posteriors depicted in the third column of Figures* 6 *and* 7, *and the second column of Figure* 8, *respectively, with* $\alpha = 0.5$. *The parameters are presented in Table* 3. *The color maps represent scaled intensities. A variation in the level of intensity is observed for all of them compared to their respective cases due to the added term in the posterior intensity. The posterior intensities in the first and second columns exhibit the estimation of the hole with higher variability as compared to the respective figures in Case*-I *and Case*-II. *The posteriors in the third column demonstrate dominance of the prior relative to their corresponding figures in Case*-III, *especially when one examines those for informative, weakly informative, and bimodal uninformative priors.*

detect the hole clearly, in (f) we observe a mode with higher variance, and in (i) and (l) a tail towards the high persistence point, implying the presence of a hole. It should be noted that with the informative prior the posterior intensity identifies the hole closer to the mode of the prior as we increase the variance in $\sigma^{\mathcal{D}_{Y_O}}$.

Case-IV: Last, in this case we concentrate on the effect of α. The rest of the parameters used for this case remain the same and are listed in Table 3. We decrease α to 0.5 to model the scenario that a feature in $\mathcal{D}_X$ has equal probability to appear or vanish in observed $\mathcal{D}_Y$. The columns of Figure 9 correspond to the parameters of the observed PD $\mathcal{D}_Y$ used in computing the posteriors depicted in the third column of Figure 6, third column of Figure 7, and second column of Figure 8, respectively. By comparing them with their respective cases, we notice a change in the intensity level in all of these due to the first term of the posterior intensity on the right-hand side of (3.6). Comparing with the respective figures in *Case*-I, we observe that the posterior intensities are estimating the hole with higher variability for the weakly informative and unimodal uninformative priors. For the bimodal prior, we observe bimodality in the posterior. Next, for *Case*-II, the existence of a hole is evident for informative and weakly informative priors with higher uncertainty when compared to their previous cases. The unimodal and bimodal uninformative priors lead to bimodal and trimodal posteriors, respectively. We observe that the posterior resembles the prior intensity more closely when we compare them to the respective figures in *Case*-III. One can especially see this with the informative, weakly informative, and bimodal uninformative priors, which have significantly increased intensities at the location of the modes of the prior.

4. Classification. The Bayesian framework introduced in this paper allows us to explicitly compute the posterior intensity of a PD given data and prior knowledge. This lays the foundation for supervised statistical learning methods in classification. In this section, we build a Bayes factor classification algorithm based on notions discussed in section 3 and then apply it on materials data, in particular on measurements for spatial configurations of atoms.

We commence our classification scheme with a PD D belonging to an unknown class. We assume that D is sampled from a Poisson PP $\mathcal{D}$ in $\mathbb{W}$ with the prior intensity $\lambda_{\mathcal{D}}$ having the form in (M2′). Consequently, its probability density has the form

$$(4.1) \qquad p_{\mathcal{D}}(D) = \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \lambda_{\mathcal{D}}(d) = \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \sum_{i=1}^{N} c_i^{\mathcal{D}} \mathcal{N}^*(d; \mu_i^{\mathcal{D}}, \sigma_i^{\mathcal{D}} I),$$

where $\lambda = \int_{\mathbb{W}} \lambda_{\mathcal{D}} = \mathbb{E}(|\mathcal{D}|)$, with probability α as in (M2′). Next, suppose we have two training sets $T_Y := D_{Y_{1:n}}$ and $T_{Y'} := D_{Y'_{1:m}}$ from two classes of random diagrams $\mathcal{D}_Y$ and $\mathcal{D}_{Y'}$, respectively. The likelihood densities of the respective classes take the form of (3.4). We then follow (3.6) to obtain the posterior intensities of $\mathcal{D}$ given the training sets T_Y and $T_{Y'}$ from the prior intensities and likelihood densities. In particular, the corresponding posterior

probability density of $\mathcal{D}$ given the training set T_Y is

$$(4.2) \quad p_{\mathcal{D}|\mathcal{D}_Y}(D|T_Y) = \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \lambda_{D|T_Y}(d)$$

$$= \frac{e^{-\lambda}}{|D|!} \prod_{d \in D} \left[(1-\alpha)\lambda_{\mathcal{D}}(d) + \frac{\alpha}{n} \sum_{y_j \in T_Y} \sum_{i=1}^{N} C_i^{d|y_j} \mathcal{N}(d; \mu_i^{d|y_j}, \sigma_i^{d|y_j} I) \right],$$

and the posterior probability density given T_Y' is given by an analogous expression. The Bayes factor defined by

$$(4.3) \qquad\qquad BF(D) = \frac{p_{D|\mathcal{D}_Y}(D|T_Y)}{p_{D|\mathcal{D}_{Y'}}(D|T_{Y'})}$$

provides the decision criterion for assigning D to either $\mathcal{D}_Y$ or $\mathcal{D}_{Y'}$. More specifically, for a threshold c, $BF(D) > c$ implies that D belongs to $\mathcal{D}_Y$ and $BF(D) < c$ implies otherwise. We summarize this scheme in Algorithm 4.1.

Algorithm 4.1. Bayes factor classification of PDs.

1: **Input 1**: Prior intensities $\lambda_{\mathcal{D}_Y}$ and $\lambda_{\mathcal{D}_{Y'}}$ for two classes of diagrams $\mathcal{D}_Y$ and $\mathcal{D}_{Y'}$, respectively; a threshold $c > 0$.
2: **Input 2**: Two training sets T_Y and $T_{Y'}$ sampled from $\mathcal{D}_Y$ and $\mathcal{D}_{Y'}$, respectively.
3: **for** D_Y and, $D_{Y'}$ **do**
4:     Compute $p_{\mathcal{D}|\mathcal{D}_Y}(D|T_Y)$ and $p_{\mathcal{D}|\mathcal{D}_{Y'}}(D|T_{Y'})$.
5: **end for**
6: Compute $BF(D)$ as in (4.3)
7: **if** $BF(D) > c$ **then**
8:     assign D to $\mathcal{D}_Y$.
9: **else**
10:     assign D to $\mathcal{D}_{Y'}$.
11: **end if**

4.1. Atom probe tomography data. Our goal in this section is to use Algorithm 4.1 to classify the crystal lattice of a noisy and sparse materials dataset, where the unit cells are either body-centered cubic (BCC) or face-centered cubic (FCC); recall Figure 1. The BCC structure has a single atom in the center of the cube, while the FCC has a void in its center but has atoms on the centers of the cubes' faces (Figure 1(b)–(c)). Despite notable differences in the physical configurations of each class, sparsity and noise do not allow the crystal structure to be revealed. For high-entropy alloys, our object of interest, atom probe tomography (APT), provides the best atomic level characterization possible. Due to the sparsity and noise in the resulting data, there are only a few algorithms for successfully determining the crystal structure; see [22, 43]. These algorithms, designed for APT data, rely on knowing the global structure a priori (which is not the case for high entropy alloys (HEAs)) and seek to discover small-scale structure within a sample.

To bypass this restriction, the neural network architecture of [62] provides a way to classify the crystal structure of a noisy or sparse dataset by looking at a diffraction image. In particular, the authors therein employ a convolutional neural network for classifying the crystal structure by examining a computer-generated diffraction pattern. The authors suggest their method could be used to determine the crystal structure of APT data. However, the synthetic data considered in [62] is not a realistic representation of experimental APT data, where about 65% of the data is missing and furthermore corrupted by observational noise. Most importantly, their synthetic data is either sparse or noisy, not a combination of both. The algorithm is also not publicly available, so a side-by-side comparison of our method with theirs using HEAs is not feasible.

It is natural to consider PDs in this setting because they distill salient information about the materials patterns with respect to connectedness and empty space (holes) within cubic unit cells; i.e, we can differentiate between atomic unit cells by examining their homological features. In particular, after storing both types of spatial configurations as point clouds, we compute their Rips filtrations (see section 2), collecting resultant 1-dimensional homological features into PDs; see Figure 10. The dataset had 200 diagrams from each class. To perform classification with Algorithm 4.1, we started by specifying priors for each class, $\lambda_{\mathcal{D}_{BCC}}$ and $\lambda_{\mathcal{D}_{FCC}}$. Two scenarios were considered, namely using separate priors (Prior-1 in Table 4) and the same prior (Prior-2 in Table 4) for both the BCC and the FCC classes. In particular, for Prior-1 we superimpose 50 PDs from each class and find the highly clustered areas by using K-means clustering. The centers of the clusters from K-means are then used as the means in Gaussian mixture priors; see (3.6). In this manner, we produce different priors for BCC and FCC classes. On the other hand, for Prior-2 we choose a flat prior with a higher variance level than that of Prior-1 for both of the classes. The parameters for these two prior intensities are in Table 4. For all cases, we set $\sigma^{\mathcal{D}_{Y_O}} = 0.1$ and $\lambda_{\mathcal{D}_{Y_S}}(x) = 5\mathcal{N}^*(x; (0,0), 0.2I)$. We chose a relatively high weight for $\lambda_{\mathcal{D}_{Y_S}}$ because the nature of the data implied that extremely low persistence holes were rare events arising from noise. To perform 10-fold cross-validation, we partitioned PDs from both classes into training and test sets. During each fold, we took the training sets from each class, T_{BCC} and T_{FCC}, and input them into Algorithm 4.1 as T_Y and $T_{Y'}$, respectively. Next, we computed the Bayes factor $BF(D) = \frac{p_{\mathcal{D}_{BCC}}(D|T_{BCC})}{p_{\mathcal{D}_{FCC}}(D|T_{FCC})}$ for each diagram D in the test sets. After this, we used the Bayes factors to construct receiver operating characteristic (ROC) curves and computed the resulting areas under the ROC curves (AUCs). Finally, we used the AUCs from 10-fold cross-validation to build a bootstrapped distribution by resampling 2000 times. Information about these bootstrapped distributions is summarized in Table 4, which shows that our scoring method almost perfectly distinguishes between the BCC and FCC classes using the Bayesian framework of section 3. Also, it exemplifies the robustness of our algorithm, as two different types of priors produce near perfect accuracy.

5. Discussion. This work is the first approach to introduce a generalized Bayesian framework for persistent homology. This toolbox gives experts the opportunity to incorporate their prior belief about data in conjunction with topological data analysis notions when faced with research questions. Our framework is entirely predicated upon modeling random PDs with Poisson PPs because of their nice qualities that behave well with Bayesian formulations. Specifically, since they are characterized entirely by their intensity measures, they allow us to
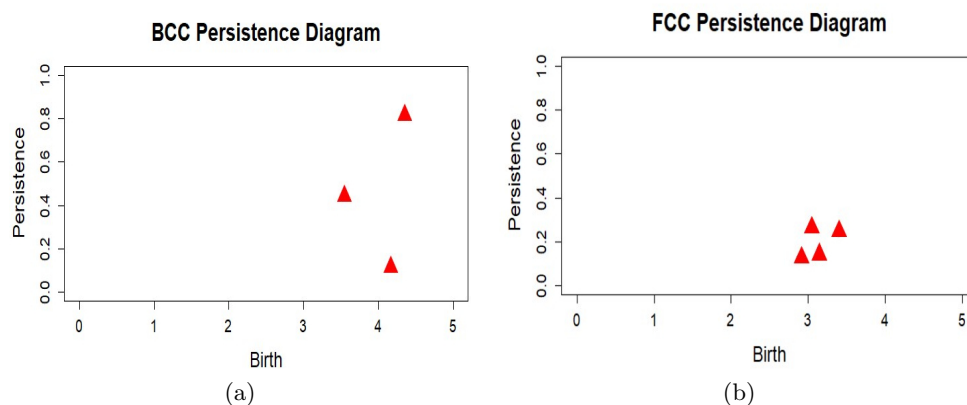
Figure 10. *PDs for members of the BCC and FCC classes.*

Table 4

Parameters for the prior intensities used in cross-validation of materials science data. Each prior λ is indexed by its corresponding class for Prior-1 or U in the case of the Prior-2. The summary of AUCs across 10-folds for materials science data after scoring with Algorithm 4.1 is presented in the last three columns.

Priors		Parameters for prior intensities			Summary of AUC		
		$\mu_i^{\mathcal{D}}$	$\sigma_i^{\mathcal{D}}$	$c_i^{\mathcal{D}}$	5th percentile	Mean	95th percentile
Prior-1	λ_{BCC}	(0.5,0.24) (3.6,3.6) (3.7,0.65)	2 2 2	1 1 1	0.931	0.941	0.958
	λ_{BCC}	(0.4,0.27) (2.8,1.2) (2.9,3)	2 2 2	1 1 1			
Prior-2	λ_U	(1,1)	20	1	0.928	0.94	0.951

quantify prior uncertainty with presupposed intensity functions, and allow for efficient computation of posterior intensities if we regard observed PDs as noisy observations described by marks in a marked Poisson PP. Interestingly, recent works [3, 37] could also be used to devise an alternative, parallel PP-based Bayesian framework for persistent homology. In particular, one could directly use Bayes rule with prior distributions constructed from [37] and then obtain posteriors with the methodology from [3], which outlines a procedure for Monte Carlo estimation of Choquet integrals. This is a worthwhile future direction for research of this nature. It should be noted that our Bayesian model considers PDs, which are summaries of the data at hand, for defining a substitution likelihood rather than using the underlying point cloud data. This does not adhere to a strict Bayesian viewpoint, as we model the behavior of the PDs without considering the underlying data (materials data in our example) used to create it; however, our paradigm incorporates prior knowledge and observed data summaries to create posterior probabilities, analogous to the notion of substitution likelihood detailed in [27]. The general relationship between the likelihood models related to point cloud data and those of their corresponding PDs remains an important open problem. Our paper also

introduces a conjugate-like family of prior intensities and stochastic kernels (our likelihood analogue), which can be used to obtain a closed form for posterior intensities. A detailed example is presented to illustrate the qualities of posterior intensities arising under several interesting parameter choices in our model. This example establishes evidence that our Bayesian framework updates prior uncertainty with new observations in a manner similar to that for standard RVs. Thus, the Bayesian inference developed herein can be reliably used for machine learning and data analysis techniques *directly* on the space of PDs. Indeed, a classification algorithm is derived and successfully applied on materials science data to assess the capability of our Bayesian framework.

Appendix A. Proof of Theorem 3.1.

Proof. By Theorem 2.10, we decompose $\lambda_{\mathcal{D}_X|D_{Y^{1:m}}}$ to write

$$(A.1) \qquad \lambda_{\mathcal{D}_X|D_{Y^{1:m}}} = \lambda_{\mathcal{D}_{X_V}|D_{Y^{1:m}}} + \lambda_{\mathcal{D}_{X_O}|D_{Y^{1:m}}} = (1 - \alpha(x))\lambda_{\mathcal{D}_X} + \lambda_{\mathcal{D}_{X_O}|D_{Y^{1:m}}},$$

where the second equality follows because $\mathcal{D}_{X_V}$ is independent of $\mathcal{D}_Y$. Theorem 2.10 allows us to express $\lambda_{\mathcal{D}_{X_O}}$ as the average of intensity functions $\lambda_{\mathcal{D}_{X_O^i}}$ for $i = 1, \ldots, m$, where the $\mathcal{D}_{X_O^i}$ are independent and equal in distribution to $\mathcal{D}_{X_O}$. That is, $\lambda_{\mathcal{D}_{X_O}} = \frac{1}{m}\sum_{i=1}^m \lambda_{\mathcal{D}_{X_O^i}}$, and by conditioning we have

$$(A.2) \qquad \lambda_{\mathcal{D}_{X_O}|D_{Y^{1:m}}} = \frac{1}{m}\sum_{i=1}^m \lambda_{\mathcal{D}_{X_O^i}|D_{Y^i}}.$$

So to expand (A.1) it suffices to compute $\lambda_{\mathcal{D}_{X_O^i}|D_{Y^i}}$ for fixed i. First, we express the finite PP $(\mathcal{D}_X, \mathcal{D}_Y)$ as a marked Poisson PP. To this end, we adopt a construction from [53], the augmented space $\mathbb{W}' := \mathbb{W} \cup \{\Delta\}$, where Δ is a dummy set used for labeling points in $\mathcal{D}_{Y_S}$. Next, we define the random set $\mathcal{H} = \mathcal{H}_{\mathbb{W}} \cup \mathcal{H}_\Delta$ such that

$$(A.3) \qquad \mathcal{H} := \Big\{(x,y) \in (\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})\Big\} \bigcup \Big\{(\Delta, y) : y \in \mathcal{D}_{Y_S}\Big\}.$$

One can observe that $\mathcal{H}$ is the superposition of two marked Poisson PPs $\mathcal{H}_{\mathbb{W}}$ and $\mathcal{H}_\Delta$, taking values in $\mathbb{W} \times \mathbb{W}$ and $\Delta \times \mathbb{W}$, respectively. Moreover, it directly follows from (M2) and (M3)(i) that $\mathcal{H}_{\mathbb{W}}$ has marginal intensity function $\alpha(x)\lambda_{\mathcal{D}_X}(x)$ on $\mathbb{W}$ and stochastic kernel density $\ell(y|x)$ while (M3)(ii) shows that $\mathcal{H}_\Delta$ has marginal intensity function $\lambda_{D_{Y_S}}(\mathbb{W})$ on $\{\Delta\}$ with stochastic kernel density $\frac{\lambda_{\mathcal{D}_{Y_S}}(y)}{\lambda_{\mathcal{D}_{Y_S}}(\mathbb{W})}$. By Theorem 2.12, the intensity functions for $\mathcal{H}_{\mathbb{W}}$ and $\mathcal{H}_\Delta$ are $\alpha(x)\lambda_{\mathcal{D}_X}(x)\ell(y|x)$ and $\lambda_{\mathcal{D}_{Y_S}}(y)$, respectively. Hence, applying Theorem 2.10 to (A.3) reveals that the intensity function for $\mathcal{H}$, $\lambda_{\mathcal{H}}$, is given by

$$(A.4) \qquad \lambda_{\mathcal{H}}(x,y) = \alpha(x)\lambda_{\mathcal{D}_X}(x)\ell(y|x)\mathbb{1}_{x \in \mathbb{W}} + \lambda_{\mathcal{D}_{Y_S}}(y)\mathbb{1}_{x \in \Delta}.$$

Let $\mathcal{H}_Y := \{y : (x,y) \in \mathcal{H}\}$, $\mathcal{H}_X := \{x : (x,y) \in \mathcal{H}\}$ be the projections of $\mathcal{H}$ onto its first and second coordinates, respectively. It immediately follows from Theorem 2.11 that $\mathcal{H}_Y$ is a Poisson PP on $\mathbb{W}$ since it is the image of $\mathcal{H}$ under a projection. Therefore, by treating the

first coordinates of $\mathcal{H}$ as marks, we may express $\mathcal{H}$ as a marked Poisson PP having intensity function $\lambda_{\mathcal{H}_Y}$ on $\mathbb{W}$ and stochastic kernel density $p(x|y)$ from $\mathbb{W}$ to $\mathbb{W}'$. Another application of Theorem 2.12 then implies

$$(A.5) \qquad \lambda_{\mathcal{H}}(x,y) = \lambda_{\mathcal{H}_Y}(y)p(x|y).$$

From (A.4) and (A.5), we obtain the identity

$$(A.6) \qquad p(x|y) = \frac{\alpha(x)\lambda_{\mathcal{D}_X}(x)\ell(y|x)\mathbb{1}_{x\in\mathbb{W}} + \lambda_{\mathcal{D}_{Y_S}}(y)\mathbb{1}_{x\in\Delta}}{\lambda_{\mathcal{H}_Y}(y)}, \quad \lambda_{\mathcal{H}_Y}(y) \neq 0.$$

Equation (A.6) describes the probability density of $\mathcal{H}$ at $x \in \mathbb{W}'$ for $y \in \mathbb{W}$ fixed. Substituting (A.6) for the Janossy density in (2.1) and applying Corollary 2.15 gives the intensity function for the PP $\mathcal{H}_X|D_{Y^i}$ whenever $\lambda_{\mathcal{H}_Y}(y) \neq 0$ for any $y \in D_{Y^i}$:

$$(A.7) \qquad \lambda_{\mathcal{H}_X|D_{Y^i}}(x) = \sum_{y\in D_{Y^i}} \frac{\alpha(x)\lambda_{\mathcal{D}_X}(x)\ell(y|x)\mathbb{1}_{x\in\mathbb{W}} + \lambda_{D_{Y_S}}(y)\mathbb{1}_{x\in\Delta}}{\lambda_{\mathcal{H}_Y}(y)}, \quad \lambda_{\mathcal{H}_Y}(y) \neq 0.$$

Restricting (A.4) and (A.5) to $\mathbb{W} \times \mathbb{W}$, we obtain $p(x|y)\lambda_{\mathcal{H}_Y}(y) = \ell(y|x)\alpha(x)\lambda_{\mathcal{D}_X}(x) = 0$, whenever $\lambda_{\mathcal{H}_Y}(y) = 0$, from which we conclude that $\lambda_{\mathcal{H}_Y}(y) \neq 0$ a.s. Hence restricting (A.7) to $\mathbb{W} \times \mathbb{W}$ yields

$$(A.8) \qquad \lambda_{\mathcal{D}_{X_O}|D_{Y^i}}(x) = \sum_{y\in D_{Y^i}} \frac{\alpha(x)\lambda_{\mathcal{D}_X}(x)\ell(y|x)}{\lambda_{\mathcal{H}_Y}(y)} \quad \text{a.s.}$$

Notice that $\mathcal{H}_Y$ is the same PP as $\mathcal{D}_{Y_O} \cup \mathcal{D}_{Y_S}$. Theorem 2.11 implies that $\mathcal{D}_{Y_O}$ is a Poisson PP, and $\mathcal{D}_{Y_S}$ is a Poisson PP by (M3), so by Theorem 2.10, $\lambda_{\mathcal{H}_Y} = \lambda_{\mathcal{D}_{Y_O}} + \lambda_{\mathcal{D}_{Y_S}}$, where $\lambda_{\mathcal{D}_{Y_O}}(y) = \lambda_{(\mathcal{D}_{X_O}, \mathcal{D}_{Y_O})}(\mathbb{W} \times y) = \int_{\mathbb{W}} \alpha(u)\lambda_{\mathcal{D}_{X_O}}(u)\ell(y|u)du$ by Theorem 2.12. Employing (A.8), one gets that

$$(A.9) \qquad \lambda_{\mathcal{D}_{X_O}|D_{Y^i}}(x) = \alpha(x) \sum_{y\in D_{Y^i}} \frac{\ell(y|x)\lambda_{\mathcal{D}_X}(x)}{\lambda_{\mathcal{D}_{Y_S}}(y) + \int_{\mathbb{W}} \ell(y|u)\alpha(u)\lambda_{\mathcal{D}_X}(u)du},$$

which proves Theorem 3.1 after substituting into (A.1). ∎

REFERENCES

[1] 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*, J. Mach. Learn. Res., 18 (2017), pp. 218–252.

[2] A. ADCOCK, E. CARLSSON, AND G. CARLSSON, *The ring of algebraic functions on persistence bar codes*, Homology Homotopy Appl., 18 (2016), pp. 381–402, https://doi.org/10.4310/HHA.2016.v18.n1.a21.

[3] H. AGAHI, H. MEHRI-DEHNAVI, AND R. MESIAR, *Monte Carlo integration for Choquet integral*, Int. J. Intell. Syst., 34 (2019), pp. 1348–1358, https://doi.org/10.1002/int.22112.

[4] A. Babichev and Y. Dabaghian, *Persistent memories in transient networks*, in Emergent Complexity from Nonlinearity, in Physics, Engineering and the Life Sciences, Springer Proc. Phys. 191, Springer, Cham, 2017, pp. 179–188.

[5] P. Bendich, J. S. Marron, E. Miller, A. Pieloch, and S. Skwerer, *Persistent homology analysis of brain artery trees*, Ann. Appl. Stat., 10 (2016), pp. 198–218, https://doi.org/10.1214/15-AOAS886.

[6] C. Biscio and J. Møller, *The accumulated persistence function, a new useful functional summary statistic for topological data analysis, with a view to brain artery trees and spatial point process applications*, J. Comput. Graph. Statist., 28 (2019), pp. 671–681, https://doi.org/10.1080/10618600.2019.1573686.

[7] O. Bobrowski, S. Mukherjee, and J. E. Taylor, *Topological consistency via kernel estimation*, Bernoulli, 23 (2017), pp. 288–328, https://doi.org/10.3150/15-BEJ744.

[8] T. Bonis, M. Ovsjanikov, S. Oudot, and F. Chazal, *Persistence-based pooling for shape pose recognition*, in Computational Topology in Image Context, A. Bac and J. L. Mari, eds., Springer, New York, 2016, pp. 19–29.

[9] P. Bubenik, *Statistical topological data analysis using persistence landscapes*, J. Mach. Learn. Res., 16 (2015), pp. 77–102.

[10] P. Bubenik, *The Persistence Landscape and Some of Its Properties*, preprint, https://arxiv.org/abs/1810.04963, 2018.

[11] G. Carlsson, T. Ishkhanov, V. D. Silva, and A. Zomorodian, *On the local behavior of spaces of natural images*, Int. J. Comput. Vis., 76 (2008), pp. 1–12.

[12] M. Carrière, S. Y. Oudot, and M. Ovsjanikov, *Stable topological signatures for points on 3D shapes*, Comput. Graph. Forum, 34 (2015), pp. 1–12.

[13] F. Chazal and B. Michel, *An Introduction to Topological Data Analysis: Fundamental and Practical Aspects for Data Scientists*, preprint, https://arxiv.org/abs/1710.04019, 2017.

[14] J. A. Chisholm and S. Motherwell, *A new algorithm for performing three-dimensional searches of the Cambridge Structural Database*, J. Appl. Crystallogr., 37 (2004), pp. 331–334.

[15] M. K. Chung, J. L. Hanson, J. Ye, R. J. Davidson, and S. D. Pollak, *Persistent homology in sparse regression and its application to brain morphometry*, IEEE Trans. Med. Imaging, 34 (2015), pp. 1928–1939.

[16] D. J. Daley and D. Vere-Jones, *An Introduction to the Theory of Point Processes, Volume* I: *Elementary Theory and Methods*, 2nd ed., Probab. Appl. (N. Y.), Springer-Verlag, New York, 2003.

[17] P. Dłotko, R. Ghrist, M. Juda, and M. Mrozek, *Distributed computation of coverage in sensor networks by homological methods*, Appl. Algebra Eng. Commun. Comput., 23 (2012), pp. 29–58.

[18] H. Edelsbrunner, *Computational Topology: An introduction*, American Mathematical Society, Providence, RI, 2010.

[19] B. D. Fabio and M. Ferri, *Comparing persistence diagrams through complex vectors*, in Proceedings of the International Conference on Image Analysis and Processing, 2015, pp. 294–305, https://doi.org/10.1007/978-3-319-23231-7_27.

[20] B. T. Fasy, F. Lecci, A. Rinaldo, L. Wasserman, S. Balakrishnan, and A. Singh, *Confidence sets for persistence diagrams*, Ann. Statist., 42 (2014), pp. 2301–2339, https://doi.org/10.1214/14-AOS1252.

[21] M. Gameiro, Y. Hiraoka, S. Izumi, M. Kramar, K. Mischaikow, and V. Nanda, *A topological measurement of protein compressibility*, Jpn. J. Ind. Appl. Math., 32 (2015), pp. 1–17.

[22] B. Gault, M. P. Moody, J. M. Cairney, and S. P. Ringer, *Atom probe crystallography*, Mater. Today, 15 (2012), pp. 378–386.

[23] R. Ghrist, *Barcodes: The persistent topology of data*, Bull. Amer. Math. Soc. (N.S.), 45 (2008), pp. 61–75, https://doi.org/10.1090/S0273-0979-07-01191-3.

[24] D. Hicks, C. Oses, E. Gossett, G. Gomez, R. H. Taylor, C. Toher, M. J. Mehl, O. Levy, and S. Curtarolo, *AFLOW-SYM: Platform for the complete, automatic and self-consistent symmetry analysis of crystals*, Acta Crystallogr. A, 74 (2018), pp. 184–203.

[25] J. D. Honeycutt and H. C. Andersen, *Molecular dynamics study of melting and freezing of small Lennard-Jones clusters*, J. Phys. Chem., 91 (1987), pp. 4950–4963.

[26] D. P. Humphreys, M. R. McGuirl, M. Miyagi, and A. J. Blumberg, *Fast estimation of recombination rates using topological data analysis*, GENETICS, 211 (2019), pp. 1194–1201, https://doi.org/10.1534/genetics.118.301565.

[27] H. JEFFREYS, *Theory of Probability*, Clarendon Press, Oxford, UK, 1961.

[28] T. F. KELLY, M. K. MILLER, K. RAJAN, AND S. P. RINGER, *Atomic-scale tomography: A 2020 vision*, Microsc. Microanal., 19 (2013), pp. 652–664.

[29] F. A. KHASAWNEH AND E. MUNCH, *Chatter detection in turning using persistent homology*, Mech. Syst. Signal Process., 70–71 (2016), pp. 527–541.

[30] J. F. C. KINGMAN, *Poisson Processes*, Clarendon Press, Oxford, UK, 1993.

[31] G. KUSANO, K. FUKUMIZU, AND Y. HIRAOKA, *Persistence weighted Gaussian kernel for topological data analysis*, in Proceedings of the 33rd International Conference on Machine Learning, Vol. 48, 2016, pp. 2004–2013.

[32] P. M. LARSEN, S. SCHMIDT, AND J. SCHIØTZ, *Robust structural identification via polyhedral template matching*, Model. Simul. Mater. Sci. Eng., 24 (2016), 055007.

[33] R. MAHLER, *Statistical Multisource-Multitarget Information Fusion*, Artech House, Boston, MA, 2007.

[34] A. MARCHESE AND V. MAROULAS, *Topological learning for acoustic signal identification*, in Proceedings of the 19th International Conference on Information Fusion (FUSION), 2016, pp. 1377–1381.

[35] A. MARCHESE AND V. MAROULAS, *Signal classification with a point process distance on the space of persistence diagrams*, Adv. Data Anal. Classif., 12 (2018), pp. 657–682, https://doi.org/10.1007/s11634-017-0294-x.

[36] A. MARCHESE, V. MAROULAS, AND J. MIKE, *K-means clustering on the space of persistence diagrams*, in Wavelets and Sparsity XVII, Proc. SPIE 10394, International Society for Optics and Photonics, Bellingham, WA, 2017.

[37] V. MAROULAS, J. L. MIKE, AND C. OBALLE, *Nonparametric estimation of probability density functions of random persistence diagrams*, J. Mach. Learn. Res., 20 (2019), pp. 1–49.

[38] V. MAROULAS AND A. NEBENFÜHR, *Tracking rapid intracellular movements: A Bayesian random set approach*, Ann. Appl. Stat., 9 (2015), pp. 926–949, https://doi.org/10.1214/15-AOAS819.

[39] N. W. MCNUTT, O. RIOS, V. MAROULAS, AND D. J. KEFFER, *Interfacial Li-ion localization in hierarchical carbon anodes*, Carbon, 111 (2017), pp. 828–834, https://doi.org/10.1016/j.carbon.2016.10.061.

[40] J. MIKE, C. D. SUMRALL, V. MAROULAS, AND F. SCHWARTZ, *Nonlandmark classification in paleobiology: Computational geometry as a tool for species discrimination*, Paleobiology, 42 (2016), pp. 696–706.

[41] Y. MILEYKO, S. MUKHERJEE, AND J. HARER, *Probability measures on the space of persistence diagrams*, Inverse Problems, 27 (2011), 124007, https://doi.org/10.1088/0266-5611/27/12/124007.

[42] M. K. MILLER, T. KELLY, K. RAJAN, AND S. RINGER, *The future of atom probe tomography*, Mater. Today, 15 (2012), pp. 158–165, https://doi.org/10.1016/S1369-7021(12)70069-X.

[43] M. P. MOODY, B. GAULT, L. T. STEPHENSON, R. K. MARCEAU, R. C. POWLES, A. V. CEGUERRA, A. J. BREEN, AND S. P. RINGER, *Lattice rectification in atom probe tomography: Toward true three-dimensional atomic microscopy*, Microsc. Microanal., 17 (2011), pp. 226–239.

[44] J. MOYAL, *The general theory of stochastic population processes*, Acta Math., 108 (1962), pp. 1–31, https://doi.org/10.1007/BF02545761.

[45] M. M. P. NICOLAU, A. J. LEVINE, AND G. E. CARLSSON, *Topology based data analysis identifies a subgroup of breast cancers with a unique mutational profile and excellent survival*, Proc. Natl. Acad. Sci. USA, 108 (2011), pp. 7265–7270.

[46] V. PATRANGENARU, P. BUBENIK, R. L. PAIGE, AND D. OSBORNE, *Topological Data Analysis for Object Data*, preprint, https://arxiv.org/abs/1804.10255, 2018.

[47] C. M. M. PEREIRA AND R. F. MELLO, *Persistent homology for time series and spatial data clustering*, Expert Syst. Appl., 42 (2015), pp. 6026–6038.

[48] J. REININGHAUS, S. HUBER, U. BAUER, AND R. KWITT, *A stable multi-scale kernel for topological machine learning*, in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 4741–4748, https://doi.org/10.1109/CVPR.2015.7299106.

[49] A. ROBINSON AND K. TURNER, *Hypothesis testing for topological data analysis*, J. Appl. Comput. Topol., 1 (2017), pp. 241–261, https://doi.org/10.1007/s41468-017-0008-7.

[50] L. SANTODONATO, Y. ZHANG, M. FEYGENSON, C. M. PARISH, M. C. GAO, R. J. K. WEBER, J. C. NEUEFEIND, Z. TANG, AND P. K. LIAW, *Deviation from high-entropy configurations in the atomic distributions of a multi-principal-element alloy*, Nat. Commun., 6 (2015), 5964, https://doi.org/10.1038/ncomms6964.

[51] I. SGOURALIS, A. NEBENFÜHR, AND V. MAROULAS, *A Bayesian topological framework for the iden-*

tification and reconstruction of subcellular motion, SIAM J. Imaging Sci., 10 (2017), pp. 871–899, https://doi.org/10.1137/16M1095755.

[52] V. D. Silva and R. Ghrist, *Coverage in sensor networks via persistent homology*, Algebr. Geom. Topol., 7 (2007), pp. 339–358, https://doi.org/10.2140/agt.2007.7.339.

[53] S. S. Singh, B.-N. Vo, A. Baddeley, and S. Zuyev, *Filters for spatial point processes*, SIAM J. Control Optim., 48 (2009), pp. 2275–2295, https://doi.org/10.1137/070710457.

[54] A. E. Sizemore, J. E. Phillips-Cremins, R. Ghrist, and D. S. Bassett, *The importance of the whole: Topological data analysis for the network neuroscientist*, Netw. Neurosci., 3 (2019), pp. 656–673, https://doi.org/10.1162/netn_a_00073.

[55] A. Spannaus, V. Maroulas, D. Keffer, and K. J. Law, *Bayesian point set registration*, in 2017 Matrix Annals, Springer, Cham, 2019, pp. 99–120.

[56] A. Togo and I. Tanaka, `Spglib`*: A Software Library for Crystal Symmetry Search*, preprint, https://arxiv.org/abs/1808.01590, 2018.

[57] K. Turner, S. Mukherjee, and D. M. Boyer, *Persistent homology transform for modeling shapes and surfaces*, Inf. Inference, 3 (2014), pp. 310–344, https://doi.org/10.1093/imaiai/iau011.

[58] V. Venkataraman, K. N. Ramamurthy, and P. Turaga, *Persistent homology of attractors for action recognition*, in Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), 2016, pp. 4150–4154, https://doi.org/10.1109/ICIP.2016.7533141.

[59] L. Wasserman, *Topological data analysis*, Annu. Rev. Stat. Appl., 5 (2018), pp. 501–535.

[60] K. Xia, X. Feng, Y. Tong, and G. W. Wei, *Persistent homology for the quantitative prediction of fullerene stability*, J. Comput. Chem., 36 (2014), pp. 408–422, https://doi.org/10.1002/jcc.23816.

[61] Y. Zhang, T. T. Zuo, Z. Tang, M. C. Gao, K. A. Dahmen, P. K. Liaw, and Z. P. Lu, *Microstructures and properties of high-entropy alloys*, Prog. Mater. Sci., 61 (2014), pp. 1–93.

[62] A. Ziletti, D. Kumar, M. Scheffler, and M. Ghiringhelli, *Insightful classification of crystal structures using deep learning*, Nat. Commun., 9 (2018), 2775, https://doi.org/10.1038/s41467-018-05169-6.