Taylor & Francis
Taylor & Francis Group

Journal of Computational and Graphical Statistics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ucgs20

Persistence Flamelets: Topological Invariants for Scale Spaces

Tullia Padellini & Pierpaolo Brutti

To cite this article: Tullia Padellini & Pierpaolo Brutti (2022): Persistence Flamelets: Topological Invariants for Scale Spaces, Journal of Computational and Graphical Statistics, DOI: 10.1080/10618600.2022.2074427

To link to this article: https://doi.org/10.1080/10618600.2022.2074427

Persistence Flamelets: Topological Invariants for Scale Spaces

Tullia Padellini*

Department of Statistical Sciences, Sapienza - Università di Roma

Directorate General for Economics, Statistics and Research, Bank of Italy

and

Pierpaolo Brutti

Department of Statistical Sciences, Sapienza - Università di Roma

*tullia.padellini@bancaditalia.it

Abstract

In recent years there has been noticeable interest in the study of the "shape of data". Among the many ways a "shape" could be defined, topology is the most general one, as it describes an object in terms of its connectivity structure: connected components (topological features of dimension 0), cycles (features of dimension 1) and so on. There is a growing number of techniques, generally denoted as *Topological Data Analysis*, or `TDA` for short, aimed at estimating topological invariants of a fixed object; when we allow this object to change, however, little has been done to investigate the evolution in its topology. In this work we define the *Persistence Flamelet*, a multiscale version of one of the most popular tool in `TDA`, the Persistence Landscape. We examine its theoretical properties and we show its performance as both an exploratory and inferential tool. In addition, we provide open source implementation of the objects and methods presented in the `R`-package `pflamelet`.

Keywords: Topological Data Analysis; Scale space methods; Bandwidth Exploration; Time Series

Introduction

Topological data analysis (TDA) is a new and expanding branch of statistics devoted to recovering the shape of the data, focusing in particular on their connectivity. At its core, TDA comprises a set of tools for detecting and recovering some structure in the data such as connected blocks or clusters (i.e. 0-dimensional topological features) and cycles or loops (i.e. 1-dimensional topological features), while also providing a measure of the importance of each of these elements. As these "topological" features can be defined for very different types of data, from standard vectors in Euclidean spaces to networks or functions, TDA has gained popularity as a unified framework to describe arbitrarily complex objects through easily interpretable features such as their peaks, loops and voids (Wasserman, 2018).

Summaries of the topology of data have been exploited in both explorative (Chazal, Guibas, Oudot and Skraba, 2013; Bendich et al., 2016) and inferential settings (Le and Yamada, 2018; Atienza et al., 2018; Padellini and Brutti, 2017), when the object of interest is supposed to have only one resolution. This assumption can however be limiting, as data can have multiple resolution either because of the way they are defined (this is the case for example of time series, whose resolution parameter is time) or because of the way we represent them (this is the case for example when we impose smoothing on the data, and we obtain different scales corresponding to different levels of smoothness).

In general, due to the ever–growing complexity of data, being able to examine it at different resolutions, hence obtaining different insights, has become a crucial feature of statistical tools. Building on the TDA's toolbox, we thus introduce a new topological summary, the *Persistence Flamelet*, which is able to characterize the evolution of the topological structure of an object across different scales and can be used for both exploratory and inferential purposes. This informative yet interpretable summary, has in fact probabilistic properties that make it suitable for statistical inference. At the same time, the Persistence Flamelet is an effective visualization tool, which allows for exploration of arbitrarily high dimensional data.

This work is structured as follows: in Section 1 we briefly review TDA and its tools for a fixed scale. In Section 2 we introduce the Persistence Flamelet as a topological summary of a scale space and we investigate its theoretical properties. Finally,

Section 3 shows how the Persistence Flamelets can be exploited in applications, with a special emphasis on two famous scale parameters: the *time* in dynamical point–clouds and the *bandwidth* in kernel density estimation. Implementations of the methods described in this work are freely available in the `R`-package `pflamelet`.

1 The shape of fixed-scale data

The mathematical backbone of `TDA` is the notion of *Persistent Homology* (Edelsbrunner and Harer, 2010). Roughly speaking, Persistent Homology provides a characterization of the topological structure of any arbitrary function of data f by building a filtration on it (typically its sub- or super-levelsets, f_ε and f^ε respectively).

The reason why it is necessary to investigate the topology of some function f rather than that of data directly is that data itself often presents only a trivial topological structure. This is especially true when our data is a set of points $\mathbb{X} = \{X_i \in \mathbb{R}^d, i = 1, \dots, n\}$, an object usually referred to as *point cloud* in the `TDA` literature.

As can be seen in Figure 1, in fact, when we look at the observed points (in black) we can see no connection between them. Every point X_i is "connected" only to itself, thus we have as many connected components as there are observations but no higher dimensional topological structures such as circles or voids. In order to recover the clear circular structure on which data lay, it is necessary to connect the points through an additional function f, in this case a distance function (whose sub-levelsets are shown in grey). The link between Persistent Homology and the "shape of the data" is that for some choice of f, sub- (or respectively super-) levelset filtrations are topologically equivalent to the space data $\mathbb{X}$ was sampled from, say $\mathcal{M}$, which in the following we will assume to be a compact manifold with no boundary embedded in $\mathbb{R}^d$. We will now show the construction of Persistent Homology for two classes of functions for which this equivalence holds: *distances* and *kernel density estimators*. While we focus on these two example for the ease of interpretation, the result presented in the following are not limited to these functions, and, depending on the application of interest, other choices of f may be better suited (Atienza et al., 2019; Chintakunta et al., 2015).

Distance functions

The most common choice for analysing the topological structure of $\mathcal{M}$ is to investigate the Persistent Homology of the sub-levelset filtrations of a distance function. At each level ε, the ε-sub-levelset of the distance d, d_ε, is defined as

$$d_\varepsilon = \bigcup_{i=1}^{n} B(X_i, \varepsilon),$$

where $B(X_i, \varepsilon) = \{x \mid d_{\mathbb{X}}(x, X_i) \leq \varepsilon\}$ denotes a ball of radius ε and center X_i, and $d_{\mathbb{X}}$ is an arbitrary distance function.

The topology of d_ε can be recovered by computing its Homology Groups; Homology groups of dimension 0, $H_0(d_\varepsilon)$, represent connected components of d_ε, $H_1(d_\varepsilon)$ represent its loops, and so on. d_ε is topologically more interesting than the original point cloud, but it is extremely sensible to the radius ε. For each value of ε, in fact, we obtain a different estimate d_ε, with a different topological structure: for small values of ε, the topology of d_ε is close to the one of the point–cloud itself. As ε grows more and more points start to be connected, until eventually the corresponding d_ε is homeomorphic to a point.

The key feature of encoding data into a filtration is that as ε grows, different sub-levelsets $d_{\varepsilon_1}, d_{\varepsilon_2}$ with $\varepsilon_1 < \varepsilon_2$ are related, so that if a feature is present in both we can say that it remains alive in the interval $[\varepsilon_1, \varepsilon_2]$. Persistent Homology then allows to see how features appear and disappear at different scales. Values $\varepsilon_b < \varepsilon_d$ of ε corresponding respectively to when two components are connected for the first time (*birth–step*) and when they connect to some other larger component (*death–step*) are the generators of a Persistent Homology Group (Figure 1).

In the statistical literature, d_ε is often known as the *Devroye–Wise support estimator* (Devroye and Wise, 1980). The consistency of the Devroye-Wise estimator justifies and motivates the use of the distance function: as d_ε is a consistent estimator of $\mathcal{M}$, the topology of d_ε is a sensible approximation of the topology of $\mathcal{M}$.

Kernel Density estimators

The second way of linking levelset filtrations to the topology of $\mathcal{M}$, the support of the distribution generating the data, is that the super–levelsets of a density function *p* can be topologically equivalent to the support of the distribution itself (Fasy et al., 2014). More formally, if the data are sampled from a distribution *P* supported on $\mathcal{M}$, and if the density *p* of *P* is smooth and bounded away from 0, then there is an interval $[\eta, \delta]$ such that the super–levelset $p^{\varepsilon} = \{x \mid p(x) \geq \varepsilon\}$ is homotopic (i.e. topologically equivalent) to $\mathcal{M}$, for $\eta \leq \varepsilon \leq \delta$.

Since the true generating density *p* is most often unknown, it is typically approximated by a kernel density estimator $\hat{p}$. A naïve way to estimate the topology of $\mathcal{M}$ is then to compute topological invariants of the super-levelset of the kernel density estimator $\hat{p}$:

$$\hat{p}^{\varepsilon} = \{x \mid \hat{p}(x) \geq \varepsilon\}.$$

The super-levelsets $\hat{p}^{\varepsilon}$, with $\varepsilon \in [0, \max \hat{p}]$, form a decreasing filtration, which means that $\hat{p}^{\varepsilon} \subset \hat{p}^{\delta}$ for all $\delta \leq \varepsilon$. As in the case of distances, for each element in the filtration, i.e. for each value ε, we obtain a different estimate $\hat{p}^{\varepsilon}$, whose topology can be characterized by its Homology Groups. Since in practice it is not possible to determine the interval $[\eta, \delta]$ in which the topology of $\hat{p}^{\varepsilon}$ is closest to that of $\mathcal{M}$, we analyse the evolution of the topology over the whole filtration. Once again, Persistent Homology allows to analyze how those Homology Groups change with ε. Persistent loops in $\hat{p}^{\varepsilon}$ naturally represent circular structures in $\hat{p}$, Persistent Homology Groups of dimension 2 indicate holes in $\hat{p}$ and so on.

An example of this construction can be seen in Figure 2. When ε is close to $\max \hat{p}$, the super-levelsets of $\hat{p}$ (highlighted in grey), are disjoint, as shown in the left panel of the Figure. When ε decreases, the local peaks that may be disconnected at the beginning (as the smaller one in the Figure), merged into the same super-levelset. Finally, for $\varepsilon = 0$, $\hat{p}^{0}$ is formed by two disjoint components, corresponding to the two main peaks of the distribution. Since 0 is the smallest value the function $\hat{p}$ can take, it is also the “last” value for which the peaks can be defined, hence it is taken to be

their “time of death”. It is worth noticing that topological features of dimension 0, or connected components have a relevant interpretation in terms of “bumps”; connected components in the filtration $\hat{p}^{\varepsilon}$ are in fact local maxima of $\hat{p}$; this is true for any super-levelset filtration. When the filtration is defined in terms of sub-levelset instead, as in the case of the distance function, connected components represent local minima.

1.1 Persistence Diagram

Persistent Homology Groups can be summarized by the *Persistence Diagram*, a multiset $D = \{z_i = (b_i, d_i)\}_{i=1}^{m}$ whose generic element (b_i, d_i) is the i^{th} generator of the Persistent Homology Group. Features with a long “lifetime” (or *persistence* $\text{pers} = b - d$) are those which can be found at many different resolution of the filtration, and are informative of the topology of $\mathcal{M}$. Points that are close to the diagonal instead represent short–lived features, which may be only noisy artifacts and can be neglected (Fasy et al., 2014).

The space of Persistence Diagrams $\mathcal{D}$ is a metric space when endowed with the *Bottleneck distance*, which, given two Persistence Diagrams D and D', is defined as

$$d_B(D, D') = \inf_{\gamma} \sup_{x \in D} \| x - \gamma(x) \|_{\infty},$$

where the infimum is taken over all bijections $\gamma : D \mapsto D'$.

Several other metrics have been proposed to compare Persistence Diagrams, for example the Wasserstein distance (Mileyko et al., 2011) or the Fisher Information metric (Le and Yamada, 2018). One of the main advantages of adopting the Bottleneck distance, is that it allows to prove *stability*, arguably one of the most important properties of Persistence Diagrams, under relatively mild assumptions (Chazal et al., 2016).

Theorem 1.1 (Stability). *Let f and g be two functions on a triangulable space* $\mathbb{X}$ *and let* D_f, D_g *be the Persistence Diagram built on their respective sub- (or super-) levelset filtrations, then*

$$d_B(D_f, D_g) \leq \|f - g\|_\infty ,$$

where $\|f\|_\infty = \sup_x |f(x)|$ *is the* L^∞ *–norm.*

In the special case of $f = d_{\mathbb{X}}$ and $g = d_{\mathbb{Y}}$ two distance functions defined on two point–clouds $\mathbb{X}$ and $\mathbb{Y}$ respectively, the stability result can be written in a more easily interpretable way:

$$d_B\left(D_{\mathbb{X}}, D_{\mathbb{Y}}\right) \leq 2 d_H\left(\mathbb{X}, \mathbb{Y}\right),$$

where $d_H(\mathbb{X}, \mathbb{Y})$ is the *Hausdorff* distance between two topological spaces $\mathbb{X}$ and $\mathbb{Y}$. Roughly speaking this means that if the two point clouds $\mathbb{X}$ and $\mathbb{Y}$ are close, their persistence Diagrams will be as well, and can be interpreted in two ways:

- *the Persistence Diagram is a topological signature:* stability reassures us that if two point clouds $\mathbb{X}, \mathbb{Y}$ are similar their Persistence Diagrams will be as well, and is therefore instrumental for using them in statistical tasks such as classification or clustering;
- *the Persistence Diagram is statistically consistent:* stability reassure us that if we are using a point–cloud $\mathbb{X}_n$ to estimate the topology of an unknown object $\mathbb{X}$, if $\mathbb{X}_n \to \mathbb{X}$ as $n \to \infty$, then $D_{\mathbb{X}_n}$ converges to $D_{\mathbb{X}}$ as well.

Stability is also key to assess statistical significance of topological features. Building on the core idea that features that are close to the diagonal are more likely to be noise than those that are far away from it, Fasy et al. (2014) proposes a way to define a bootstrap confidence band around the diagonal. Points of the Diagram laying outside the band are taken to be significant, as they are most likely signal, whereas those inside the band may be just noise.

1.2 Persistence Landscape

Persistence Diagrams are defined in spaces endowed with only a metric structure, which can be limiting in data analysis. A collection of Persistence Diagrams $D_1, \ldots, D_n$ in fact does not have a unique mean, nor a satisfying measure of variability (Turner

et al., 2014). More critically, although it is possible to define a probability distribution on the space of Persistence Diagrams $\mathcal{D}$, (Mileyko et al., 2011), it is still not clear how to explicitly derive it (if it is possible to derived it at all). In order to overcome these issues and to work with more statistics-friendly spaces, several tools have been developed to convert Persistence Diagrams into functional objects, the most famous being the Persistence Landscape (Bubenik, 2015) and the Persistence Silhouette (Chazal et al., 2014). These topological summaries are built by mapping each point $z = (b, d)$ of a Persistence Diagram *D* to a piecewise linear function called the "triangle" function T_z, which is defined as:

$$T_z(y) = (y - b + d)1_{[b-d,b]}(y) + (b + d - y)1_{(b,b+d]}(y) \qquad y \in [0, Y],$$

where $Y = \max_{z \in D}(b + d)/2$ and $1_A(x)$ is the standard indicator function: $1_A(x) = 1$ if $x \in A$ and $1_A(x) = 0$ otherwise. Informally a triangle function links each point of the Diagram to the diagonal with segments parallel to the axes, and then rotates them of 45 degrees.

The triangles T_z can be combined in many different ways. If we take their k-max, i.e. the k^{th} largest value in the set $T_z(y)$, we obtain the k^{th} *Persistence Landscape*

$$\lambda_D^k(y) = k\text{-}\max_{z \in D} T_z(y) \qquad y \in [0, Y], \in k = 1, \ldots, m.$$

The Persistence Landscape λ_D is a representation of the Persistence Diagram *D* as a collection $\{\lambda_D^1, \ldots, \lambda_D^K\}$ of piecewise linear functions, indexed by the order of the maximum to be considered in defining the Landscape, *k*, which can be any number between 1 and *m*, the cardinality of the Persistence Diagram *D*. If we take the weighted average of the functions $T_z(y)$, we have the *Power Weighted Silhouette*

$$\psi_p(y) = \frac{\sum_{z \in D} w_z^p T_z(y)}{\sum_{z \in D} w_z^p} \qquad y \in [0, Y].$$

Figure 3 shows a point cloud with its corresponding Persistence Diagram and Landscape. The two circles in the data, clearly picked up by the Persistence Diagram, correspond to the two peaks of the Persistence Landscape. Interestingly,

the Landscape also retains information about the persistence of these two loops or cycles, with the peak on the left, corresponding to the smaller circle, being slightly shorter than the one on the right.

While the space of Persistence Diagrams $\mathcal{D}$ is only a metric space, Persistence Landscapes are defined in a much richer Banach space $\mathcal{L}$, endowed with the following norm

$$\left\|\lambda_D\right\|_p^p = \sum_k \left\|\lambda_D^k\right\|_p^p,$$

where $\left\|\lambda^k\right\|_p$ is the L^p–norm

$$\left\|\lambda^k\right\|_p = \left(\int \lambda^k \mathrm{d}\mu\right)^{1/p}.$$

It is not possible to go back from Persistence Landscapes to Persistence Diagrams, meaning that there is a loss of information in going from Persistence Diagrams to Persistence Landscapes. However the Persistence Landscape is still informative, since stability still holds (Bubenik, 2015).

Theorem 1.2. *Let f, g be two functions on* $\mathbb{X}$ *and let* D_f *and* D_g *be the Persistence Diagrams built from their super- (or sub-) levelsets, then*

$$d_\Lambda\left(\lambda_{D_f}, \lambda_{D_g}\right) \leq \left\|f - g\right\|_\infty,$$

where $d_\Lambda\left(\lambda_{D_f}, \lambda_{D_g}\right) = \left\|\lambda_{D_f} - \lambda_{D_g}\right\|_\infty$ *is the* L^∞ *–distance in the space of Persistence Landscapes,* $\mathcal{L}$.

Persistence Landscapes are piece–wise linear functions, which makes it possible to define a (unique) mean and a variance for any collection of them. The main advantage of the Persistence Landscape over the Persistence Diagram is that it is defined in a Banach Space, which is instrumental in statistical learning as it allows for a full characterization of the Persistence Landscape as a random variable. More details can be found in the Supplementary Material.

2 The Persistence Flamelet

Persistence Diagrams and Persistence Landscapes give us a full characterization of a function f in terms of the topology of its sub–levelset (or super–levelset) filtration, however they do not show changes in its structure as f varies. We now focus on the case where rather than one function f we are dealing with a family of functions $\mathcal{F} = \{ f_\sigma, \sigma \in S \}$, indexed by some parameter σ, which represent the resolution or the scale of the object f_σ. This is a very common setting in statistics, as the presence of multiple scales in the analysis can arise both from the nature of the data themselves (this is the case for example when σ is the time in time series or the spatial resolution in images or geo-referenced data) and from the algorithm used to perform the analysis, where σ typically represent a tuning parameter.

Although traditional methods focus on selecting a single optimal scale σ^*, inspired by scale space theory (Lindeberg, 1994), we adopt the idea that different scales yield different information, hence all of them must be simultaneously taken into account in order to get a better understanding of the phenomenon under analysis. We restrict ourselves to the case where the $\mathcal{F} = \{ f_\sigma, \sigma \in S \}$ is continuously indexed by the scale parameter σ, defined in a bounded interval S. For the sake of simplicity we will assume $\sigma \in [0,1]$, as every bounded interval can be rescaled to $[0,1]$. Two notable examples that we will explore more thoroughly in the following are kernel smoothers, for which the resolution σ is given by the bandwidth parameter h, and time–varying processes, whose scale σ is the time, t.

Previous attempts at encoding a multi-resolution family $\mathcal{F} = \{ f_\sigma, \sigma \in [0,1] \}$ into the `TDA` framework focused on considering the Persistence Diagram itself as a function of the scale parameter σ. The family of Persistence Diagrams $\mathbb{D} = \{ D_\sigma, \sigma \in [0,1] \}$ corresponding to $\mathcal{F}$, is known as *Persistence Vineyards* (Cohen-Steiner et al., 2006) and is a stable and continuous representation of the topology of the whole $\mathcal{F}$ (Munch, 2013). Despite their theoretical relevance, however, Persistence Vineyards suffer from several drawbacks that hinder their popularity in applications. Comparing two of them, for example, is indeed computationally very intensive, due to the fact that matching distances need to be computed for all the Persistence Diagrams

comprising the scale-space. As opposed to the Persistence Diagram, which is praised for being an invaluable visualization tool, graphical representation of Vineyards may at times be cumbersome to interpret. Moreover, they share all the drawbacks and limitations of Persistence Diagrams, more specifically they lack a unique average and a measure of variability for a group of them (Turner et al., 2014), and, once again, since their probabilistic behavior is not well understood, the use of Persistence Vineyard in statistical inference is severely compromised.

We thus introduce a new representation, based on the Persistence Landscape, that overcomes most of these issues and provides a functional representation with very favorable probabilistic properties, which can be exploited for statistical inference using well known tools of Functional Data Analysis, thus greatly extending the potential application of topological summaries. It is worth noticing that although in the following we focus on Persistence Landscapes, the same results hold for Persistence Silhouettes as well. In order to explicitly take into account the multiple resolutions of $\mathcal{F}$, we consider the Persistence Landscapes λ_{D_σ} corresponding to the family $\mathcal{F} = \{ f_\sigma, \sigma \in [0,1] \}$ as a function of the scale parameter σ. Visually we can think of such function as a “flow” of landscapes, one for each resolution, smoothly moving and resembling a tiny fire (see, for example, Figure 7).

Definition 2.1 (Persistence Flamelet). Given a Persistence Vineyard $\mathbb{D}$, that is a collection of Persistence Diagrams $\mathbb{D} = \{D_\sigma, \sigma \in [0,1]\}$, continuously indexed by some parameter $\sigma \in [0,1]$, and $k \in \mathbb{N}^+$, we define the k^{th} *Persistence Flamelet* as the function

$$\Lambda^k(\sigma, y) = \lambda^k_{D_\sigma}(y) \qquad \forall \sigma \in [0,1],\ y \in [0, Y],\ k \in \mathbb{N}^+.$$

As the Landscape itself, the *Persistence Flamelet* Λ is also a collection $\Lambda = \{\Lambda^k, k \in \mathbb{N}^+\}$ indexed by the order of the $\max$ we consider.

The theoretical reassurance that the Persistence Flamelet is a meaningful topological summary is its stability, which we will prove in the following. Before doing so, however, we need to introduce a notion of *proximity* between Persistence Flamelets.

Definition 2.2 (Integrated Landscape distance). Let $\mathbb{D} = \{D_\sigma, \sigma \in [0,1]\}, \mathbb{G} = \{G_\sigma, \sigma \in [0,1]\}$ two Persistence Vineyards and $\Lambda_\mathbb{D}, \Lambda_\mathbb{G}$ the corresponding Persistence Flamelets. We define the *Integrated Landscape distance* between $\Lambda_\mathbb{D}$ and $\Lambda_\mathbb{G}$ as

$$I_\Lambda(\Lambda_\mathbb{D}, \Lambda_\mathbb{G}) = \int_0^1 d_\Lambda(\lambda_{D_\sigma}, \lambda_{G_\sigma}) \mathrm{d}\sigma.$$

As functional objects, Persistence Flamelets can be compared by a natural extension of the usual L^p metrics, which is way less computationally demanding than the matching distances needed for Persistence Diagrams and Vineyards.

Theorem 2.1. *Let $\mathbb{D} = \{D_\sigma, \sigma \in [0,1]\}, \mathbb{G} = \{G_\sigma, \sigma \in [0,1]\}$ two Persistence Vineyards and $\Lambda_\mathbb{D}, \Lambda_\mathbb{G}$ the corresponding Persistence Flamelets, then:*

1. *$\Lambda_\mathbb{D}$ and $\Lambda_\mathbb{G}$ are continuous with respect to the Bottleneck distance;*
2. *$I_\Lambda(\Lambda_\mathbb{D}, \Lambda_\mathbb{G}) \leq I_B(\mathbb{D}, \mathbb{E})$*

where $I_B(\mathbb{D}, \mathbb{G}) = \int_0^1 d_B(D_\sigma, G_\sigma) \mathrm{d}\sigma$ is the Integrated Bottleneck distance for Persistence Vineyards as defined in Munch (2013).

Proof.

These statements are a direct consequence of the Stability Theorem for Persistence Landscapes (Theorem 1.2) and the continuity of Persistence Vineyards, in fact:

1. For a fixed σ, consider D_σ and $D_{\sigma+\varepsilon}$ (same applies for $\mathbb{G}$). By Theorem 1.2 and the continuity of $\mathbb{D}$ we have

   $$0 \leq \lim_{\varepsilon \to 0} d_\Lambda\left(\lambda_{D_\sigma}, \lambda_{D_{\sigma+\varepsilon}}\right) \leq \lim_{\varepsilon \to 0} d_B\left(D_\sigma, D_{\sigma+\varepsilon}\right) = 0.$$

2. Since for a fixed σ we have, by Theorem 1.2 we have

   $$d_\Lambda\left(\lambda_{D_\sigma}, \lambda_{G_\sigma}\right) \leq d_B\left(D_\sigma, G_\sigma\right)$$

   integrating both terms is enough to prove the result.

□

The Persistence Flamelet is also a random variable defined in a Banach space. In analogy with what Bubenik (2015) has done for Persistence Landscapes, we define a norm for Persistence Flamelets, more specifically

$$\|\Lambda\|_p^p = \int_0^1 \sum_k \left\|\lambda_{D_\sigma}^k\right\|_p^p \mathrm{d}\sigma.$$

Then, following Ledoux and Talagrand (2013), we can extend the Law of Large Numbers and the Central Limit Theorem to this new object.

Corollary 2.1.1 (Strong Law of Large Numbers). *Let* $\{\Lambda_n\}_{n\in\mathbb{N}}$ *be a sequence of independent copies of* Λ *and, for a given n, let* $S_n = \Lambda_1 + \ldots + \Lambda_n$*, where the sum is defined pointwise.*

$$\frac{S_n}{n} \to \mathbb{E}(\Lambda) \quad \text{almost surely} \Leftrightarrow \mathbb{E}\|\Lambda\| < \infty.$$

Corollary 2.1.2 (Central Limit Theorem). *Assume* $\mathcal{B}$ *has type 2 in the sense of Hoffmann-Jorgensen et al. (1976). If* $\mathbb{E}(V) = 0$ *and* $\mathbb{E}(\|\Lambda\|^2) < \infty$ *then* $\frac{S_n}{\sqrt{n}}$ *converges weakly to a Gaussian random variable* $G(\Lambda)$ *with the same covariance structure as* Λ.

Proofs follow from Theorem A.1 and Theorem A.2 of the Supplementary Material.

2.1 Confidence Band for Persistence Flamelets

In order to strengthen the role of the Persistence Flamelet as a statistical tool, we now show that it is possible to build confidence bands on the *mean Persistence Flamelet* by means of bootstrapping, in analogy with what has been done for Persistence Landscapes (Chazal, Fasy, Lecci, Rinaldo, Singh and Wasserman, 2013). What follows applies to any level *k* of a Flamelet, hence we omit any explicit reference to it in order to ease the notation.

Let $\mathbb{L}_n = \{\Lambda_1, \dots, \Lambda_n\}$, be a sequence of Persistence Flamelets independently sampled from some probability distribution $\mathcal{P}$. Then the *mean Persistence Flamelet*, μ, is defined as

$$\mu(\sigma, y) = \mathbb{E}_p[\Lambda(\sigma, y)] \qquad \sigma \in [0,1],\ y \in [0, Y].$$

Theorem 2.2 (Consistency of the Bootstrap). *Let* $\mathcal{G} = \{g_{\sigma,t}, \sigma \in [0,1], y \in [0, Y]\}$ *be the class of functions defining the Persistence Flamelet, that is, for any Persistence Vineyard* $\mathbb{D}$ *, the function* $g_{\sigma,t}$ *is given by as*

$$g_{\sigma,y}(\mathbb{D}) = \Lambda(\sigma, y).$$

If the family $\mathcal{F} = \{f_\sigma, \sigma \in [0,1]\}$ *characterizing the Flamelet is Lipschitz in its scale argument, that is*

$$\left\| f_\sigma - f_\tau \right\|_2 \leq K\,|\sigma - \tau|$$

then the class $\mathcal{G}$ *is Donsker, and the bootstrap procedure defined in Algorithm 1 is valid.*

Algorithm 1: Bootstrap Bands for Persistence Flamelets

Input:

- $\mathbb{L}_n = \{\Lambda_1, \dots, \Lambda_n\}$, i.i.d. sample from some probability distribution $\mathcal{P}$ over the space of Persistence Flamelets
- α confidence level
- B number of Bootstrap repetitions

1 Compute $\bar{L}_n(\sigma, t) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_i(\sigma, y)$;

2 **for** *j in* $1:B$ **do**

3 Sample (with replacement) n elements $\Lambda_1^j, \dots, \Lambda_n^j$ from $\mathbb{L}_n$;

4 Compute the bootstrapped sample mean $\bar{L}_n^j(\sigma, t) = \frac{1}{n}\sum_{i=1}^{n} \Lambda_i^j(\sigma, y)$;

5 Compute $\theta_j = \sup_{\sigma,y} |\sqrt{n}(\bar{L}_n^j(\sigma,y) - \bar{L}_n(\sigma,y))|$;

6 **end**

7 Define $q_\alpha = \inf\{q \mid \frac{1}{B}\sum_{i=1}^{B} \mathbb{I}(\theta_i > q) \leq \alpha\}$;

Output:

$$CB_n(\sigma,t) = [\bar{L}_n(\sigma,y)) - q_\alpha / \sqrt{n}, \bar{L}_n(\sigma,y)) + q_\alpha / \sqrt{n}]$$

Corollary 2.2.1. *The band CB_n obtained from Algorithm 1 is an asymptotic confidence band for the Persistence Flamelet at confidence level* $1-\alpha$.

As any topological summary, the Persistence Flamelet may be affected by noise and highlight spurious features, and confidence bands may be used to assess whether or not a topological feature is statistically significant. If we take the empty Flamelet, i.e. the constant Flamelet located on 0, to represent the case of no significant topological structure, the procedure illustrated before can in fact be used to define a confidence band on the noise component of the topological structure. Comparing this band with the observed Flamelet allows us to detect which parts (if any) of the Flamelet are to be considered noisy artifacts resulting from the procedure adopted to build the topological summaries, or significantly different than zero and can be thus considered *topological signal*.

Alternatively, in order to check whether or not a feature is to be judged as relevant, we suggest to exploit the fact that, in practice, the Persistence Flamelet is computed over a finite set of values $S = \{\sigma_1, \dots, \sigma_J\}$ of the scale parameter σ defining the family $\mathcal{F}$. For each of the Persistence Diagrams corresponding to elements of S, it is possible to define a simplification scheme that retains only significant topological feature filtering out the noise. One strategy to implement this idea is to *reshape* the Diagram so that the persistence of its “relevant” features is maximized, as suggested in Atienza et al. (2019). Here, instead, in order to preserve the original structure of the Diagram, we adopt the approach introduced in Fasy et al. (2014), which consists in building a confidence band around the diagonal of the Diagram via bootstrap:

points of the Diagram that lay within this band are not significantly different from the diagonal itself, and can thus be removed from the Diagram. The Persistence Flamelet built using Diagrams "denoised" in this fashion, then contains only significant topological features; in the following we will refer to this technique as " Diagram cleaning". It is worth noticing that since this procedure requires building multiple confidence bands, possibly using the same data, it may be necessary to adopt a multiplicity correction in order to ensure the proper coverage level. With respect to the denoising based on the Confidence Band for the Flamelets, this is more of a global approach, in the sense that it does not depend on the choice of a functional representation of the Persistence Diagram, nor on the tuning parameter k and p of, respectively, Landscape and Silhouettes; nevertheless, this procedure is not explicitly tailored for the Persistence Flamelet and may be affected by the choice of the grid S.

It is worth mentioning that Confidence Bands are one way of simplifying the Persistence Diagrams (or Persistence Flamelets) "a posteriori". As shown in (Chintakunta et al., 2015), however, it is also possible to directly build the Diagram so that the information it contains is "significant". The construction of the Persistence Flamelet is agnostic with respect to the way the family of Diagrams comprising the Persistence Vineyard are computed, hence both strategies are viable to ensure that the topological noise it may contain is minimized.

3 Applications

In this last section, we illustrate how the Persistence Flamelets, so far only a rather abstract object, may be encountered and fruitfully used as both an inferential and exploratory tool.

3.1 Time Series / EEG Dynamic Point–Clouds

The easiest way to understand the need for topological characterization of a continuously varying space is to consider the case where the scale parameter σ is time, t. The Persistence Flamelets allows in fact for a characterization of a time–varying system $\mathcal{F} = \{ f_t, t \in [0,1] \}$ in terms of its topology (Munch et al., 2015; Munch, 2013) by allowing us to simultaneously study the shape of any

time–dependent function f_t and how it evolves with time t. Again, although this framework is general enough to cover any arbitrary function f_t, as long as it is continuous with respect to time, we are especially interested in the case where f_t is a function of data.

Assume that at each time t we observe a sample $\mathbb{X}(t) = \{X_1(t), \dots, X_k(t)\}$ drawn from some distribution P_t. The Persistence Flamelet Λ built on distance functions or kernel density estimators estimate the topology of the whole continuous–time generating process $\{P_t, t \in [0,1]\}$. The trace of the sample in the time interval $\{\mathbb{X}(t), t \in [0,1]\}$, usually called *Dynamic Point Cloud*, is just a high dimensional time series, hence Persistence Flamelets can be exploited as a tool to extract a new type of insights on time series of arbitrarily high dimension.

In the special case of dynamic point–clouds, the stability result of Theorem 2.1 can be restated as follows.

Corollary 3.0.1. *Let* $\{\mathbb{X}(t), \mathbb{Y}(t)\}$ *with* $t \in (0,1)$ *two continuous dynamic point clouds,* $\Lambda_{\mathbb{X}}$ *and* $\Lambda_{\mathbb{Y}}$ *their corresponding Persistence Flamelets, then:*

$$I_{\Lambda}(\Lambda_{\mathbb{X}}, \Lambda_{\mathbb{Y}}) \leq I_H(\mathbb{X}, \mathbb{Y}),$$

where $I_H(\mathbb{X}, \mathbb{Y}) = \int_0^1 d_H\big(\mathbb{X}(t), \mathbb{Y}(t)\big)\,dt$ *is the Integrated Hausdorff distance for dynamic point–clouds, as defined in Munch (2013).*

Figure 4 shows two Persistence Flamelets built from electroencephalography (EEG) tracks, freely available on the `UCI Machine Learning Repository`. EEG are electric impulses recorded at a very high frequency (256 *Hz*) through multiple electrodes (64 in this study), located in different areas of the skull. At each time t, topological features represent dependency structure in the signal, which is relevant information per se, but since it is also important to assess whether or not these connection persist in time, this kind of data fits perfectly in our framework.

We compare the EEGs of 10 alcoholic and 10 control patient, all subject to the same stimulus. For each of them we have 5 trials of 1 second; EEG are typically very noisy

hence we average them across repetition before computing their topological summaries. Features of both dimension 1 and dimension 0 seem to be concentrated in the same area of the Persistence Diagrams, we thus build the Persistence Flamelet using as a base summary the Persistence Silhouette, which takes into account all points in the Diagram, as well as the Persistence Landscapes, for which the selection of the order of the *k*-max is non-trivial in this application.

The Persistence Flamelet highlights a topological difference in behavior of the two groups; as shown in Figure 4 the signal from the control group, in fact, appears to be characterized by a few persistent features. In the alcoholic patient instead there is less structure; the number of features is higher than in the control patient, yet they all have a smaller persistence, and could thus be interpreted as noise. In order to assess whether there is statistical substance to this claim, we perform a two sample test to compare the average silhouettes of the two groups. As the sample size is rather small, exploiting the asymptotic normality of the average Flamelet is not advisable, hence we resort to the permutation test detailed in Algorithm 2.

Algorithm 2: Two-Sample Permutation Tests for Persistence Flamelets

---

Input:

- $\mathbb{L}^{(1)} = \{\Lambda_1^{(1)}, \dots, \Lambda_{n_1}^{(1)}\}, \mathbb{L}^{(2)} = \{\Lambda_1^{(2)}, \dots, \Lambda_{n_2}^{(2)}\}$ i.i.d. samples of Persistence Flamelets
- *B* number of Bootstrap repetitions

1 Compute $\overline{L}^{(j)}(\sigma, t) = \frac{1}{n_j} \sum_{i=1}^{n_j} \Lambda_i^{(j)}(\sigma, y)$ for $j = 1, 2$;

2 Compute $\theta^{(1,2)} = \sup_{\sigma, y} |(\overline{L}^{(1)}_{n_1}(\sigma, y) - \overline{L}^{(2)}_{n_2}(\sigma, y))|$;

3 Define the joint sample $\mathbb{L}_n = \{\Lambda_1^{(1)}, \dots, \Lambda_{n_1}^{(1)}, \Lambda_1^{(2)}, \dots, \Lambda_{n_2}^{(2)}\}$;

4 **for** *i in* $1 : B$ **do**

5 Sample (without replacement) n_1 elements $\Lambda_1^{(1,i)}, \dots, \Lambda_{n_1}^{(1,i)}$ from $\mathbb{L}_n$;

6 Define $\mathbb{L}_i^{(1)} = \{\Lambda_1^{(1,i)}, \dots, \Lambda_{n_1}^{(1,i)}\}$ and $\mathbb{L}_i^{(2)} = \{\Lambda_1^{(2,i)}, \dots, \Lambda_{n_2}^{(2,i)}\}$ its complement with respect to $\mathbb{L}_n$;

7 Compute the bootstrapped sample means $\overline{L}_i^{(1)}(\sigma, y)$ and $\overline{L}_i^{(2)}(\sigma, y)$ for the two samples;

8 Compute $\theta_i = \sup_{\sigma, y} |(\overline{L}_i^{(1)}(\sigma, y) - \overline{L}_i^{(2)}(\sigma, y))|$;

9 **end**

Output:

$$\text{p-value} = \frac{1}{B}\sum_{i=1}^{B} \mathbb{I}(\theta_i > \theta^{(1,2)});$$

Let m be the maximum number of elements per Diagram in $\mathbb{D}$ and J be the number of elements of the finite grid used to in practice to compute the Persistence Flamelet ($S = \{\sigma_1, \dots, \sigma_J\}$). In both Algorithm 2 and Algorithm 1 the most intensive step, the computation of θ_i, is $O(mJ)$, as follows naturally from the results of Bubenik and Dłotko (2017). If we wanted to compare in the same fashion two Persistence Vineyards, however, the computational cost would be a considerably larger $O(m^{1.5}\log(m)J)$ (Kerber et al., 2017).

As the Flamelet allows to isolate topological invariants of different dimensions, it can be used to retrieve qualitative differences on the topology, as well as quantitative ones. In addition to assessing the presence of a topological discrepancy, in fact, it is also informative on which kind of features are responsible for it. Results shown in Table 1, in fact, highlight that the distinction between the two groups does not depend on the sole presence of connected components (i.e. electrodes that are in some sense “close”) but it is determined by the nature of the association within these components, which is formalized by whether or not these sets of electrodes form cyclical structures. We consider two different orders of Landscapes for each dimension to show that the pattern we observe depend on the dimension and not on the order of the landscape we consider. We take the two largest order k of the landscape, which in the case of dimension 0 corresponds to $k = 2, 3$, as $k = 1$ is a constant triangle equal to the diameter of the data and it is not discriminative.

Interestingly, this behavior is robust with respect to the base function used to build the Flamelet, as inferential conclusions when considering the Persistence Landscape (with different choices of *k*) are coherent with those obtained using the Persistence Silhouette.

3.2 Data Smoothing / Kernel Density Estimation

In the statistical literature, scale–space ideas have been especially popular in the context of *data smoothing*. In its broader definition, data smoothing is a family of methods aimed at recovering some structure in the data. Depending on their scale, however, smoothing methods may enhance noise or neglect relevant features, so that it is crucial to understand the impact of the smoothing level on the smoothed object. The Persistence Flamelet can be used to summarize and evaluate the evolution of the whole smoothing process, by tracking (and visualizing) the appearance and disappearance of feature of arbitrary dimension.

Among all the smoothing methods, we focus on Kernel Density Estimation (KDE) (Scott, 2015), for which the role of topological features (especially that of 0^{th} dimensional Homology Groups) is a well established problem (Chaudhuri and Marron, 1999). Features affected by the smoothing process such as local peaks (or, in topological terms, 0^{th} dimensional Homology Groups), are in fact especially meaningful in the case of KDE; local modes of a density and their basin of attraction represent for example one way of defining clusters (Comaniciu and Meer, 2002). Persistence Flamelets allows us to explore also higher dimensional features, such as cycles or voids, which have been noticeably neglected.

Given a sample $\{X_1 \dots, X_n\}$, drawn from some smooth density *p*, a Kernel Density Estimator $\hat{p}_h$ is defined as

$$\hat{p}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i),$$

where $K_h(x - y) = \frac{1}{h} K(\frac{x - y}{h})$ is a scaled kernel, *h* is the bandwidth parameter and $K(\cdot)$, the kernel, is a non-negative, symmetric function that integrates to 1.

While any kernel function $K(\cdot)$ may be used without compromising the performance of the estimator, the bandwidth parameter represent the level of smoothing and needs to be finely tuned. In the scale-space approach, given some bounded range of bandwidths $H \subset \mathbb{R}^+$, all the estimators $\hat{p}_h$ are simultaneously considered, so that the object of interest becomes the family of smooths $\mathcal{F} = \{\hat{p}_h : h \in H\}$. Since K_h is continuous with respect to h by definition, it is immediate to see that the Persistence Flamelets can be used to investigate and characterize $\mathcal{F}$.

In the exploration framework, the first attempt at investigating the relation between the bandwidth of a kernel density estimator and its topology `SiZer` (Chaudhuri and Marron, 1999). Roughly speaking, given a sample $\{X_1, \ldots, X_n\}$ drawn from a univariate density p, `SiZer` (SIgnificant ZERo crossings of derivatives) is a map showing where in space, x, and scale, h, the kernel density estimator $\hat{p}_h(x)$ is significantly increasing or decreasing. Since local peaks of a curve can be thought of as points where its derivative changes sign, the basic idea of `SiZer` is assess where this change happens, by testing whether the sign of the derivative $\hat{p}'_h(x)$ for each couple of values (x, h) is positive or negative. Values (x, h) corresponding to significantly positive derivatives are shown in red and significantly negative are shown in black, as in Figure 7.

`SiZer` is intrinsically 1–dimensional and even though it has been extended to 2–dimensional densities, especially in the context of image analysis, (Godtliebsen et al., 2004) the features it hunts for are always and only local modes. The Persistence Flamelet provides a further extension in two different directions:

- it can be used to investigate topological features of any dimension, rather than only feature of dimension 0, i.e. local peaks;
- it does not depend on the dimension of the data and can thus be used to investigate kernel densities for very high dimensional data.

Like `SiZer`, the Persistence Flamelet is able to assess the significance of each peak, by exploiting the tools shown in Section 2.1, but, in addition, it also provides a measure of the relevance of each feature: its persistence.

3.2.1 Bandwidth Exploration

We now show two real–data applications. In the first univariate one we quickly compare the Persistence Flamelet with `SiZer` and show that, when both are available they yield similar insights. The second is a bivariate example, which motivates investigating higher dimensional features and highlights the potential of the Persistence Flamelet when other tools are not available.

Eartquakes I / Depth

In our first example we consider a classical dataset in kernel density estimation, the depth of the 512 earthquakes beneath the Mt. St. Helens volcano in the months before the eruption of 1982 (Scott, 2015). Figure 5 shows the 1^{st} and the 2^{nd} Persistence Flameles for the 0 dimensional topological feature of the density estimator $\hat{p}$ built with the Gaussian Kernel:

$$\hat{p}_h(x) = \frac{1}{n}\sum_{i=1}^{n} K_h(x - X_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{\sqrt{2\pi h}} \exp\left\{\frac{1}{2h}(x - X_i)^2\right\}.$$

In order to retain only significant topological information, before computing the Flamelets we built confidence bands on the diagonal of each of the Diagrams used to define it, and features whose persistence was smaller than the upper limit of such bands, where then discarded from the Diagram. As the same data are used to build the Diagrams corresponding to the different bandwidths, we adopt a Bonferroni correction on the confidence level to ensure proper coverage over the whole Flamelet.

The 1^{st} Persistence Flamelet consists of only one peak, representing the global maximum, which, as we can expect, always persists. This is not very informative, and when analyzing dimension 0 topological features, it is thus advisable to consider 2^{nd} Persistence Flamelet, which represents the most relevant local peaks. In this case we can see that the two peaks appearing in the 2^{nd} Persistence Flamelet correspond to the two points in the Diagram (which in turn correspond to the two bumps we can see in the KDE in Figure 7). As we can see from Figure 5, the 2^{nd} Persistence Flamelet behaves differently than 1^{st} Persistence Flamelet; when the bandwidth grows in fact, the two secondary peaks are smoothed away. Figure 7

shows the comparison with `SiZer`, and it is easy to see that the two approaches lead to very similar conclusions. In order to make this comparison possible, we highlight here that we cleaned the Diagrams following the procedure introduced in Section 2.1 and we removed from the Persistence Diagrams the noisy features falling inside the confidence band before computing the Flamelet, so that the topological features displayed on it are all statistically significant. The three peaks appear for $h = 0.05$, then one of them disappear at around $h = 0.25$, one other around $h = 0.35$ and, the last one always survives (in the given range of bandwidths).

Earthquakes II / Locations

For our second example we consider earthquake data coming from the `USGS catalog`. Our sample consists of the locations, expressed in latitude and longitude, of 6500 events with magnitude higher than 5, taking place between June 2013 and June 2017. The 2–dimensional density p generating the data $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ can still be estimated using the kernel density estimator with a Gaussian Kernel:

$$\hat{p}_h(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{\boldsymbol{H}}(\boldsymbol{x} - \boldsymbol{X}_i) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2\pi |\boldsymbol{H}|^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{X}_i)^t \boldsymbol{H}^{-1}(\boldsymbol{x} - \boldsymbol{X}_i)\right\}.$$

Notice that in the multivariate case, the bandwidth is not a scalar but rather a matrix $\boldsymbol{H}$, however we chose an isotropic Gaussian Kernel, which corresponds to imposing a spherical structure to the covariance matrix

$$\boldsymbol{H} = h\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \qquad h \in \mathbb{R}^+,$$

so that the kernel density estimator expression can be simplified as follows:

$$\hat{p}_h(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{2\pi h} \exp\left\{-\frac{1}{2h^2}(\boldsymbol{x} - \boldsymbol{X}_i)^t(\boldsymbol{x} - \boldsymbol{X}_i)\right\}.$$

Earthquakes are concentrated around circular structures, also known as *plates*. According to Plate Tectonics, in fact, the Earth's lithosphere is broken into 7 main plates, plus a number of minor ones. Since earthquakes are caused by the movements of neighboring plates, the density p naturally inherits the Earth's plates

structure. In terms of topology, plates can be thought as loops, or dimension 1 Homology Groups.

The dimension 1 Persistence Flamelet of the kernel density estimator $\hat{p}$ can be employed to assess whether or not kernel density estimators are able to recover these loops. The Persistence Flamelet shown in Figure 8 presents 7 crests, each of them representing one persistent loop in $\mathcal{F}$; this seems to suggest that at, different resolution, the kernel density estimator is able to recover all the 7 main plates. Notice that as opposed to the 0^{th} dimensional case, where there is always one feature, the global maximum, dominating all the others, when analysing loops we can limit our analysis to the 1^{st} Persistence Flamelet.

In this example specifically, the Persistence Flamelet shows that there is one loop that persists noticeably more than all the others; as persistence is a measure of the importance of a feature, this suggests that there is one plate which is more neatly detected than all others. This highly persistent loop closely resembles the contour of the Philippine plate, which is not surprising, since more than 26% of the seismic activity in the given time interval was concentrated in the area between Philippine and Japan.

3.2.2 Bandwidth Selection

We conclude by showing that, even though the Persistence Flamelet is intrinsically related to the scale-space principle, according to which no bandwidth candidate is taken to be more informative than the others, our topological summary also plays a role in the context of *bandwidth selection*, and we can use it to heuristically choose a "topologically–aware" bandwidth.

In the literature on bandwidth selection the topological structure of the KDE is usually ignored (with the exception of local modes (Genovese et al., 2016)). However, as standard approaches to the task (most noticeably cross-validation) have proven to fail when the density is concentrated around lower dimensional structures (Genovese et al., 2016), we believe that taking into account topological invariants whose dimension is larger than 0 (local modes) yet smaller than the ambient space, may be beneficial.

Intuitively, since persistence can be interpreted as a measure of the importance of each feature, bandwidths corresponding to peaks in the Persistence Flamelet result in estimators that highlight the most prominent features in the density. By selecting the value of *h* that maximise the Persistence Flamelet, the *topologically–aware* $\hat{h}_{\mathrm{TA}}$, we are forcing the density estimator to retain the most relevant topological treats.

Let us consider again the **Earthquake II** example. By choosing the value of *h* that maximise the Persistence Flamelet, we are forcing the density estimator to emphasize the most persistent loop. The kernel density estimator $\hat{p}_{h_{\mathrm{TA}}}$, shown in Figure 9, is in fact concentrated around the Philippine plate, as we could expect. To understand why such a topologically–aware bandwidth selection heuristic may be useful, let us compare it with more established methods for bandwidth selection: Silverman's Normal Rule and a Plug–in bandwidth selection criterion. We intentionally ignore cross validation methods because, as shown in Genovese et al. (2016), they are known to display poor behaviour in this setting. When the density is singular, which is the case when the support of the distribution is concentrated on an object whose dimension is smaller than the ambient dimension, cross validation will in fact select 0 as optimal bandwidth, and it is thus not informative.

The first alternative we consider is an extension of Silverman Normal Rule, one of the most famous "rule of thumb" for bandwidth selection, to the case of densities with singular features, as detailed in Chacón et al. (2011). More specifically, given a sample $\{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n\} \in \mathbb{R}^D$, from some distribution *P*, the optimal bandwidth *h* for recovering the *d*–dimensional features is

$$\hat{h}_s^2 = \left(\frac{4}{n(d+2)} \right)^{\frac{2}{4+d}} s,$$

where $s = D^{-1} \sum_{j=i}^{D} s_j^2$ and s_j^2 is the variance of the j^{th} variable. Despite the fact that we set $d = 1$, in order to take into account the loop structure, the density estimator, shown in Figure 10, does not seem to recover any of the plates at all.

The second approach we consider is a plug–in bandwidth estimator $\boldsymbol{H}_{\mathrm{PI}}$, obtained by minimizing the AMISE (Asymptotic Mean Integrated Square Error) with respect to the bandwidth *h*; details are given in Chacón et al. (2011). Since limiting the case of scalar bandwidths, as we did until here, may seem too restrictive, in this final example we relax the hypothesis of spherical covariance and do not impose any structure on the bandwidth matrix $\boldsymbol{H}$. The additional complexity of the estimator does not however result in a better estimation: as we can see in Figure 11, the plates structure of the true density is still not recognizable.

4 Conclusion and Future Developments

In this work we introduced a new multiscale topological summary, the Persistence Flamelet, we characterized it in a probabilistic framework and we proved that it retains the topological information contained in the original scale spaces, and thus it is a meaningful topological summary. The very different nature of the two examples we considered (dynamic point clouds and kernel smoothers) show the versatility of this representation, which allows to account for the presence of multiple scales in the data and also in the tools we typically use to analyse them, and illustrate how the Persistence Flamelet allows to effectively summarize non-linear dependencies within an object.

So far we exclusively focused on comparing objects of the same kind, i.e. Flamelets built on the same type of data, such as EEG recordings. One of the main features of the Persistence Flamelets, however, is that they allow to compare data structures of potentially different types (e.g. networks and functional data), hence providing a unified framework for analysing complex data. In the future we plan to exploit this property to match different sources of neuroimaging data, more specifically EEG recordings, which are functional objects, and functional/structural networks obtained from fMRI data. We also wish to extend the use of the Flamelet to the supervised setting, especially for problems related to neuroimaging data, and see whether this topological summary is informative enough to help in predicting different neurological conditions.

Finally, we plan to investigate further the properties of Persistence Flamelets–related heuristics for bandwidth selection. We have already seen how picking the bandwidth that maximises the "persistency" seems to be promising. We plan to investigate it even further also considering the use of the Persistence Flamelet to select a bandwidth that reflects some previous knowledge on the topology of the object of interest. In addition, since features that appear at many different resolution can be thought as the most relevant ones, it may also be interesting to explore persistence in bandwidth ranges as an additional measure of relevance for topological traits.

References

Atienza, N., Gonzalez-Diaz, R. and Rucco, M. (2019), 'Persistent entropy for separating topological features from noise in vietoris-rips complexes', *Journal of Intelligent Information Systems* **52**(3), 637–655.

Atienza, N., González-Díaz, R. and Soriano-Trigueros, M. (2018), 'On the stability of persistent entropy and new summary functions for tda', *arXiv preprint arXiv:1803.08304*.

Bendich, P., Marron, J. S., Miller, E., Pieloch, A. and Skwerer, S. (2016), 'Persistent homology analysis of brain artery trees', *The annals of applied statistics* **10**(1), 198.

Bubenik, P. (2015), 'Statistical topological data analysis using persistence landscapes', *The Journal of Machine Learning Research* **16**(1), 77–102.

Bubenik, P. and Dłotko, P. (2017), 'A persistence landscapes toolbox for topological statistics', *Journal of Symbolic Computation* **78**, 91–114.

Chacón, J. E., Duong, T. and Wand, M. (2011), 'Asymptotics for general multivariate kernel density derivative estimators', *Statistica Sinica* pp. 807–840.

Chaudhuri, P. and Marron, J. S. (1999), 'Sizer for exploration of structures in curves', *Journal of the American Statistical Association* **94**(447), 807–823.

Chazal, F., de Silva, V., Glisse, M. and Oudot, S. (2016), *The Structure and Stability of Persistence Modules*, SpringerBriefs in Mathematics, Springer International Publishing.

Chazal, F., Fasy, B. T., Lecci, F., Rinaldo, A., Singh, A. and Wasserman, L. (2013), ' On the bootstrap for persistence diagrams and landscapes', *Model. Anal. Inform. Sist.* **20**(6), 111–120.

Chazal, F., Fasy, B. T., Lecci, F., Rinaldo, A. and Wasserman, L. (2014), Stochastic convergence of persistence landscapes and silhouettes, *in* 'Proceedings of the thirtieth annual symposium on Computational geometry', ACM, p. 474.

Chazal, F., Guibas, L. J., Oudot, S. Y. and Skraba, P. (2013), 'Persistence-based clustering in riemannian manifolds', *Journal of the ACM (JACM)* **60**(6), 41.

Chintakunta, H., Gentimis, T., Gonzalez-Diaz, R., Jimenez, M.-J. and Krim, H. (2015), 'An entropy-based persistence barcode', *Pattern Recognition* **48**(2), 391–401.

Cohen-Steiner, D., Edelsbrunner, H. and Morozov, D. (2006), Vines and vineyards by updating persistence in linear time, *in* 'Proceedings of the twenty-second annual symposium on Computational geometry', ACM, pp. 119–126.

Comaniciu, D. and Meer, P. (2002), 'Mean shift: A robust approach toward feature space analysis', *IEEE Transactions on pattern analysis and machine intelligence* **24**(5), 603–619.

Devroye, L. and Wise, G. L. (1980), 'Detection of abnormal behavior via nonparametric estimation of the support', *SIAM Journal on Applied Mathematics* **38**(3), 480–488.

Edelsbrunner, H. and Harer, J. (2010), *Computational topology: an introduction*, American Mathematical Soc.

Fasy, B. T., Lecci, F., Rinaldo, A., Wasserman, L., Balakrishnan, S., Singh, A. et al. (2014), 'Confidence sets for persistence diagrams', *The Annals of Statistics* **42**(6), 2301–2339.

Genovese, C. R., Perone-Pacifico, M., Verdinelli, I. and Wasserman, L. (2016), 'Non-parametric inference for density modes', *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **78**(1), 99–126.

Godtliebsen, F., Marron, J. S. and Chaudhuri, P. (2004), 'Statistical significance of features in digital images', *Image and Vision Computing* **22**(13), 1093–1104.

Hoffmann-Jorgensen, J., Pisier, G. et al. (1976), 'The law of large numbers and the central limit theorem in banach spaces', *Annals of Probability* **4**(4), 587–599.

Kerber, M., Morozov, D. and Nigmetov, A. (2017), 'Geometry helps to compare persistence diagrams', *Journal of Experimental Algorithmics (JEA)* **22**, 1–20.

Le, T. and Yamada, M. (2018), Persistence fisher kernel: A riemannian manifold kernel for persistence diagrams, *in* 'Advances in Neural Information Processing Systems', pp. 10007–10018.

Ledoux, M. and Talagrand, M. (2013), *Probability in Banach Spaces: isoperimetry and processes*, Springer Science & Business Media.

Lindeberg, T. (1994), 'Scale-space theory: A basic tool for analyzing structures at different scales', *Journal of applied statistics* **21**(1-2), 225–270.

Mileyko, Y., Mukherjee, S. and Harer, J. (2011), 'Probability measures on the space of persistence diagrams', *Inverse Problems* **27**(12), 124007.

Munch, E. (2013), Applications of persistent homology to time varying systems, PhD thesis, Duke University.

Munch, E., Turner, K., Bendich, P., Mukherjee, S., Mattingly, J., Harer, J. et al. (2015), 'Probabilistic fréchet means for time varying persistence diagrams', *Electronic Journal of Statistics* **9**(1), 1173–1204.

Padellini, T. and Brutti, P. (2017), 'Supervised learning with indefinite topological kernels', *arXiv preprint arXiv:1709.07100*.

Scott, D. W. (2015), *Multivariate density estimation: theory, practice, and visualization*, John Wiley & Sons.

Turner, K., Mileyko, Y., Mukherjee, S. and Harer, J. (2014), 'Fréchet means for distributions of persistence diagrams', *Discrete & Computational Geometry* **52**(1), 44–70.

Wasserman, L. (2018), 'Topological data analysis', *Annual Review of Statistics and Its Application* **5**(1), 501–532.

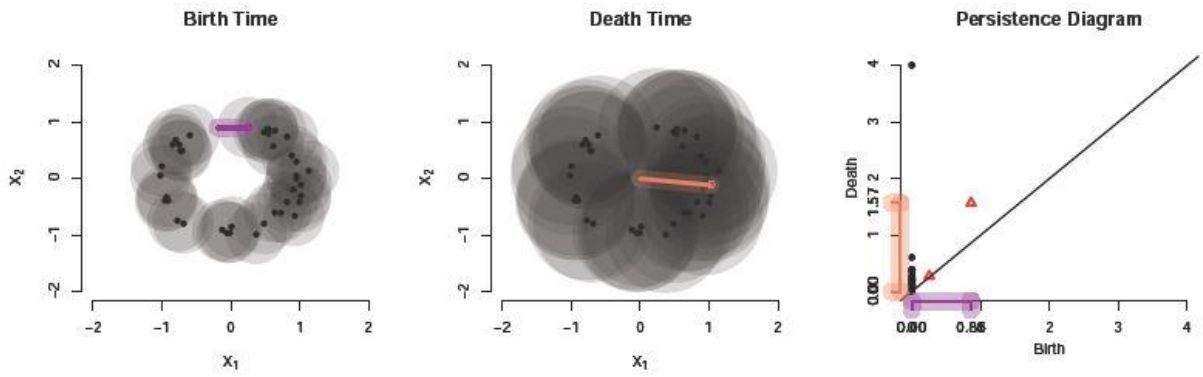

Fig. 1 From left to right: birth of the circle in the filtration, death of the circle and summarizing Persistence Diagram.

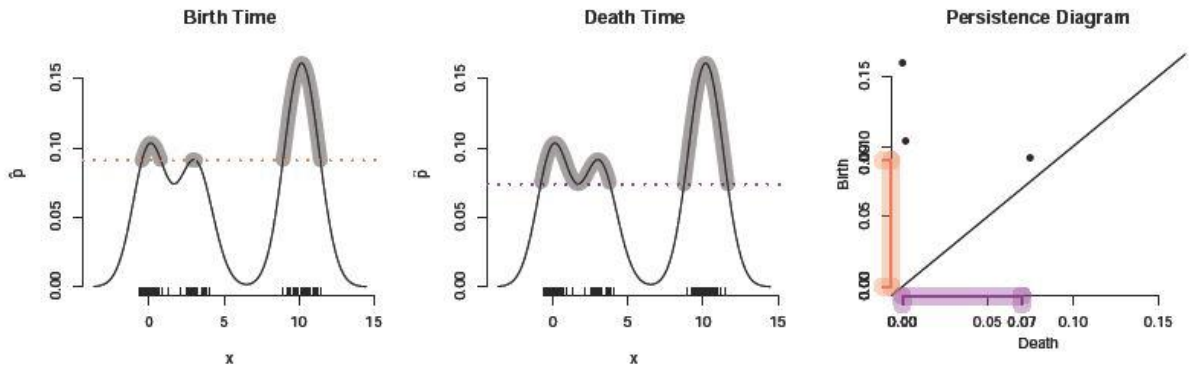

Fig. 2 From left to right: birth of the smallest peak in the filtration, $\hat{p}^{b}$, death of the smallest peak in the filtration $\hat{p}^{d}$ and summarizing Persistence Diagram.

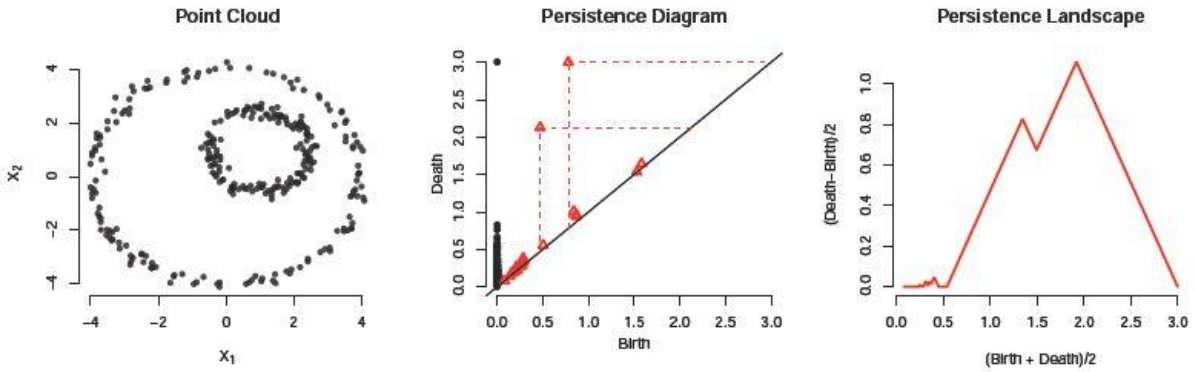

Fig. 3 A point cloud (left) with its corresponding Persistence Diagram (centre) and Persistence Landscape (right).

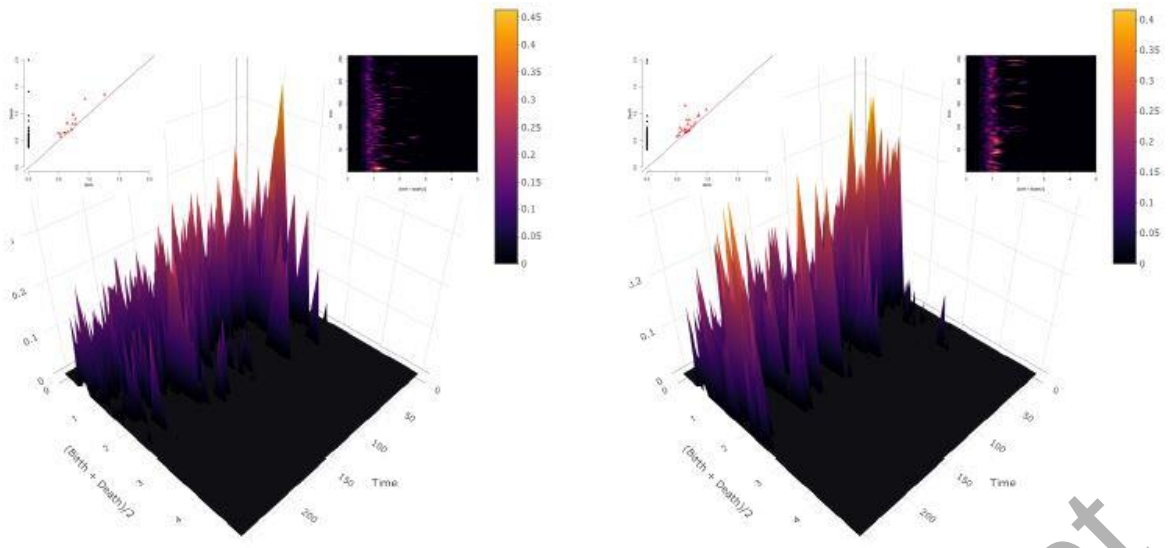

Fig. 4 Persistence Flamelets of Dimension 1 for the EEG data of one alcoholic (left) and one control (right) subject.

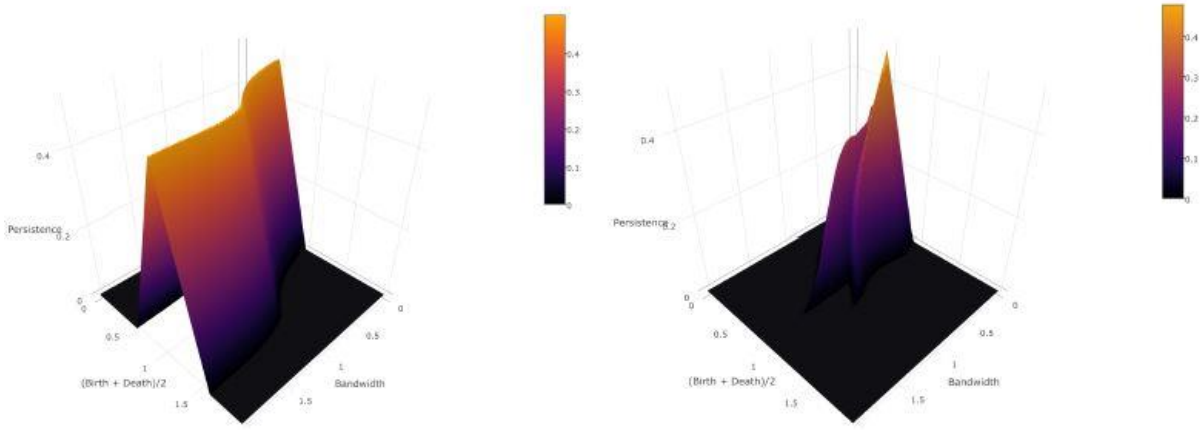

Fig. 5 1^{st} (left) and 2^{nd} (right) Persistence Flamelets of dimension 0.

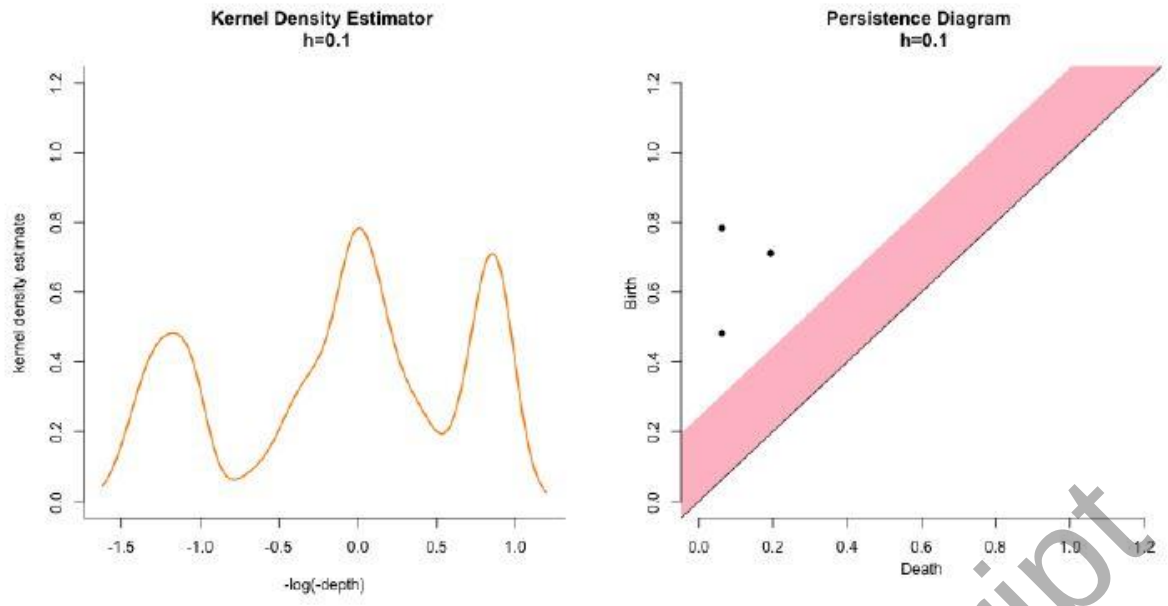

Fig. 6 From left to right: Kernel Density Estimator of the Mt. St. Helens dept data (with $h = 0.1$) and corresponding Persistence Diagram. Highlighted in pink is the confidence band around the diagonal.

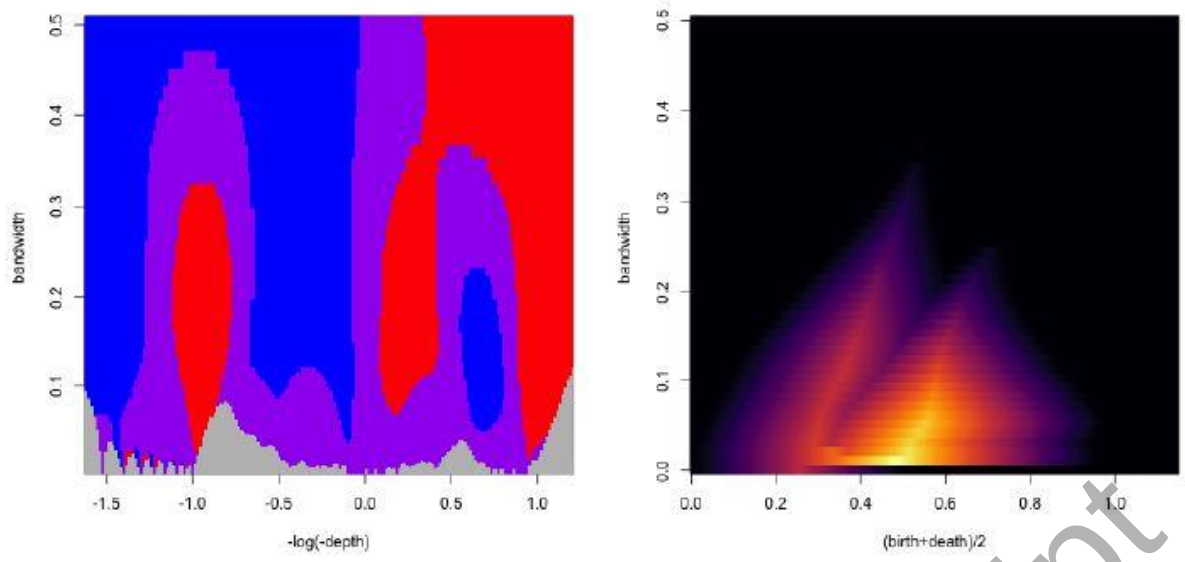

Fig. 7 `SiZer`, the 1^{st} and 2^{nd} Persistence Flamelets of dimension 0. In order to facilitate the comparison with `SiZer`, the Persistence Flamelet is projected and represented as a matrix.

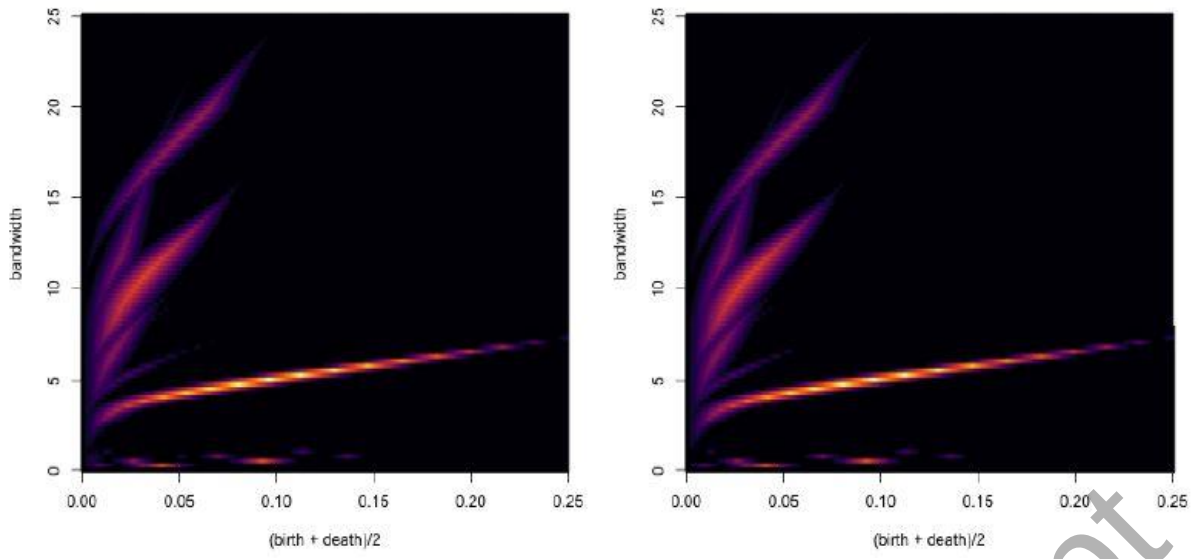

Fig. 8 Dimension 1 Persistence Flamelets for earthquakes locations KDE before (left) and after (right) cleaning each Diagram with a confidence band.

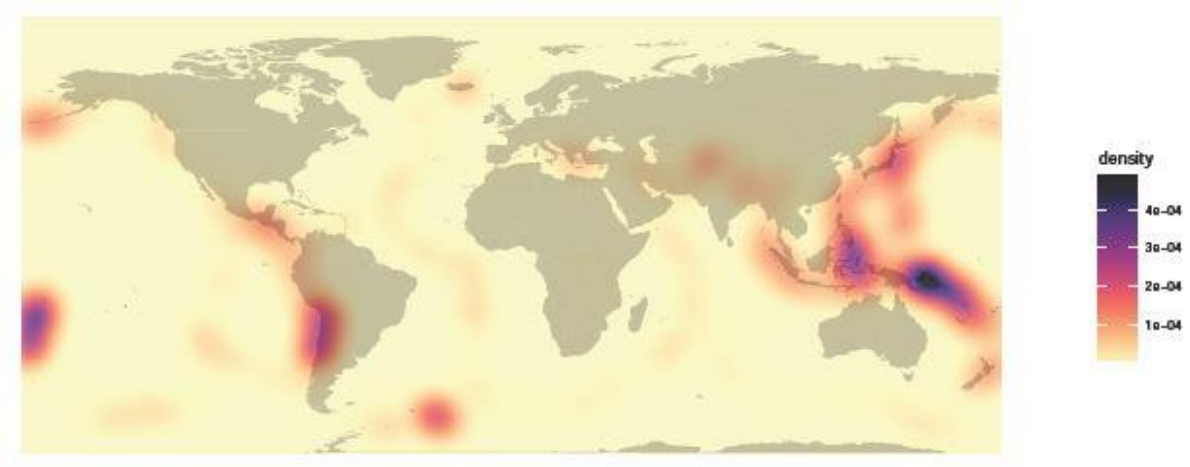

Fig. 9 Density estimation with the topologically aware bandwidth $\hat{h}_{\mathrm{TA}}$.

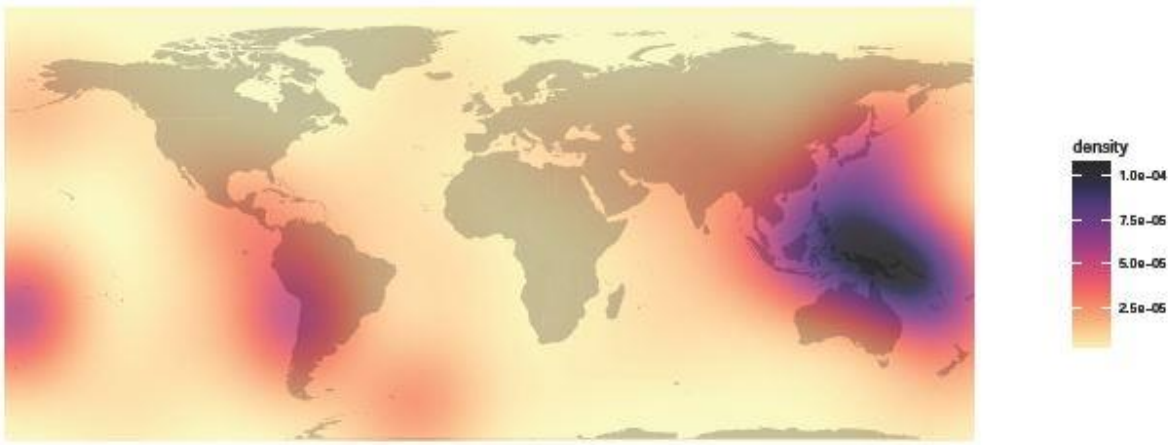

Fig. 10 Density estimation with extended Silverman Normal bandwidth $\hat{h}_s$.

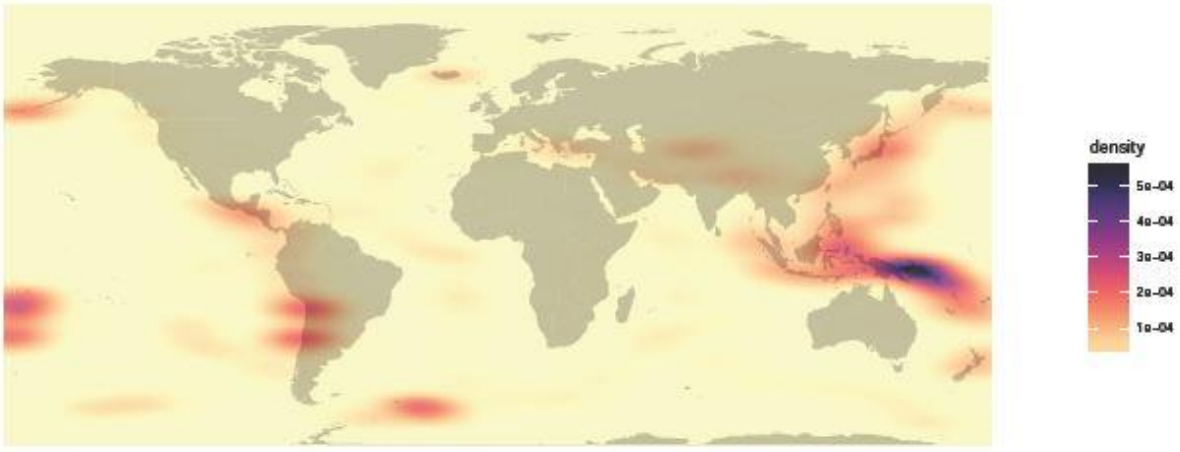

Fig. 11 Density estimation with anisotropic Plug–in bandwidth matrix $\boldsymbol{H}_{\mathrm{PI}}$.

Table 1 Bootstrapped p-value for the Two-Sample test

	Persistence Landscape		Persistence Silhouette
Dimension 0	0.200 ($k = 2$)	0.895 ($k = 3$)	0.199
Dimension 1	0.048 ($k = 1$)	0.032 ($k = 2$)	0.020