Time topological analysis of EEG using signature theory

Stéphane Chrétien, Ben Gao, Rémi Vaucher, Astrid Thébault Guiochon

Time topological analysis of EEG using signature theory

Stéphane Chrétien [1] & Ben Gao [2] & Rémi Vaucher [3] & Astrid Thébault Guiochon [4]

[1] *Laboratoire ERIC, Université Lyon 2, France, stephane.chretien@univ-lyon2.fr*
[2] *Halias Technologie, Grenoble, France, ben.gao@halias.fr*
[3] *Halias Technologies & Laboratoire ERIC, université Lyon 2, remi.vaucher@halias.fr*
[4] *Laboratoire EMC, Université Lyon 2, a.thebaultguiochon@univ-lyon2.fr*

Résumé. La détection d'anomalies dans les signaux multivariés est une tâche de première importance dans de nombreuses disciplines (épidémiologie, finance, sciences cognitives et neuro-sciences, cancérologie etc). Dans cette optique, l'analyse topologique des données (TDA) offre une batterie d'invariants "de forme" qui peuvent être exploités pour la mise en œuvre d'un schéma de détection efficace. Notre contribution consiste à étendre les construction présentées dans [5] sur la construction de complexes simpliciaux à partir des Signatures des signaux et leur capacités prédictives, plutôt que de l'utilisation d'une distance générique comme dans [12]. La théorie des Signatures est une nouvelle thématique en Machine Learning [3] issue des travaux récents sur les notions de Chemins Rugueux (Rough Paths) développés par Terry Lyons et son équipe [9] sur la base du formalisme introduit par Chen [2]. Nous explorons en particulier la détection des changements de topologie, sur la base du suivi de l'évolution de la persistance homologique et des nombres de Betti associés au complexe introduit dans [5]. Nous appliquons nos outils pour l'analyse de signaux cérébraux de type EEG afin de détecter des phénomènes précurseurs aux crises d'épilepsies.

Mots-clés. Analyse de données topologique, Séries Temporelles, Statistiques Appliquées

Abstract. Anomaly detection in multivariate signals is a task of paramount importance in many disciplines (epidemiology, finance, cognitive sciences and neurosciences, oncology, etc.). In this perspective, Topological Data Analysis (TDA) offers a battery of "shape" invariants that can be exploited for the implementation of an effective detection scheme. Our contribution consists of extending the constructions presented in [5] on the construction of simplicial complexes from the Signatures of signals and their predictive capacities, rather than the use of a generic distance as in [12]. Signature theory is a new theme in Machine Learning [3] stemming from recent work on the notions of Rough Paths developed by Terry Lyons and his team [9] based on the formalism introduced by Chen [2]. We explore in particular the detection of changes in topology, based on tracking the evolution of homological persistence and the Betti numbers associated with the complex introduced in [5]. We apply our tools for the analysis of brain signals such as EEG to detect precursor phenomena to epileptic seizures.

Keywords. Topological and Geometric Data Analysis, Time Series, Applied Statistics

1 Introduction

Topological data analysis [1] is a rapidly growing field that has emerged recently, based on the intriguing observation that data come with shape-like properties. In general, existing topological structures that are built from data, such as Cech or Vietoris Rips complexes, make essential use of a metric that may not be fully suitable because inherited from the space where our data are imbedded instead of the geodesic distance on the manifold where our data truly live. In a recent paper [5], we proposed a novel way of buiding simplicial complexes based on performance in prediction rather than distance. The main tool for performing prediction is the signature transform [3], that easily extract meaningful features from multidimensional signals. In this setting, each node represents a component of the signal and a node belongs to a simplex if the signature of all the nodes from the simplex accurately explains the considered node, statistically speaking. Similarly, a face belongs to a simplex if the signature of the signals represented by the face is accurately predicted by the adjacent faces of various orders in the simplex. In this approach, Signatures faithfully represent faces and simplices together with a natural orientation and make the topological construction better motivated and more statistically meaningful. From a computational viewpoint, selecting faces that explain another one using their respective signatures can be done efficiently using the LASSO, in the spirit of the celebrated method proposed in [10] for estimating graphical models. More recently, [12] solved this problem more precisely by considering a signal correlation matrix. This method allows to create high-dimensionnal structures.

In the present paper, we examine a more refined aspect of our topological construct : the dynamic evolution of the topology as a function of time as complexes undergo potential structural transformations at specific change points in time, reflecting the apperance of certain phenomena. In the area of neuroscience, this approach will be instrumental for detecting change points at which electroencephalograms reflects known "neuroscientific" behaviors.

1.1 Signature of rough paths (in a nutshell)

Consider a d-dimensionnal path $X = (X^1, ..., X^d) : \mathbb{R} \to \mathbb{R}^2$. In the following, we will note $S_I^{(k)}(X)$ the k-th degree signature applied to X on an time interval $I = [a, b]$. The **signature** of X is given by the tensors sequence $S_I^{(k)}(X) \in \mathbb{R}^{\overbrace{d \times d \times \cdots \times d}^{k \text{ times}}}$ for all $k \in \mathbb{N}$. We will use the notion of truncated signature of order K to define the sequence of signature tensor $S_I^{(k)}, k \leq K$. These tensors are given, for $\{i_1, ..., i_k\} \subset \{1, ..., k\}$ by :

$$\left(S_I^{(k)}(X) \right)_{i_1, ..., i_k} = S_I^{i_1, ..., i_k}(X) = \int \cdots \int_{a < t_1 < t_2 < \cdots < t_k < b} dX_{t_1}^{i_1} dX_{t_2}^{i_2} \ldots dX_{t_k}^{i_k}$$

The signature of a rough path serves as a powerful geometric feature extractor. To compute the signature of a multidimensional signal that is discretely sampled, we must consider linear interpolation between each consecutive time point, thus invoking Chen's theorem :

Theorem 1.1 (Chen's identity)**.** *Consider* $X : [a, b] \to \mathbb{R}^d$ *and* $Y : [b, c] \to \mathbb{R}^d$. *Define :*

$$X * Y = \begin{cases} X_t, & si \ t \in [a, b] \\ Y_t - Y_0 + X_b, & si \ t \in [b, c] \end{cases}$$

as the **concatenation** *of* X *and* Y*. Then :*

$$S_I^{(k)}(X * Y) = S_I^{(k)}(X) \otimes S_I^{(k)}(Y)$$

To center/rescale the signals, one has to consider the following properties

— For any constant $\gamma \in \mathbb{R}^d$, $S_I^{(k)}(X + \gamma) = S_I^{(k)}(X)$

— For any constant $\lambda \in \mathbb{R}$, $S_i^{(k)}(\lambda X) = \lambda^k S_I^{(k)}(X)$

Furthermore, to ensure a "pseudo" unicity (because it is computationally infeasible to create the whole signature) we will consider the next result :

Theorem 1.2. *If* $\exists i \in \{1, ..., k\}$ *such that* X^i *strictly monotonic, then* $S_I(X)$ *uniquely defines* X.

Finally, defining $\mathcal{C}_{\mathrm{mon}}^{1-var}(I, \mathbb{R}^d)$ the space of d-dimensionnal signals (linearly interpolated on a subdivision $D(I)$ of I) with $|X|_{1-var} = \sup_{t_i \in D(I)} \sum_i d(X_{t_i}, X_{t_{i+1}}) < \infty$ and such that $\exists i \in \{1, \dots, d\}$ with X^i strictly monotonic, then the signature transform $S_I : \mathcal{C}_{\mathrm{mon}}^{1-var}(I, \mathbb{R}^d) \to S_I(\mathcal{C}_{\mathrm{mon}}^{1-var}(I, \mathbb{R}^d))$ is an homeomorphism. [7]

1.2   Simplicial complex

Consider a set of d vertices $V = \{X^1, ..., X^d\}$.

Definition 1.1.   — *For* $k < d$*, a* k-simplex σ_k *is a* $n + 1$ *set and all its subsets [14].*
   — *A geometric realization of* σ_k *is given by the convex hull* E_k *of its* $n+1$ *points (eventually embedded) in* $\mathbb{R}^d$ *such that* $\dim(E_k) = k$ *(so* $d \geq k$*)*

With this tool, one can build a new structure on V[1].

Definition 1.2. *An abstract simplicial complex* $\mathcal{C}$ *on a finite vertex set* V *is a collection* $\{\sigma_k, k < d\}$ *of simplices such that :*

   — $v \in \mathcal{C}$ *if* $v \in V$
   — $\tau \subseteq \sigma \in \mathcal{C} \Rightarrow \tau \in \mathcal{C}$

The **dimension** *of* $\mathcal{C}$ *is the highest* k *such there exists a* k-simplex *in* $\mathcal{C}$*.*

A simplex $\sigma \in \mathcal{C}$ is called a **face** of $\mathcal{C}$. The upcoming definition is the most important for our algorithm :

Definition 1.3. *Consider* $k \in N^*$ *and* σ *a* k-simplex *in a simplicial complex* $\mathcal{C}$*. Its* **link** $Lk(\sigma, \mathcal{C})$ *in* $\mathcal{C}$ *is the set of all faces* $\tau \subset \mathcal{C}$ *such that :*

3

— $\sigma \cap \tau = \varnothing$

— $\sigma \cup \tau$ *is a face of* $\mathcal{C}$

Remarks :

— The link of a vertex is called **neighborhood** in graph theory.

— In the following algorithm, for a vertex v and a fixed k, we build the subset of all the k-simplex in $\text{Lk}(v, \mathcal{C})$ that we call k-dimensional Link (for simplicity).

1.3 Some tools for simplicial complex analysis.

The main goal of this section is to introduce some useful topological invariants, the first of these being the **Betti numbers**[1]. Informally, each b_k count the number of $k+1$-dimensional holes in $\mathcal{C}$. In our experiments, we focus on 2-dimensional complexes, and thus only consider b_0 (the number of connected components) and b_1 (the number of 2-dimensional holes).

The second invariant that we used is the **Persistence Diagram**. To get a numerical summary of this, we compute the notion of **Persistence Entropy**[13, 4]. A more robust notion of Persistence Summary presented in [6] could be put to work, but the implementation raises much more complex issues that we will not address in the present work.

Definition 1.4. *Consider a k-dimensional simplicial complex $\mathcal{C}$. A **filtration** over $\mathcal{C}$ is a sequence $(\mathcal{C}_i)_{0 \leq i \leq n}$ of simplicial complexes such that :*

— *For all $1 \leq i \leq j \leq n$, $\mathcal{C} \subset \mathcal{C}_j$*

— $\mathcal{C}_n = \mathcal{C}$

A natural way of building a filtration is by defining an increasing (non necessary strictly) sequence of time of arrival $(b_i)_{1 \leq i \leq n}$ for each simplex in $\mathcal{C}$. In Cech complex, the time of arrival correspond to the radius r at wich the simplex appear in $\mathcal{C}$.

Definition 1.5. *Consider a simplicial complex $\mathcal{C}$ and his persistence diagram DP*

$$DP = \{[b_i, d_i), i \in \{1, ..., \#\mathcal{C}\}\}$$

with b_i and d_i respectively birth and death times for sub-simplex of $\mathcal{C}$.
Define

$$p_i = \frac{b_i - d_i}{\sum\limits_j b_j - d_j}$$

***Persistence entropy** (PE) is given by the Shannon entropy of $\{pi\}$:*

$$PE = -\sum_i p_i \log(p_i)$$

Simplistically , the persistence entropy measures the similarity (or dissimilarity) between closure speed of k-dimensional holes : high level PE indicates that all holes are filled at same speed, in opposition to a PE close to 0. Since a quickly filled hole is likely just noise, PE quantifies how many significatives sub-structures lies in $\mathcal{C}$.

Beginning with section 2, we will consider these structures and invariants as evolving as a function of time : for any timestamp $t \in I$, we will consider the simplicial complex $\mathcal{C}(t)$, its associated betti numbers $b_0(t), ..., b_n(t)$ and its related persistence entropy denoted by $PE(t)$.

Remark : many other features may be extracted from the filtration we introduce for $\mathcal{C}$. More will be studied in a long version of this contribution.

1.4   Our simplicial complex construction algorithm

1.4.1   The algorithm

Our algorithm build the simplicial complex on $X = \{X^1, ..., X^d\}$ by constructing the k-dimensional link (iteratively on k) of a channel X^i. Here is a simplified version of the algorithm [5].

Data: Fix the degree deg of signature and a computation interval I.
Result: The simplicial complex of interaction among each time series.
for k *from 1 to* $d-1$ **do**
    for i *from 1 to* d **do**
        Compute $S^{deg}(\tilde{X}^i)$ (where $\tilde{X}^i(t) = (t, X^i(t)))$.;
        For every word $i_1...i_k$ of $\{1, ..., d\} \setminus \{i\}$, compute $S_I^{deg}(\tilde{X}^{i_1...i_k})$ where
        $\tilde{X}^{i_1...i_k}(t) = (t, X^{i_1}, ..., X^{i_k})$;
        Predict $S_I^{deg}(\tilde{X}^i)$ from $\left\{ S_I^{deg}(\tilde{X}_{i_1...i_k}) \right\}_{i_1...i_k \in (\{1,...,d\} \setminus \{i\})^k}$ with LASSO;
        if $R^2 > 0,67$ **then**
           | Select all non-zero $\beta_{i_1...i_k}$
        else
           | Fix $\beta_{i_1...i_k} = 0$ for all $i_1...i_k$.
        end
        for *Every* $\beta_{i_1...i_k} \neq 0$ **do**
           | Create the simplex whose vertices are $\{X^i, X^{i_1}, ..., X^{i_k}\}$
        end
    end
end

Algorithm 1: Sequential construction of a simplicial complex on $(X^1, ..., X^d)$

1.4.2 The induced filtration :

In order to retrieve the persistence diagram, we need to construct a filtration and so birth time.

— Take all the weights $(\beta_\sigma)_{\sigma \in \mathcal{C}}$ attached to each simplex.
— Create $b_\sigma = 1 - \frac{\beta_\sigma}{\sum\limits_{\sigma \in \mathcal{C}} \beta_\sigma}$.
— For each $\tau, \sigma \in \mathcal{C}$ such that $\tau \subset \sigma$ and $b_\tau > b_\sigma$, then fix $b_\tau = b_\sigma$.

This (non necessarily strictly) increasing sequence ensures that a highly significative simplex appears early in the filtration. In the following section, we show how to exploit this filtration.

2 Empirical study

We performed a set of computational experiements using EEG signals from the CHB-MIT Scalp Database [8]. The main benefit of using real-world signals is to test the stability of our algorithm against a uncontrolled noise.

2.1 Importance of hyperparameter values

We will evaluate the impact of the choice of the main parameters based on a one-hour trajectory, averaged every second. The signals reflect the occurence of an epilepsy seizure which t lasts less than one minute (on the interval $[49', 51']$). We focus our exploration on the LASSO parameters and the sliding Window size for the computation of the Signature. For the sake of simplicity, we will specify the maximal dimension of $\mathcal{C}(t)$ as equal to 2.

LASSO parameters For every channel X^i, we solve the following minimization problem :

$$\min_{\beta \in \mathbb{R}} \|S_I^{deg}(\tilde{X}^i) - \sum_{j \neq i} \beta_j S_I^{deg}(\tilde{X}^j)\|_2^2 + \lambda_1 \sum_{j \neq i} |\beta_j| \tag{1}$$

LASSO allows us to select which signatures (and then by homeomorphism which X^j) lies in the neighbourhood of X^i. This selection depends entirely on λ_1. However, this minimization only creates the 1-dimensional neighbourhood of X^i. to creates the 2-dimensional NH, we solve :

$$\min_{\beta \in \mathbb{R}} \|S_I^{deg}(\tilde{X}^i) - \sum_{j_1, j_2 \neq i, j_1 \neq j_2 i} \beta_{j_1 j_2} S_I^{deg}(\tilde{X}^{j_1 j_2})\|_2^2 + \lambda_2 \sum_{j \neq i} |\beta_{j_1 j_2}| \tag{2}$$

This naturally brings a second parameter.

Experimental results not included in the present manuscript but available in the Appendix show that fixing λ_1 and varying λ_2 does not change b_1 nor the PE trajectory.

Then, fixing λ_2 and varying λ_1 does not give the same result : on the one hand, similarly to the precedent case, persistence entropy is not much impacted by how λ_1 varies. On the other hand, b_1 's trajectory becomes noisier when one enforces sparsity in the LASSO.

Sliding window size for computing the Signature : At each time t we build a simplicial complex $\mathcal{C}_t$, and for a path P defined on I, we compute $S^k_{[t-L,t]}(P)$ (under the condition $[t-L,t] \subset I$). Length L of I is then a hyperparameter that can impact the topological structure of $\mathcal{C}(t)$. First results show that choosing L is key to safe detection of the sought for patterns.

2.2 Experiments with multiple EEG trajectories.

In the proposed experiments, we focus on a multivariate EEG signal sampled during 15 consecutives hours, and that includes 3 seizures. The method's hyperparameters are specified as follow :

— $\lambda_1 = 1$ and $\lambda_2 = 1$

— $L = 50$.

The main goal is to retrieve pre-critical and/or critical behaviour based on the Betti numbers and the Persistence Entropy, our main topological invariants of interests. According to [11],

Pre-critical behaviour : are caracterized by "area partitioning" (loss of connectivity between neuronal areas, loss of synchrony). Our main hope is that this behaviour is reflected in the trajectory of b_1 : a loss of connectivity should result in poorly interconnected cortical areas, resulting in several stabilized 2-dimensional holes. An expected dynamics is then to observe the steady growth of b_1 in time. It would be relevant to extend the dimension to $k = 3$ in order to see if this fact is confirmed with $b_2(t)$. Another idea would be to caracterise every hole in $\mathcal{C}$ and track the stable ones among them.

Critical behaviour : Neuronal populations will abruptly synchronise throughout the duration of the crisis, leading to hypersynchrony before returning to a more 'local' level of synchrony.

In Figure 1 below, one observes the global behavior of b_1 and PE on 1 hour of time (second trajectory). We computed an average measure on a clean trajectory (the green line) for both b_1 and PE. The grey area represent the interval $[x(t) - \hat{\sigma}_x, x(t) + \hat{\sigma}_x$ with $x = b_1$ or PE and $\hat{\sigma}_x$ computed on $[t-h,t]$.
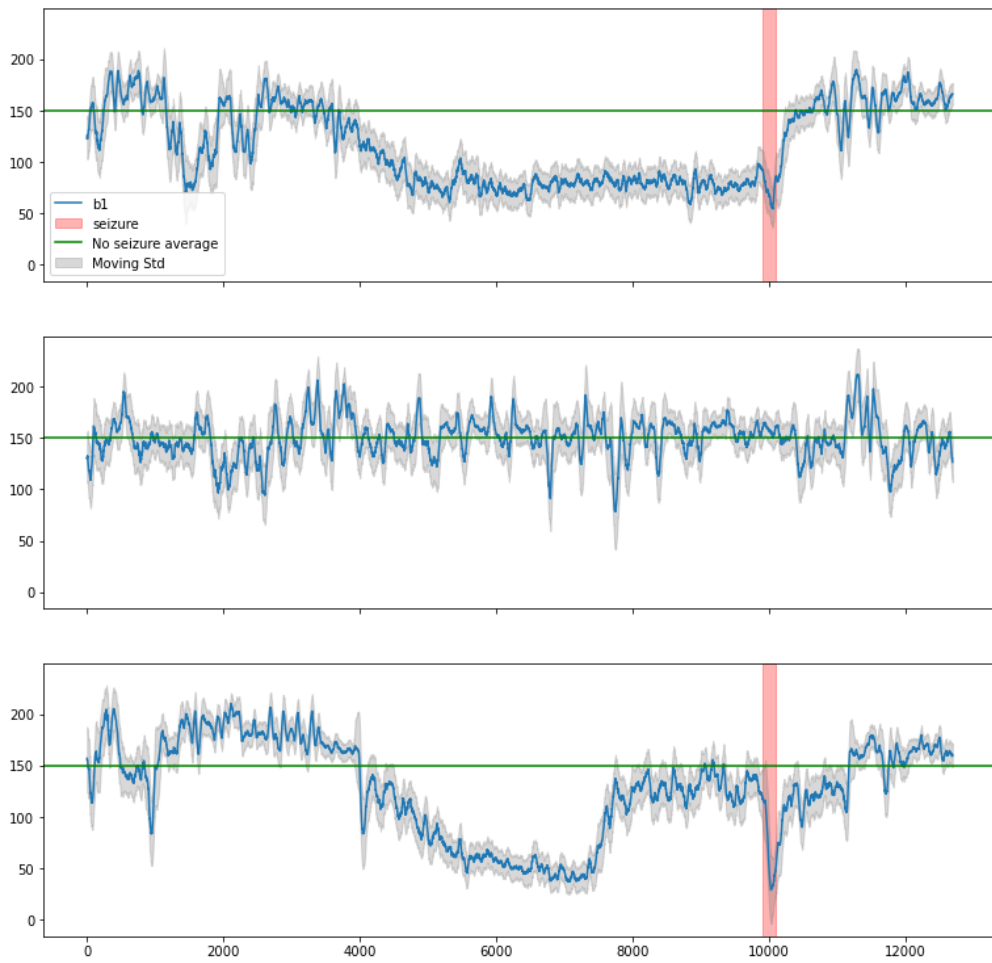
b1 time variation through different configuration

FIGURE 1 − $b_1(t)$ on 3 trajectories

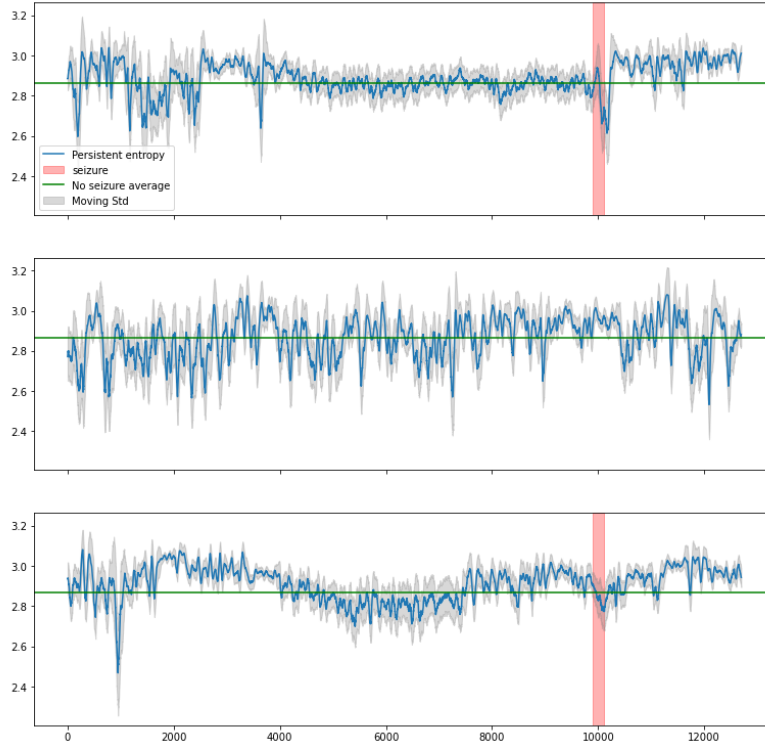Persistent entropy time variation through different configuration

FIGURE 2 – $PE(t)$ on 3 trajectories

General behavior (second trajectory) : We see a volatility around the average value along time, with small perturbations that we cannot explain. The volatility goes from 100 to 200 for b_1 and from 2.6 to 3 for PE.

Critical behavior (red area in first and third trajectory) : The critical behavior is caracterized by an abrupt diminution to 50 in b_1's trajectory (more connected structure in the simplicial complex before getting back to near the average value. Unfortunately, it cannot really be distinguished from PE.

Pre-critical behavior (6000 times before critical) : The precrital behavior is characterised by a steep decrease of b_1 (between 50 and 100) on the same time interval for the two seizure trajectories. An sudden increase in noticeable just before the next seizure, whilst remaining under the average value.

Heuristically, we can observe a decrease of the PE's standard deviation during the precrital stage. We plan to extend this study in a future work.

2.3   Discussion

Many improvements can still be made to the proposed methodology, in particular (i) using Knockoff filters to guarantee that the faces are selected with confidence, (ii) using SLOPE instead of the LASSO, (iii) using higher-order sub-complexes, etc. One may also use the path

$(b_1(t), PE(t))$ and its signature as well for extracting more sensitive detection pipelines.

Références

[1] Frédéric Chazal and Bertrand Michel. An introduction to topological data analysis : fundamental and practical aspects for data scientists. *Frontiers in artificial intelligence*, 4 :108, 2021.

[2] Kuo-Tsai Chen. Integration of paths, geometric invariants and a generalized baker-hausdorff formula. *Annals of Mathematics*, 65(1) :163–178, 1957.

[3] Ilya Chevyrev and Andrey Kormilitzin. A primer on the signature method in machine learning. *arXiv preprint arXiv :1603.03788*, 2016.

[4] Harish Chintakunta, Thanos Gentimis, Rocio Gonzalez-Diaz, Maria-Jose Jimenez, and Hamid Krim. An entropy-based persistence barcode. *Pattern Recognition*, 48(2) :391–401, 2015.

[5] Stéphane Chrétien, Ben Gao, Astrid Thébault Guiochon, and Rémi Vaucher. Leveraging the power of signatures for the construction of topological complexes for the analysis of multivariate complex dynamics.

[6] Yu-Min Chung and Austin Lawson. Persistence curves : A canonical framework for summarizing persistence diagrams. *Advances in Computational Mathematics*, 48(1) :6, 2022.

[7] Peter K Friz and Nicolas B Victoir. *Multidimensional stochastic processes as rough paths : theory and applications*, volume 120. Cambridge University Press, 2010.

[8] Ary L Goldberger, Luis AN Amaral, Leon Glass, Jeffrey M Hausdorff, Plamen Ch Ivanov, Roger G Mark, Joseph E Mietus, George B Moody, Chung-Kang Peng, and H Eugene Stanley. Physiobank, physiotoolkit, and physionet : components of a new research resource for complex physiologic signals. *circulation*, 101(23) :e215–e220, 2000.

[9] Terry Lyons and Zhongmin Qian. *System control and rough paths*. Oxford University Press, 2002.

[10] Nicolai Meinshausen and Peter Bühlmann. High-dimensional graphs and variable selection with the lasso. 2006.

[11] V Navarro, M Le Van Quyen, S Clemenceau, C Adam, C Petitmengin, F Dubeau, J Gotman, J Martinerie, and M Baulac. Seizure prediction : from myth to reality. *Revue Neurologique*, 167(3) :205–215, 2010.

[12] Giovanni Petri, Paul Expert, Federico Turkheimer, Robin Carhart-Harris, David Nutt, Peter J Hellyer, and Francesco Vaccarino. Homological scaffolds of brain functional networks. *Journal of The Royal Society Interface*, 11(101) :20140873, 2014.

[13] Matteo Rucco, Filippo Castiglione, Emanuela Merelli, and Marco Pettini. Characterisation of the idiotypic immune network through persistent entropy. In *Proceedings of ECCS 2014 : European Conference on Complex Systems*, pages 117–128. Springer, 2016.

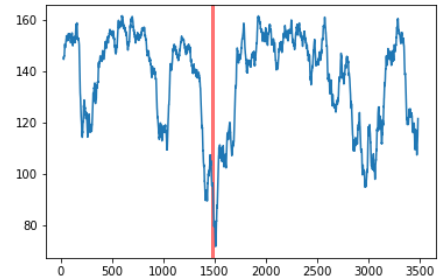[14] Hal Schenck. *Algebraic Foundations for Applied Topology and Data Analysis*, volume 1. Springer Nature, 2022.

A Appendix
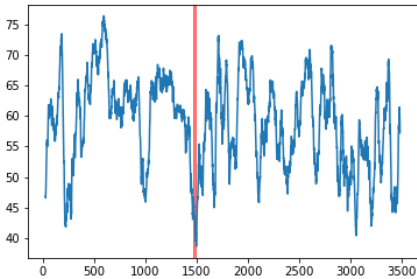
A.1 Impact of the hyperparameters for $b_1(t)$

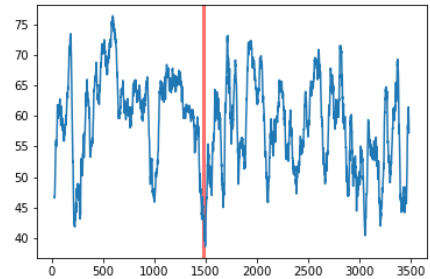### A.1.1 λ_i :
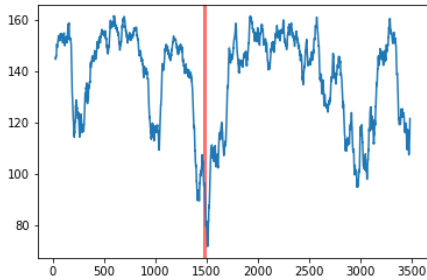
a) $\lambda_1 = \lambda_2 = 1$.

$\lambda_1 = 1,\ \lambda_2 = 10^4$
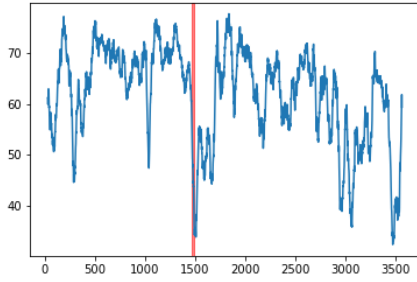
c) $\lambda_1 = 10^4,\ \lambda_2 = 1$
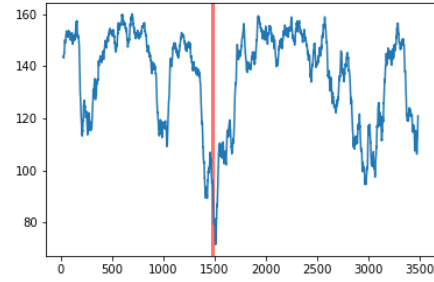
c) $\lambda_1 = 10^4,\ \lambda_2 = 10^4$

FIGURE 3 – Study of the effect of λ_1 and λ_2 on $b_1(t)$ (with a seizure in red)

.

As can be observed, when λ_2 increases, $b_1(t)$ remains the same on the full interval. On the other hand, a noisier trajectory results from changing the value of λ_1. One suspects that only the 1-dimensional edges that are not subsets of 2-dimensional faces (dis)appear as we move λ_i.

11

A.1.2 L :

a) $L = 20s$.

$L = 100s$
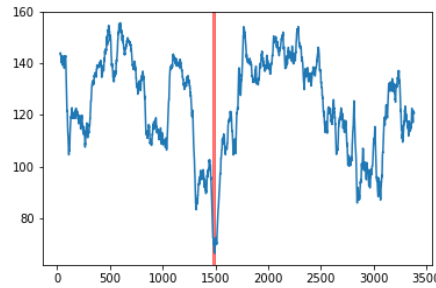
c) $L = 200s$

FIGURE 4 – Study of the influence of L on $b_1(t)$ (with a seizure in red)
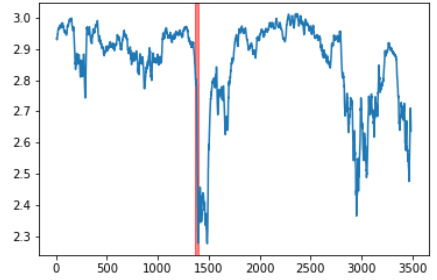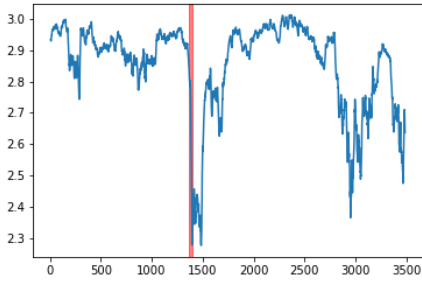
.

As can be expected from varying L, keeping L at a low level make the signal more noisy. Surprisingly, the amplitude changes as L is increased, implying that either more edges are created or less 2-dimensional faces are created (or both). This fact will be studied more precisely in a longer version of the paper.

A.2   Parameters study for $PE(t)$
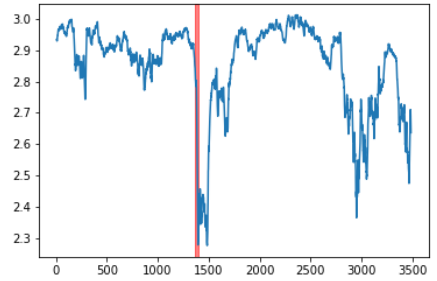
A.2.1 λ_i :
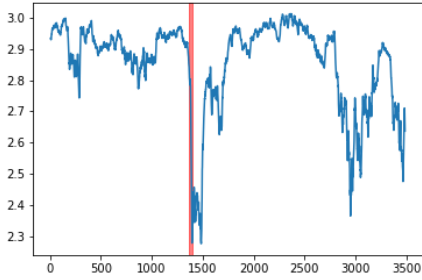
a) $\lambda_1 = \lambda_2 = 1$.

$\lambda_1 = 1, \ \lambda_2 = 10^4$
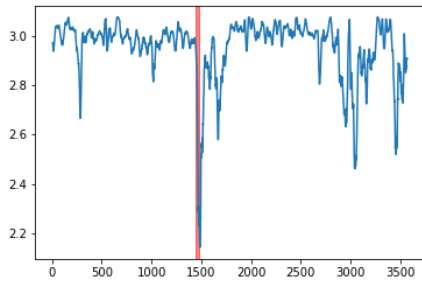
c) $\lambda_1 = 10^4, \ \lambda_2 = 1$
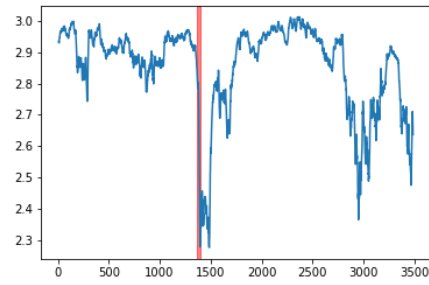
c) $\lambda_1 = 10^4, \ \lambda_2 = 10^4$

FIGURE 5 – Study of the effect of λ_1 and λ_2 on $PE(t)$ (with a seizure in red)
.

It is clear from the numerical experiments that persistence entropy is not impacted by varying λ_1. Related to the variations of b_1, one can deduce that the created edges did not influence persistence entropy.
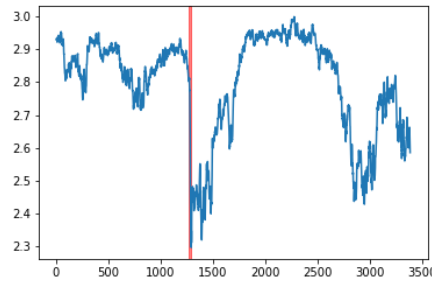
A.2.2 L :

a) $L = 20s$.

$L = 100s$

c) $L = 200s$

FIGURE 6 – Study of the effect of L on $PE(t)$ (with a seizure in red)

.

Clearly, lower level of the window size ensures the stability of persistent entropy. With this stability comes shorter events (like at $t = 300$ or $t = 3000$). Higher levels of window size brings longer events, and earlier distinguishable (as the small event at $t = 500$ show)