Severe Slugging Flow Identification from Topological Indicators

Simone Casolo*
Cognite, Oksenøyveien 10, 1366 Lysaker, Norway
NorwAI, Norwegian Research Center for AI Innovation and

(Dated: January 6, 2023)

In this work, topological data analysis is used to identify the onset of severe slug flow in offshore petroleum production systems. Severe slugging is a multiphase flow regime known to be very inefficient and potentially harmful to process equipment and it is characterized by large oscillations in the production fluid pressure. Time series from pressure sensors in subsea oil wells are processed by means of Takens embedding to produce point clouds of data. Embedded sensor data is then analyzed using persistent homology to obtain topological indicators capable of revealing the occurrence of severe slugging in a condition-based monitoring approach. A large data set of well events consisting of both real and simulated data is used to demonstrate the possibility of automatizing severe slugging detection from live data via topological data analysis. Methods based on persistence diagrams are shown to accurately identify severe slugging and to classify different flow regimes from pressure signals of producing wells with supervised machine learning.

I. INTRODUCTION

Multiphase flow is a complex phenomenon in chemical engineering and it is still very challenging to fully understand. In particular, gas-liquid multiphase flow is ubiquitous in the chemical industry e.g. in heat exchangers or in process industries, where fluids are transported through piping systems to process facilities. In petroleum engineering mutliphase flow is ubiquitous [1]. Gas wells may also produce condensate and water while oil wells may also produce gas, water and suspended solids. Gas can also be re-injected downhole to reduce the fluid density and viscosity, increasing production. In certain process conditions, a two-phase flow can transition to more complex patterns where large bubbles form in the liquid phase [2]. As the flow velocity increases, bubbles can then coalesce into trains of large bullet-shape gas pockets (Taylor bubbles) that occupy almost the entire piping cross sectional area [3], separated by slugs of liquid. This flow regime is known as hydrodynamic slugging or slug flow [4, 5].

Differently from slug flow, severe slugging has a long liquid slug body that might be longer than the riser length [6]. Severe slugging in offshore petroleum production systems can be triggered at locations where the pipe inclination changes from a downward angle, such as in the pipeline, followed by an upward angle or a vertical pipe, such as in risers. These are pipes connecting subsea wells to a topside processing facility and have a J-shape due to seabed topography. The phenomenon is triggered when the liquid produced by the wells accumulates at the bottom of the riser, blocking the gas passage and compressing the incoming gas at the pipeline (slug formation). As the liquid accumulates, its level increases to reach the top of the riser while the gas pressure reaches a maximum and there is only liquid flowing in the riser (slug production). The liquid accumulation front in the pipeline is then pushed ahead until it reaches the bottom of the riser, leading to the gas penetration into the riser. The fluid column becomes lighter, decreasing the pressure and then increasing the gas flow. When gas reaches the top of the riser its passage is then free (blow down stage), causing a violent expulsion and a rapid decompression that brings the process to slug formation again (Figure 1, [7, 8]).

Different types of severe slugging were defined depending on the flow conditions in the pipeline-riser system [9]: (SS1) Severe slugging 1 is characterized by a liquid slug length greater to or equal to one riser length, (SS2) Severe slugging 2 has a liquid slug length shorter than one riser length, and (SS3) Severe slugging 3 occurs when there is continuous gas penetration at the bottom of the riser, leading to cyclic variations of smaller periods and amplitudes when compared to SS1 and SS2.

Severe slugging leads to alternating long liquid slugs followed by high gas rates, causing poor phase separation, loss of production and large vibrations which may induce fatigue load on the processing facilities [10]. Given such detrimental effects on production and operation, a great deal of research was dedicated to predicting its occurrence. Schmidt et al. [11, 12] performed an experimental and modeling study of the phenomenon and proposed a prediction criterion based on the flow pattern in the pipeline and flow stability in the riser. Bøe [13] and Pots et al. [14] proposed a stability criterion based on a force balance between the hydrostatic pressure in the riser and the gas pressure in the pipeline. The Taitel [15] criterion is instead based on a similar force balance with the inclusion of a gas perturbation stability criterion. The prediction of the severe slugging phenomenon was also explored by using simplified transient multiphase flow models, such as the ones proposed by Fabre et al. [16], Sarica and Shoham [17], Baliño et al. [7], and Nemoto and Baliño [18].

* simone.casolo@cognite.com, ORCID: 0000-0003-2945-8931

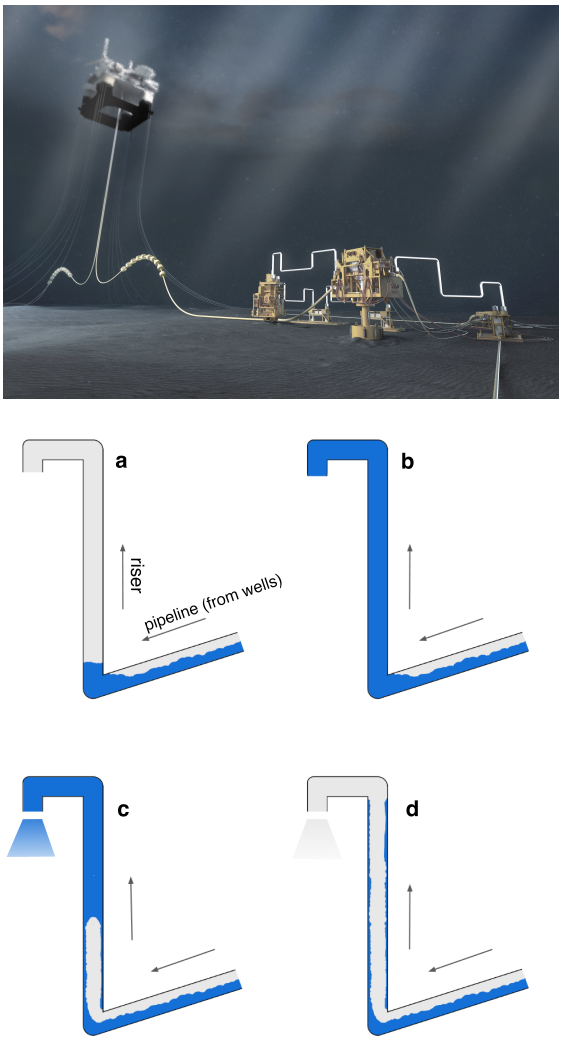

FIG. 1. Top: J-riser connecting an offshore floating topside installation to a subsea well. Image courtesy of Aker Solutions AS. Bottom: mechanism of slug formation in vertical risers. *a*: blockage at the riser base and slug formation, *b*: slug growth and movement towards the separator, *c*: blowout and fast liquid production; *d*: gas production and liquid fallback.

The literature reports different techniques to mitigate severe slugging such as (a) back-pressure increase, represented by an increase in the separator pressure, as reported by Yocum [8], Schmidt [12], and Baliño [7], (b) choking, i.e. reducing the choke valve opening, as reported by Jansen [19], Yocum [8], and Schmidt [11], (c) gas lift at the base of the riser, as reported by Jansen [19] and Schmidt [12], and (d) subsea separation, forcing gas and liquid to be transported in separate flowlines, as reported by Song and Kouba [20]. More recently, different authors (Pedersen [5], Havre and Dalsmo [21], Pedersen [22]) have proposed the use of control techniques, such as active feedback control of the topside choke valve, to avoid flow instabilities while maintaining desirable oil production. Di Meglio [23] proposes different control strategies to attenuate slugging, classifying them as model-free strategies, based on the implementation of, for instance, PI-controllers using process variables, and model-based strategies, which utilizes simplified flow models in association with control techniques. Different simplified models were proposed for vertical risers, such as the models from Storkaas and Skogestad [24], Tuvnes [25], Kaasa [26], Silva and Nydal [27] and Jahanshahi and Skogestad [28, 29], and for S-shaped risers, such as the model proposed by Nemoto [30].

In this work we apply topological data analysis (TDA), to identify and characterize instabilities in the flow of offshore oil and gas wells. This approach is entirely based on signal analysis from well's pressure sensor data and does not require information on the industrial asset (pipe height, shape, etc.) fluid characteristics nor thermodynamics.

Recently, TDA has been successfully used in physics [31–33], biosciences [34], astronomy [35], and chemical engineering [36], as a powerful tool for detecting emerging patterns in data and physical systems. This method is based on an embedding of process time series and on their analysis by means of persistent homology, a branch of algebraic topology. Time series are transformed into point clouds in a high-dimensional space where the information of the physical process is encoded onto their shape and extracted via persistent homology. We will show how this approach can be used for condition monitoring of well events by detecting the signatures of production instabilities as they appear in both real and simulated offshore sensors' data and for their automatic classification via simple and explainable machine learning methods. In this work we are going to focus on presenting an intuitive picture of the methodology without focusing on formalism or mathematical details and rather focusing on its applications. A more formal and comprehensive introduction to the subject can be found in references [36] and [37, 38].

II. THEORETICAL METHODS

A. Time series embedding

The evolution of a physical dynamical system, such as that of a multiphase fluid, depends on a large set of parameters (fluid characteristics, reservoir and gas lift pressures, choke valve openings, etc.) and can be fully described by a multivariate time series $\mathbf{x}(t) \in \mathbb{R}^N$, *i.e.* a trajectory in an N-dimensional phase space. Here we assume that the topology of the phase space contains all the information of the physical system and, by having access to that, one could forecast the dynamics of the physical system. For a deterministic dynamic system, by knowing all the system parameters and their derivatives their evolution could be computed exactly. Unfortunately, most of these N parameters are hidden or unknown in practical situations as meters and sensors in the field measure only a few process indicators (temperature,

pressures, flow rates), each effectively mapping $\mathbf{x}(t)$ onto each sensor. The sensors' time series can be considered as sampling of the phase space and its analysis is then a proxy to access the space's topology *i.e.* its physical properties.

Here we assume that the trajectory in phase space is evolving smoothly in time and that it is confined to a manifold $\mathcal{M} \subset \mathbb{R}^N$ known as an attractor, where $\dim(\mathcal{M}) < N$. The trajectory's evolution can then be written in terms of finite time intervals τ by a flow function ϕ at a given time t, $\phi : \mathcal{M} \times \mathbb{R} \to \mathcal{M}$ such as:

$$\mathbf{x}(t+\tau) = \phi(\mathbf{x}(t), \tau) = \phi_\tau(\mathbf{x}(t)) \tag{1}$$

and for $k \in \mathbb{N}$, $\mathbf{x}(t+k\tau) = \phi_\tau^k(\mathbf{x}(t))$.

The measurement process of the physical system, can be represented by the measurement function $\Phi : \mathbb{R}^N \to \mathbb{R}$ applied at regular intervals τ to generate the sensor time series $\{x(t_i)\}_{i\in\mathbb{N}} = \{\Phi(\mathbf{x}(i\cdot\tau))\}_i$. In order to study the full dynamical system is necessary to "reconstruct" the trajectory $\mathbf{x}(t)$ given only the measured time series $\{x(t_i)\}$. This can be achieved by a time-delay embedding:

$$F(\mathbf{x}(t)) = \begin{bmatrix} x(t_i) \\ x(t_{i+1}) \\ \vdots \\ x(t_{i+d}) \end{bmatrix} = \begin{bmatrix} \Phi(\mathbf{x}(i\tau)) \\ \Phi(\mathbf{x}((i+1)\tau)) \\ \vdots \\ \Phi(\mathbf{x}((i+d)\tau) \end{bmatrix} = \tag{2}$$

$$= \begin{bmatrix} \Phi(\mathbf{x}(i\tau)) \\ \Phi(\phi_\tau^1(\mathbf{x}(i\tau))) \\ \vdots \\ \Phi(\phi_\tau^d(\mathbf{x}(i\tau))) \end{bmatrix}$$

Here, $F(\mathbf{x}(t))$ is a delay coordinate map from $\mathcal{M}$ to the so-called reconstruction space $\mathbb{R}^{d+1}$ build by stacking repeated measurements of the system. Takens theorem [39] sets the sufficient conditions on the spaces and the measurement function for which $F(\mathbf{x}(t))$ is diffeomorphic to $\mathcal{M}$, *i.e.* a trajectory reconstruction space preserves the topology of the dynamical system $\mathbf{x}(t) \in \mathcal{M}$.

In practice, the embedding of the sensor data is done by choosing the embedding dimension d and the time delay τ with which to sample the time series d times. The product $d\tau$ can be seen as a window in time where each time-delayed sampled value in the window becomes the element of a vector, $\{x(t), x(t+\tau), \ldots, x(t+d\tau)\}$. By sliding this window forward in time the procedure generates a series of vectors and therefore a cloud of points following a trajectory in $\mathbb{R}^{d+1}$.

The shape of the point cloud depends on the sliding window size and, in turn, the embedding dimension d and the time delay τ have to be chosen correctly in order to guarantee a faithful embedding of the time series and avoid non-physical situations such as trajectory crossings. In this study, the optimal time delay was chosen by determining the first minimum of the mutual information function which is a measure of how much information consecutive points share [40]. The embedding dimension was computed instead via the false nearest neighbor algorithm of Abarbanel, Kennel and Brown [41].

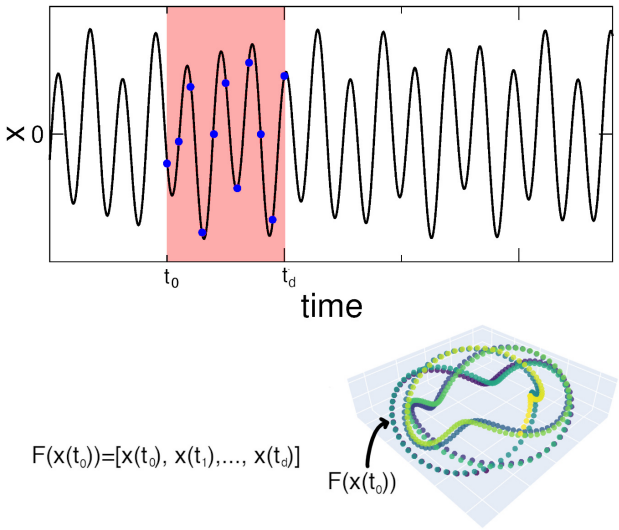

FIG. 2. Top: quasi-periodic signal undergoing Takens embedding in the window shaded in red in the interval (t_0, t_d=t_0+$d\tau$) to produce one point $F(\mathbf{x}(t_0))$ in a $d+1$-dimensional space. The embedding relies on sampling $d+1$ points spaced by τ time. Bottom: the point cloud obtained by sampling repeatedly the signal by sliding the sampling window ahead of one time step.

B. Simplicial Complexes and Filtrations

For studying properties of the shape of a point cloud of data one needs a set of tools for extracting topological invariants from the shape. These tools are defined in the discipline of algebraic topology, which defines a set of procedures to translate geometrical properties of a shape (boundaries, holes, etc.) into linear algebra. Topology, *per se* does not apply to sets of points, rather to continuous shapes. Considering the point cloud obtained from the time series embedding as a discrete sampling of an underlying continuous topological space, it is then possible to build a combinatorial representation of such space based on simplicial complexes.

Connected points in space may form simplices: a k-simplex is defined as the convex hull of the set $\mathcal{S}_k = \{x_1, x_2, ..., x_{k+1}\}$ of independent points. For example, a single point in space is a 0-simplex, two points connected to each other make a 1-simplex, three connected points forming a filled triangle make a 2-simplex, four points forming a filled tetrahedron make a 3-simplex, etc. generalized to the smallest solid shape with $k+1$ vertices in k dimensions. In terms of graphs, a simplex is made of a set of $k+1$ vertices building a complete graph where each pair of vertices is connected by an edge. When selecting a point $x_i \in \mathcal{S}$, its complement $\mathcal{S}/x_i$ form a *face* of the

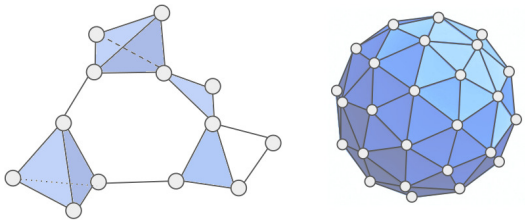

FIG. 3. Left: a three dimensional simplicial complex consisting of two tetrahedra (3-d simplices), two triangles (2-d simplices) and thirteen connected points (0-d simplices). The complex shows thirteen topological features (generators) of the homology class H_0 and two to H_1. Right: a triangulated sphere as an example of two-dimensional simplicial complex.

simplex and the union of the faces defines the boundary of $\mathcal{S}_k$, denoted by $\partial_k \mathcal{S}_k$ where ∂_k is known as the boundary operator. It is easy to visualize how a simplex could be continuously deformed to a circle or a ball; indeed a k-complex is homeomorphic to the k-ball B^k, while its boundary is homeomorphic to the k-sphere[42].

A *simplicial complex* $\mathcal{K}$ is a set of simplices such as every face of a simplex part of $\mathcal{K}$ is also in $\mathcal{K}$ and where the intersection of any two simplices in $\mathcal{K}$ is either empty or a simplex of dimension lower or equal to the highest dimensional simplex they share. In other words, a 3-dimensional simplicial complex is a set of simplices stuck together to include at least one tetrahedron or possibly more sharing faces or a vertex, but cannot contain any 4-dimensional simplex. Simplicial complexes are then a tool for encoding a continuous topological space into a combinatorial object (a graph), that can then be analyzed by linear algebra.

The construction of the complex for a given point cloud is performed via a *filtration* process. Given a point cloud of data, it is then possible to build a geometrical object made of complexes, but these are not uniquely defined as the rationale for which points to connect or not; it is still arbitrary. Suppose the set of points $\mathcal{S}_k$ is part of a metric space and choose a real number $\alpha \geq 0$. The Vietoris-Rips complex $\alpha(\mathcal{S}_k)$ is the set of simplices $\{x_1, x_2, ..., x_{k+1}\}$ such that the Gromov-Hausdorff distance $d_X(x_i, x_j) \leq \alpha$ for all pairs of points $\{x_i, x_j\}$. In other words, by building a closed ball of radius $d_X/2$ around each point, the d-simplices are built by connecting points where balls have pairwise intersections. By continuously extending the radius of the balls from zero to a given cutoff value one obtains a nested family of complexes known as a filtration [43]. Other criteria are available for the filtration such as the Čech complex, defined as the set of simplices such that the closed balls have a non-empty intersection. For both these definitions, it can be proven (nerve theorem) that the simplices obtained from the filtration share the same topological features of the underlying vector space of points (*i.e.* they are homotopy equivalent) [44].

C. Simplicial Homology

Homology theory is a concept in algebraic topology useful to study the topological properties of an object such as the number of connected components, the number of holes, etc. by means of algebra and relating them to Abelian groups. For any dimension k, the k-homology group H_k is a vector space with dimension equal to the number of k-dimensional holes. Therefore, given a point cloud constituting a vector space, connected via a filtration to form simplicial complexes, the zero-dimensional homology group H_0 represents the connected components of the complex, H_1 represents the one-dimensional loops, the H_2 represents the two-dimensional holes, etc.

A k-chain c_k is defined as the set of all k-simplices in $\mathcal{K}$ and can be written as:

$$c_k = \sum_{i=1}^{p} \epsilon_i x_i \quad \epsilon_i \in \mathbb{Z}_2 \tag{3}$$

all the elements c_k form a group $C_k(\mathcal{K})$, the chain group, a vector space with coefficients in $\mathbb{Z}_2$ (set of integers modulo 2). It can be shown that the boundary of the chain group is the $k-1$ chain group $\partial_k C_k(\mathcal{K}) = C_{k-1}(\mathcal{K})$. In other words, C_0 is the group consisting of linear combination of the complex's vertices (a zero-chain), C_1 consists of linear combination of the edges (1-chain), C_2 of the faces (2-chain), etc. In this framework, the boundary operator is a homomorphism mapping $C_k \to C_{k-1}$ and we can define a *cycle* t as an element of C_k with zero boundary $t \in C_k : \partial_k(t) = 0$, *i.e.* the group $Z_k = \ker \partial_k \subset C_k$ is the group of cycles, which are boundary-less. The group obtained from the image of the boundary operator $B_k = \mathrm{im}\partial_{k+1}$ consists instead of the boundaries of simplicies of dimension $k+1$. This implies that repeated application of the boundary operator eventually leads to computing the boundary a one-dimensional chain (edge), which is necessarily zero.

$$\ldots \xrightarrow{\partial_{k+2}} C_{k+1} \xrightarrow{\partial_{k+1}} C_k \xrightarrow{\partial_k} C_{k-1} \ldots \xrightarrow{\partial_1} C_0 \xrightarrow{\partial_0} 0 \tag{4}$$

Note that for a k-dimensional simplicial complex $C_n = 0, \forall n > k$. The homology group H_k is then defined as the quotient between the group of cycle and boundary groups of cycles

$$H_k(\mathcal{K}) = Z_k(\mathcal{K})/B_{k+1}(\mathcal{K}) \tag{5}$$

and identifies cycles in a cycle group by dividing them in classes that differ by a boundary. An element in H_k is known as a homology class.

At its core, the homology group a formal way to identify the group of cycles that enclose a hole; whenever cycles do not bound higher-dimensional simplices, we have a hole. Holes are cycles that are not boundaries at the same time. As an example, take a triangle which is the face of a filled tetrahedron: this is a two-dimensional cycle that is also the boundary of the three-dimensional simplex ($H_2 = 0$).

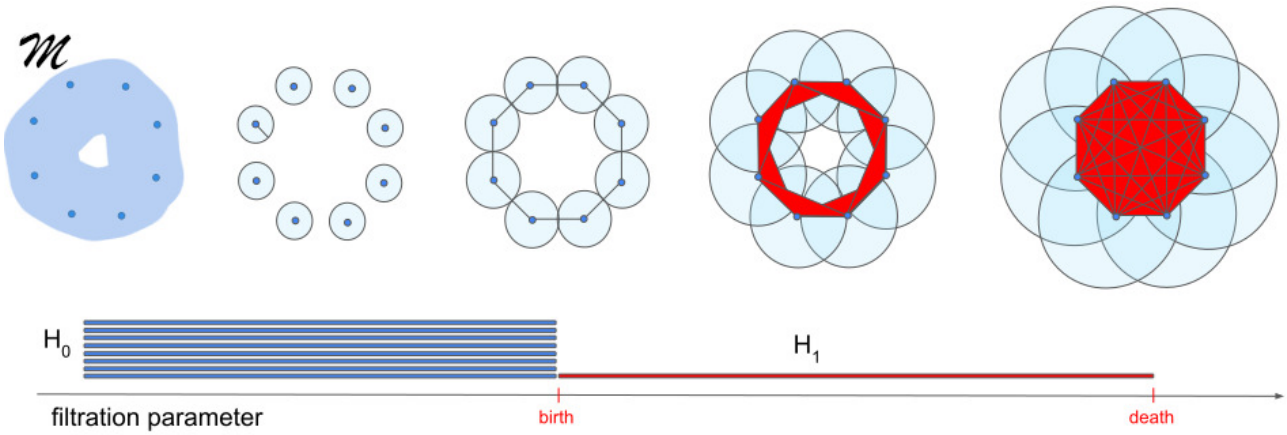

FIG. 4. Sketch of the filtration process with the associated barcode diagram, showing the birth and death of topological features. Around each of the eight points sample the manifold $\mathcal{M}$ Vietoris-Rips complexes are built from the intersection of progressively expanding ball with radius ϵ (filtration parameter). Zero-dimensional components corresponding to the homology class H_0 are 'born' at $\epsilon = 0$ and 'die' at the value at which a one dimensional H_1 loop is born. The loop has a death value when balls intersect each other to fill the loop to a disk.

An empty tetrahedron (or a triangulated 2-sphere) is instead a two-dimensional simplicial complex, and in this case the triangle encloses a hole ($H_2 = \mathbb{Z}$). Spaces with the same homology group have the same amount of holes (are homotopy equivalent) hence share the same topology. When the point cloud is the sampling of a phase space associated with a physical process, the hope (often justified by Niyogi-Smale-Weinberger theorem [45]) is that if X is a (slightly) noisy sample from a submanifold $\mathcal{M} \subset \mathbb{R}^N$, the union of the balls around points of X would have the same topology as the phase space. This makes it possible to perform inference on the phase space and, in turn, on the physical process.
The rank of the homology group is known as the Betti number (β_k) and "counts" the number of k-dimensional holes of a space: β_0 represents connected components, β_1 circles, β_2 voids, etc. As an example, for a two dimensional circle the set of Betti numbers $\{\beta_0, \beta_1, \beta_2\}$ are $\{1, 1, 0\}$, for a filled disk $\{1, 0, 0\}$, a hollow sphere $\{1, 0, 1\}$, for a filled ball $\{1, 0, 0\}$, for a torus $\{1, 2, 1\}$, etc.

D. Persistent Homology

Persistent homology is a technique used to show the evolution of the homology of a point cloud through a filtration. As the filtration parameter grows, the points get connected to form a nested sequence of Vietoris-Rips complexes, each of them having a corresponding homology class for each dimension.

$$\mathcal{K}_1 \subseteq \mathcal{K}_2 \subseteq \cdots \subseteq \mathcal{K}_n \Rightarrow$$
$$\mathrm{H}_k(\mathcal{K}_1), \mathrm{H}_k(\mathcal{K}_2), \ldots, \mathrm{H}_k(\mathcal{K}_n) \tag{6}$$

Along the filtration, loops and holes form only at given filtration parameters, and merge to larger structures further along the process. In this context a homology class is said to be born at the point b if it is not part of the map $\mathrm{H}_k(\mathcal{K}_{b-1}) \to \mathrm{H}_k(\mathcal{K}_b)$, and it is said to die at the point d if it is found in the map $\mathrm{H}_k(\mathcal{K}_d)$, but not in $\mathrm{H}_k(\mathcal{K}_{d+1})$. In this way, each homology class has associated a birth and a death, $\{b, d\} \in \mathbb{R}^2$, that can be represented in the so-called persistence diagram, showing the persistence (defined as $d - b$) for classes in each dimension. Persistence diagrams are then defined only above the diagonal axis as $b > d$, and the further a point in the diagram is from the diagonal, the more persistent is the topological feature along the filtration. Low persistence points are then associated with short-lived holes, *i.e.* noise.
Point clouds and persistence diagram from homology classes of dimension $k = \{0, 1\}$ for both a white noise time series and a $\cos(t)$ signal are shown in Figure 5. Zero-dimensional homology classes, H_0, correspond to connected components, that are therefore born at the beginning of the filtration and cluster at the edge of the diagram. One-dimensional classes, H_1 represent loops (one-dimensional holes), and have a very low persistence for the noise sample, as the point cloud does not contain any real loops. For the trigonometric signal, it can be shown (see Appendix A) that the embedding of a periodic signal produces ellipses in the point cloud. The loop size depends on the scale of the signal, therefore, high-persistence H_1 homology classes are expected when analyzing time series with periodic or quasi-periodic components.

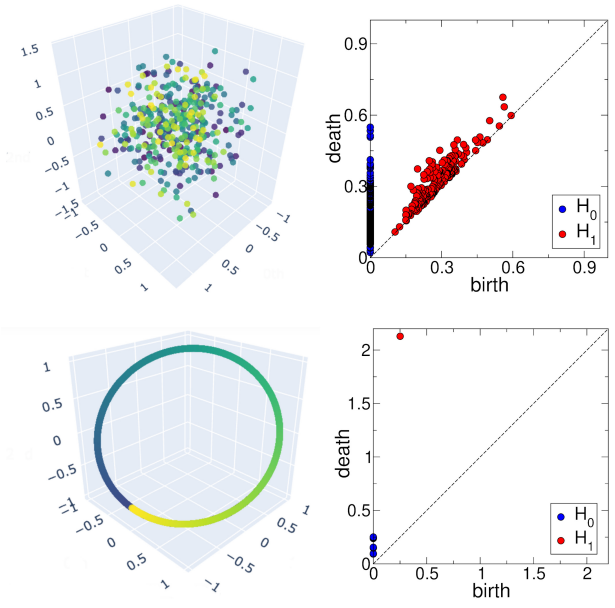

FIG. 5. Top: point cloud from Takens embedding of a white noise time series with standard deviation $\sigma = 0.5$ (left) and its corresponding persistent diagram based on Vietoris-Rips complexes (right). Bottom: same for the time series $\cos(t)$. Colors in the point clouds represent the starting time for them embedding of each point. Note the one-dimensional loop resulting from the embedding of $\cos(t)$ is a generator of the corresponding high persistence H_1 point in the diagram, while noise gives rise to points with low persistence, clustering close to the diagonal.

III. COMPUTATIONAL TOOLS AND DATA SETS

Embeddings and topological data analysis and persistent homology in this work were computed with the packages Giotto-TDA [46] and Scikit-TDA [47], while machine learning classification were performed with the Scikit-Learn library [48]. In order to visualize high-dimensional point clouds, the set of points is then processed by principal component analysis (PCA) and only the three highest contributions are displayed.

The data set considered in this study consists of two main sources. The condition monitoring study used two time series consisting of four days of continuous data measured at a 10 s resolutions (34560 data points), collected for a subsea oil and gas production asset in the North Sea portion of the Norwegian continental shelf. The data consist of: downhole pressure, downhole temperature and flow-line pressure measured for a single producing well. The data set has been anonymized.

The data set used for the machine learning automatic classification of slugging events was taken from a public repository known as 3W data set: a combination of more than 2000 multi-variate time series, either real or simulated with the OLGA software [49], of undesired rare well events collected from 21 wells operated by Petrobras offshore Brazil between 2012 and 2019 [50]. This work utilizes only the 3W time series labeled as "normal", "severe slugging" and "flow instabilities" (1040 total time series), consisting of downhole pressure (P-PDG), wellhead temperature (T-TPT) and wellhead pressure (P-TPT), available at a 1 s resolution. Severe slugging is typically classified in type 1, type 2 and type 3, depending on ratio between the liquid slug length and the riser length [9]. However the 3W dataset does not differentiates between these three regimes, labelling the data as (a) normal, with steady stable flow, (b) flow instabilities, in which at least one of the monitored variables undergoes relevant oscillations but with tolerable amplitudes, and (c) severe slugging, a critical type of instability with well-defined periodicity (around 30, 45, or 60 min) and amplitudes sufficiently high to be detected by sensors along the entire production line. Albeit TDA likely to classify type 1, 2 and 3 severe slugging, these should be verified using a dedicated and balanced data set: therefore this work will follow the simplified 3W classification. Behavior labeled as "flow instabilities" is intended in 3W as minor slugging events in which the multiphase flow shows jittering pressure oscillations with amplitude below 1 bar and frequency is in the order of seconds or few minutes, in contrast to severe slugging in with oscillations of several bars and frequency of tens of minutes. The 3W time series was pre-processed to remove time series with pressure differences between the initial and final 30 minutes average value higher than 5 bar, to remove transient flow situations. In addition, signals were down-sampled to restrict the analysis to less than 3000 points, in order to keep the computational cost to a manageable level. No outlier or individual point was removed from the sensors' signal.

IV. TOPOLOGICAL INDICATORS

The shape of data can be inferred from the persistence diagram (D) and represented by a series of topological indicators. Each of the points in the diagram is defined by its birth (b) and death (d) coordinate. A measure of the complexity and size of the topological features of H_k is the p-norm, defined as:

$$\mathcal{P}_p^{H_k}(D_{H_k}) = \left(\sum_{\{b,d\} \in D} |d - b|^p \right)^{1/p} \tag{7}$$

In particular, the infinity norm $\mathcal{P}_\infty^{H_k}$ measures the largest lifetime (persistence) of a given feature. This is a useful indicators as noise gives rise to points in D with short lifetime, while relevant features of the points cloud (*e.g.* loops) are expected to have high persistence.

Another measure of complexity is the persistence entropy [51, 52], $\overline{E}_{H_k}(D)$, a measure of the distribution of points

along the diagram based on Shannon's entropy formula.

$$\overline{\mathrm{E}}_{\mathrm{H}_k}(D) = -\frac{1}{\log_2 \mathcal{S}(D)} \sum_{\{b,d\}\in D_{\mathrm{H}_k}} \frac{|d-b|}{\mathcal{S}(D)} \log_2 \left(\frac{|d-b|}{\mathcal{S}(D)} \right)$$

where

$$\mathcal{S}(D_{\mathrm{H}_k}) = \sum_{\{b,d\}\in D} |d-b| \tag{8}$$

Another typical topological indicator are Betti curves (β_k), which count the amount of k-dimensional topological features (*i.e.* the Betti number [53]) at each value of the filtration parameter ϵ. More formally, $\beta_k(\epsilon)$: $\mathbb{R} \to \mathbb{N}$ whose values on $\epsilon \in \mathbb{R}$ is the number of points $(b_i, d_i) \in D$ such as $b_i \leq \epsilon < d_i$. In formulae, for a given filtration parameter

$$\beta_k(\epsilon) = \sum_{(b,d)\in D} w(b,d)\mathbf{1}_{\epsilon\in D} \tag{9}$$

where $w(b,d)$ is a uniform weight and $\mathbf{1}$ the indicator function [54]. Roughly speaking this is akin to a Lebesgue integral for the persistence diagram. The value of Betti curve is therefore a measure of the density of points in the persistence diagram, with domain extending to the value of the point with highest persistence for a given dimension k. The β_k values are therefore dependent on the amount of points in a diagram and, in turn, on the signal length and the embedding parameters. It then follows that while it is not possible to compare Betti curves obtained from different embeddings, it is still possible to see how the Betti curve for slug flow extends to much higher filtration parameters, implying a more sparse persistence diagram.

V. STATE CHANGE IDENTIFICATION VIA TOPOLOGICAL INDICATORS

A single severe slugging event in a vertical riser system was studied by means of topological data analysis. The following analysis was applied to the bottom hole pressure signal, but the same procedure applied to other available sensor data (flowline pressure and flowline temperature) produced the same qualitative results. Steady-state flow at a pressure of 164 bar transitioned to a severe slugging flow regime, showing pressure oscillations with a period of approximately 12 minutes, amplitude of about 6 bar which lasted for about 1.5 hours before decaying back to steady-state flow (Fig.6.a).

Takens embedding was applied to the signal on a time window of 4500 s in both steady-state and slug flow to identify the main topological features in each flow regime. The embedding of steady-state flow (d=6, τ=9) produced a point cloud close to spherical in shape and with a radius of about 0.2 (Fig.6.b). Albeit presenting some structure, this pattern is similar to that of white noise as shown in Figure 2 and the associated persistence diagram shows points clustered mostly near the birth-death diagonal.

TABLE I. Topological indicators for white noise generated with σ=0.5, steady-state flow and severe slug flow from the signal shown in Figure 6a.

	Noise	Normal Flow	Slug Flow
$\mathcal{P}_\infty(\mathrm{H}_0)$	1.06	0.15	0.20
$\mathcal{P}_\infty(\mathrm{H}_1)$	0.29	0.19	7.09
$\overline{\mathrm{E}}_{\mathrm{H}_0}$	1.11	1.45	1.39
$\overline{\mathrm{E}}_{\mathrm{H}_1}$	1.59	1.83	1.05

The embedding of the slug flow portion of the time series (d=11, τ=14) shows instead clear periodic orbits with a gyration radius about one order of magnitude larger that in the previous case. Loops are indeed expected as they are a signature of periodicity in the signal, typical of the cases of severe slugging. The persistent diagram for slugging shows high persistence points for the homology class H_1, corresponding to one-dimensional loops and some degree of residual noise.

Betti curves are shown in Figure 6.g and f. The number of points in the two diagrams is different, therefore the Betti curves have different magnitudes, but it is apparent that steady-state flow shows some overlap between β_0 and β_1, while they are spread over a much larger domain.

In order to find suitable classifiers for the different flow regimes, some topological indicators were computed for the two flow regimes and are summarized in Table I. When considering the maximum persistence $\mathcal{P}_\infty$ for each dimension, H_1 components show a higher persistence. For slugging flow, while H_0 components show only a marginal increase in persistence, H_1 components show much larger value than in steady-state flow, due to the appearance of loops originated from the periodic pressure oscillations. Such patterns also have an effect on the persistence entropy of each diagram. Slug flow shows lower entropy for both H_0 and H_1 indicating a high degree of information.

It is then natural to test persistence entropy, the maximum persistence and Betti curves along the whole continuous signal, as indicators for identifying and monitoring the state change of the flow conditions in the well from steady-state to slug flow and *vice versa*. Imagining a system in which sensor data is accumulated continuously it would then be possible to apply repeatedly at time t_0 topological data analysis to a time-window of size $n, (t_{-n}; t_0)$, shifting this time window ahead in time every time a new signal data is received from the sensor. Whereas this shift is larger than one time-step, this is known as *stride* (δ) and it acts as a delay between successive time windows analysis, allowing for the time of running the analysis and storing the results. Each of these time windows is then analyzed by applying repeatedly Takens embedding and generating a point cloud and generating persistence diagrams as described in Section

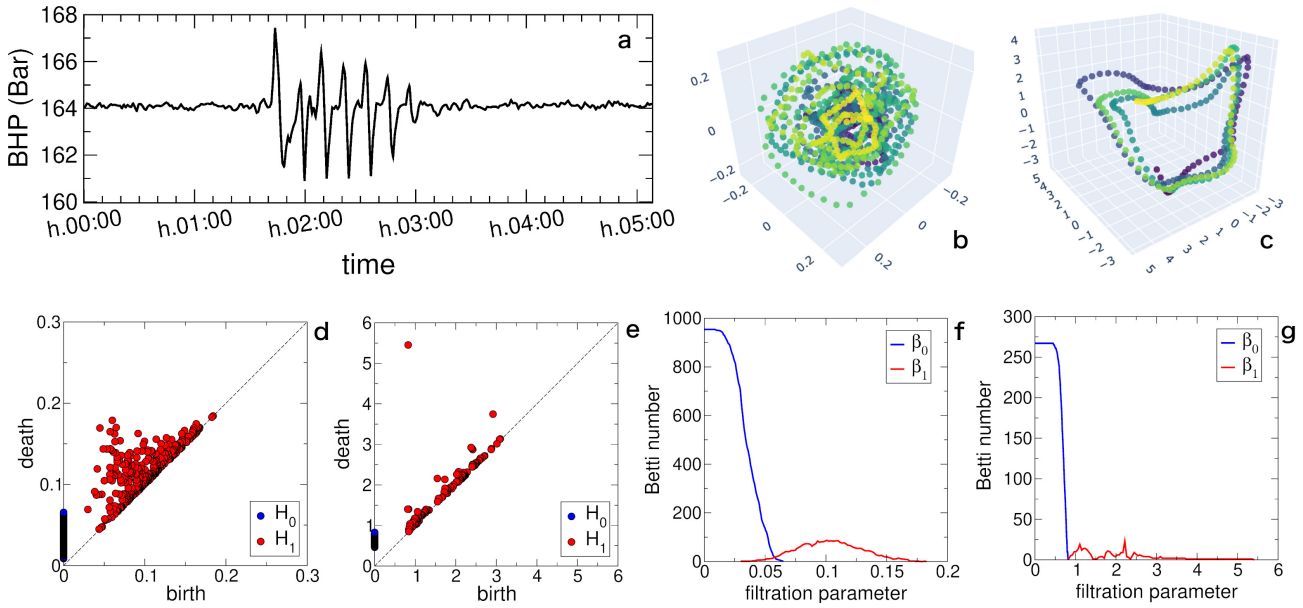

FIG. 6. a: time series for the bottom hole pressure (BHP) along the approximately 1.5 hours long slugging flow event. b-c: point cloud for steady-state flow (b, ca. at time h.00:30) and for slug flow (c, ca. at time h.02:00), d-e: persistence diagram for steady-state flow (d), and for slug flow (e). Betti number curves for steady-state flow (f) and slug flow (g).

II.

The time series in Figure 6.a was then analyzed by sampling a sliding time window of 8000 s, slightly larger than the expected slugging event. Instead of optimizing the embedding parameters following the heuristics of Section II, parameters were kept fixed along the whole time series analysis to allow for a better comparison between persistence diagrams and Betti curves taken at different times along the signal time series. The embedding parameters are $d = 8, \tau = 42, \delta = 200$ s, giving rise to a series of 96 persistence diagrams across 7 hours of sensor data. Following this approach the first 17 time windows include steady-state flow only, windows 41-60 include slug flow and windows 18-40 and 61-80 include a transient flow regime where one regime transitions to the other.

Results are shown in Figure 7. The maximum persistence of diagrams for both the homology dimensions (Fig.7.a) show an initial stable value corresponding to steady-state flow, then developing a peak right at the transition with a maximum around the middle of the analyzed interval where they fall back to steady-state values. This behavior is in agreement with the values in Table I, where the maximum persistence increases when transitioning from regular to slug flow conditions. Looking at Betti curves, it is convenient displaying the mean of the values $\langle \beta_k \rangle$ after normalizing over the filtration parameters as this is a measure of how many topological features are found in each persistence diagram, and it is shown in Figure 7.c,d. Both $\langle \beta_0 \rangle$ and $\langle \beta_1 \rangle$ are symmetric and clearly deviate from the steady-state value at the beginning of the transient conditions, similarly for $\mathcal{P}_\infty$.

Normalized persistence entropies show a decrease to a plateau in correspondence with the slug flow pattern. For

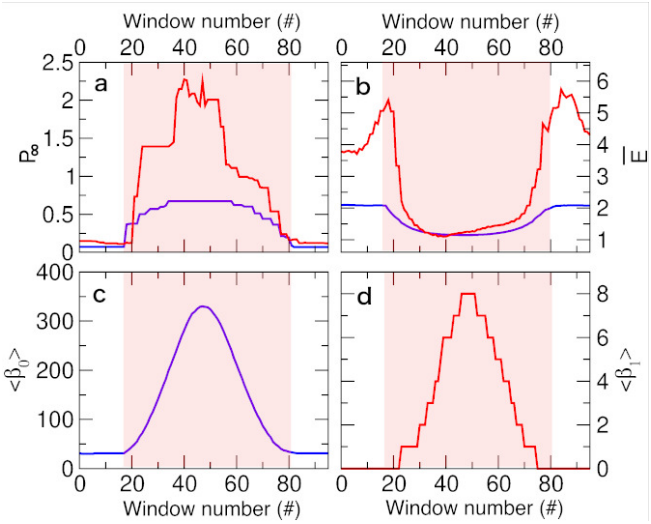

FIG. 7. Topological indicators computed at progressive sliding time windows. *a*: Maximum persistence, $\mathcal{P}_\infty$, *b*: normalized persistence entropy $\overline{E}_{H_k}$, *c* and *d*: mean of the Betti curves $\langle \beta_k \rangle$. Results for homology class H_0 (blue line) and H_1 (red line). The shaded area shows the time windows that include, even partially, signal with oscillations due to severe slugging.

the H_0 class the transition is sharp, while for H_1 entropy initially increases to a peak at the transition points before to fall to values similar to those of $\overline{E}_{H_0}$ (Fig.7b).

All the topological indicators above could be used for the the identification of slug flow by introducing, for example, simple thresholds systems, but none of them is an

actual distance function between persistence diagrams. As a further topological condition indicator we introduce the 2-Wasserstein distance [55], defined as:

$$W_q^{\mathrm{H}_k}(X,Y) = \left[\inf_{\eta: X \to Y} \sum_{x \in X} ||x - \eta(x)||_\infty^q \right]^{1/q}, \quad q = 2$$

where and $||(x,y)||_\infty$ is the L_∞-norm, and η is the set of all bijections (pairs of points) between the two persistence diagrams X and Y. Figure 8 shows the matrix graphical representation of the distance between the i-th and j-th persistence diagrams, produced from time snapshots (time windows) of the pressure signal taken at progressively increasing time.

First of all, the matrices in the figure show that persistence diagrams for the time series are stable, *i.e.* small shifts in the signal, produce small changes in the Wasserstein distance between diagrams [56]. This is an important property when comparing persistence diagrams, guaranteeing that flow regimes with radically different signals would lay at large W_2 distance from each other. While the matrix of distances between diagrams for H_0 look very symmetric, those for H_1 are not. Curiously, the H_0 persistence for time windows at the beginning and the end of the slugging event are very close in W_2 distance, as shown by the presence of two minima spanning diagonals of the plot, albeit the shape of the signal is not symmetric in time. The H_1 persistence is instead less symmetric but still clearly shows a large distance between diagrams corresponding to steady-state flow and slug flow. This demonstrates that the 2-Wasserstein distance can also be used as an effective indicator for identifying slug flow by simply setting a threshold on the maximum distance between diagrams taken a different times along a signal as long as the diagram for steady-state flow (normal conditions) is correctly identified. Other metrics such as the bottleneck distance [55], and the 2-norm of persistence landscapes [57] were also tested and found to produce similar results but with larger degrees of noise.

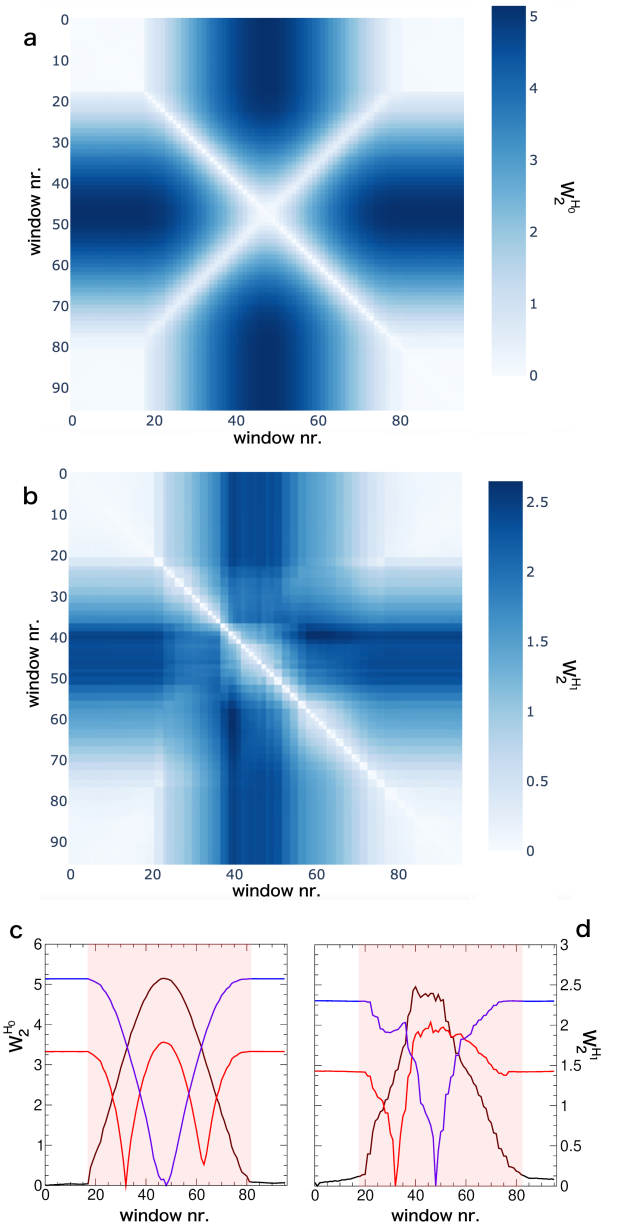

FIG. 8. Wasserstein distance between persistence diagrams from different time windows across the signal. *a*: $\mathrm{W}_2^{\mathrm{H}_0}(i,j)$, *b*: $\mathrm{W}_2^{\mathrm{H}_1}(i,j)$, *c*: black line: $\mathrm{W}_2^{\mathrm{H}_0}(1,j)$, distance between time window #1 (steady-state flow) and other windows. Red line: $\mathrm{W}_2^{\mathrm{H}_0}(32,j)$ distance between transient flow and other windows. Blue line $\mathrm{W}_2^{\mathrm{H}_0}(48,j)$ distance between slug flow and other windows. *d*: same as b but for $\mathrm{W}_2^{\mathrm{H}_1}(i,j)$. The shaded area shows the time windows that include, even partially, slug flow oscillations.

VI. MACHINE LEARNING CLASSIFICATION OF FLOW REGIMES VIA TOPOLOGICAL INDICATORS

The aim of this section is to automatically classify sections of sensor time series corresponding to severe slugging flow behavior on a system with multiple oil and gas wells. For this purpose the wellhead pressure time series from the 3W data set were analyzed. These data were found to be the most reliable after removing faulty sensor data. The same analysis performed on other sensors' data such as bottom hole pressure and flowline temperature was found to give the same qualitative results.

At first, the data were analyzed to confirm that the same conclusions drawn in Section V hold true for the 3W data set as well. Takens embedding and persistence diagrams were computed for prototypical cases of each of the three flow regime labels: normal flow, flow instabilities and severe slugging. Flow instabilities are considered to be flow fluctuations of moderate amplitude, not leading to a spontaneous blowdown as for severe slugging [50]. Takens embedding was applied to a time window of 3000 s at the optimal dimension and time delay parameters. Results are shown in Figure 9.

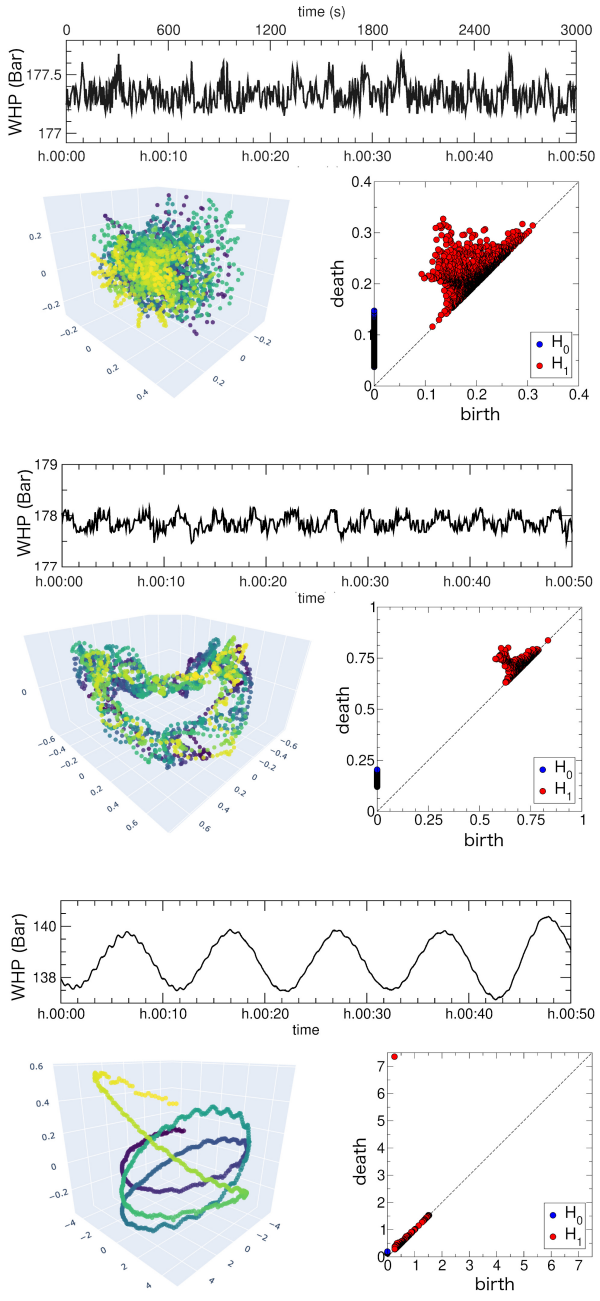

FIG. 9. Topological data analysis for time series from the 3W data set. Time series of signal for wellhead pressure (WHP), point clouds from Takens embedding and persistence diagram for, from top to bottom: normal operation flow (d=6, τ=36 s), minor flow instabilities (d=11, τ=129 s), severe slugging flow (d=12, τ=90 s).

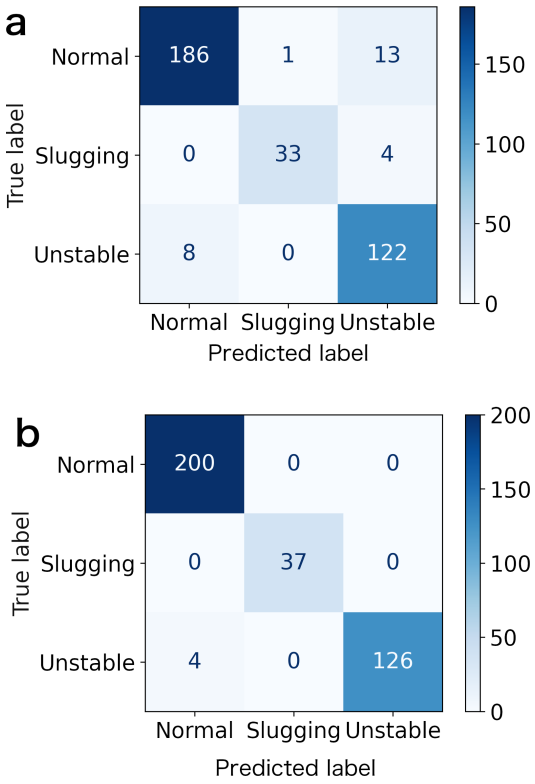

FIG. 10. Confusion matrices for classification of flow regimes based on topological indicators $\overline{\mathrm{E}}_{\mathrm{H}_k}$, $\mathcal{P}_\infty(\mathrm{H}_k)$, $\langle\beta_k\rangle$ for $k \in \{0,1\}$. *a*: logistic regression classifier. *b*: Random forest classifier

As for the state change case, the embedding of normal steady-state flow produces a point cloud close to spherical in shape and with a radius of about 0.2 (Fig.9) determined by the noise's amplitude mostly clustered at the birth-death diagonal.

For flow instabilities, the time series shows oscillations of about 0.5 bar but not a clear periodic signature. The embedding produced an interesting toroidal pattern and the persistence diagram shows a cluster for H_1 but no high persistence points. The case of severe slugging is radically different, as the pressure signal shows oscillation with amplitude of about 2 bar and a period of about 10 minutes. The embedding shows instead periodic orbits with limited amount of noise, resulting in a H_1 point at very high persistence.

Machine learning classification of the flow regimes can be performed with the following different approaches. One is to label time series using a supervised learning classifier algorithm on the ensemble of the eight scalar-valued topological indicators, which proved useful for the state change detection of Section V (persistence entropy, normalized persistence entropy, maximum persistence, average of the Betti curve, each for H_0 and H_1 homology classes). These indicators were obtained from processing the persistence diagram and should capture enough information about the shape of the embedded signals to classify them correctly. For the classification a 65-35% train-test split was used. For the optimization of flow efficiency, a correct identification of severe slugging is crucial. We therefore consider methods that label slugging cases as normal operations or unstable flow as more severe failure than the opposite case.

Results for multi-modal logistic classifier on the indica-

TABLE II. Relative feature importance for the random forest classifier applied on identifying slug flow and unstable flow from topological indicators on the 3W data set.

Feature	Relative Importance
$\langle\beta_0\rangle$	0.302
$\mathcal{P}_\infty(\mathrm{H}_1)$	0.274
$\mathcal{P}_\infty(\mathrm{H}_0)$	0.158
$\overline{\mathrm{E}}_{\mathrm{H}_0}$	0.153
$\overline{\mathrm{E}}_{\mathrm{H}_1}$	0.075
$\langle\beta_1\rangle$	0.040

tors listed above are shown in the confusion matrices in Figure 10. Logistic regression without any penalty term manages to classify flow regimes with an accuracy score of 92.9% (recall: 92.0%, precision: 93.6%, F_1: 92.7%), but still labels 11% of the severe slugging cases as unstable flow and 13% of unstable flow time cases as normal operations (Figure 10.*a*). The inverse of the regularization strength parameter C was optimized (C_{max}=0.01 and C_{min}=100) to maximize the classification accuracy. Regularization terms L_1 and L_2 were found to have little to no effect on the classification accuracy.

Moving to a different classification algorithm, a random forest classifier dramatically improves the results (accuracy 99.2%, recall: 99.2%, precision: 99.5%, F_1: 99.4%, Figure 10.*b*). All severe slugging events were classified correctly and the only error being 3% of flow instabilities labeled as normal operations. This is somewhat expected as unstable flow conditions are not well defined, therefore their boundary with normal operating conditions, especially in transient conditions, can be subjective. Feature importance for the Gini-classifier is based on the mean decrease in feature impurity [58], is shown in Table II.

Another approach is to classify directly an encoding of the whole persistence diagram. Persistence images [59] are built by adding a Gaussian at each point in a persistence diagram and replacing the death coordinate with the persistence (death-birth). Persistent images for one severe slugging time series are shown in Figure 11.a and 11.b for H_0 and H_1 respectively. The images from each homology dimension are then added to form a single image and then classified with logistic regression and random forest algorithms. The results for the two approaches are identical and shown in Figure 11.c (accuracy 99.5%, recall: 99.5%, precision: 99.7%, F_1: 99.6%). Classification of the images performed generally better than that of the topological indicators as all severe slugging events were classified correctly and fewer unstable flow conditions were wrongly assigned to normal condition. By looking at the coefficients of the logistic classifier for the severe slugging label (Fig.11.d), it is possible to notice that the areas of the images with a larger weight in the classification are those with low or zero-birth time, *i.e.* points corresponding to the homology class H_0. This is in agreement with the feature importance in Table II where the highest importance features all belonged to the H_0 class.

VII. SUMMARY AND CONCLUSIONS

Topological data analysis is a useful tool to extract information from complex data sets by inferring the shape of data in a high-dimensional embedded space. In particular, TDA proved to be effective in identifying periodic patterns in data as these appear as holes and loops in the embedding space. Typically in chemical engineering periodic patterns in sensor data are associated with slug flow, where large gas bubbles induced by the piping height profile move in the fluid phase inducing regular spikes in pressure. Severe slugging events in multiphase flow are a major problem in chemical engineering as they are a hallmark of inefficient mass and momentum transport but may also lead to damage to the plant equipment. Sensor data from offshore oil and gas wells was analyzed with TDA in a data pipeline consisting in Takens embedding followed by persistent homology. Persistence diagrams were processed to produce persistence images and to extract topological indicators. Such indicators show a stable value when the pressure is at steady-state and have a clear transition in correspondence with the onset of slug flow. As well, when introducing a proper concept of distance between persistence diagrams the two dynamical states, these appear clearly separated and distinct. The distance and topological indicators can then be used for the identification of dynamic state change (steady-state - severe slugging) by setting up a simple threshold-alert system.

Machine learning-based automatic classification of flow regimes was performed on a large, public data set of undesired well events known as the 3W data set. From the 3W set, pressure time series labeled as either "normal operation", "flow instabilities", and "severe slugging" were classified with high accuracy classification by means of simple linear classifiers such as logistic regression and random forest algorithms. The results of the classification based on a limited number of topological indicators such as persistence entropy, maximum persistence and Betti curves, were found dependent on the algorithm, while classification based on persistence images were found more accurate and method independent.

The results above suggest that condition monitoring and dynamics state change tasks to identify severe slugging could be performed efficiently on industrial assets by employing TDA-based methods. A public repository with the code used in this work can be found in reference [60].

ACKNOWLEDGMENTS

The work presented in this paper is associated with the center SFI NorwAI, where Cognite AS is one of the funding partners. The center is partly funded by the Research Council of Norway, within the Centers for Research-based

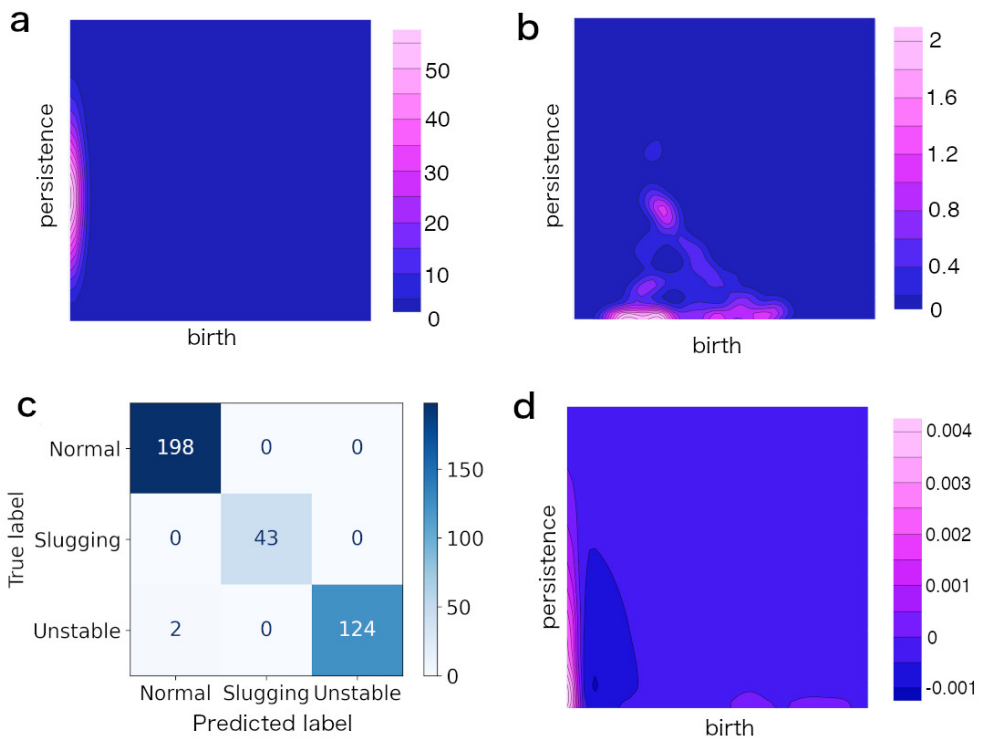

FIG. 11. *a*: H_0-persistence image for a severe slugging event, *b*: H_1-persistence image for a severe slugging event, *c*: confusion matrix for the logistic regression classifier based on the sum of H_0 and H_1 persistence images, *d*: graphical representation of the coefficients from the logistic classifier showing the importance of each pixel in the classification of severe slugging events.

Innovation Scheme (SFI), project number 309834. The author thanks Dr. Rafael H. Nemoto, Jaijith Sreekantan, Dr. Alexander Stasik (SINTEF), Hayden Hohns and Dr. Andris Piebalgs for enlightening discussions. Cognite AS is kindly acknowledged for providing resources for this research.

Appendix A: Analysis of periodic signals

In order to show how a signal containing a periodic or quasi-periodic component is characterized in topological data analysis, we follow the approach of Perea and Harer [38], assuming a generic (noiseless) signal can be represented by a trigonometric function. For example, defining the dimension of the embedding d and the time delay τ, one could embed an infinite periodic signal in $\mathbb{R}$ such as $f(t) = \cos(nt)$, $n \in \mathbb{N}$ to $\mathbb{R}^{d+1}$ by constructing the vector

$$\begin{aligned} F_{d,\tau}(f(t)) = & \begin{bmatrix} \cos(nt) \\ \cos(n(t+\tau)) \\ \vdots \\ \cos(n(t+d\tau)) \end{bmatrix} = \begin{bmatrix} \cos(nt) \\ \cos(nt+n\tau) \\ \vdots \\ \cos(nt+dn\tau) \end{bmatrix} \\ = & \cos(nt) \begin{bmatrix} 1 \\ \cos(n\tau) \\ \vdots \\ \cos(dn\tau) \end{bmatrix} - \sin(nt) \begin{bmatrix} 0 \\ \sin(n\tau) \\ \vdots \\ \sin(dn\tau) \end{bmatrix} \\ = & \cos(nt)\mathbf{u} + \sin(nt)\mathbf{v} \end{aligned} \tag{A1}$$

Note that the choice of d and τ implies the definition of a time window of size $(t_0, t_0 + d\tau)$ that is sampled d times every τ time steps. Repeated applications of such embedding to a collection of progressively increasing times $F_{d,\tau}(f(t))$ at $t_0, t_{0+\Delta}, \ldots, t_{0+M\Delta}$ produce M vectors, hence a cloud of M points in $\mathbb{R}^{d+1}$.

The embedded signal $F_{d,\tau}(f(t))$ describes a point in the sliding window point cloud. The equation A1 describes an ellipse in $span(\mathbf{u}, \mathbf{v})$ as long as the vectors $\mathbf{u}$ and $\mathbf{v}$ are linearly independent. The minor and major axes length is corresponding to the square roots of the two eigenvalues $\lambda_\pm$ of the matrix $\mathbb{A}$ [38]:

$$\mathbb{A} = \begin{pmatrix} ||\mathbf{u}||^2 & -\mathbf{u} * \mathbf{v} \\ -\mathbf{v} * \mathbf{u} & ||\mathbf{v}||^2 \end{pmatrix} \tag{A2}$$

$$\lambda_\pm = \frac{(d+1) \pm \sqrt{\frac{\sin(n(d+1)\tau)}{\sin(n\tau)}}}{2} \tag{A3}$$

and that the ellipse turns into a circle whenever the minor axis is maximised, i.e. whenever the time window satisfies the following condition [38]:

$$d\tau = \left(\frac{d}{d+1}\right)\frac{2\pi}{n} \tag{A4}$$

i.e. whenever the window size is a multiple of the frequency of the time series periodicity. By scaling the original time series to $f_{scaled}(t) = K\cos(nt)$ while keeping constant the embedding parameters, such as when for changing units of measures of the signal, it follows that the $\mathbb{A}$ matrix and its eigenvalues gets scaled accordingly by a factor K^2. The distance between each point in the $\mathbb{R}^{d+1}$ cloud is then scaled, preserving the shape of the point cloud's topology. The same holds for offsetting the time series to a different mean value, $f(t) = \cos(nt) + \eta$. Instead, by increasing the values of the oscillation frequency n, the difference between the two eigenvalues in eq.A3 increases its linear eccentricity. This also implies that for periodic signals with increasing oscillation frequency n (at constant window size) the ellipse cloud-shape roundness is expected to change. Therefore for the Takens embedding, the topology of the embedded data is invariant under scaling and shifting the time series, but not under changes in frequency.

The effect of the periodic signal on the overall shape of the point cloud is less straightforward but can be deduced by representing the signal by a Fourier series. This is reasonable for smooth signals as integrable functions are guaranteed to have vanishing Fourier coefficients by the Riemann-Lebesgue lemma. In this case it can be proven that for signals $f(t)$ where $||f(t)||_2$=1 and

$$\int f(t)dt = 0, \quad f(t) = f\left(\frac{2\pi t}{n}\right) \quad \forall n \in \mathbb{N} \tag{A5}$$

the sliding window approach produces an elliptic point cloud, similarly to Eq.A1 with maximum roundness as in Eq.A4 [38].

In case of more complex patterns in the time series, such as for commensurate combination of trigonometric functions such as $f(t) = \cos(nt) + \cos(mt)$, $n/m \in \mathbb{Q}$, the embedding gives rise to a "wavy" ellipse, with oscillations on dimensions perpendicular to the ellipse's plane. For quasi-periodic time series (where the ratio n/m is irrational) it can instead be shown the the embedding gives rise under specific conditions to to a hyper-torus [37] (see Figure 2).

The arguments above show that the choice of the embedding parameters do influence the shape of the point cloud, and therefore these could be tuned to optimize the sensitivity of the data analysis when *e.g.* the target is to identify periodic components in a signal such as for severe slugging in multiphase flow. In this case while the periodicity of the slugging signal is typically not known *a priori*, its known that its order of magnitude is in the range of several minutes and this would allow to "filter" out from the analysis periodic components in the signal occurring on a much shorter time scale such as minor hydrodynamic slugging and other flow instabilities.

It is also worth mentioning that both Nyquist and Takens theorems set a fundamental limit for reproducing the information contained in the time series, where as the embedding dimension should be chosen to be at least twice as large as the largest signal frequency [38].

We note here that traditional approaches based on Fourier transform (FT) methods have proven to be very efficient in detecting severe slugging. The main advantage of topological methods over FT methods lays in the embedding procedure, which requires in principle only one oscillation period in order to produce a loop, i.e. a

point in the persistence diagram, while the height and sharpness of frequency peaks in FT methods is dependent on the ratio between the sampling rate and the wave length.

[1] J. Bomba, D. Chin, A. Kak, and W. Meng, Flow assurance engineering in deepwater offshore - past, present, and future, in *OTC Offshore Technology Conference*, Vol. 04 (2018).

[2] A. O. Morgado, J. M. Miranda, J. D. P. Araújo, and J. B. L. M. Campos, Review on vertical gas–liquid slug flow, Int. J. Multiphase Flow **85**, 348 (2016).

[3] R. M. Davies and G. I. Taylor, The mechanics of large bubbles rising through extended liquids and through liquids in tubes, Proc. R. Soc. Lond. A **200**, 375–390 (1950).

[4] D. Barnea, a unified model for predicting flow-pattern transitions for the whole range of pipe inclinations, Int. J. Multiphase Flow **13**, 1 (1987).

[5] S. Pedersen, P. Dudevic, and Z. Yang, Review of slug detection, modelling and control techniques for offshore oil and gas production processes, in *IFAC-PapersOnLine*, Vol. 48-6 (2015) p. 089.

[6] S. G. Nnabuife, H. Tandoh, and J. F. Whidborne, Slug flow control using topside measurements: A review, Chem. Eng. J. Advances **9**, 100204 (2022).

[7] J. L. Baliño, K. P. Burr, and R. H. Nemoto, Modeling and simulation of severe slugging in air–water pipeline–riser systems, Int. J. Multiphase Flow **36**, 643–660 (2010).

[8] B. T. Yocum, Offshore riser slug flow avoidance: mathematical models for design and optimization, in *SPE European Meeting* (1973).

[9] V. Tin and M. Sarshar, An investigation of severe slugging characteristics in flexible risers., in *Proc. The 6th. International Conference on Multiphase Production* (1993).

[10] S. Mokhatab, *Severe Slugging in Offshore Production Systems* (Nova Science Publishers, 2010).

[11] Z. Schmidt, J. P. Brill, and H. D. Beggs, Experimental study of severe slugging in a two-phase-flow pipeline-riser system., SPE Journal **20**, 407 (1980).

[12] Z. Schmidt, D. R. Doty, and K. Dutta-Roy, Severe slugging in offshore pipeline riser-pipe systems., SPE Journal **25**, 27 (1985).

[13] A. Bøe, Severe slugging characteristics; part i: flow regime for severe slugging; part ii: Point model simulation study, Selected Topics in Two-Phase Flow, NTH, Trondheim, Norway (1981).

[14] B. F. M. Pots, I. G. Bromilow, and M. J. W. F. Konijn, Severe slug flow in offshore flowline/riser systems, SPE Production Engineering , 319 (1987).

[15] Y. Taitel, Stability of severe slugging, Int. J. Multiphase Flow **12**, 203 (1986).

[16] J. Fabre, L. Peresson, J. Corteville, R. Odello, and T. Bourgeois, Severe slugging in pipeline/riser systems, SPE Production Engineering **5**, 299 (1990).

[17] C. Sarica and O. Shoham, A simplified transient model for pipeline-riser systems., Chem. Eng. Sci. **46**, 2167 (1991).

[18] R. H. Nemoto and J. L. Baliño, Modeling and simulation of severe slugging with mass transfer effects., Int. J. Multiphase Flow **40**, 144 (2012).

[19] F. Jansen, O. Shoham, and Y. Taitel, The elimination of severe slugging—experiments and modeling, Intl. J. of Multiphase Flow **22**, 1055 (1996).

[20] S. Song and G. Kouba, Fluids transport optimization using seabed separation, J. Energy Resour. Technol. **122**, 105 (2000).

[21] K. Havre and M. Dalsmo, Active Feedback Control as the Solution to Severe Slugging, in *SPE Annual Technical Conference and Exhibition*, Vol. SPE-71540-MS (2001).

[22] S. Pedersen, P. Durdevic, and Z. Yang, Learning control for riser-slug elimination and production-rate optimization for an offshore oil and gas production process, IFAC Proceedings Volumes **47**, 8522 (2014).

[23] F. D. Meglio, G.-O. Kaasa, N. Petit, and V. Alstad, Stabilization of slugging in oil production facilities with or without upstream pressure sensors, J. of Process Control **22**, 809 (2012).

[24] E. Storkaas and S. Skogestad, A low-dimensional dynamic model of severe slugging for control design and analysis., in *Proceeding of the 11th International Conference on Multiphase Flow. BHR Group, San Remo, Italy* (2003).

[25] H. Tuvnes, Severe well slugging and reservoir-well interactions, modeling and simulations., in *Thesis, Norwegian University of Science and Technology.* (2008).

[26] G. O. Kaasa, V. Alstad, J. Zhou, and O. M. Aamo, Attenuation of slugging in unstable oil wells by nonlinear control., in *Proceeding of the 17th IFAC World Congress. IFAC, Seoul, Korea.* (2008).

[27] C. M. D. Silva, F. Dessen, and O. J. Nydal, Dynamic multiphase flow models for control., in *Proceeding of the 7th North American Conference on Multiphase Techonology. BHR Group, Banff, Canada* (2010).

[28] E. Jahanshahi and S. Skogestad, Simplified dynamical models for control of severe slugging in multiphase risers, IFAC Proceedings Volumes **44**, 1634 (2011), 18th IFAC World Congress.

[29] E. Jahanshahi and S. Skogestad, Simplified dynamic models for control of riser slugging in offshore oil production, Oil and Gas Facilities **3**, 080 (2014).

[30] R. H. Nemoto, S. Abrol, and G. Becquin, Simplified Model for Control of Severe Slugging in S-Shaped Risers, in *Offshore Technology Conference Brasil*, Vol. Day 2 Wed, October 28 (2015).

[31] Y. Hiraoka, T. Nakamura, A. Hirata, E. G. Escolar, K. Matsue, and Y. Nishiura, Hierarchical structures of amorphous solids characterized by persistent homology, Proc. Natl. Acad. Sci. USA **113**, 7035 (2015).

[32] A. Cole, G. J. Loges, and G. Shiu, Quantitative and interpretable order parameters for phase transitions from persistent homology, Phys. Rev. B **104**, 104426 (2021).

[33] J. Arnold, F. Schäfer, M. Žonda, and A. U. J. Lode, Interpretable and unsupervised phase classification, Phys. Rev. Research **3**, 033052 (2021).

[34] C. F. Loughrey, P. Fitzpatrick, N. Orr, and A. Jurek-Loughrey, The topology of data: opportunities for cancer research, Bioinformatics **37**, 3091 (2021).

[35] T. Sousbie, The persistent cosmic web and its filamentary structure – I. Theory and implementation, Mon. Not. R. Astron. Soc. **414**, 350 (2011).

[36] A. D. Smith, P. Dłotko, and V. M. Zavala, Topological data analysis: Concepts, computation, and applications in chemical engineering, Comput. Chem. Eng. **146**, 107202 (2021).

[37] J. A. Perea, Persistent homology of toroidal sliding window embeddings, in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (IEEE, 2016) pp. 6435–6439.

[38] J. A. Perea and J. Harer, Sliding windows and persistence: as application of topological methods to signal analysis, Found. of Comput. Mathematics **15**, 799 (2015).

[39] F. Takens, Detecting strange attractors in turbulence, in *Dynamical Systems and Turbulence, Lecture Notes in Mathematics*, Vol. 898, edited by D. A. Rand and L.-S. Young (Springer-Verlag, 1981) pp. 366–381.

[40] H. S. A.M. Fraser, Independent coordinates for strange attractors from mutual information, Phys. Rev. A **33**, 1134–1140 (1986).

[41] H. Abarbanel, M. Kennel, and R. Brown., Determining embedding dimension for phase-space reconstruction using a geometrical construction, Phys. Rev. A **45**, 3403–3411 (1992).

[42] H. Edelsbrunner and J. Harer, *Computational Topology. An introduction* (American Mathematical Society, 2010).

[43] H. Edelsbrunner, D. Letscher, and A. Zomorodian, Topological persistence and simplification, in *Proceedings 41st Annual Symposium on Foundations of Computer Science*, Vol. 208 (2000) pp. 454–463.

[44] Vietoris–Rips complexes also provide topologically correct reconstructions of sampled shapes, Computational Geometry **46**, 448 (2013), 27th Annual Symposium on Computational Geometry (SoCG 2011).

[45] P. Niyogi, S. Smale, and S. Weinberger, Finding the homology of submanifolds with high confidence from random samples, Discrete & Comput. Geom. **39**, 419 (2008).

[46] J. B. Pérez, S. Hauke, U. Lupo, M. Caorsi, and A. Dassatti, giotto-ph: A Python library for high-performance computation of persistent homology of Vietoris-Rips filtrations (2021), arXiv:2107.05412.

[47] N. Saul and C. Tralie, Scikit-TDA: Topological data analysis for python (2019).

[48] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, Scikit-learn: Machine learning in Python, J. of Mach. Learn. Res. **12**, 2825 (2011).

[49] K. H. Bendiksen, D. Maines, R. Moe, and S. Nuland, The dynamic two-fluid model OLGA: Theory and application, SPE Prod. Eng. **6**, 171 (1991).

[50] R. E. V. Vargas, C. J. Munaro, P. M. Ciarelli, A. G. Medeiros, B. G. do Amaral, D. C. Barrionuevo, J. C. D. de Araújo, J. L. Ribeiro, and L. P. Magalhães, A realistic and public dataset with rare undesirable real events in oil wells, J. Pet. Sci. Eng. **181**, 106223 (2019).

[51] N. Atienza, R. Gonzalez-Diaz, and M. Rucco, Persistent entropy for separating topological features from noise in Vietoris-Rips complexes, J. Intell. Inf. Syst. **52**, 637 (2019).

[52] N. Atienza, R. Gonzalez-Díaz, and M. Soriano-Trigueros, On the stability of persistent entropy and new summary functions for topological data analysis, Pattern Recognit. **107**, 107509 (2020).

[53] A. Hatcher, *Algebraic Topology* (Cambridge University Press, 2002).

[54] F. Chazal and B. Michel, An introduction to topological data analysis: Fundamental and practical aspects for data scientists, Front. Artif. Intell. **4**, 667963 (2021).

[55] M. Kerber, D. Morozov, and A. Nigmetov, Geometry helps to compare persistence diagrams, ACM J. Exp. Algorithmics **22**, 10.1145/3064175 (2017).

[56] D. Cohen-Steiner, H. Edelsbrunner, and J. Harer, Modeling and simulation of severe slugging in air–water pipeline–riser systems, Discrete Comput. Geom. **37**, 103 (2007).

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

[58] L. Breiman, Random forests, J. Mach. Learn. **45**, 5 (2001).

[59] H. Adams, T. Emerson, M. Kirby, R. Neville, C. Peterson, P. Shipman, S. Chepushtanova, E. Hanson, F. Motta, and L. Ziegelmeier, Persistence images: A stable vector representation of persistent homology, J. Mach. Learn. Res. **18**, 218–252 (2017).

[60] S. Casolo, Github repository (2022), https://github.com/simonecasolo/TDA-for-Slugging.