

Physics based Pathline Clustering and Exploration

Duong B. Nguyen^{†1}, Lei Zhang^{‡1}, Robert S. Laramée^{§2}, David Thompson^{¶3}, Rodolfo Ostilla Monico^{||1}, and Guoning Chen^{**1}

¹University of Houston, USA

²Swansea University, Wales, UK

³Mississippi State University, USA

Abstract

Most existing unsteady flow visualization techniques concentrate on the depiction of geometric patterns in flow, assuming the geometry information provides sufficient representation of the underlying physical characteristics, which is not always the case. To address this challenge, this work proposes to analyze the time-dependent characteristics of the physical attributes measured along pathlines which can be represented as a series of time activity curves (TAC). We demonstrate that the temporal trends of these TACs can convey the relation between pathlines and certain well-known flow features (e.g., vortices and shearing layers), which enables us to select pathlines that can effectively represent the physical characteristics of interest and their temporal behavior in the unsteady flow. Inspired by this observation, a new TAC-based unsteady flow visualization and analysis framework is proposed. The center of this framework is a new similarity measure that compares the similarity of two TACs, from which a new spatio-temporal, hierarchical clustering that classifies pathlines based on their physical attributes, and a TAC-based pathline exploration and selection strategy are proposed. A visual analytic system incorporating the TAC-based pathline clustering and exploration is developed, which also provides new visualizations to support the user exploration of unsteady flow using TACs. This visual analytic system is applied to a number of unsteady flow in 2D and 3D to demonstrate its utility. The new system successfully reveals the detailed structure of vortices, the relation between shear layer and vortex formation, and vortex breakdown, which are difficult to convey with conventional methods.

Keywords: unsteady flow visualization, time activity curves, pathline clustering

CCS Concepts

• **Human-centered computing** → **Visualization techniques; Scientific visualization;**

1. Introduction

Vector field visualization is a ubiquitous technique that is employed to study a wide range of dynamical systems ranging from automobile and aircraft engineering, to climate study, combustion dynamics, earthquake engineering, and medicine. Many effective approaches have been developed to visualize such complex data [ELC*12, LHZP07, PPF*11, SWJS08]. Among these techniques, the geometric-based approaches [MLP*10, ELC*12, SLC19] are commonly applied due to their intuitive representation of flow behavior. Examples of geometric-based techniques include various integral curve/surface based representations and integral

curve clustering that concentrate on the geometric characteristics of the flow (e.g., the shape or curvature of integral curves).

However, there are two limitations with the existing geometric-based methods. First, important physically relevant features are not always captured. For example, vector field topology –an abstract representation of the geometric characteristics of steady flow, only encodes hyperbolic features in the flow. Other physically relevant information, e.g., vortices, shearing, etc., is not always captured. Similarly, clustering methods that select integral curves to ensure sufficient spatial coverage and to reduce cluttering typically do not consider physical importance. Second, a geometric representation may not intuitively reveal the physical behavior of the flow, as shown in Figure 1. To incorporate physics into the visualization of flows, Zhang et al. [ZNT*17] introduced a Lagrangian accumulation framework that can be used to characterize integral curves by inspecting their respective overall attribute behaviors (i.e., each integral curve is assigned a value by accumulating the values of an attribute of interest along the curve). That technique was inspired by the pathline attributes introduced by Shi et al. [STH*09]. In the

[†] duongnguyenbinh@gmail.com

[‡] zhanglei.later@gmail.com

[§] R.S.Laramee@swansea.ac.uk

[¶] dst@cavs.msstate.edu

^{||} rostillia@central.uh.edu

^{**} chengu@cs.uh.edu

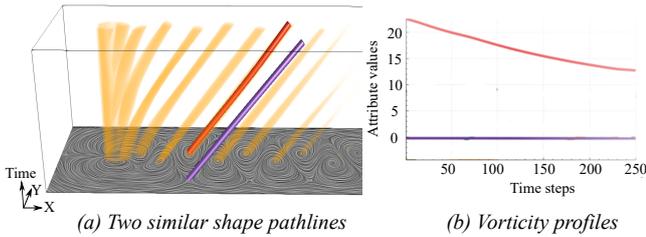


Figure 1: (a) Two pathlines with similar geometry derived from the flow behind a cylinder and the vorticity field shown by the volume rendering. (b) The corresponding profiles of the vorticity measured along the two pathlines show different physical behaviors.

meantime, Lee et al. proposed a visualization framework to analyze time-varying data sets with a time activity curve (TAC) based distance field [LS09b], which is used to highlight features. All the previously mentioned research looks at the overall characteristics (or the sum/average) of local attributes, which may not capture all defining characteristics, e.g., the diffusion behavior of Q concentration [HWM88] during the advection of vortices over time (Figure 8), due to the suppression of local information.

To address the above limitations and incorporate more detailed physics into the analysis and visualization of unsteady flow, we propose a novel visual analysis framework based on the temporal behavior of local, physical attributes of interest measured along individual pathlines. Similar to Lee et al. [LS09b], we refer to the temporal profile of the attribute along a pathline as a time activity curve (TAC). Unlike Shi et al. [STH*09] and other similar methods that also compute the pathline attributes, our framework takes into account the arbitrary movement of the observer. That is, the physical features in our framework are computed in the optimal reference frames; hence the geometric representation aligns with the corresponding physical attribute field and the TAC profiles do not vary under reference frame transformations [GT20]. The benefits of analyzing the flow behavior based on TACs are two-fold. First, they are 1D plots that are independent of the flow dimension (i.e., applicable to both 2D and 3D data); hence processing them is easier. Second, the geometric characteristics of TACs (e.g., ascending, descending, peaks, valleys, etc.) reveal the interaction of flow particles with physical features over time. This enables us to explain the geometric characteristics of the corresponding pathlines and vice versa. For instance, in the 2D flow behind cylinder (Figure 8), pathlines seeded in the core area of a vortex exhibit large positive Q values and slowly change over time, matching the flat geometry of those pathlines. In contrast, pathlines seeded outside the core of a vortex have large variation of Q values over time, matching the helical behavior of the pathlines. In addition, the attribute profile provides additional information that the geometry of pathlines cannot convey, such as the decaying of the Q concentration over time, indicating the loss of the rotation momentum of vortices during transport. Such an in-depth coupling of pathline characteristics and physical attributes has not been studied previously.

The above observations and benefits of TACs lead us to believe that TAC analysis may serve as the cornerstone of an exploration framework that reveals relevant flow behavior more effec-

tively than inspecting the geometric characteristics of the integral curves alone. Our contributions are summarized as follows.

- We propose to couple the pathline characterization with the attribute profiles (i.e., TACs) measured along the pathlines to provide a more informative exploration of unsteady flow. Different from previous work, we take into account all the sampled attribute values along each pathline during their characterization.
- We introduce a number of analysis and exploration techniques based on TACs, including a new spatio-temporal, hierarchical clustering of pathlines based on their respective TACs (Section 4.2) and a TAC-based pathline selection and exploration. Central to these techniques is a comprehensive similarity measure for the comparison of two TACs, which we refer to as a TAC similarity measure (TSM) that incorporates the global correlation of pair-wise TACs and the spatio-temporal distances between them (Section 4.1).
- We develop a visual exploration system (Section 5) that integrates the aforementioned TAC-based analysis and exploration techniques with a number of novel visualizations to support an effective user exploration of the pathline behaviors based on their respective TACs, including a modified edge-bundling visualization of TAC clusters and 2D stack plot for TAC behavior summarization and exploration (Section 4.3.3).

We have applied our TAC-based exploration system to a number of 2D and 3D unsteady flows. Our framework effectively reveals the two-layer configuration of a vortex and its decay over time in vortex shedding, which is difficult to reveal via conventional methods. We also facilitate interpretation of the temporal behavior of vortex rings in both 2D and 3D simulations, including its interaction with a wall and its breakdown. In particular, our system successfully selects pathlines to effectively depict the temporal behaviors of small-scale vortices in a number of 2D unsteady flows that was difficult to achieve previously. Furthermore, we demonstrate the flexibility of our framework by combining it with other analysis techniques (e.g., shearing layer criterion and FTLE computation [Hal01]) to help enhance their characterizations.

2. Related Work

There is a large body of literature on the analysis and visualization of flow data. Interested readers are referred to recent surveys for dense and texture-based visualization techniques [LHD*04], geometric-based methods [ELC*12, MLP*10], illustrative visualization [BCP*12], clustering-based approaches [SLC19], topology-based methods [LHZP07, PPF*11], and partition-based techniques [SWJS08]. In this section, we focus on the most closely-related work.

Topological vector field analysis. Vector field topology provides a streamline classification strategy based on the origin and destination of the individual streamlines. Since its introduction to the visualization community [HH89], vector field topology has received extensive attention. A large body of work has been introduced to identify different topological features, including fixed points [PP03, TSH01] and periodic orbits [CML*07, TWS04, WS01]. Recently, Chen et al. [CMLZ08] studied the instability of trajectory-based vector field topology and, for the first time, proposed Morse decomposition for vector field topology computation, which leads to

a more reliable interpretation of the resulting topological representation of vector fields. Szymczak et al. [SZ12] introduced a new approach to converting the input vector field to a piecewise constant (PC) vector field and computing the Morse decomposition on a triangulated manifold surface.

For the topological analysis of unsteady flow, *Lagrangian Coherent structures (LCS)*, i.e., curves (2D) or surfaces (3D) in the domain across which the flux is negligible, were introduced to identify separation structures in unsteady flow. The computation of LCS was first introduced by Haller [Hal01] by computing the *Finite-Time Lyapunov Exponent (FTLE)*, whose ridges indicate the LCS. FTLE has been compared with the separatrices in the steady case [SP07], and its computational performance has been improved substantially [GWT*08]. Recently, Fuchs et al. [FKS*10] presented an extended critical point concept to adapt the notion of vector field topology to unsteady flows. Sadlo and Weiskopf introduced a streakline-based topology based on generalized streaklines [SW10]. It successfully characterizes the saddle type of hyperbolic features and has been extended to study 3D unsteady flow topology [ÜSE13].

Integral curve attributes. Salzbrunn and Scheuermann introduced *streamline predicates* that classify streamlines by interrogating them as they pass through user-specified features, e.g., vortices [SS06]. Later, this approach was extended to classifying pathlines [SGSM08]. At the same time, Shi et al. [STH*09] presented a data exploration system to study the characteristics of pathlines based on various attributes, including winding angle. Recently, a statistics-based method was proposed to help select the proper set of pathline attributes to improve interactive flow analysis [PLMH12]. More recently, McLoughlin et al. [MJL*13] introduced the idea of a streamline signature based on a set of curve-based attributes including curvature and torsion. This streamline signature is used as a measure of the similarity between streamlines, pathlines, and helps domain experts place and filter streamlines for the creation of an informative and uncluttered depiction of 3D flow. Zhang et al. [ZCL*16] extended Lagrangian accumulation to define an attribute field based on the accumulated values along integral curves. This attribute field employs an Eulerian representation of Lagrangian information in a similar fashion to texture-based techniques. It conveys a continuous representation of the variation associated with integral curve behavior to some extent.

Time-varying series analysis and visualization. TACs have been studied in scientific visualization in recent years [WYM08, WS09a, WFMF00, GRCR03, WS09b, LS09a, FMHC07]. Lee et al. proposed a visualization framework to analyze time-varying data sets with a TAC-based distance field [LS09b]. This field provides a visualization to highlight the position of the features; however, it still does not provide certain details about an individual TAC, especially the temporal occurrence and period of an interesting feature. Wei et al. introduced a dual-space method to analyze turbulent combustion particle data, starting by clustering the time series curves in the phase space of the data, and then visualizing the corresponding trajectories of each cluster in the physical space [WYG*11]. The 2D time series curves are constructed using the correlation between temperature and mixture fraction. These curves are then clustered using the statistical model-based method. For spatio-temporal vi-

ualization of vortex features, Ferrari et al. [FHM20] combined the vortex core lines extracted from the maxima score correlation of the two attributes λ_2 and vorticity along the time dimension to create an evolution surface of vortices. Ferstl et al. proposed a time-hierarchical clustering approach for analyzing the temporal growth of the uncertainty in ensembles of weather forecasts [FKRW17]. For a thorough overview of approaches for the time-varying data, please refer to the surveys [EA12, Lia05]. In this work, we opt for the AHC clustering for a consistent multi-level abstract representation of the flow. We apply a new similarity-based measure over the statistical model as in [WYG*11] driven by the unique requirements of our problem. Parts of this work have been presented as an IEEE Visualization 2019 short paper [NZL*19]. In comparison with that short paper, the work presented here uses a different similarity measure for TAC comparison and for pathlines selection, applies different strategies to choose candidate cut points for temporal clustering, combines the proposed clustering framework with other analysis (like FTLE), and provides additional experimental results to fully demonstrate the usefulness of the proposed framework.

3. Vector Field and TAC Background

Consider a spatio-temporal domain \times where \subset^d is a d -manifold ($d = 2, 3$) and \subset , a general vector field can be expressed as an ordinary differential equation (ODE) $\dot{\mathbf{x}} = V(\mathbf{x}, t)$. For an *unsteady* (or *time-dependent*) vector field $V(\mathbf{x}, t)$, the trajectory of a particle starting at \mathbf{x}_0 and at time t_0 is called a *pathline*, denoted by $\mathbf{x}_{\mathbf{x}_0, t_0}(t) = \mathbf{x}_0 + \int_{t_0}^t V(\mathbf{x}_{\mathbf{x}_0, t_0}(\tau), t_0 + \tau) d\tau$.

3.1. Local Attributes in Vector Field

There are a number of local attributes that are of interest to domain experts. Given a steady vector field \mathbf{v} , its spatial gradient $\nabla_{\mathbf{x}}\mathbf{v}$ is referred to as its *Jacobian*, denoted by \mathbf{J} . \mathbf{J} can be decomposed as $\mathbf{J} = \mathbf{S} + \mathbf{R}$, where $\mathbf{S} = \frac{1}{2}[\mathbf{J} + (\mathbf{J})^T]$ and $\mathbf{R} = \frac{1}{2}[\mathbf{J} - (\mathbf{J})^T]$ are the symmetric and antisymmetric components of \mathbf{J} , respectively. A number of flow attributes can be derived from \mathbf{v} , \mathbf{J} , \mathbf{S} and \mathbf{R} [PLMH12]. In the examples shown in the paper, we utilize the following local attributes, \mathbf{A}_i , for experimentation.

- \mathbf{A}_1 : vorticity magnitude, $\|\nabla \times \mathbf{v}\|$.
- \mathbf{A}_2 : λ_2 , computed as the second largest eigenvalue of the tensor $\mathbf{S}^2 + \mathbf{R}^2$ [JH95].
- \mathbf{A}_3 : $Q = \frac{1}{2}(\|\mathbf{R}\|^2 - \|\mathbf{S}\|^2)$. [Hun87]
- \mathbf{A}_4 : local shear rate, defined as the Frobenius norm of \mathbf{S} .
- \mathbf{A}_5 : norm of \mathbf{J} , $\sqrt{\sum_{ij} J_{ij}^2}$. [GT18]

The attributes selected in this work are widely used in fluid mechanics to characterize different physical properties of the flow. For example, Q , λ_2 and vorticity are the common attributes for vortex characterization. Local shear rate and the norm of Jacobian are used to study the local divergence and separation behavior in the flow. Note that although not included in the above list, other attributes provided in a given simulation (e.g., acceleration, kinetic energy, pressure and dye) can also be used with our framework. The correlation and dependency of these attributes has been studied in [BMLC19]. One important aspect about the derived attributes is that they change when the reference frames change. As a result, TACs are variant under the movement of reference frames. To make

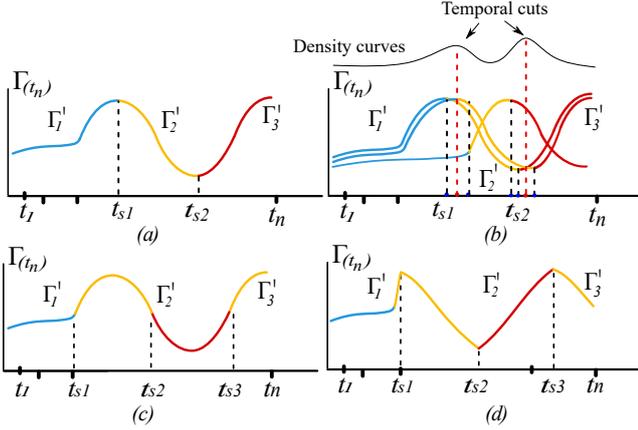


Figure 2: (a) An example TAC with three events. t_{s1} and t_{s2} are two split points defined at the extrema. (b) An example of identifying temporal cuts for temporal clustering. Cuts are selected at the maxima of the density curve. (c) A similar TAC as in (a), but the split points are defined at the maxima and minima of the derivative curve (d).

the TACs independent from the reference frame transformation, we utilize the hyper-objective measure from Günther et al. [GT20] to find the optimal local reference frame in which the flow appears near-steady. Under this reference frame, the velocity and Jacobian become objective which means they are invariant under any smooth affine transformations (rotation, translation and uniform scale). The attributes computed in the near-steady field are also affine invariant. For the complete explanation about the changes of reference frames, please refer to [GT20].

3.2. Time Activity Curve (TAC)

In this section, we describe how TACs are computed based on a given local attribute. We also introduce the concept of split points, which are used to segment a TAC into multiple intervals.

Definition of TAC: Given a local attribute \mathbf{A} , a Lagrangian TAC along a pathline \mathcal{C} of a particle, seeded at \mathbf{x} at time t , can be expressed as:

$$\Gamma_{\mathbf{A},\mathcal{C}}[i] = \mathbf{A}(\mathcal{C}(\mathbf{x}, t_i), t_i) | i = 1, 2 \dots n \quad (1)$$

where $t_1, t_2 \dots t_n$ are the sample times within the time window T and $\mathcal{C}(\mathbf{x}, t_i)$ is the location of \mathbf{x} on the pathline \mathcal{C} at time t_i . Similarly, Eulerian TACs of attributes can be measured at a fixed location over time. For the rest of the discussion, we focus on Lagrangian TACs unless specified otherwise.

To simplify the notation, we denote a Time Activity Curve as $\Gamma = \{\Gamma[i] | i = 1, 2 \dots n\}$ where $\Gamma[i]$ is the local attribute value at time t_i . Figure 2(a) illustrates a TAC where the x axis indicates time indexes and the y axis shows the local attribute values at the corresponding times. The length of a TAC is the number of values in the TAC, indicating the lifespan of the corresponding particle.

Split points: Since TACs are time series data, we want to analyze

their behaviors in different time intervals. To facilitate the temporal characterization, we split a TAC into multiple sub-TACs by using either the extrema or inflection points as shown in Figure 2(a)(c). These split points are later used in the temporal clustering to segment all Γ in identical time intervals (Section 4.3.1).

4. TAC-based Flow Exploration Framework

Overview. Our pipeline consists of two main phases: computation and exploration (Section 5). We concentrate on the computation phase in this section. First, we densely and uniformly sample the particles in the flow domain and compute pathlines. The user can also manually select seed points in the region of interest to compute pathlines. Depending on the attributes of interest, the corresponding TACs are derived and segmented into multiple time intervals (Section 4.3.1). Next, we perform a hierarchical clustering based on the characteristics of the entire TACs (Section 4.2). Based on the global clustering result, we perform a hierarchical temporal clustering of TACs to capture the level-of-detail characterization of their temporal behavior (Section 4.3.2). From the spatio-temporal clustering result, flow exploration (Section 5) is conducted from the following three perspectives: flow space, attribute space and temporal space. In the following we detail our TAC-based clustering.

4.1. TAC-based Similarity Measure

To assist the spatio-temporal clustering of TACs, we first describe our similarity measure for TACs. To compare the difference in the characteristics of two TACs, the similarity measure takes both the temporal trends and magnitude of TACs into account. The traditional distance metrics, such as the Euclidean distance and the Pearson correlation coefficient, concentrate on either the trend or the magnitude of the TACs and cannot satisfy our needs. Figure 3 (left column of (a) and (b)) illustrates the limitations of Euclidean distance and Pearson correlation coefficient in characterizing the difference in a number of representative TACs. Another metric for measuring the similarity of two time series is Dynamic Time Warping (DTW) [LS09b]. DTW considers both shift and deformation of the time series. However, the time stamp for each sample in the TACs has specific meaning which requires us to align the TACs based on the time stamps, making DTW less suitable in our cases.

To address the limitations of the existing similarity measures, we introduce a new TAC Similarity Measure (TSM) to calculate the similarity of TACs based on their spatial and correlation differences. The proposed measure requires linear time to compute; thus, it is practical to apply the measure to large data sets. TSM is defined as follows:

$$D_{tsm}(\Gamma_1, \Gamma_2) = (1 + P_c D_{corr}(\Gamma_1, \Gamma_2)) D_e(\Gamma_1, \Gamma_2) \quad (2)$$

$$D_{corr}(\Gamma_1, \Gamma_2) = 0.5 - \frac{cov(\Gamma_1, \Gamma_2)}{2\sigma_{\Gamma_1}\sigma_{\Gamma_2}} \quad (3)$$

where cov is the covariance and σ_{Γ} is the standard deviation of Γ .

$$D_e(\Gamma_1, \Gamma_2) = \sqrt{\sum_{i=1}^n (\Gamma_1[i] - \Gamma_2[i])^2} \quad (4)$$

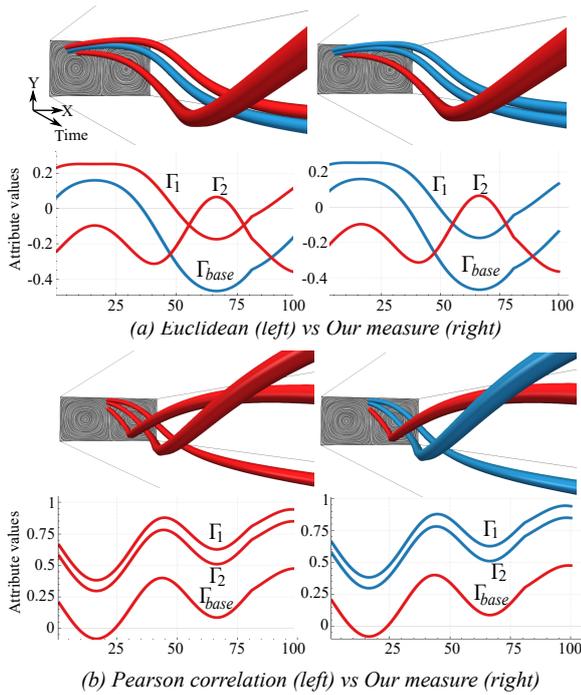


Figure 3: Comparison of our TAC Similarity Measure (TSM) and (a) Euclidean distance metric and (b) Pearson correlation, respectively. The TACs are computed based on the λ_2 attribute on the Double Gyre simulation. Colors represent clusters. In both cases, the difference of TACs cannot be accurately measured by Euclidean distance or Pearson correlation (left column). (a) $D_e(\Gamma_{base}, \Gamma_1) = 42.32 > D_e(\Gamma_{base}, \Gamma_2) = 38.87$, using the Euclidean distance. (b) $D_p(\Gamma_1, \Gamma_2) = D_p(\Gamma_2, \Gamma_{base}) = 1$ using Pearson correlation, resulting in all of them belonging to the same group. TSM can differentiate the behavior of TACs more accurately in both cases (right column): (a) $D_{tsm}(\Gamma_{base}, \Gamma_1) = 42.32 < D_{tsm}(\Gamma_{base}, \Gamma_2) = 72.45$ (b) $D_{tsm}(\Gamma_1, \Gamma_2) = 18.87 < D_{tsm}(\Gamma_2, \Gamma_{base}) = 32.61$.

In the above definition, $D_e(\Gamma_1, \Gamma_2)$ represents the Euclidean spatial distance between the two TACs Γ_1 and Γ_2 . $D_e(\Gamma_1, \Gamma_2)$ addresses the challenges illustrated in the left column of Figure 3(a) where TACs exhibit similar trends but a different spatio-temporal distance measure. $D_{corr}(\Gamma_1, \Gamma_2)$ measures the global correlation between TACs Γ_1 and Γ_2 . If $D_{corr}(\Gamma_1, \Gamma_2) = 1$, then the first term in Equation 3 ($1 + P_c \times D_{corr}(\Gamma_1, \Gamma_2)$) = 2 which means Γ_1 and Γ_2 have opposite trends. In contrast, if $D_{corr}(\Gamma_1, \Gamma_2) = 0$, Γ_1 and Γ_2 have the same trend and the value of the first term is 1. $D_{corr}(\Gamma_1, \Gamma_2)$ aims to resolve the ambiguity illustrated in Figure 3(a) where Γ_1 and Γ_2 have similar Euclidean distance, but different correlation distance relative to Γ_{base} . By using $D_{corr}(\Gamma_1, \Gamma_2)$, the TSM measure can differentiate Γ_1 and Γ_2 ; hence it groups Γ_1 to Γ_{base} rather than Γ_2 illustrated in the right column of Figure 3(a). $D_{corr}(\Gamma_1, \Gamma_2)$ introduces a penalty factor P_c , which represents a user-assigned importance for the spatial difference and the global correlation, respectively. The higher value of P_c , the more weight given to the global correlation. By default, we set $P_c = 1$.

We use multiplication instead of addition to combine the first and second terms in Eq. 2 due to the relation between the two terms, and their value range difference. The first term is equal to 1 when the two TACs contain similar trends. In this case, the distance between two TACs is completely based on the second term (i.e., Euclidean distance). If the two TACs have opposite trends, then the distance between the TACs is expected to be large. By multiplying, we magnify the second term by a maximum of two when the TACs have inverse trends.

Figure 3 illustrates the advantages of TSM (right column) over the Euclidean distance and Pearson correlation. In both cases, TSM can differentiate the behavior of TACs more accurately.

With the above similarity measure, we can develop a TAC-based pathline selection and exploration strategy which allows us to highlight pathlines whose TACs exhibit certain specific characteristics (i.e., the distance of their TACs to a reference TAC is smaller than a threshold). We will defer the discussion of this functionality until the results.

4.2. TAC-based Clustering

In order to provide different levels of detail for flow behavior w.r.t. the local attributes, we perform the clustering of TACs using the new similarity measure over all temporal samples, coupled with the popular agglomerative hierarchical clustering (AHC). The linkage type used in this work is the complete linkage since it is better for finding compact clusters of approximately equal diameter [Def77]. To reduce the traditional cubic time complexity, we implement the parallel, locally-ordered AHC proposed by Walter et al. [WBKP08], which runs in sub-quadratic time.

AHC gives rise to a hierarchical tree with each node representing a cluster and each bifurcation representing a merging (see an example shown in view 2 of Figure 6). The different heights of the hierarchical tree which correspond to the distance values of the two clusters indicate the merging order. The lower the level, the sooner the leaf nodes will be merged. A sample merging order can be found in Figure 7 where the hierarchical tree with six leaf nodes is reduced to three nodes. As the number of clusters represents the level of abstraction, increasing the number of clusters results in more details to be revealed as shown in Figure 7 (top row). Choosing a suitable number of clusters of pathlines to reveal the most interesting flow behavior is not trivial, and it is a trade off between the details of flow behavior and the clearness of the cluster structures. In practice, it is an exploration process. Since the hierarchical tree is a natural product of AHC, a user can interactively select the number of clusters to show after the clustering computation (i.e., no re-clustering is needed).

4.3. TAC-based Temporal AHC

Two TACs that belong to two clusters may possess local segments having similar behavior (Figure 2(c-d)), which cannot be captured in the above global clustering along the entire time range. To address this, we propose a hierarchical clustering algorithm in the temporal dimension, i.e., a temporal AHC.

We aim to address following technical challenges to achieve temporal AHC: (1) identify the appropriate temporal partitioning; (2)

perform the AHC within each time interval obtained from step one; (3) handle the transition of AHC results between consecutive time intervals. Solving these problems is not trivial. First, all TACs have different temporal behavior. Even the TACs that belong to the same cluster may exhibit slightly different behavior, which makes the selection of cutting points (or cuts) for temporal partitioning difficult. Second, the AHC performed on individual time intervals and along the time axis should be consistent in terms of the error threshold and similarity characterization of the clusters. Third, the AHC results obtained in consecutive time intervals may not be identical. It is important to keep track of their transition relation (i.e., bifurcate or merge) across the cuts. In the following, we detail our solutions to these challenges.

4.3.1. Time Interval Segmentation of TAC

To study TACs in a level-of-detail fashion, we apply time interval segmentation to a group of TACs. The time intervals that segment TACs must preserve TAC characteristics. In other words, one primitive trend of a TAC is not expected to be segmented into two time intervals, which causes fragmentation. For an individual TAC, we can simply apply 1D Morse decomposition to generate the temporal sequences of TAC segments, as shown in Figure 2(b). However, for a group of TACs, it is not guaranteed that the segment split points are identical. To address this, we utilize a 1D Gaussian kernel density estimation (KDE) and choose the point with the highest estimated density as the split point. Specifically, we first identify the inflection points for each TAC. Let x_1, x_2, \dots, x_n be a set of 1-dimensional inflection points on \mathbb{R} and let H be a positive definite bandwidth value. The univariate fixed bandwidth kernel estimator is defined as [Sil86]: $f(x) = \frac{1}{nH} \sum_{i=1}^n K(\frac{x-x_i}{H})$, where K is the Gaussian kernel $K(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$. Selection of the bandwidth value H is important in KDE as it can make the density estimate smoother or noisier. However, in our case, the value of H does not affect the position of the point with the highest estimated density. Thus, by default we set H to 1.

The cutting points obtained from the above KDE segment the entire time period T into a number of intervals of varying length, referred to as $T = \langle T_1, T_2, \dots, T_m \rangle$. In this way, all TACs are segmented by these splitting points which attempt to preserve the most common characteristics of all TACs. An example of time interval segmentation is illustrated in Figure 2(c).

4.3.2. Temporal Hierarchical Clustering

After performing the temporal partitioning and obtaining the local time intervals, we now perform AHC within each interval. We apply the proposed similarity measure (Eq. 2) for temporal clustering.

Assigning the cluster number for each time interval is difficult because, on one hand, the number of time intervals in time-hierarchical clustering varies, while on the other hand, the cluster distances in different time intervals may be different. To show the consistent changes across time intervals, the same treatment needs to be applied uniformly. Therefore, we use the distance threshold ϵ for the global clustering to guide the clustering within individual time intervals. Specifically, the distance threshold ϵ_i for time interval T_i is determined by the time range of the interval, i.e., $\epsilon_i = \frac{|T_i|}{T} \epsilon$.

In this way, it is foreseeable that there are more clusters generated in the time intervals where the TACs behave more diversely, i.e., when TACs have larger dissimilarities.

The goal of temporal clustering is to build up a hierarchical tree of the input m time intervals obtained in the previous temporal partitioning, i.e., m leaf nodes of the tree, so that the level-of-detail of a TAC's behavior can be observed in the temporal dimension. In contrast to the spatial hierarchical clustering, in which any two clusters can be selected for a merging operation, in temporal hierarchical clustering only two clusters that are contiguous in time can be merged together, which makes the merging operation simpler. In our implementation, starting from the initial m leaf nodes (i.e., m initial time intervals), a distance array $D \in \mathbb{R}^{(m-1)}$ is created. Each entry indicates the dissimilarity after a pair of consecutive time intervals are merged into one. $D[i]$ can be computed as follows.

$$D[i] = \eta(T_i) + \eta(T_{i+1}) - \eta(T_i \cup T_{i+1}) \quad (5)$$

where $\eta(T_k)$ is the average pairwise dissimilarity within a time interval T_k , $T_i \cup T_{i+1}$ is the new time interval obtained by merging T_i and T_{i+1} .

$$\eta(T_k) = \sqrt{\frac{\sum_{i=1}^{|T_k|} \sum_{j=i+1}^{|T_k|} (D_{tsm}(\Gamma_i, \Gamma_j))^2}{|T_k|(|T_k| - 1)}} \quad (6)$$

$\eta(T_k)$ reflects the compactness of the TACs in the time interval T_k . The larger the value of $\eta(T_k)$, the further the TACs in T_k are located from the centroid.

In spatial AHC, the two clusters with the smallest distance are selected for merging. Similarly, in temporal hierarchical clustering, the two time intervals with the smallest dissimilarity changes are merged together first. In other words, time interval T_k and its neighboring time interval T_{k+1} that satisfies $D[k] \leq D[i], \forall 1 \leq i \leq M$, are first merged together to generate a new time interval $T_k + T_{k+1}$ and then removed from the node list. Consequently, a new $m - 2$ dimension distance array $D_{(m-2)}$ is generated with the remaining $m - 1$ nodes. The above merge process is iterated until only one time interval, i.e., the entire time period, remains as the root of the temporal hierarchical tree. The height of the temporal hierarchical tree built on m time intervals is $m - 1$. On the i^{th} level of the tree, i.e., the height is i , there are $m - i$ time intervals. This temporal clustering strategy can be applied to all TACs or a subset of TACs grouped based on the global clustering results from Section 4.2. For the latter, different numbers of temporal cuts (and temporal segments) may be resulted for different global clusters, depending on the overall TAC behavior of each cluster (see Figure 10(c) for an example.)

4.3.3. Visualization of TAC Clusters

An improved edge bundling technique. To visualize TAC clusters with less cluttering, we adapt the edge bundling technique for parallel coordinate plot visualization by Palmas et al. [PBO*14]. Details of our adaption can be found in the supplemental document. To ensure the color consistency for the temporal clustering visualization of TACs, we assign a color to a cluster C_p based on its *main source cluster*, i.e., the cluster from which most TACs in C_p

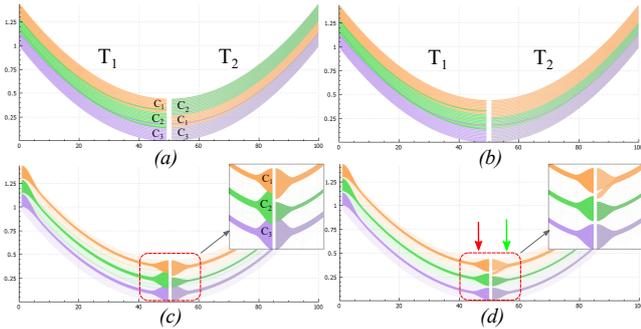


Figure 4: Visualization of transition between time intervals. Results before (a) and after (b) cluster ID adjustment, respectively. (c) edge-bundling visualization of the result. (d) modified edge-bundling visualization. Magnified views show the transition between two time intervals.

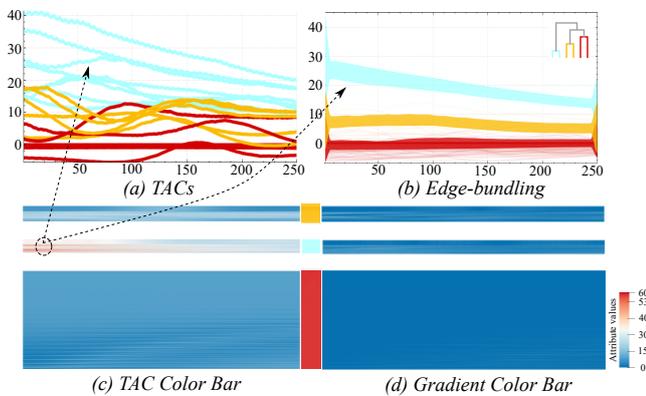


Figure 5: The visual comparison of edge-bundling with 1D TAC color plot. (a) Actual TACs exhibit occlusion. (b) Edge-bundling visualization provides an overview of each group of TACs. (c) The smooth 1D stacked color plot means that the TACs are grouped effectively. (d) The uniform blue color in the gradient color plot indicates the similarity among neighboring TACs in each cluster.

originate in the previous time interval. For example, in Figure 4(a), cluster C_1 in T_1 is the main source of cluster C_2 in T_2 . Thus, the color of C_2 in T_2 will be set in a manner consistent with C_1 in T_1 (Figure 4(b)).

Visual overlapping persists at two ends of edge bundle as shown in Figure 4(c). To address this limitation, we offset proportionally to clusters' size, whose heads or tails are overlapping. As illustrated by the red arrow in Figure 4(d), the minimum value of C_2 at the tail end is increased and the maximum value of C_3 at the tail end is decreased, eliminating the overlapping between C_2 and C_3 while preserving the relative range size simultaneously. Removing overlapping at the tail of T_{k-1} makes the boundaries of source clusters clear. To fully resolve the connections among time intervals, we visualize both main and minor sources at the head of a cluster. From Figure 4(d), we can easily ascertain the transition of clusters between two time intervals.

A 2D stack plot. Although edge-bundling visualization is an effective way to provide an overview about a group of TACs, the detailed behavior of individual TACs and the difference between TACs in the group is not conveyed effectively (see the cyan TAC cluster shown in Figure 5(b)). To address this, we visualize each TAC using a 1D bar, whose colors are determined by attribute values of the TAC over time. We then stack these 1D color plots to form a 2D color plot (Figure 5(c)). Note that TACs belonging to the same cluster are rendered next to each other. With this condensed representation, one can easily assess the clustering quality. That is, if the color in this 2D plot is smooth, it means that the neighboring TACs have similar characteristics, indicating a good clustering result. The distance between two neighboring TACs is also converted to the 1D color bar to create the gradient plot. If neighboring TACs have similar patterns, the gradient between them is small, then the plot exhibits mostly uniform color.

5. TAC-based Flow Visualization System

A multiple coordinated view system is developed to enable visualization and exploration of unsteady flows using TACs. Users can flexibly switch between different views to perform analysis and comparison. Our system provides visualizations in the temporal space, attribute space and the original flow domain. Figure 6 illustrates the user interface. ① The top left view has multiple tabs showing TACs with temporal and global clustering results. ④ The center view is the control panel that provides functions for user interaction. ③ The bottom view shows a Line Integral Convolution (LIC) [CL93] image generated for the first time step, and enables users to select the region of interest (ROI). ⑤ The top right view visualizes pathlines in the space-time domain and the detailed TACs (⑥) in the selected region of interest. View ② shows the hierarchical clustering tree. Users can interact with the system and explore the clustering results in four ways: (1) The user can select a region of interest in view ③ and inspect the behaviors of the particles seeded within this region. A volume rendering coupled with iso-surfacing is used to visualize the behavior of the pathlines, which are colored based on the spatio-temporal clustering in view ⑤. (2) The user can choose a specific cluster in view ① to analyze its TACs' behaviors and highlight specific TAC and its corresponding pathline (in view ⑤). (3) The user can choose a temporal cluster to analyze TACs' behavior within a specific time interval. (4) The user can manipulate the error threshold or number of clusters in view ② to inspect the abstract flow behavior with different levels of details (see Figure 7 for an example).

5.1. TAC-based Exploration

In addition to the TAC-based clustering framework introduced above, which is suitable for providing the overview of the behaviors of the unsteady flows. Our visual analytic system also supports a number of exploration functionality based on TACs.

TAC-based pattern search. Similar to the pattern search from a set of integral curves based on some template curve and its geometric characteristics, the TSM measure for TACs can be applied to perform the pattern search based on TACs' behaviors. Unlike clustering which considers all pairwise TAC distances, the pattern

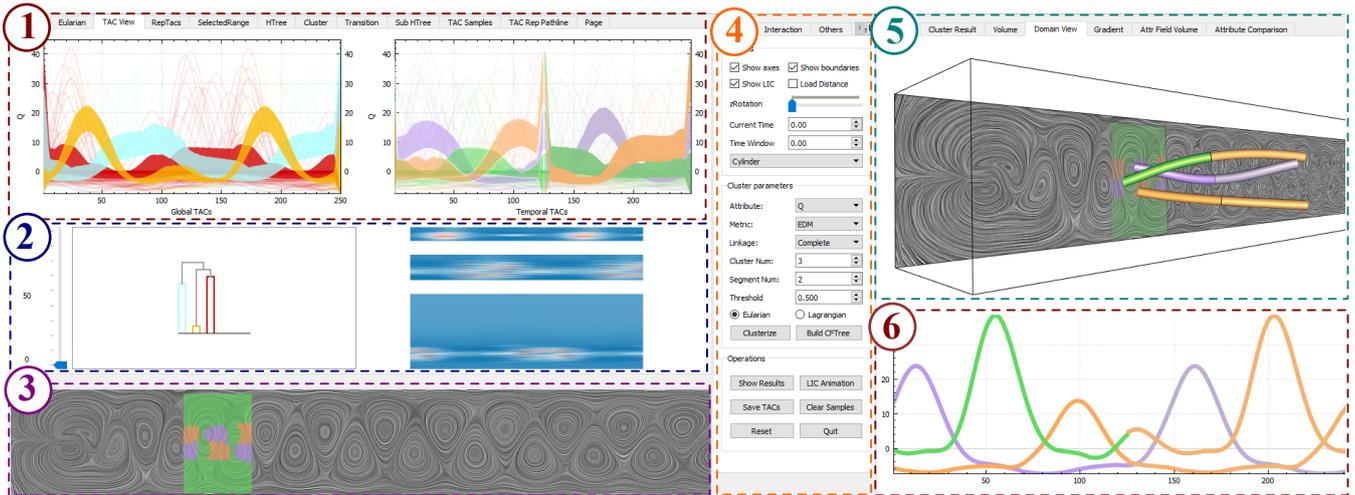


Figure 6: User interface of our system. (1) Multiple tab views show TACs with temporal and global clustering results. (2) Hierarchical Tree generated from the AHC and the TAC color bars. (3) the Line Integral Convolution (LIC) image generated for the first time step, allowing users to select the region of interest (ROI). (4) The control panel providing functions for user interaction. (5) Pathlines in the flow domain and (6) the detailed TACs in the selected region of interest.

search starts with a reference TAC that can be selected by domain experts. The system then returns the most similar TACs and their corresponding pathlines. This feature provides the freedom to users to customize the TAC characteristics interesting to them (Figure 12(b)).

Input filtering. Instead of using a set of densely placed pathlines in the first time step of the data sets, the experts may wish to focus on a specific subset of pathlines based on some prior knowledge about the flow. For instance, in the study of the relation between shear layer and the vortex formation, the users may select a subset of pathlines seeded within the shear layer of the flow (i.e. regions with negative Q values) and study the characteristics of their corresponding TACs in hope with finding the pathlines starting from the shear layer that may participate in the formation of vortices (i.e. entering vortex regions) in a later time (Figure 13(b)). On the other hand, the experts may be interested in certain features arising in a later time and wish to see the origin of the particles that enter these features. In this case, rather than computing the pathlines using forward tracing, backward tracing can be used to compute the pathlines starting from the feature areas. In all these ad-hoc exploration, our system can perform clustering and visualization on only a subset of pathlines.

6. Applications

We have utilized our TAC-based exploration framework to help our experts from aerospace engineering and mechanical engineering analyze vortex structures in different situations (Sections 6.1) and help reveal subtle difference in the seemingly symmetric flow behavior (Section 6.2).

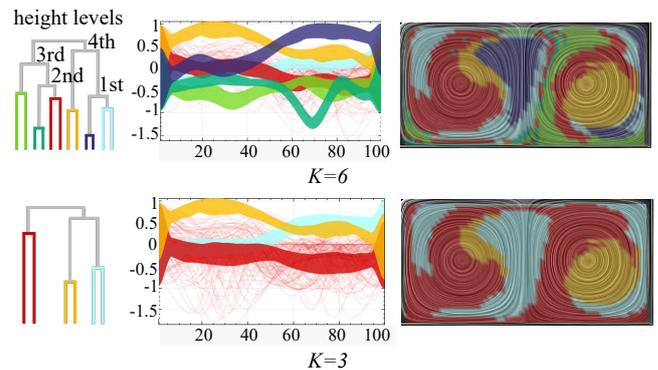


Figure 7: Clustering results of the Double Gyre flow using Lagrangian TACs of attribute curl with different numbers of clusters ($K=6, 3$ from top to bottom, respectively), specified by the user in the hierarchical tree view (left column). The height levels of the hierarchical tree indicate the merging order. The lower level, the sooner the leaf nodes will be merged. The middle two columns show the TAC clusters, stacked plots and their representative TAC curves. The right column shows the clusters in the flow domain. The time window is $T = 100 \times 0.01$ with 100×50 sampling points.

6.1. Vortex Structure Analysis

In the following, we demonstrate how we apply the proposed TAC-based clustering to help study vortex behavior in a number of 2D and 3D unsteady flows. Vortices are one of the most important dynamics in flow that often relate to energy/material transport and mixing [Lei16]. The attributes applied in the following studies are mostly Q , λ_2 , and vorticity (or curl in 2D). Although these attributes have different temporal trends, they often result in similar

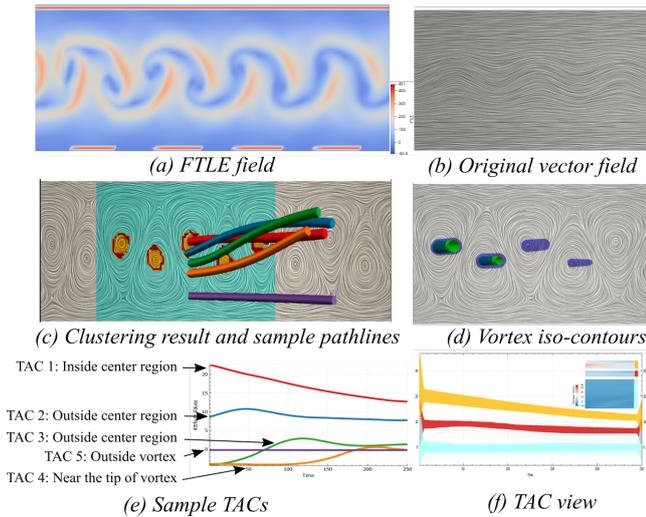


Figure 8: Clustering result of the 2D cylinder flow using TACs of attribute Q reveals a three-layered configuration of the vortex system ($p_c = 100$). (a) FTLE field shows the Lagrangian coherent structures. (b) The LIC texture computed from the original velocity field does not reveal the vortex structures. (c) Pathlines sampled along a vertical line passing the center of a vortex and (e) their corresponding TACs. The TAC (red) of the pathline seeded in the center region decreases monotonically over time, indicating the diffusion of the concentrated vorticity, which gradually increases the vorticity in the outer regions of the shedded vortices (TACs 2-4). TAC 5 corresponds to the pathline seeded outside of vortex region, which exhibits stable characteristic. (d) The two iso-contours (blue and green) with Q values of 17 and 24, respectively, cannot fully capture the vortex configuration. The LIC texture is computed from the velocity in the optimal reference frame (f) TAC profiles of our clustering results.

clustering results as shown in the supplemental document. In the following, we only provide the representative result for each data set using one of these attributes. In practice, the user should explore different attributes to identify the attribute that best reveals the flow behavior of interest.

2D flow behind a cylinder. Next, we apply our technique to a 2D simulation of the flow behind a square cylinder with a Reynolds number of 160 [WT10]. The simulation covers a subset of the spatio-temporal domain, $[-0.5, 7.5] \times [-0.5, 0.5] \times [15, 23]$, where the vortex shedding is fully formed. According to the domain experts, the core region of a vortex in this flow has a motion close to that of a rigid body rotation, which helps to preserve the shape of the vortex. However, the concentrated vorticity in the vortex cores will diffuse due to viscosity (i.e., friction) and the absence of an external forces to maintain the rotation [Lei16]. The diffused vorticity will reach the outer layer of vortices where it interacts with vorticity from other vortices, thus losing the coherent character. The overall structure of the vortices is stable due to interleaving and somewhat symmetric configuration of the counter-rotating vortices.

We choose the first 250 time steps of this simulation and use

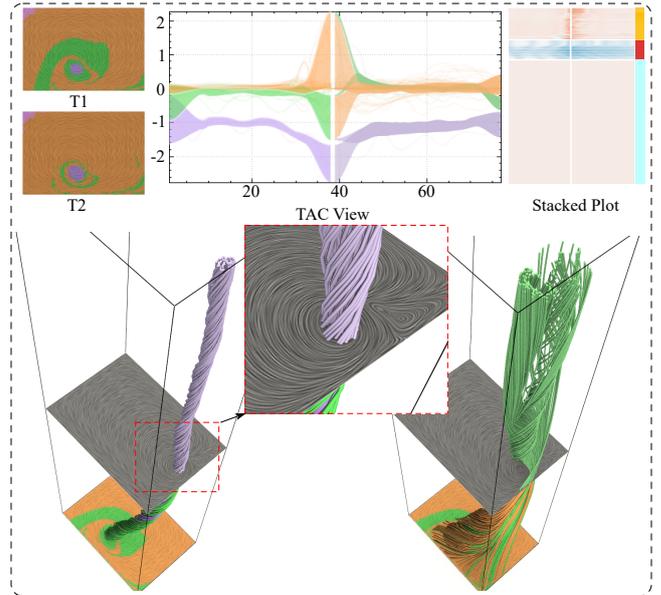


Figure 9: Temporal clustering results of the 2D vortex ring ($p_c = 1$). TACs are computed by using the λ_2 attribute. The split points for the KDE computation are selected at the extrema values. The temporal cut indicates the moment when the main vortex impacts the wall.

a spatial resolution of 1200×150 to compute pathlines and measure the attributes along them. As demonstrated in Figure 8(a), our framework identifies three regions using the TACs of the Q attribute without significant user intervention: the viscous vortex core where the vorticity is concentrated, the outer layer of the vortices where vorticity diffuses and grows and the region outside of the vortices where the flow is irrotational. In addition, the TACs' visualization in Figure 8(d) informatively characterizes the attribute behaviors within different flow regions. Specifically, the decay of the rotational momentum of the vortex core as expected by the experts is clearly depicted by the monotonic decrease of the orange TAC that illustrates the vortex core behavior. In contrast, the traditional iso-contouring (or iso-surfacing in the space-time domain) has a difficult time to depict this configuration. For instance, Figure 8 (c) shows two iso-surfaces computed with two different Q values. Due to the decrease of Q concentration from left to right in space, the selected thresholds may not lead to iso-surfaces to depict the behavior of vortices in the far right of the flow, whose Q concentration may be similar to other regions without a vortex.

In a detailed study of the above behavior of vortices, we sample 5 pathlines along a vertical line passing the center of a vortex (Figure 8 (a-b)). Clearly, we see three different types of TAC behaviors: (1) the decaying of Q concentration over time along the core (the red TAC/pathline); (2) the increasing and shifting of the peak Q values of the TACs corresponding to the pathlines seeded at locations gradually moving away from the vortex core (blue, green, and orange TACs/pathlines); and (3) a flat TAC/pathline (purple). While types (1) and (3) are easily understandable, the behavior of the TACs in group (2) is interesting. On the one hand, one can see

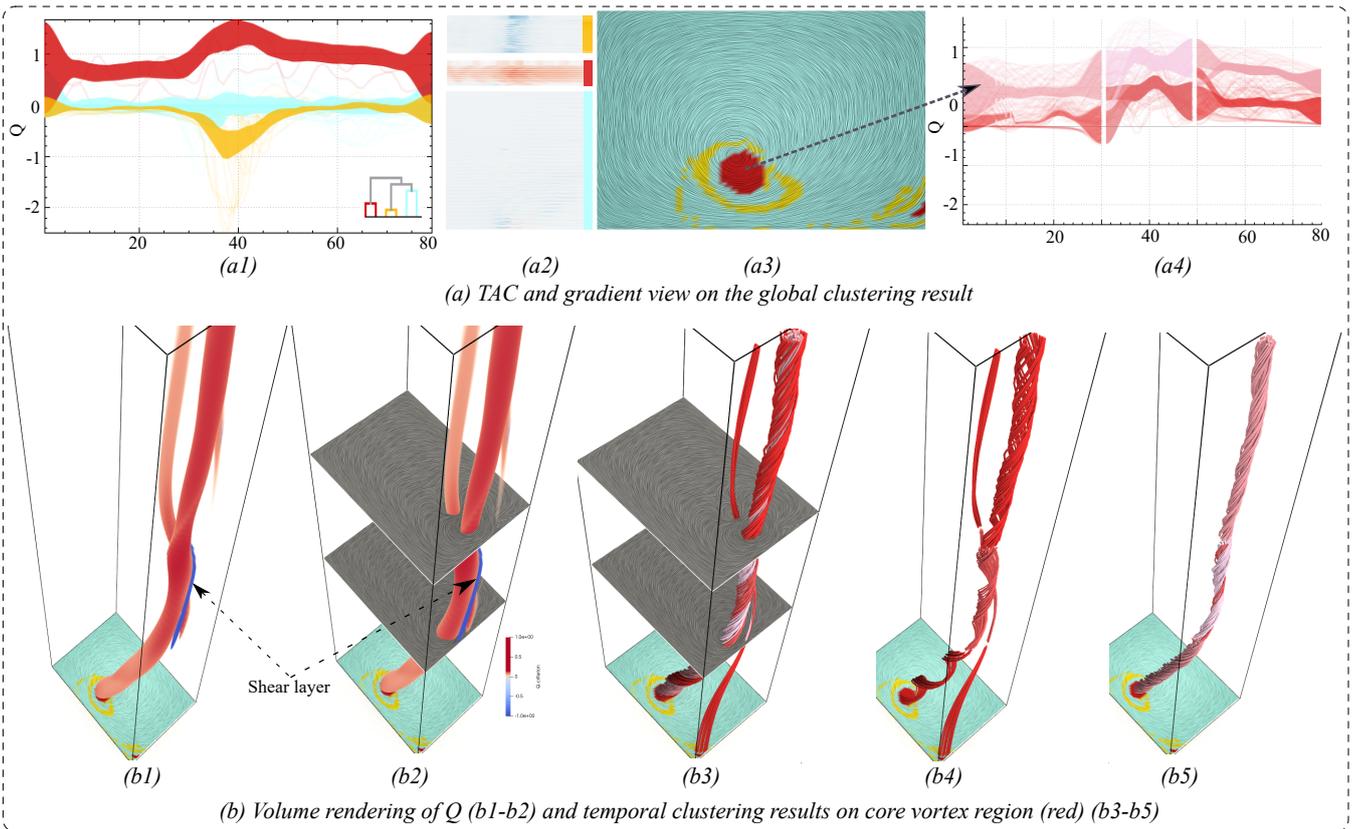


Figure 10: Clustering results of the vortex ring ($p_c = 1$). TACs are computed by using the Q attribute. (a) Global clustering result. (a1) - TAC view, (a2) - Stacked plot (a3) - LIC texture view (a4) - The temporal clustering results on the core vortex region (red). (b1) Original volume rendering of Q . (b2) Volume rendering of Q with two temporal cuts generated by our methods. The second temporal segment aligns with the shear layer when the main vortex collides and sticks with the boundary wall before the secondary vortex is created. (b3-b5) Pathline view of the temporal clustering results computed based on pathlines and their respective TACs in the core vortex region (a4) reveals the behaviors of the two main vortices in the simulation.

the correlation of the peak locations of these TACs with the changing direction (or turning) of their respective pathlines. This is important, as it associates the geometric characteristics of pathlines with relevant physics. On the other hand, the shifting of the peaks in part indicates the propagation of the rotation momentum outwardly from the vortex core. The increase from negative Q values to positive Q values for the green and orange TACs also associate the shearing layer ($Q < 0$) with its corresponding vortex region. Such a detailed behavior, though known by experts, has not been studied for the flow behind cylinder in the visualization community and cannot be easily obtained with other methods alone.

2D vortex ring. The next 2D data set simulates a vortex ring hitting a wall with a Reynolds number of 2000. During the interaction, the vortex ring approaches the wall and causes a boundary layer to appear. As the vortex slides against the wall, the boundary layer becomes unstable and is lifted up as a secondary vortex, which in turn lifts up the primary vortex. This data set helps us analyze the role of coherent structures interacting with boundaries, and the generation of turbulence in wall-bounded flows. Our temporal segmentation result using the λ_2 attribute is shown in Figure 9, which

demonstrates that our method can detect the moment when the vortex impacts the wall, and automatically generates a temporal cut at that time. The candidate cuts for each TAC are determined based on the extrema of the TAC. The global clustering results with the Q attribute are shown in Figure 10(a1). The 2D stack color plot in Figure 10(a2) allows us to assess the quality of the clustering results, which is hidden in the edge-bundling visualization. The patterns of TACs in each cluster looks alike. In particular, the red group has stronger positive Q values indicating the vortex core area, while the yellow group corresponds to the particles having negative Q . Also, the peak location of the yellow group indicates the appearance of the shear layer when the secondary vortex is induced and lifted from the wall.

To separate the two vortices in this simulation, we apply the temporal clustering to the dark red group which corresponds to the vortex core region. The results shown in Figure 10(b) illustrate that we can extract the main vortex (pink) and the secondary vortex (red) thanks to the difference in their respective TAC profiles. In comparison with traditional techniques such as volume rendering (b1-b2) or iso-surfacing, our method requires fewer selections of

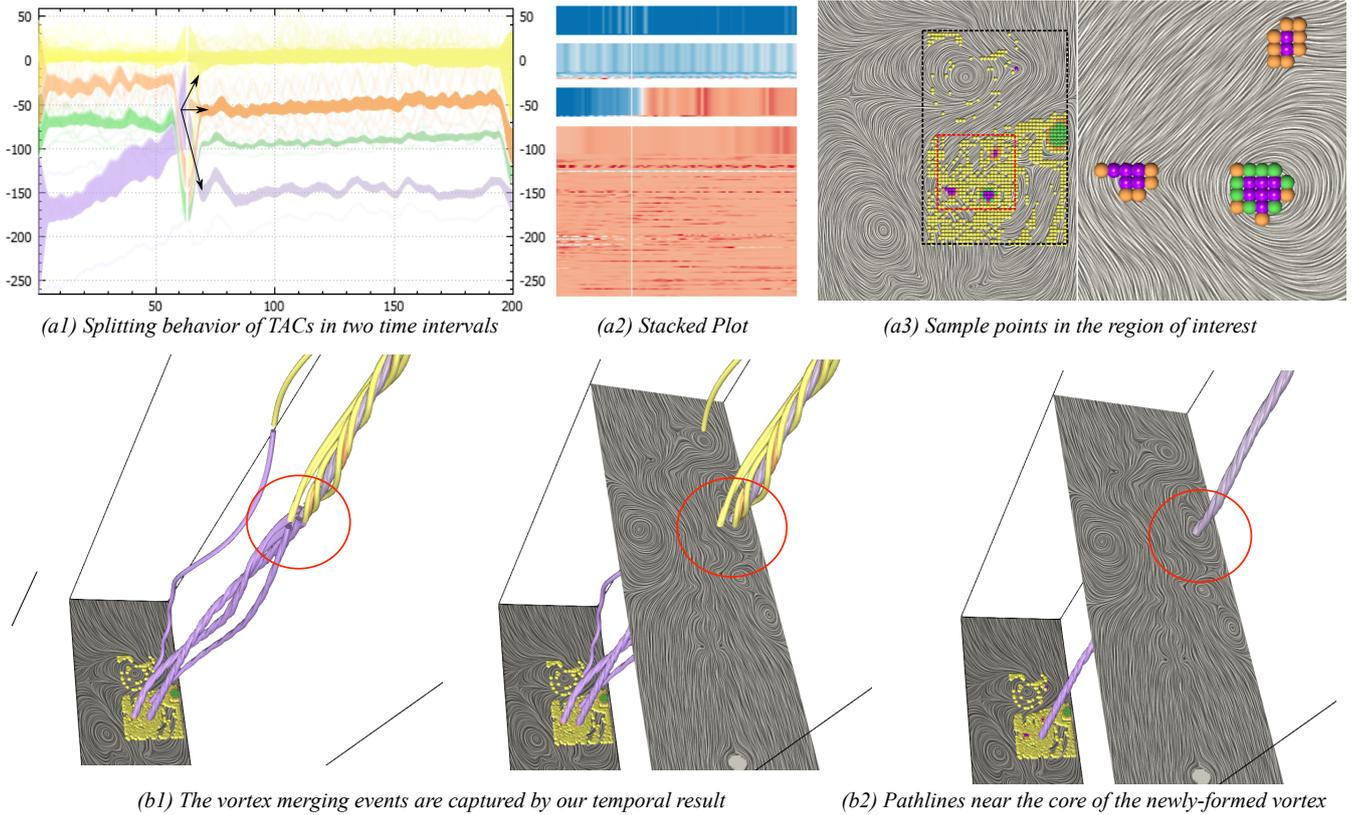


Figure 11: Temporal clustering result of the Boussinesq flow using Lagrangian TACs of the attribute λ_2 ($p_c=1$). (a1) The TAC view shows the splitting behavior of TACs in the two contiguous time intervals. (a2) The stacked plot. (a3) Sample points from four clusters in the region of interest. The close up view on the right focus on three clusters enclosing the vortex regions. (b1) Pathline visualizations show the vortex merging events. (b2) Pathlines near the core of the newly-formed vortex.

thresholds to produce the similar result. In addition, pathlines can provide more detail behaviors of the vortices (e.g., the rotation degree and direction (b4)). It also illustrates the origins of the particles that involve in the generation of the secondary vortex (i.e., the small cluster of pathlines in Figure 10 (b3,b4)), which has not been shown previously for this flow. Knowing this is important to understand the dynamics of boundary shear layer of the flow. The three generated temporal segments (of TACs) in Figure 10(a4,b) reveal three main physical events that domain experts care about: before/during/after the main vortex hitting the wall. Here, the temporal cuts were generated based on the split points determined by the derivative curve of the TACs (Figure 2(c-d)). In particular, the event of impact starts with the occurrence of the shear layer (shown in (b1-b2) as well as indicated by the yellow cluster in (b1)) and ends with this shear layer becomes the secondary vortex.

Boussinesq Figure 11 shows the temporal clustering results of TACs based on the λ_2 attribute for the Boussinesq flow [GT20]. This flow has numerous vortices with varying sizes that in part rotate around and collide with each other. The simulation has a dimension of 150×450 , and we used 1000 time slices which correspond to one half of the time window. Figure 11(a1) shows two temporal segmentations of four TAC clusters. The multi-layered

vortex structure is also revealed here. According to the λ_2 -criterion, a region is considered as a vortex if its λ_2 value is smaller than zero [GT18]. Thus, the green and violet clusters correspond to the core regions of the vortices, while the orange cluster captures the outer vortex regions. The yellow TACs are outside of the vortex region. For the small-scale vortices, the core areas are encoded in the violet cluster because of their smaller negative λ_2 values. The positions of the vortices are highlighted in (a3). Three nearby vortices in the violet clusters coalesce into a new vortex in a later time. Our temporal segmentation can not only capture the moment the event happens as shown in (b1), but also reveal the physical transportation of particles in the simulation. In particular, the splitting behavior of TACs in (a1) indicates that some particles from the core region (violet) remain in the core, others move to the outer vortex core region (orange) or exit the vortex area (yellow). Such a detailed temporal behavior of small-scale vortices is not easy to obtain with the previous techniques (e.g, geometric-based pathline clustering or thresholding).

3D vortex tube simulation. We also performed experiments using a 3D vortex tube simulation, which simulates two parallel, counter-rotating vortex tubes at a circulation-based flow with $Re = 3500$ and a distance of 2.5 radii apart. The two vortex tubes undergo

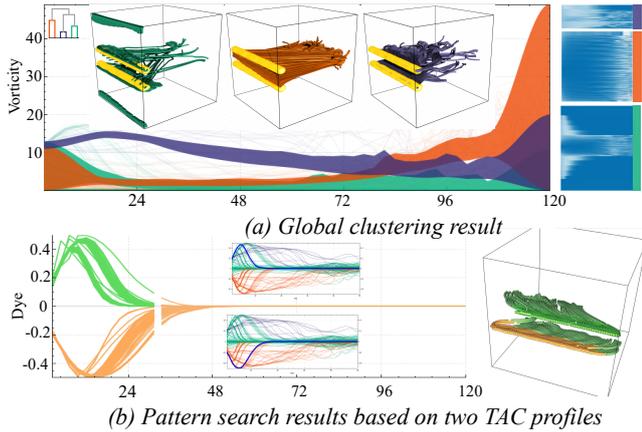


Figure 12: (a) Global clustering result of the 3D vortex tube simulation with TACs of the vorticity attribute that reveals a three-layered structure ($p_c=1$). (b) Pattern search results: Given two groups of TACs with opposite trends, we can find a group of similar TACs and their corresponding pathlines that have symmetric geometric configuration. The attribute used in (b) is dye, a tracer simulated as a passive scalar with a Schmidt number of unity [BMLC19].

an elliptical instability [SLD10] that ends with a vortex disintegration. The two vortices interact with each other mainly through the strain produced by the differential velocities induced. The simulation has dimensions of $360 \times 360 \times 360 \times 120$ in a volume of $[0, 2] \times [0, 2] \times [0, 2]$. The total size of this simulation is 36GB. Again, the pathlines are seeded at the left boundary plane, and vorticity is used here. The results shown in Figure 12 reveal a three-layered structure. The blue cluster includes particles with higher vorticity residing in the two vortex core areas. The green cluster involves particles residing in the outer layer of the two vortices. Their respective TACs are relatively stable. The orange cluster corresponds to particles seeded in the region between two vortices which becomes turbulent at a later time. Accordingly, the vorticity values for this group of particles increase substantially at later times when the flow becomes turbulent.

6.2. Exploration of Other Flow Features

In addition to the above analysis of the temporal behaviors of vortex structures within various 2D and 3D unsteady flows, we also apply our framework to study other flow behaviors.

Study subtle difference in axis-symmetric flows. Figure 13 shows the clustering results using TACs of the attribute Q for the 3D simulated flow behind a cylinder [vFWTS08] with dimensions of $192 \times 64 \times 48$ in the volume of $[-12, 20] \times [-4, 4] \times [0, 6]$ with 101 time steps. Considering the transitional nature of this flow, we select a seeding plane near the left boundary (i.e., $X = -11$) with 64×48 uniform samples. Pathlines that leave the domain earlier are discarded. From the result, we see that the symmetric pathline behavior is captured by the clustering (i.e., the orange and fern clusters in Figure 13 (a1-a2)). In addition, their TACs reveal a clear

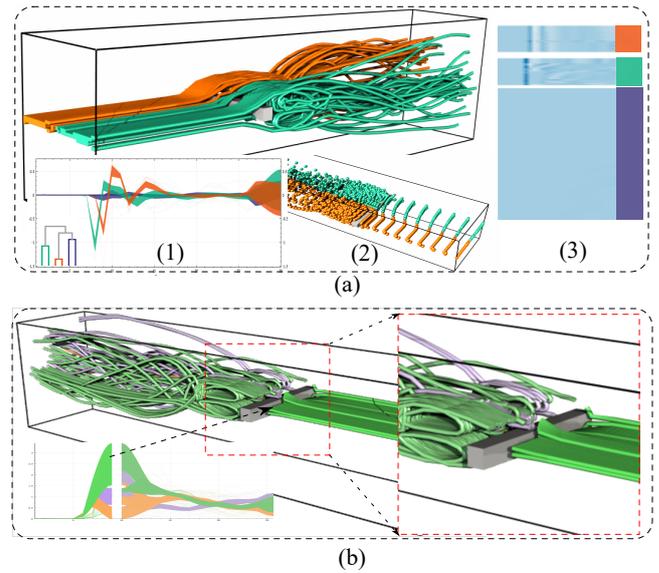


Figure 13: Clustering result of the 3D Cylinder flow ($p_c = 1$). (a) Global clustering result with TACs computed from the Q attribute. The TAC profiles corresponding to the two symmetry pathline groups at the center location show a shifting, indicating particles in these two regions exhibit vortex shedding at different times. (b) Temporal clustering result with TACs computed from the local shear rate. The pathline view focuses on the green cluster whose corresponding pathlines have different behaviors in the second time interval (i.e., splitting into three clusters). From the physics point, the vortex regions (i.e., with high positive Q values) usually have little or no shearing flow (i.e., with low local shear rate), and vice versa. This negative correlation can be also observed with our TACs in which the moments that TACs have negative Q are identical to the moments that they have high positive local shear rate.

shifting in their temporal trends, indicating that their similar pathline configurations occur at different times. This information cannot be easily obtained by inspecting only pathlines.

In the example shown in Figure 12(b), our domain experts observe two interesting TACs that have opposite trends. They select two sample TACs and perform the pattern search. Note that the pattern search is a commonly used operation in the integral curve analysis [WESW17] as it helps to locate similar features in a dense and visual cluttered set of curves. The obtained results show an intrigued pattern. The two set of TACs are symmetric and the directions of their corresponding pathlines are opposite. Figure 15 demonstrates another use case with the 3D flow behind a cylinder where our users can find a group of particles that move stably and form loops after the flow collides with the cylinder.

Studying separation behaviors in flow. Flow separation in unsteady flows is an important dynamic that experts are interested. Finite Time Lyapunov Exponent (FTLE) [Hal01] is typically used for highlighting the locations where flow separation is the strongest. However, it cannot reveal the cause of the separation. To demonstrate how our framework can help explain this to some extent, we

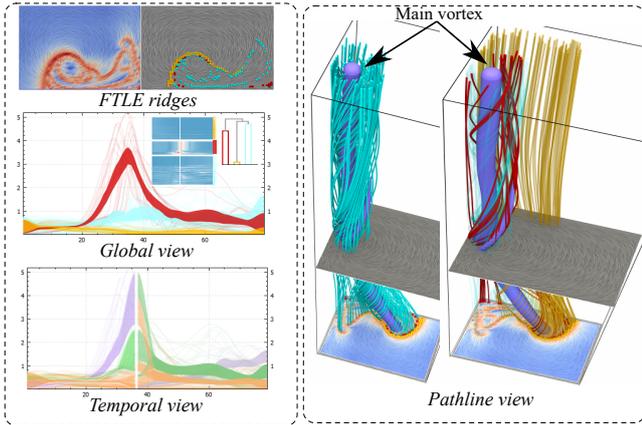


Figure 14: FTLE filtering on the 2D vortex ring simulation. The seeding points are placed in the FTLE ridges and TACs are generated by using the local shear rate attribute. It can be seen from both global and temporal views that the TAC analysis results based on our clustering method highlight three groups of particles in the high FTLE value region. The yellow particles have high separation degrees near the wall boundary. The light blue group indicates the separation around the boundary of vortices. The red group has large shear values indicating the separation between the main and secondary vortices. Here, the blue iso-surface of the main vortex is provided as a reference.

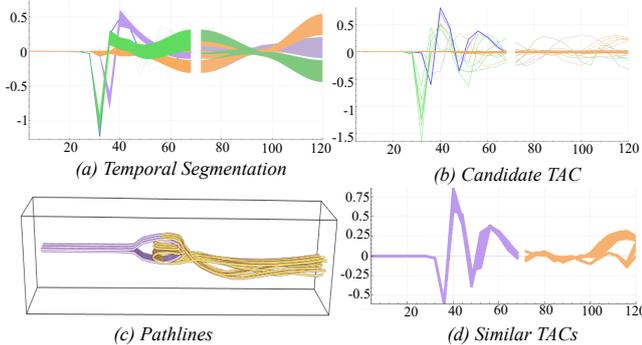


Figure 15: Based on the segmentation result (a), users can choose an interesting TAC (b) and perform the pattern search which results in the similar TACs (d) and their corresponding pathlines (c).

compute the FTLE field on the 2D vortex ring simulation, and use the obtained FTLE ridge (i.e., with top 20% FTLE values) to select a set of pathlines for study. Specifically, we perform both the global and temporal clustering of this set of pathlines based on their TACs of the local shear rate attribute. Figure 14 shows the results, which highlight three groups of particles having high degrees of separation: the red group indicates the separation between the main and secondary vortices, the light blue particles move to the vortex boundaries, while the yellow particles transport to the physical wall boundary locations. That said, there are at least three different pathline behaviors that cause the same FTLE ridge. Our technique can help reveal this more effectively.

Table 1: Performance of AHC clustering on four datasets

Simulations	Numbers of TACs	Time steps	Running Time
Flow behind Cylinder 2D [WT10]	20000	500	48.6s
Vortex Ring [OV93]	16384	80	12.4s
Boussinesq 2D [GT20]	67000	1000	144.8s
Flow behind Cylinder 3D [vFWTS08]	30720	102	30.5s
Tube 3D [SLD10]	129600	52	182.7s

Studying relation between shearing layer and vortex formation.

To reveal the relation between shearing layer and vortex formation, we apply the $Q < 0$ criterion on the 3D cylinder flow to select a set of pathlines that have negative Q values in the earlier time. We then perform clustering on these pathlines based on their TACs of the local shear rate attribute. The results are shown in Figure 13(b). As can be seen, there is a strong correlation between the negative Q and local shear rate since the highest shearing values align with the lowest Q values right after the flow encounters the cylinder, where a shearing layer is formed. In a later time, pathlines with strong shear rate participate the vortex shedding formation behind cylinder (i.e., the green and violet groups).

6.3. Performance and Comparison with Other Methods

Performance. All numerical experiments are carried out on a PC with an Intel Core i7-3537U CPU and 128GB RAM with a NVIDIA Quadro 4000 graphic card. The most time consuming task in the system is the AHC clustering. The detailed average running time of AHC clustering on five unsteady flow simulations is reported in Table 1.

Comparison with the model-based clustering. For time-series data (e.g., TACs) clustering, both similarity-based methods and model-based methods are usually applied. Our approach belongs to similarity-based strategy. In contrast to similarity-based methods, model-based methods assume that the data is generated by a mixture of underlying probability distributions, from which expectation-maximization (EM) learning algorithms can be derived for clustering, eliminating the need of a distance metric [GGW07]. However, model-based methods may not capture the important difference between two time-series. Figure 16 shows the comparison between the two clustering methods. From the comparison, we see that the model-based method tends to classify TACs only based on the magnitude while AHC considers both magnitude and shapes of TACs. For our purposes both the trend (or shape) and the magnitude of the TACs are important in characterizing their similarity. The proposed AHC method successfully capture both aspects. For more detailed discussion between these two different clustering approaches, please refer to the supplemental document.

7. Conclusion

In this work, we propose an interactive visualization framework for analysis and exploration of unsteady flow based on TACs. Given a vector field, we first compute the TACs over the entire flow domain and apply time interval segmentation to all TACs. To describe the behavior of a TAC, a sub-TAC extraction method is introduced to identify one or more interesting temporal trends. To measure

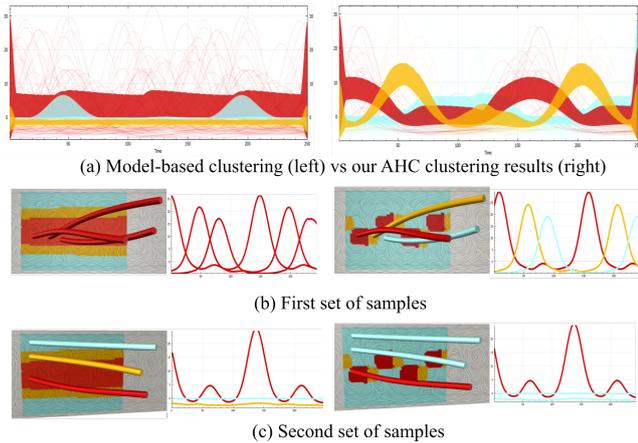


Figure 16: Comparison between AHC and model-based clustering. Model-based method classifies TACs only based on the magnitude while AHC considers both magnitude and shapes of TACs.

the similarity of two TACs properly, we introduce a new similarity measure, called the TAC Similarity Measure (TSM) to calculate the dissimilarity of TACs based on their events.

We implement the Agglomerative Hierarchical Clustering algorithm with the new TSM measure for the clustering of Lagrangian TACs on different temporal intervals. The clustering results provide different levels of details for flow behavior in both space and time, which facilitates data exploration. We also improve an edge-bundling technique to better represent the general behavior of TACs in a cluster and the connection of clusters among different time intervals. We introduce a 2D stack plot to visualize the TAC clusters without occlusion. Our framework has been evaluated on multiple unsteady flow simulations, and helps domain experts analyze vortex structures and other flow features.

Limitations: There are a number of limitations of our current system. First, the clustering computation is the most time consuming task as we haven't fully optimized the AHC algorithm. Second, our current TAC-based framework concentrates on scalar attributes. However, it may be extended to other attribute types, such as vector-valued and tensor-valued attributes. The supplemental document provides a couple examples on the extension to vector-valued attributes (e.g., velocity vectors). Nonetheless, the visualization of the clustering results in the TAC space needs to be addressed. Third, the 2D stack plots can provide a summary view of the global clustering results, but they cannot properly visualize the transition between neighboring temporal clusters, which we plan to address. Fourth, our framework has been evaluated via the vortex structure analysis and the exploration of other relevant features such as shearing layers and symmetric behavior, and the flows shown in this paper have relatively simple configurations (except for the Boussinesq flow). In the future, it is important to apply our framework to more complex turbulent flows for the study of energy transport to further evaluate its scalability. Finally, it would be interesting to extend our framework for the clustering of path surfaces to provide a more informative visualization for the study of 3D un-

steady flow behaviors. However, to achieve that, effective path surface seeding and placement to achieve sufficient spatial coverage while reducing overlap needs to be addressed, as well as the design of an effective similarity measure for the comparison of two surfaces, which we plan to explore in the future.

Acknowledgment

We would like to thank the anonymous reviewers for their valuable suggestions to our submission. This research was supported by NSF IIS 1553329.

References

- [BCP*12] BRAMBILLA A., CARNECKY R., PEIKERT R., VIOLA I., HAUSER H.: Illustrative Flow Visualization: State of the Art, Trends and Challenges. In *Eurographics 2012 - State of the Art Reports* (2012), Cani M.-P., Ganovelli F., (Eds.), The Eurographics Association, pp. 75–94. doi:10.2312/conf/EG2012/stars/075-094. 2
- [BMLC19] BERENJKOUB M., MONICO R. O., LARAMEE R. S., CHEN G.: Visual analysis of spatio-temporal relations of pairwise attributes in unsteady flow. *IEEE transactions on visualization and computer graphics* 25, 1 (2019), 1246–1256. 3, 12
- [CL93] CABRAL B., LEEDOM L. C.: Imaging vector fields using line integral convolution. In *Proceedings of the 20th annual conference on Computer graphics and interactive techniques* (1993), ACM, pp. 263–270. 7
- [CML*07] CHEN G., MISCHAIKOW K., LARAMEE R. S., PILARCZYK P., ZHANG E.: Vector field editing and periodic orbit extraction using Morse decomposition. *IEEE Transactions on Visualization and Computer Graphics* 13, 4 (Jul./Aug. 2007), 769–785. 2
- [CMLZ08] CHEN G., MISCHAIKOW K., LARAMEE R. S., ZHANG E.: Efficient Morse decompositions of vector fields. *IEEE Transactions on Visualization and Computer Graphics* 14, 4 (Jul./Aug. 2008), 848–862. 2
- [Def77] DEFAYS D.: An efficient algorithm for a complete link method. *The Computer Journal* 20, 4 (1977), 364–366. doi:10.1093/comjnl/20.4.364. 5
- [EA12] ESLING P., AGON C.: Time-series data mining. *ACM Computing Surveys (CSUR)* 45, 1 (2012), 12. 3
- [ELC*12] EDMUNDS M., LARAMEE R. S., CHEN G., MAX N., ZHANG E., WARE C.: Surface-based flow visualization. *Computers & Graphics* 36, 8 (2012), 974–990. 1, 2
- [FHM20] FERRARI S., HU Y., MARTINUZZI R. J.: Evolution surfaces for spatiotemporal visualization of vortex features. *Canadian Journal of Electrical and Computer Engineering* 43, 1 (2020), 30–42. 3
- [FKRW17] FERSTL F., KANZLER M., RAUTENHAUS M., WESTERMANN R.: Time-hierarchical clustering and visualization of weather forecast ensembles. *IEEE Transactions on Visualization and Computer Graphics* 23, 1 (2017), 831–840. 3
- [FKS*10] FUCHS R., KEMMLER J., SCHINDLER B., WASER J., SADLO F., HAUSER H., PEIKERT R.: Toward a Lagrangian vector field topology. *Computer Graphics Forum* 29 (2010), 1163–1172. 3
- [FMHC07] FANG Z., MÖLLER T., HAMARNEH G., CELLER A.: Visualization and exploration of time-varying medical image data sets. In *Proceedings of Graphics Interface 2007* (New York, NY, USA, 2007), GI '07, ACM, pp. 281–288. URL: <http://doi.acm.org/10.1145/1268517.1268563>, doi:10.1145/1268517.1268563. 3
- [GGW07] G. GAN C. M., WU J.: *Model-Based Clustering Algorithms*. SIAM, 2007, pp. 227–242. doi:10.1137/1.9780898718348.ch14. 13
- [GRCR03] GUO H., RENAUT R., CHEN K., REIMAN E.: Clustering huge data sets for parametric pet imaging. *BioSystems* 71 (2003), 81–92. 3

- [GT18] GÜNTHER T., THEISEL H.: The state of the art in vortex extraction. *Computer Graphics Forum* 37, 6 (2018), 149–173. doi:10.1111/cgf.13319. 3, 11
- [GT20] GUNTHER T., THEISEL H.: Hyper-objective vortices. *IEEE Transactions on Visualization Computer Graphics* 26, 03 (mar 2020), 1532–1547. doi:10.1109/TVCG.2018.2868760. 2, 4, 11, 13
- [GWT*08] GARTH C., WIEBEL A., TRICOCHÉ X., JOY K. I., SCHEUERMANN G.: Lagrangian visualization of flow-embedded surface structures. *Computer Graphics Forum* 27, 3 (2008), 1007–1014. 3
- [Hal01] HALLER G.: Lagrangian coherent structures and the rate of strain in two-dimensional turbulence. *Phys. Fluids A* 13 (2001), 3365–3385. 2, 3, 12
- [HH89] HELMAN J. L., HESSELINK L.: Representation and display of vector field topology in fluid flow data sets. *IEEE Computer* 22, 8 (August 1989), 27–36. 2
- [Hun87] HUNT J.: Vorticity and vortex dynamics in complex turbulent flows. *Transactions of the Canadian Society for Mechanical Engineering* 11, 1 (1987), 21–35. doi:10.1139/tcsme-1987-0004. 3
- [HWM88] HUNT J. C., WRAY A., MOIN P.: Eddies, streams, and convergence zones in turbulent flows. In *Studying Turbulence Using Numerical Simulation Databases*, 2 (1988), vol. 1, pp. 193–208. 2
- [JH95] JEONG J., HUSSAIN F.: On the identification of a vortex. *Journal of fluid mechanics* 285 (1995), 69–94. 3
- [Lei16] LEISHMAN J. G.: *Principles of Helicopter Aerodynamics*. Cambridge University Press, New York, NY, USA, December 2016. 8, 9
- [LHD*04] LARAMEE R. S., HAUSER H., DOLEISCH H., POST F. H., VROLIJK B., WEISKOPF D.: The state of the art in flow visualization: dense and texture-based techniques. *Computer Graphics Forum* 23, 2 (June 2004), 203–221. URL: <http://www.VRVis.at/ar3/pr2/star/>. 2
- [LHZP07] LARAMEE R., HAUSER H., ZHAO L., POST F. H.: Topology based flow visualization: the state of the art. In *Topology-Based Methods in Visualization (Proceedings of Topo-in-Vis 2005)* (2007), Mathematics and Visualization, Springer, pp. 1–19. 1, 2
- [Lia05] LIAO T. W.: Clustering of time series data — a survey. *Pattern recognition* 38, 11 (2005), 1857–1874. 3
- [LS09a] LEE T., SHEN H.: Visualization and exploration of temporal trend relationships in multivariate time-varying data. *IEEE Transactions on Visualization and Computer Graphics* 15, 6 (Nov 2009), 1359–1366. doi:10.1109/TVCG.2009.200. 3
- [LS09b] LEE T.-Y., SHEN H.-W.: Visualizing time-varying features with tac-based distance fields. In *Visualization Symposium, 2009. PacificVis'09. IEEE Pacific* (2009), IEEE, pp. 1–8. 2, 3, 4
- [MJL*13] MCLOUGHLIN T., JONES M. W., LARAMEE R. S., MALKI R., MASTERS I., HANSEN C. D.: Similarity measures for enhancing interactive streamline seeding. *IEEE Transactions on Visualization and Computer Graphics* 19, 8 (2013), 1342–1353. 3
- [MLP*10] MCLOUGHLIN T., LARAMEE R. S., PEIKERT R., POST F. H., CHEN M.: Over two decades of integration-based, geometric flow visualization. In *Computer Graphics Forum* (2010), vol. 29, Wiley Online Library, pp. 1807–1829. 1, 2
- [NZL*19] NGUYEN D. B., ZHANG L., LARAMEE R. S., THOMPSON D., MONICO R. O., CHEN G.: Unsteady flow visualization via physics based pathline exploration. In *2019 IEEE Visualization Conference (VIS)* (2019), IEEE, pp. 286–290. 3
- [OV93] ORLANDI P., VERZICCO R.: Vortex rings impinging on walls: axisymmetric and three-dimensional simulations. *Journal of Fluid Mechanics* 256 (1993), 615–646. doi:10.1017/S0022112093002903. 13
- [PBO*14] PALMAS G., BACHYNSKYI M., OULASVIRTA A., SEIDEL H. P., WEINKAUF T.: An edge-bundling layout for interactive parallel coordinates. In *Visualization Symposium (PacificVis), 2014 IEEE Pacific* (2014), IEEE, pp. 57–64. 6
- [PLMH12] POBITZER A., LEZ A., MATKOVIC K., HAUSER H.: A statistics-based dimension reduction of the space of path line attributes for interactive visual flow analysis. In *PacificVis* (2012), pp. 113–120. 3
- [PP03] POLTHIER K., PREUSS E.: Identifying vector fields singularities using a discrete hodge decomposition. In *Mathematical Visualization III* (2003), Ed: H.C. Hege, K. Polthier, pp. 112–134. 2
- [PPF*11] POBITZER A., PEIKERT R., FUCHS R., SCHINDLER B., KUHN A., THEISEL H., MATKOVIC K., HAUSER H.: The state of the art in topology-based visualization of unsteady flow. *Computer Graphics Forum* 30, 6 (September 2011), 1789–1811. URL: <http://dx.doi.org/10.1111/j.1467-8659.2011.01901.x>. 1, 2
- [SGSM08] SALZBRUNN T., GARTH C., SCHEUERMANN G., MEYER J.: Pathline predicates and unsteady flow structures. *The Visual Computer* 24, 12 (2008), 1039–1051. 3
- [Sil86] SILVERMAN B. W.: *Density Estimation for Statistics and Data Analysis*. Chapman & Hall, London, 1986. 6
- [SLC19] SHI L., LARAMEE R. S., CHEN G.: Integral curve clustering and simplification for flow visualization: A comparative evaluation. *IEEE Transactions on Visualization and Computer Graphics* (2019), prePrint. doi:10.1109/TVCG.2019.2940935. 1, 2
- [SLD10] SCHAEFFER N., LE DIZÈS S.: Nonlinear dynamics of the elliptic instability. *Journal of Fluid Mechanics* 646 (2010), 471–480. 12, 13
- [SP07] SADLO F., PEIKERT R.: Efficient visualization of Lagrangian coherent structures by filtered amr ridge extraction. *IEEE Transactions on Visualization and Computer Graphics* 13, 6 (2007), 1456–1463. doi:http://dx.doi.org/10.1109/TVCG.2007.70554. 3
- [SS06] SALZBRUNN T., SCHEUERMANN G.: Streamline predicates. *IEEE Transactions on Visualization and Computer Graphics* 12, 6 (2006), 1601–1612. doi:http://doi.ieeecomputersociety.org/10.1109/TVCG.2006.104. 3
- [STH*09] SHI K., THEISEL H., HAUSER H., WEINKAUF T., MATKOVIC K., HEGE H.-C., SEIDEL H.-P.: Path line attributes - an information visualization approach to analyzing the dynamic behavior of 3D time-dependent flow fields. In *Topology-Based Methods in Visualization II* (Grimma, Germany, 2009), Hege H.-C., Polthier K., Scheuermann G., (Eds.), Mathematics and Visualization, Springer, pp. 75–88. doi:10.1007/978-3-540-88606-8_6. 1, 2, 3
- [SW10] SADLO F., WEISKOPF D.: Time-dependent 2-d vector field topology: An approach inspired by Lagrangian coherent structures. *Computer Graphics Forum* 29, 1 (2010), 88–100. 3
- [SWJS08] SALZBRUNN T., WISCHGOLL T., JÄNICKE H., SCHEUERMANN G.: The state of the art in flow visualization: Partition-based techniques. In *In Simulation and Visualization 2008 Proceedings* (2008), Hauser H., Strassburger S., Theisel H., (Eds.), SCS Publishing House, pp. 75–92. 1, 2
- [SZ12] SZYMCAK A., ZHANG E.: Robust Morse decompositions of piecewise constant vector fields. *Visualization and Computer Graphics, IEEE Transactions on* 18, 6 (2012), 938–951. 3
- [TSH01] TRICOCHÉ X., SCHEUERMANN G., HAGEN H.: Continuous topology simplification of planar vector fields. In *Proceedings of IEEE Visualization 2001* (2001), pp. 159–166. 2
- [TWS04] THEISEL H., WEINKAUF T., SEIDEL H.-P.: Grid-independent detection of closed stream lines in 2D vector fields. In *Proceedings of the Conference on Vision, Modeling and Visualization 2004 (VMV 04)* (Nov. 2004), pp. 421–428. 2
- [ÜSE13] ÜFFINGER M., SADLO F., ERTL T.: A time-dependent vector field topology based on streak surfaces. *IEEE Trans. Vis. Comput. Graph.* 19, 3 (2013), 379–392. 3
- [vFWTS08] VON FUNCK W., WEINKAUF T., THEISEL H., SEIDEL H.-P.: Smoke surfaces: An interactive flow visualization technique inspired by real-world flow experiments. *IEEE Transactions on Visualization and*

- Computer Graphics* 14, 6 (November - December 2008), 1396–1403.
URL: <http://tinoweinkauf.net/>. 12, 13
- [WBKP08] WALTER B., BALA K., KULKARNI M., PINGALI K.: Fast agglomerative clustering for rendering. In *2008 IEEE Symposium on Interactive Ray Tracing* (2008), pp. 81–86. 5
- [WESW17] WANG Z., ESTURO J. M., SEIDEL H.-P., WEINKAUF T.: Stream line-based pattern search in flows. *Computer Graphics Forum* 36, 8 (2017), 7–18. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1111/cgf.12990>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1111/cgf.12990>, doi:10.1111/cgf.12990. 12
- [WFMF00] WONG K.-P., FENG D., MEIKLE S., FULHAM M.: Segmentation of dynamic pet images using cluster analysis. *IEEE Nuclear Science Symposium* 3 (200), 18/126–18/130. 3
- [WS01] WISCHGOLL T., SCHEUERMANN G.: Detection and visualization of closed streamlines in planar fields. *IEEE Transactions on Visualization and Computer Graphics* 7, 2 (2001), 165–172. 2
- [WS09a] WOODRING J., SHEN H.-W.: Multiscale time activity data exploration via temporal clustering visualization spreadsheet. *IEEE Transactions on Visualization and Computer Graphics* 15, 1 (2009), 123–137. 3
- [WS09b] WOODRING J., SHEN H.-W.: Semi-automatic time-series transfer functions via temporal clustering and sequencing. In *Proceedings of the 11th Eurographics / IEEE - VGTC Conference on Visualization* (Chichester, UK, 2009), EuroVis'09, The Eurographics Association & John Wiley & Sons, Ltd., pp. 791–798. doi:10.1111/j.1467-8659.2009.01472.x. 3
- [WT10] WEINKAUF T., THEISEL H.: Streak lines as tangent curves of a derived vector field. *IEEE Transactions on Visualization and Computer Graphics (Proceedings Visualization 2010)* 16, 6 (November - December 2010), 1225–1234. 9, 13
- [WYG*11] WEI J., YU H., GROUT R. W., CHEN J. H., MA K.-L.: Dual space analysis of turbulent combustion particle data. In *Pacific Visualization Symposium (PacificVis), 2011 IEEE* (2011), IEEE, pp. 91–98. 3
- [WYM08] WANG C., YU H., MA K.-L.: Importance-driven time-varying data visualization. *IEEE Transactions on Visualization and Computer Graphics* 14, 6 (2008), 1547–1554. 3
- [ZCL*16] ZHANG L., CHEN G., LARAMÉE R. S., THOMPSON D., SESCO A.: Flow visualization based on A derived rotation field. In *Visualization and Data Analysis 2016, San Francisco, California, USA, February 14-18, 2016* (2016), pp. 1–10. 3
- [ZNT*17] ZHANG L., NGUYEN D., THOMPSON D., LARAMÉE R. S., CHEN G.: Enhanced Vector Field Visualization via Lagrangian Accumulation. *Computer and Graphics (special issue of CAD/Graphics)* (2017), 224–234. 1