Uncertainty Visualization of 2D Morse Complex Ensembles Using Statistical Summary Maps

Tushar M. Athawale, Dan Maljovec, Lin Yan, Chris R. Johnson, Valerio Pascucci, Bei Wang

Abstract—Morse complexes are gradient-based topological descriptors with close connections to Morse theory. They are widely applicable in scientific visualization as they serve as important abstractions for gaining insights into the topology of scalar fields. Data uncertainty inherent to scalar fields due to randomness in their acquisition and processing, however, limits our understanding of Morse complexes as structural abstractions. We, therefore, explore uncertainty visualization of an ensemble of 2D Morse complexes that arises from scalar fields coupled with data uncertainty. We propose statistical summary maps as new entities for quantifying structural variations and visualizing positional uncertainties of Morse complexes in ensembles. Specifically, we introduce three types of statistical summary maps – the probabilistic map, the significance map, and the survival map – to characterize the uncertain behaviors of gradient flows. We demonstrate the utility of our proposed approach using wind, flow, and ocean eddy simulation datasets.

Index Terms—Morse complexes, uncertainty visualization, topological data analysis

1 INTRODUCTION

Visualization play an integral role for effective data storytelling and decision making. Understanding the effects of data uncertainty on visualizations has been recognized as one of the top research challenges [1], [2], [3], [4]. In this paper, we focus on the notion of aleatoric uncertainty, which arises due to randomness in data acquisition and processing, and “can not be reduced or removed by model improvements or increases in measurement accuracy” [5]. Uncertainty visualization focuses on improving our ability to reason about the data by communicating their aleatoric uncertainties [5], and it has been shown to be effective in practice [6]. A common practice to mitigate the effects of uncertainty is to combine multiple simulations of a phenomenon (e.g., with varying parameters and/or different instruments) into an ensemble dataset; see [7] for a survey on ensemble visualization.

In this paper, we investigate uncertainty in Morse complexes for an ensemble of 2D scalar fields. Morse complexes and Morse-Smale complexes [8] are topological descriptors based on Morse theory [9], [10] that provide abstract representations of the gradient flow behavior of scalar fields [11]. Morse complexes [8] are the building blocks for Morse-Smale complexes, which have shown great utility in numerous scientific applications, from identifying burning regions in combustion experiments [12] to counting bubbles in mixing fluids [13]. They also appear in partition-based regression [14], [15] and statistical inference [16].

Given a Morse function $f$ defined on a manifold $M$, $f: M \rightarrow \mathbb{R}$, the Morse complex (and Morse-Smale complex) of $f$ decomposes $M$ into regions (referred to as cells) with uniform gradient behavior (see Sec. 2 for definitions). Morse and Morse-Smale complexes have been extensively studied under both piecewise-linear (PL) and combinatorial settings (see Sec. 3). However, visualization of these topological descriptors in the face of uncertainty remains challenging.

The uncertainties of Morse complexes capture information about their accuracy, reliability, and variability [1]. In terms of accuracy, Gyulassy et al. [17] have introduced algorithms that improve upon the geometric quality of Morse-Smale complexes. Their algorithms are shown to produce the correct results on average, and the standard deviation approaches zero with increasing mesh resolution. In terms of variability, Thompson et al. [18] have briefly mentioned a Monte Carlo sampling method to quantify variations in the boundaries of Morse cells.

Motivated by limited prior work in encoding uncertainty of topological descriptors [19], [20], we study the uncertainty in Morse complexes for an ensemble of 2D scalar fields. Suppose $n$ ensemble members are given as scalar fields defined on a shared 2D domain, $f_1, \ldots, f_n : M \rightarrow \mathbb{R}$, where $M \subset \mathbb{R}^2$. We study an ensemble of Morse complexes $M_1, \ldots, M_n$ computed from these functions. We assume that each ensemble member $f_i$ is drawn from some distribution that is concentrated around a (potentially unknown) ground truth function $f$, i.e., $f_i(x) \sim f(x) \pm \epsilon_i(x)$ for any $x \in M$ (for some $\epsilon_i(x) \geq 0$).

In this work, we propose vector-valued statistical summary maps for encoding and visualizing structural variations of Morse complexes. Specifically, we introduce three types of statistical summary maps to characterize the uncertain behavior of gradient flows and to be utilized in uncertainty visualization for an ensemble: the probabilistic map, the significance map, and the survival map. We give an overview of our computational pipeline in Fig. 1 using a toy example from the Ackley function.

In our first type of statistical summary map, the probabilistic map $P : M \rightarrow \mathbb{R}^n$, we quantify the uncertainty based on the destination of gradient flows. We assume that by combining persistence simplification [21], [22] with certain labeling strategy (e.g., via k-means clustering, mandatory
critical points [23], or Morse mapping [24]), the local maxima of each ensemble member share the same set of labels, denoted as \[ l = \{1, 2, \cdots, l\} \]. For the Ackley dataset, we have \( l = 9 \) maxima after persistence simplification (see Figs. 1d-1i). Given such a shared labeling, for each point \( x \in M \) in the domain, we compute a probability distribution of its gradient destination across the ensemble. We visualize the probabilistic map using color blending (i.e. weighted mean values) in Figs. 1j-1l, where each color represents a distinct label; see Sec. 4 for details.

In our second type of statistical summary map, the significance map \( I : M \rightarrow \mathbb{R}^n \), we quantify the uncertainty based on the persistence associated with the destination of gradient flows. We assign a \textit{significance measure} to each point \( x \in M \), which is the persistence value of a maximum to which the gradient flow originating at \( x \) terminates. We visualize the significance map using point-wise mean, variance, and entropy in Figs. 1m-1o respectively, see Sec. 5 for details.

In our third type of statistical summary map, the survival map \( S : M \rightarrow \mathbb{R}^n \), we study directional changes of gradient flows as a result of persistence simplification. We assign a \textit{survival measure} for each point \( x \in M \) based on how frequently it changes its gradient destinations during a hierarchical simplification process. We visualize the survival map (Figs. 1p-1r) using its point-wise mean, variance, and entropy; see Sec. 6 for details.

\textbf{Contribution.} In summary:

- We quantify the uncertainties in gradient behaviors across a 2D Morse complex ensemble using a probabilistic map \( P \), a significance map \( I \), and a survival map \( S \).
- These vector-valued statistical summary maps capture the uncertainty based upon the variations among gradient destinations (\( P \)), the persistence of these destinations (\( I \)), and the directional changes of gradient flows (\( S \)), respectively.
- The maps \( I \) and \( S \) employ information obtained during persistence simplification of each ensemble member at a fixed scale and across all scales, respectively.
- We apply various uncertainty visualization techniques, such as mean-, variance-, and entropy-based visualization, as well as interactive probability queries [25] to our statistical summary maps, to understand the Morse complex structural uncertainty in synthetic and simulation datasets.

\section{Technical Background}

There are a couple of technical gradients for our approach, namely Morse complexes and persistence simplification.

\textbf{Morse complexes.} We focus on the construction of 2D Morse complexes. For simplicity, let \( M \subset \mathbb{R}^2 \) be a 2D smooth manifold with boundary (we further ignore the boundary condition for most of our discussion). Let \( f : M \rightarrow \mathbb{R} \) be a Morse function; \( \nabla f \) denotes its gradient. A point \( x \in M \) is considered critical if \( \nabla f(x) = 0 \); otherwise it is regular. At any regular point \( x \), the gradient is well defined, and integrating it in both ascending and descending directions...
traces out an integral line, which is a maximal path whose tangent vectors agree with the gradient [26]. Each integral line begins and ends at critical points. The descending manifold surrounding a local maximum is defined as all the points whose integral lines end at the local maximum. The descending manifolds decompose the domain into 2-cells, whereas critical points are the 0-cells, and integral lines connecting the critical points are the 1-cells. As illustrated in Fig. 2a, these cells form a complex called a Morse complex of $f$, denoted as $M = M_f$ (whenever $f$ is clear from the context). In particular, all the points inside a single 2-cell have their local gradient flows (integral lines) ending at the same local maximum (Fig. 2b).

**Perspective and persistence simplification.** Persistent homology is a tool in topological data analysis for quantifying the significance of topological features. It is widely used for data denoising through persistence simplification [21]. In visualization, persistence has been used to simplify topological structures, such as Morse and Morse-Smale complexes [8], [27]. For a 2D scalar function, we create a hierarchical Morse complex [21] by simplifying persistence pairs (in this case, maximum-saddle pairs) in the order of increasing persistence values [22]. Persistence assigned to each critical point in the complex intuitively describes the scale at which a critical point disappears through simplification. Persistence pairs can be simplified by successively canceling pairs of critical points connected in the complex with minimal persistence while avoiding certain degenerate situations (see [22] for implementational details).

Fig. 3 illustrates the process of persistence simplification for a 2D Morse complex. A saddle-maximum pair $(z, x)$ with the minimal persistence in Fig. 3a is simplified in Fig. 3c (the respective 2D views are shown in Fig. 3b and Fig. 3d). The process merges the orange cell into the green cell. The gradient flows of all points in the orange cell change their destination from $x$ to $y$.

3 RELATED WORK

**Representations of Morse-Smale complexes.** Morse and Morse-Smale complexes are defined for functions on smooth $d$-manifolds. Moving from the smooth category to the discrete category requires considerable effort to ensure structural integrity and to simulate differentiability [8]. In general, Morse and Morse-Smale complexes can be represented explicitly or implicitly [22]. The first, an explicit representation, is computed in 2D [8] and 3D [28] for piecewise linear (PL) functions defined on triangulated domains. The second, an implicit representation, originates from Discrete Morse theory [29] where a Morse-Smale complex is implicitly represented by a combinatorial gradient field [22]. We present our results for the explicit representations of Morse complexes, although our methods do not depend upon the choice of the representation. Note that in image analysis, the watershed algorithm [30] is analogous to the computation of Morse complexes in low dimensions.

**Uncertainty visualization of critical points and gradient fields.** The critical points and induced gradient field of a scalar function characterize the structure of its corresponding Morse complex. A few recent works have focused on data uncertainty and its effects on the critical points and gradient fields. Mihai and Westermann [31] proposed likelihood visualizations of the critical points for an uncertain scalar field, which was extended to Gaussian-distributed uncertain scalar fields by Liebmann and Scheuermann [32]. Hüttenberger et al. [33] exploited the idea of Pareto optimality for predicting the positions of local extrema for multifield data. Günther et al. [23] devised mandatory critical regions as a way to segment the domain of uncertain data, where at least one critical point of an unknown underlying function is guaranteed to exist within a mandatory critical region. Favelier et al. [34] developed persistence-based clustering of ensemble members followed by mandatory critical regions for visualizing positional uncertainties of critical points. In this work, we leverage the idea of mandatory critical regions in our probabilistic map (Sec. 4).

Pfaffelmoser et al. [35] analyzed the variability in gradient fields induced by uncertain scalar fields, where gradients are computed using the notion of *central differences*. Otto et al. [36], [37] proposed Monte Carlo gradient sampling for visualizing variations of pathlines in 2D and 3D uncertain vector fields. Bhatia et al. [38] studied *edge maps* for
error analysis of uncertain gradient flows. Nagraj et al. [39] proposed a measure to quantify gradient uncertainty for multifield data.

Uncertainty visualization of topological descriptors. A major challenge in visualizing topological descriptors is to encode data uncertainty. Various uncertainty visualization techniques [40], [41], [42] have been proposed to explore structural variations of contour trees for noisy data. Recent work by Lin et al. [20] studied structural averages of merge trees in the context of uncertainty visualization. The analysis and visualization of topological variations in the context of uncertain data remains an open research challenge [19].

Several studies have addressed challenges associated with level sets visualization in the face of uncertainty, including contour boxplots [43], probabilistic marching cubes [44], [45], and level set extraction from uncertain data [46], [47], [48]. Multicharts for comparative 3D ensemble visualization [49], dynamic volume lines [50], Gaussian mixture data representations [51], and statistical volume visualization [52] are a few important contributions in volume rendering for visualizing uncertainty.

4 PROBABILISTIC MAP

Using the probabilistic map \( P : \mathbb{M} \rightarrow \mathbb{R}^n \), we quantify the uncertainty based on the destinations of gradient flows. Specifically, for each point \( x \in \mathbb{M} \) in the domain, we compute a probability distribution of its gradient destination across the ensemble.

Given an ensemble of Morse complexes \( M_1, \ldots, M_n \), we combine persistence simplification with a certain labeling strategy to obtain a shared label set for local maxima of the ensemble members. Let \( [l] = \{1, 2, \ldots, l\} \) denote such a label set. In other words, for each ensemble member \( M_i \), its local maxima have labels that form a subset of \([l] \). Now fix \( M_i \), we trace the gradient flow of each point \( x \in \mathbb{M} \) toward its destination, a local maximum \( y \in \mathbb{M}_i \), and assign to \( x \) the label of \( y \); let \( \alpha_i : \mathbb{M} \rightarrow [l] \) denote such an assignment. The probabilistic map \( P : \mathbb{M} \rightarrow \mathbb{R}^n \) is defined as a discrete probability distribution of values in \((\alpha_1(x), \alpha_2(x), \ldots, \alpha_n(x))\) for each \( x \in \mathbb{M} \).

Ackley dataset. We describe our pipeline for the probabilistic map via a synthetic dataset, called the Ackley dataset. Fig. 1a visualizes the Ackley function [53] \( f \) as the ground truth. \( f \) is made into a Morse function using simulation of simplicity [54]. \( f \) contains nine (local) maxima, which produce nine 2-cells in its corresponding Morse complex in Fig. 1d. We generate an ensemble of uncertain scalar fields \( \{f_i\}_{i=1}^n \) by mixing \( f \) with a small amount of noise sampled from a uniform distribution (i.e., \( \epsilon_i(x) \sim U(0, 0.3 \times p_f) \)), where \( p_f \) is the persistence of the smallest topological feature (a maximum-saddle pair) in \( f \). Two ensemble members \( f_1(x) \sim f(x) \pm \epsilon(x) \) are shown in Fig. 1f and 1h, respectively. For comparison, we compute the mean field of the ensemble, \( f = (\sum_i f_i)/n \) and visualize its Morse complex in Fig. 1e. The Morse complex of \( f \) (Fig. 1e) appears similar to the ground truth (Fig. 1d); however, it does not capture structural variations among the boundaries of 2-cells.

Persistence simplification and labeling. First, we perform a pre-processing step that combines persistence simplification with certain labeling strategy such that the maxima of each ensemble member share the same set of labels, denoted as \([l]\) = \{1, 2, \ldots, l\}. There are three labeling strategies, via k-means clustering, mandatory maxima [23], or Morse mapping [24].

For the strategy based on k-means clustering, we use a set of \( n \) persistence graphs (Fig. 4a), each of which shows the number of maxima as a function of persistence [55], to guide the selection of a simplification scale. For a fixed \( M_\nu \), each maximum has its associate persistence value that indicates at which scale this feature would be simplified and thus represents its significance [56]. The shape of the persistence graph, in particular, a plateau, indicates a stable range of scales to separate noise from features [56]. Combining the persistence graphs and spaghetti plots of 1-cells (Fig. 4b) helps us identify a scale for which we observe an agreement in the number maxima across the ensemble. We demonstrate such a process in Fig. 4 where \( l \) is determined to be 9 with the aid of persistence graphs and spaghetti plots. We then apply k-means clustering (setting \( k = l \)) to the maxima across all ensemble members, where each maximum of a simplified \( M_i \) is assigned a label in \([l]\).

Fig. 4. The persistence simplification process for the Ackley dataset. (a) By overlaying persistence graphs, all ensemble members have the same number (i.e., 9) of maxima (dotted pink line) for a scale at 0.3 within a common stable region (dotted red line). (b) At scale 0.3, overlaid spaghetti plots of 1-cells (i.e., 2-cell boundaries) among the simplified Morse complexes across the ensemble exhibit a significant spatial variation, but the topology of these 1-cells remains consistent.

For the strategy based on mandatory maxima, we apply the technique of David et al. [23] to an ensemble of uncertain Ackley functions \( \{f_i\}_{i=1}^n \), resulting in \( l = 9 \) mandatory maxima, which are represented by different colors in Fig. 1c. Mandatory maxima are defined to be spatial regions and function ranges where local maxima have to occur across the ensemble [23]. We then apply persistence simplification to each Morse complex \( M_i \) until we are left with \( l \) maxima. Subsequently, we assign each maxima to its nearest mandatory maxima to obtain its label in \([l]\).

For the strategy based on Morse mapping, we employ a method that has been applied to the tracking of critical points [24]. It builds upon the partition of the domain provided by a Morse complex. For a pair of Morse complexes \( M_1 \) and \( M_2 \), we say a maxima \( x \in M_1 \) is weakly mapped to a maxima \( y \in M_2 \) if \( x \) belongs to the 2-cell surrounding \( y \) in \( M_2 \). If \( x \) is mapped to \( y \) and \( y \) is mapped to \( x \), then \( x \) is strongly mapped to \( y \). We choose an ensemble member with a minimum number of maxima as a pivot \( M_\nu \). For each other member, we assign labels to its maxima by computing both weak and strong mappings against \( M_\nu \).

Computing the probabilistic map. Now, for each \( M_i \), \( \alpha_i : \mathbb{M} \rightarrow [l] \) assigns each \( x \in \mathbb{M} \) the label of its
gradient destination (another maxima). The probabilistic map \( P : \mathbb{M} \rightarrow \mathbb{R}^l \) is defined to be a discrete probability distribution of values in \( (\alpha_1(x), \ldots, \alpha_n(x)) \). Let \( P_j(x) \) be the number of times \( x \) is assigned a label \( j \in [l] \) divided by \( n \). Then \( P(x) = (P_1(x), \ldots, P_l(x)) \). For a point \( x \in \mathbb{M} \), if \( P_j(x) = 1 \) (implying \( P_k = 0 \) for all \( k \neq j \)) for some \( j \), then \( x \) is a point with certainty; otherwise it is a point with uncertainty. Points with certainty are those whose gradient flows to the maxima with the same label, while points with uncertainty are those whose gradient destinations vary across ensemble members.

Visualizing \( P \) via color blending. Fig. 1j visualizes \( P \) for the Ackley dataset via color blending. Suppose each label is assigned a color, \( \{c_1, \ldots, c_l\} \), where \( c_i \in \mathbb{R} \). Point \( x \) is then assigned a color as \( \sum ic_i P_i(x) \). Fig. 1k shows the points with certainty in color. For example, all orange points have their gradients flow to the maximum with the same label. The white regions are points with uncertainty. Points with uncertainty are further visualized with color in Fig. 1l based on their proximity to the points with certainty. For a pair of adjacent regions with different labels \( i \) and \( j \) (e.g., orange vs. light green), a black contour contains all points \( x \in \mathbb{M} \) such that \( P_i(x) = 0.5 \) for some label \( i \); we refer to such black contours as expected boundaries.

Visualizing \( P \) via entropy and interactive queries. Furthermore, we compute and visualize point-wise entropy of \( P \). Given \( P : \mathbb{M} \rightarrow \mathbb{R}^l \), the Shannon entropy [57] at \( x \in \mathbb{M} \) is computed as \( E(x) = -\sum_i (P_i(x) \log_2 P_i(x)) \). Since uniform probability yields maximum uncertainty and therefore maximum entropy, points with high entropy represent positions whose gradient destinations have higher unpredictability. As illustrated in Fig. 5, points with high entropy \( (E \geq 1) \) are concentrated within the four corners, as their gradient destinations are (approximately) uniformly distributed across four maxima.

To further understand the points with uncertainty in \( P \) (Fig. 1j), we provide interactive queries based on the framework of Potter et al. [25]. Points at selected locations are queried of its corresponding distribution \( P \) (Fig. 6).

Visualize \( I \) with mean, variance and entropy. The map \( I \) can be visualized using its point-wise mean, variance, and entropy (Figs. 1m-o). The point-wise mean of \( I \), is defined as \( \bar{I} := \frac{1}{n} \sum \beta_i \). To highlight further the variability of 1-cells, we employ a quantized visualization by dividing the range of \( I \) into a number of intervals and visualize the pre-image of each interval using a miscellaneous colormap (Figs. 7c). The point-wise variance is defined as \( \text{Var}(\beta) := \frac{1}{n} \sum (\beta_i - \bar{I})^2 \). The point-wise Shanon entropy is computed by obtaining a probability distributions of the significance values \( \beta_1(x), \ldots, \beta_n(x) \) using kernel density estimation. For example, via interactive queries, such distributions at the locations with labels \( x = 0 \) and \( x = 1 \) are visualized in Figs. 7e-f, respectively.

To capture structural variations in a local neighborhood of a point, we also study variance and entropy computed in a \( 5 \times 5 \) patch centered at a given point. This is referred to as the patch-wise variance/entropy. They can be considered as smoothing filters that emphasize and enhance uncertain features in local neighborhoods, see Sec. 7.1 for details.

5 Significance Map

Using the significance map \( I : \mathbb{M} \rightarrow \mathbb{R}^n \), we quantify the uncertainty of a point based on the persistence associated with its gradient destinations at a fixed scale across the ensemble. First, we use persistence graphs [55] in combination with spaghetti plots (Fig. 4) to identify the scale of persistence simplification across the ensemble. Second, we apply persistence simplification with the identified simplification scale to each Morse complex \( M_k \). We assign a significance measure to each point \( x \in \mathbb{M} \) of a simplified ensemble member \( M_l \), which is equal to the persistence value of a local maximum in \( M_l \) to which the gradient flow originating at \( x \) terminates; let \( \beta_i : \mathbb{M} \rightarrow \mathbb{R} \) denote such an assignment. Figs. 7a-b visualize significance assignments across two ensemble members for the Ackley dataset. The significance map \( I : \mathbb{M} \rightarrow \mathbb{R}^n \) is defined to be a vector of significance measures across the ensemble \( (\beta_1(x), \beta_2(x), \ldots, \beta_n(x)) \) for each point \( x \in \mathbb{M} \).
6 Survival Map

Using the survival map $S : M \to \mathbb{R}^n$, we study directional changes of gradient flows as a result of persistence simplification across all scales. For a fixed ensemble member $M_i$, we first apply a hierarchical persistence simplification of $M_i$ using persistence as a scale parameter. In the case of a 2D Morse complex, we focus on canceling maximum-saddle pairs until only the global maximum remains. We assign a survival measure for each point $x \in M$ based on how frequently it changes its gradient destinations during the simplification process. The less frequently $x$ changes its gradient destinations, the greater is its survival measure, and vice versa. In other words, the survival measure quantifies the survivability of consistent flow behaviors. Let $\gamma_i : M \to \mathbb{R}$ denote such an assignment. The survival map $S : M \to \mathbb{R}^n$ is defined to be a vector of survival measures across the ensemble, $(\gamma_1(x), \gamma_2(x), \cdots, \gamma_n(x))$ for each $x \in M$.

Algorithmic details. To compute $\gamma : M \to \mathbb{R}$ for a particular ensemble member, we use Fig. 3 to illustrate one step of our algorithm with a toy example. Suppose we simplify the maximum-saddle pair $(x, z)$ with the lowest persistence $\lambda$. As a result, the gradient flows that terminate at $x$ (Fig. 3a) are redirected to the nearby maxima $y$ (Fig. 3b), effectively merging the orange 2-cell sounding $x$ into the green 2-cell surrounding $y$. $\gamma : M \to \mathbb{R}$ is increased by $\lambda$ for all points in the green 2-cell of Fig. 3a, and it is unchanged everywhere else in the domain. In other words, points in the green region (Fig. 3a) has “survived” the simplification without changing their gradient destinations; therefore, they are “rewarded” the amount $\lambda$. $\gamma$ therefore captures the survivability of local gradient destinations after simplification.

Following the above process, we compute $\gamma_i : M \to \mathbb{R}$ for each ensemble member. We initialize $\gamma_i$ to be zero everywhere. Let $\{\lambda_1, \cdots, \lambda_{n_i}\}$ denote the persistence of maximum-saddle pairs to be cancelled in an increasing order. We perform $n_i$ steps of persistence simplification. For each step $j$ ($1 \leq j \leq n_i$), $\gamma_i$ is incremented within a local neighborhood where the gradient flow destinations survive (remain unchanged) after simplification. The above process is repeated until the entire Morse complex is simplified into a single 2-cell surrounding the global maximum ($j = n_i$).

Visualize $S$ with mean, variance, and entropy. Figs. 1p-r visualize $S$ via its point-wise mean, variance, and entropy, respectively. For instance, the mean is defined as $\mu(x) = \frac{1}{n} \sum_{i=1}^{n} \gamma_i(x)$. The yellow region in Fig. 1p suggests the existence of a relatively tall peak, and the dark blue regions represent the existence of relatively low peaks across all ensemble members. This behavior is consistent with the ground truth Ackley function depicted in Fig. 1a.

7 Results

We demonstrate the utility of our proposed statistical summary maps for gaining insights into Morse complex uncertainty for simulated and observed scientific datasets.

7.1 Wind Dataset

We first analyze a set of 15 vector fields (velocity) from a wind dataset of the IRI/LDEO Climate Data Library. The dataset pressure_level_wind is obtained using the NCEP ensemble system with the forecast and perturbed parameters. We analyze the dataset with a pressure level at 200 hPA and a forecast hour at 0 on January 01, 2015 over a spatial range of 150W-40.5°W and 90°N-10°S. The sampling resolution for the grid is 1.5° along each spatial dimension, resulting in a 68x68 grid representing the domain of interest. We obtain an ensemble of 15 Morse complexes computed from the negation of velocity vector magnitudes. That is, we focus on features surrounding the local minima of the velocity magnitude scalar field, which correspond to the critical points of vector fields. Fig. 8a shows the mean vector field.

Fig. 8. Wind dataset. (a) Visualization of the negated mean velocity magnitude field (red means low and blue means high velocity magnitude); and its corresponding Morse complexes before (b) and after (c) persistence simplification.

Persistence simplification. Guided by persistence graphs and spaghetti plots, we first apply persistence simplification to obtain a common label set across all ensemble members. From the persistence graphs in Fig. 9a, we identify a common plateau that indicates a stable range of scales to separate features from noise. In particular, at a scale 7 (dotted red line) within the plateau, all members have 11 maxima after simplification (dotted pink line). Spaghetti plots of the simplified Morse complexes with 11 maxima are shown in Fig. 9b. Although there are significant spatial variations of 1-cells, their topology remain sufficiently consistent across the ensemble. Therefore, we analyze the ensemble at the chosen scale where each member contains 11 maxima after simplification.

For comparison, Fig. 8b visualizes the Morse complex of the mean vector field, which is also simplified to contain 11 maxima in Fig. 8c. Fig. 10 visualizes the Morse complexes for two ensemble members before and after simplification. The spatial variations of 1-cells appear to be substantial, even after simplification. Although persistence simplification of the mean field gives us a high-level view of its
gradient behavior, it does not give us insight into positional uncertainties of 1-cells (i.e., 2-cell boundaries).

**Fig. 10.** Wind dataset: Morse complexes of two ensemble members before (a, c) and after (b, d) persistence simplification.

**Labeling.** Next, we apply k-means clustering to find label correspondences across the simplified ensemble members. Fig. 11a shows a scatter plot of the maxima across simplified ensemble members. Fig. 11b illustrates label correspondences after k-means clustering (where \( k = 11 \)). For comparison, we also experiment with an alternative labeling strategy based on mandatory maxima. Fig. 11c visualizes 11 mandatory maxima computed using the framework of David et al. [23]. Fig. 11d shows label assignment of each maximum to its nearest mandatory maxima. For the wind dataset, both labeling strategies (k-means vs. mandatory critical points) provide identical results.

**Fig. 11.** Labeling with k-means clustering: (a) a scatter plot of maxima across simplified ensemble members; (b) the k-means clustering of maxima with \( k = 11 \). Labeling with mandatory maxima: (c) mandatory maxima are shown as colored regions; (d) each ensemble maximum is assigned the label of its nearest mandatory maximum.

**Probabilistic map.** The probabilistic map \( \mathcal{P} \) is computed and visualized using color blending in Fig. 12a. In Fig. 12b, \( \mathcal{P} \) is visualized based on point-wise entropy (thresholded at \( > 0.9 \)). Via interactive queries, the gradient flows passing through location \( x = 0 \) have a high probability of terminating at maxima associated with labels 4 (orange) or 9 (violet); while the gradients passing through \( x = 1 \) may terminate at maxima with labels 2 (blue), 6 (red), or 7 (green).

**Significance map.** The significance map \( I \) is visualized via its point-wise mean, patch-wise variance, and entropy, as illustrated in Fig. 13. When \( I \) is visualized with its mean in (a) and (b), it captures uncertain segmentation of the domain. In particular, the yellow region in (a) encloses points with high average significance (persistence) and highlights an important feature shared across all members. The regions with higher fluctuations in colors in the quantized visualization (b) represent positional uncertain of 1-cells.

The variance-based visualization of \( I \) captures how far a set of significant values are spread out from their mean at a local neighborhood. In particular, the yellow region in Fig. 13d highlights the points with the largest variation. The entropy-based visualization of \( I \) in Fig. 13e highlights neighborhoods that exhibit relatively high randomness in their significance values, thus capturing boundary uncertainty exceptionally well. Points with high entropy, e.g., those enclosed by the red dotted squares in (e), are shown to be consistent with the areas with large spatial variations in the spaghetti plots (c). For comparison, our entropy-based visualization (e) complements the aggregate segmentation of Thompson et al. [18] in (f), which highlights the points that are crossed frequently by the 1-cells across ensembles. It captures the frequencies of points serving as 2-cell boundaries, while the entropy captures the structural variability of boundaries in local neighborhoods.

**7.2 Navier Stokes Simulation Dataset**

We study a time-dependent flow simulation dataset [59].

The dataset originates from a direct numerical Navier Stokes simulation by Camarri et al. [59]. It is a 3D flow around a confined square cylinder where the square cylinder has been positioned symmetrically between two parallel walls. Camarri et al. used a uniformly resampled version, which

was provided by Tino Weinkauf and used by von Funck et al. [60] for smoke visualization. We consider an ensemble consisting of velocity vector fields from time steps 60-66 of the 3D simulation. We take a 2D slice perpendicular to the z-axis (z = 24) and use velocity magnitude as the underlying scalar fields.

Persistence simplification. First, we apply persistence simplification to obtain a common label set across all ensemble members, guided by persistence graphs and spaghetti plots in Fig. 14. In particular, at the selected simplification scale (dotted red line) in Fig. 14a, 5 out of 7 (70\%) members agree on the number of maxima (10) after simplification.

We illustrate three ensemble members at time steps 60, 63, and 66, in Figs. 15a-c, respectively. For each time step, we visualize, from left to right, the underlying scalar field, together with Morse complexes before and after persistence simplification. The mean field (Fig. 15d), on the other hand, misses significant number of features compared to individual members due to the spatial shift of critical points across ensemble. Furthermore, the mean field Morse complex does not give any insight into its structural uncertainty.

However, spaghetti plots of the simplified Morse complexes in Fig. 14b do not display the topological consistency of 1-cells as in Sec. 7.1. In such cases, we demonstrate below how Morse mapping may be an effective strategy for our labeling process.

Labeling. In Fig. 16, we compare the three labeling strategies in finding label correspondences across the simplified ensemble members. Since the mandatory maxima strategy extracts common denominator of maxima across all ensemble members, its results are sensitive to noise in the data. Specifically, for ensembles with relatively large noise, the number of mandatory maxima tend to be small (3), as illustrated in Fig. 16d. If we simplify each ensemble member to have 3 maxima and assign each maximum the label of its nearest mandatory maximum, we will miss most of the features in the ensemble (Fig. 16e). As illustrated Fig. 16a and Fig. 16(b-c), the Morse mapping and the k-means clustering strategies provide slightly different results. The Morse mapping strategy (Fig. 16a) does not require the simplified ensemble members to have the same number of maxima, therefore it is more flexible than the k-means clustering.

Fig. 16. Navier Stokes simulation dataset. (a) Labeling with Morse mapping. (b-c) Labeling with k-means clustering. (d-e) Labeling with mandatory maxima, which are shown as colored regions in (e); ensemble maxima are labeled with nearest mandatory maxima (f).

Probabilistic map. We further compare and contrast the k-means clustering and Morse mapping labeling strategies for the computation of the probabilistic map \( P \). The probabilistic map visualizations using either labeling strategy give insight into the positional uncertainty of the expected 2-cell boundaries that are not observable via the mean field (15d). However, as illustrated in Fig. 17b, the expected 2-cell boundaries using Morse mapping appear to be more consistent with respect to the underlying topology of individual ensemble members than those obtained using k-means clustering in Fig. 17a.

Significance map and survival map. Figs. 18a-c visualize the significance maps \( I \) via point-wise mean, patch-wise variance and entropy, respectively. The region enclosed by the red dotted box within the spaghetti plots in Fig. 14b shows a lack of spatial consistency in the 1-cells across ensemble members. The boundary of such a region is highlighted via the variance in Fig. 18b (white dotted box).

On the other hand, the survival map \( S \) visualized via its mean highlights the yellow feature immediately behind the square cylinder (white dotted box) in Fig. 18d, which has
the highest survival measure across all scales. The patch-wise variance and entropy of $S$ (Figs. 18e-f) both capture high variabilities along its boundaries (white dotted boxes).

**Persistence simplification.** First, we apply persistence simplification to obtain a common label set across all ensemble members, guided by persistence graphs and spaghetti plots in Fig. 19. In particular, at the selected simplification scale (dotted red line) in Fig. 19a, 5 out of 10 (50%) members agree on the number of maxima (11) after simplification.

We illustrate three ensemble members in Figs. 20a-c, respectively. For each ensemble member, its corresponding simplified Morse complex contains 2-cells that highlight important features of ocean eddies (white boxes). The mean field Morse complex in Fig. 20d, however, does not give any insight into the structural uncertainty, that is, the variabilities of these features across the ensemble.

On the other hand, spaghetti plots of the simplified Morse complexes in Fig. 19b do not display the topological consistency of 1-cells as in Sec. 7.1. In such cases, we demonstrate below how Morse mapping may be an effective strategy for our labeling process.

**Labeling.** In Fig. 21, we compare the three labeling strategies. As illustrated in Fig. 21d, the number of mandatory maxima is small (3) since ensemble members have large variations. Simplifying each ensemble member to have 3 maxima will miss most of the features of interest (Fig. 21e). The Morse mapping (Fig. 21a) and the k-means clustering (Fig. 21b-c) strategies, on the other hand, provide reasonable results; the Morse mapping is more flexible without requiring the same number of maxima across the ensemble.

---

Fig. 21. Red Sea dataset. (a) Labeling with Morse mapping. (b-c) Labeling with k-means clustering. (d) Mandatory maxima are shown as colored regions. (e) Labeling with nearest mandatory maxima.

Probabilistic map. We visualize the probabilistic map using both k-means clustering and Morse mapping labeling strategies. Both visualizations in Fig. 22 highlight positional uncertainty of 2-cell boundaries invisible to the mean field of Fig. 20d. However, the expected 2-cell boundaries using Morse mapping appear to be more spatially stable than those obtained via k-means clustering.

Fig. 22. Red Sea dataset: the map $P$ is visualized based on (a) k-means clustering and (b) Morse mapping strategies.

Figs. 23a-c visualize our entropy-based exploration of probabilistic map for lower entropy thresholds of 1.5, 1.25, and 1, respectively. Figs. 23d-f carve out regions in the domain, where ensemble agrees in their gradient destinations for at least 80%, 70%, and 60% members, respectively. Thus, the shared features across the ensemble are discoverable in Figs. 23d-f.

Significance map and survival map. Figs. 24a-c visualize the significance maps $I$ via point-wise mean, patch-wise variance and entropy, respectively. The significance map $I$ in Fig. 24a highlights the presence of a shared feature (white dotted box) with relatively high persistence. The same region (white dotted box) also is of high variance (Fig. 24b) and high entropy (Fig. 24c). Interestingly, the survival map $S$ appears to be low for this region (white dotted box in Fig. 24d), which requires further investigation.

8 CONCLUSION

We propose statistical summary maps as new abstractions for quantifying structural variations among ensembles of Morse complexes that arise from 2D uncertain scalar fields. We introduce three types of statistical summary maps, the probabilistic map $P$, the significance map $I$, and the survival map $S$. We take advantages of persistence simplification, mandatory critical points, or Morse mapping, to derive labeling strategies for ensemble members. We employ techniques such as color blending, entropy-based visualization, interactive queries, and quantized visualizations to understand the structural variability captured by our statistical summary maps.

Regarding labeling strategies, we demonstrate by experiments in Sec. 7.1 and the supplementary material that mandatory maxima capture almost all common features when there is a small amount of noise across the ensemble. However, when the ensemble members become quite noisy, the number of mandatory critical points is significantly smaller than the number of maxima in any ensemble member, rendering the label assignment unreliable (Sec. 7.2 and 7.3). The k-means clustering and the Morse mapping strategies are more general, with the latter being the most flexible when ensemble members are not required to have the same number of maxima.

For future work, we plan to extend our work for Morse complexes beyond 2D. While Morse complexes may be approximated in higher dimensions, visualizing uncertainties in higher dimensions will require new visual mappings.

ACKNOWLEDGMENTS

This work is supported in part by IIS-1910733, DBI-1661375, and IIS-1513616; as well as NIH grant P41 GM103545-18; and the Intel Parallel Computing Centers Program. We would like to thank Talha Bin Masood for providing pointers to example datasets and valuable inputs regarding the Morse mapping strategy.


Tushar Athawale is a postdoctoral fellow at the University of Utah's Scientific Computing & Imaging (SCI) Institute. He received Ph.D. in Computer Science from the University of Florida in May, 2015. He was a full-time employee at MathWorks, Inc., before joining the SCI institute. His primary research interests are in uncertainty visualization and statistical analysis for scientific data. Tushar completed his B.S. in Computer Engineering from the University of Pune, India.

Dan Maljovec got his Ph.D. from the School of Computing at the University of Utah where he focused on the application of topological models to high dimensional data. He received his B.S. in Computer Science from Gannon University in 2009. He was a research assistant at the SCI Institute since 2012. He is now a full-time employee at Recursion Pharmaceuticals at Salt Lake City.

Lin Yan is a Ph.D. student from the School of Computing and the Scientific Computing & Imaging (SCI) Institute at the University of Utah. She received her B.S. and M.S. from Shanghai Jiao Tong University, China (M.S. in Control Science and Engineering, and B.S. in Automation). Her research interests are topological data analysis and scientific data visualization of high-dimensional point clouds, scalar fields, vector fields, and tensor fields.

Chris R. Johnson is a Distinguished Professor of Computer Science and founding director of the Scientific Computing & Imaging (SCI) Institute at the University of Utah. With Charles Hansen, he co-edited the Visualization Handbook. He is a Fellow of AIMBE (2004), AAAS (2005), SIAM (2009), and IEEE (2014) and has received multiple research awards, including the NSF Presidential Faculty Fellow award from President Clinton, the Governor’s Medal for Science and Technology, the Utah Cyber Pioneer Award, the IEEE Visualization Career Award, the IEEE IPPDS Charles Babbage Award, the IEEE Sidney Fernbach Award, and the University of Utah’s most prestigious faculty award, the Rosenblatt Prize.

Valerio Pascucci is the John R. Parks Inaugural Endowed Chair and the founding Director of the Center for Extreme Data Management Analysis and Visualization (CEDMAV) of the University of Utah. Valerio is also a Faculty of the SCI Institute, a Professor of the School of Computing, University of Utah. His research interests include big data management and analytics, progressive multi-resolution techniques in scientific visualization, discrete topology, geometric compression, computer graphics, computational geometry, geometric programming, and solid modeling.

Bei Wang is an assistant professor at the School of Computing and a faculty member at the Scientific Computing and Imaging (SCI) Institute, University of Utah. She received her Ph.D. in Computer Science from Duke University. She is interested in the analysis and visualization of large and complex data. Her research interests include topological data analysis, data visualization, computational topology, computational geometry, machine learning and data mining.