

# Fiber Uncertainty Visualization for Bivariate Data With Parametric and Nonparametric Noise Models

Tushar M. Athawale, Chris R. Johnson, Sudhanshu Sane, and David Pugmire

**Abstract**—Visualization and analysis of multivariate data and their uncertainty are top research challenges in data visualization. Constructing fiber surfaces is a popular technique for multivariate data visualization that generalizes the idea of level-set visualization for univariate data to multivariate data. In this paper, we present a statistical framework to quantify positional probabilities of fibers extracted from uncertain bivariate fields. Specifically, we extend the state-of-the-art Gaussian models of uncertainty for bivariate data to other parametric distributions (e.g., uniform and Epanechnikov) and more general nonparametric probability distributions (e.g., histograms and kernel density estimation) and derive corresponding spatial probabilities of fibers. In our proposed framework, we leverage Green's theorem for closed-form computation of fiber probabilities when bivariate data are assumed to have *independent* parametric and nonparametric noise. Additionally, we present a nonparametric approach combined with numerical integration to study the positional probability of fibers when bivariate data are assumed to have *correlated* noise. For uncertainty analysis, we visualize the derived probability volumes for fibers via volume rendering and extracting level sets based on probability thresholds. We present the utility of our proposed techniques via experiments on synthetic and simulation datasets.

**Index Terms**—Uncertainty visualization, fiber surfaces, and probability

## 1 INTRODUCTION

In an era in which the growth of and dependence on data is increasing, understanding the inherent uncertainty is critical. Uncertainty can be introduced in many different ways [8]. Experimental and observational data are limited to the accuracy of the sensors. Accuracy in simulation data is a function of the model, numerical approximation, and parameters (e.g., grid resolution, number of particles) being used. Additional uncertainty can be introduced by data reduction (e.g., filtering, lossy compression), which is often used to address the challenges of big data. Further, the data are becoming increasingly more complex. Simulation data are often multivariate, with multiple scalar-, vector-, and/or tensor-quantities simulated over the spatial domain. Effective communication of uncertainty is, therefore, of paramount importance to perform risk-aware science and increase trust in scientific decisions [27, 30, 70]. Although there have been several advances in multivariate data visualization and analysis [16], relatively few studies have investigated the uncertainty of multivariate data [7, 44]. Here, we address the challenge of uncertainty visualization of bivariate scalar fields.

Fiber surfaces as proposed by Carr et al. [9] are a well-known bivariate data visualization technique for exploring correlations among variables. Fiber surfaces generalize the idea of isosurfaces [37] for univariate data to bivariate data. More specifically, given a bivariate function  $f : \mathbb{R}^3 \rightarrow \mathbb{R}^2$  with a three-dimensional (3D) spatial domain and a two-dimensional (2D) attribute space, fiber surfaces are a preimage of

a user-defined *fiber surface control polygon* (FSCP) or *trait* selected in the 2D attribute space (see Sect. 2.1). We use the terms FSCP and trait interchangeably. FSCP generalizes the idea of an isovalue for univariate data to higher dimensions. In this paper, we study the uncertainty in fiber positions arising from noise in bivariate data.

Our contributions in this paper are primarily motivated by the recent work by Zheng and Sadlo [69], who studied the positional uncertainty of fibers for uncertain bivariate fields. They proposed a closed-form uncertainty quantification framework for bivariate data fibers assuming independent/correlated Gaussian (parametric) noise models and rectangular FSCP selection (see Sect. 2.2). The probabilities of uncertain fiber positions for rectangular FSCP selection (referred to as *uncertain range-fibers*) were visualized with direct volume rendering and were shown to significantly deviate from the fiber positions extracted when assuming no uncertainty. One of the future directions identified by Zheng and Sadlo (see Sect. 7 of [69]) included extending their work to probability distributions other than Gaussian, which we address here.

In this paper, we expand the work by Zheng and Sadlo to other parametric distributions, such as uniform and Epanechnikov, and more general *nonparametric* noise distributions (histograms or kernel density estimation [KDE] [41]) for *arbitrary* shapes of FSCP. The nonparametric models of uncertainty can provide greater accuracy than the parametric models in the context of visualization [2, 4, 45] owing to their ability to capture more realistic shapes of underlying probability distributions. Compared with the more restrictive rectangular polygon shapes, FSCP's arbitrary shapes provide an additional flexibility to choose features from the 2D attribute space and visualize their fiber surfaces. For example, Carr et al. [9] and Klacansky et al. [32] studied fiber surface visualizations for arbitrary FSCP shapes in application domains, including chemistry and medical imaging (see Sect. 2.1).

**Contributions.** Our contributions in this paper are fourfold. First, we propose a closed-form statistical framework for uncertainty quantification of fibers of bivariate data when noise in data is modeled with *independent parametric probability distributions*, such as uniform, Epanechnikov, and Gaussian (Sect. 4). Note that our framework is applicable to FSCP with arbitrary shapes as opposed to being limited to rectangular FSCP [52, 69]. Our derivations for parametric distributions act as building blocks for our derivations relevant to nonparametric distributions. Second, we present a closed-form statistical framework to derive the positional uncertainty of fibers when noise in bivariate data is modeled with *independent nonparametric probability distributions* (e.g., histograms or Parzen window/KDE [41]) (Sect. 5.1). We present our results for noise kernels, including uniform, Epanechnikov, and Gaussian. Third, we present a nonparametric statistical framework

• Tushar M. Athawale and David Pugmire are with Oak Ridge National Laboratory. E-mail: {athawaletm | pugmire}@ornl.gov.  
• Chris R. Johnson is with Scientific Computing & Imaging (SCI) Institute at the University of Utah. E-mail: crj@sci.utah.edu.  
• Sudhanshu Sane is with Luminary Cloud, Inc. E-mail: sudhanshu.sane@gmail.com.

This manuscript has been authored by UT-Battelle, LLC under Contract No. DE-AC05-00OR22725 with the U.S. Department of Energy. The publisher, by accepting the article for publication, acknowledges that the U.S. Government retains a non-exclusive, paid up, irrevocable, world-wide license to publish or reproduce the published form of the manuscript, or allow others to do so, for U.S. Government purposes. The DOE will provide public access to these results in accordance with the DOE Public Access Plan (<http://energy.gov/downloads/doe-public-access-plan>).

Manuscript received xx xxx. 201x; accepted xx xxx. 201x. Date of Publication xx xxx. 201x; date of current version xx xxx. 201x. For information on obtaining reprints of this article, please send e-mail to: [reprints@ieee.org](mailto:reprints@ieee.org). Digital Object Identifier: [xx.xxxx/TVCG.201x.xxxxxxx](https://doi.org/10.1109/TVCG.2022.3209424)

for uncertainty quantification of bivariate data fibers when noise in variables is assumed to be *correlated* (Sect. 5.2). We present a closed-form uncertainty quantification of fibers when the noise correlation is estimated with 2D histograms, and a numerical integration technique is proposed to approximate fiber uncertainty when the noise correlation is estimated with bivariate KDE [62]. Lastly, we leverage the idea of *vertex-based classification* [4] to extract the most probable fiber surface from uncertain bivariate data (Sect. 7).

We organize our paper as follows: In Sect. 2, we define our notation and provide a brief overview of bivariate fiber surfaces (Sect. 2.1) and the state of the art in uncertainty visualization of fibers (Sect. 2.2). We then discuss relevant prior work in multivariate data and uncertainty visualization in Sect. 3. We describe our proposed frameworks for uncertainty quantification of fibers with parametric noise assumptions in Sect. 4 and nonparametric noise assumptions in Sect. 5. The memory and computational complexities of our proposed statistical techniques are discussed in Sect. 6. We describe our visualization methods in Sect. 7. Finally, we present the results of our statistical uncertainty analysis of synthetic and simulation datasets in Sect. 8 and discuss conclusions and potential future directions in Sect. 9.

## 2 BACKGROUND

### 2.1 Fiber Surfaces

We briefly summarize the fiber surface extraction technique proposed by Carr et al. [9] for bivariate scalar fields. Broadly, for a multivariate function  $f: \mathbb{R}^n \rightarrow \mathbb{R}^m$ , the spatial domain  $\mathcal{D} \subset \mathbb{R}^n$  is mapped to an attribute space  $\mathcal{A} \subset \mathbb{R}^m$ . The attribute space  $\mathcal{A}$  encodes scientific observations or simulations comprising scalar, vector, and tensor fields. In this paper, we consider bivariate scalar fields with  $\mathcal{D} \subset \mathbb{R}^3$  and  $\mathcal{A} \subset \mathbb{R}^2$ . Let  $\mathcal{A} = \{\mathcal{A}_1, \mathcal{A}_2\}$  be the attribute space with two attributes or scalar functions  $\mathcal{A}_1$  and  $\mathcal{A}_2$ . A *fiber* is then defined as the inverse image of a point  $a = (a_1, a_2) \in \mathcal{A}$  [9, 32, 49], and it corresponds to an intersection of isosurfaces [37] for isovalue  $\mathcal{A}_1 = a_1$  and  $\mathcal{A}_2 = a_2$ .

The fibers of interest can be specified with FSCP, which we denote as a trait,  $\mathcal{T} \subset \mathcal{A}$ . Each point on a trait boundary has a corresponding fiber, and such fibers carve out or represent a *fiber surface*. Consider the fiber surface example illustrated for the ethandiol dataset in Fig. 1, which is similar to the one by Carr et al. [9]. Fig. 1a visualizes a continuous scatterplot [5] of the grid vertex data with the electron density (Rho) as attribute  $\mathcal{A}_1$  on the horizontal axis and the reduced gradient (s) [28] as attribute  $\mathcal{A}_2$  on the vertical axis. The continuous scatterplot in all our examples is computed with the topology toolkit [60]. The four traits  $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4\}$  are selected as colored polygons with arbitrary shapes in an attribute space. The respective colored fiber surfaces (i.e., the set of fibers corresponding to points along the polygon boundaries) are visualized in Fig. 1b.

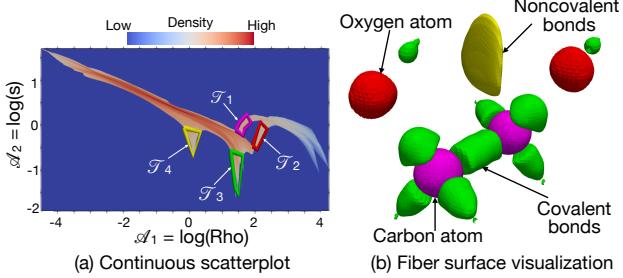


Fig. 1. Fiber surface visualization of the ethandiol dataset. (a) Visualization of a continuous scatterplot of the electron density (Rho) and reduced gradient (s) at a logarithmic scale. (b) Fiber surfaces for four traits  $\{\mathcal{T}_1, \mathcal{T}_2, \mathcal{T}_3, \mathcal{T}_4\}$  denoted by polygons in image (a).

The process of fiber surface extraction is similar to the process of isosurface extraction, which uses the marching cubes algorithm [37] but with a few tweaks [9]. Initially, each grid vertex of domain  $\mathcal{D}$  is classified as *interior* or *exterior* depending on whether the function values lie inside (interior) or outside (exterior) FSCP. Mathematically, a vertex  $v$  is classified as interior if  $(\mathcal{A}_1(v), \mathcal{A}_2(v)) \in \mathcal{T}$ ; otherwise, it is

classified as exterior. Such classification determines the marching cubes topology case within each grid cell for fiber surface reconstruction, and inverse linear interpolation may be applied in the attribute space  $\mathcal{A}$  to estimate fiber surface positions on grid edges (please refer to Algorithm 1 of [9]). Feature level sets [25] generalize the idea of fiber surfaces and extract features based on distance  $d$  from trait  $\mathcal{T}$ . For  $d = 0$ , the feature corresponds to the fiber surface itself. For  $d > 0$ , features present a spatial evolution of fiber surfaces in the vicinity of trait  $\mathcal{T}$ .

### 2.2 Fiber Uncertainty for Rectangular FSCP

The uncertainty in multivariate data can result in ambiguity as to whether a vertex should be classified as interior or exterior. The problem of uncertain vertex classification was recently addressed by Zheng and Sadlo [69] and Sane et al. [52]. Zheng and Sadlo proposed a statistical framework to compute the probability of data at point  $P \in \mathcal{D}$  being in the interior of a rectangular FSCP. The computed probabilities were then visualized via direct volume rendering and referred to as uncertain range-fibers. Their analysis assumed the independent/correlated Gaussian (parametric) noise model and a rectangular shape of trait  $\mathcal{T}$ .

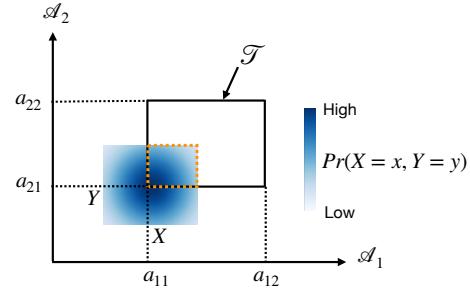


Fig. 2. Illustration of the interior probability computation for a single point of a spatial domain when FSCP is rectangular.  $X$  and  $Y$  denote the uncertain data ranges of a point, with blue denoting the joint probability distribution over the uncertain range. The black rectangle denotes trait  $\mathcal{T}$ . The probability of a point being in the interior of a fiber surface can be computed by integrating the probability density function (blue) over trait  $\mathcal{T}$  or the region enclosed by the dotted orange rectangle.

We illustrate the probability computation process proposed by Zheng and Sadlo in Fig. 2. Let random variable  $\mathcal{U} = (X, Y)$  denote uncertain bivariate data at position  $P \in \mathcal{D}$ , where  $X \subset \mathcal{A}_1$  and  $Y \subset \mathcal{A}_2$  are the two random variables that indicate uncertain data ranges for two attributes. Let  $\text{pdf}_X(x)$  and  $\text{pdf}_Y(y)$  denote the probability distributions of random variables  $X$  and  $Y$ , respectively. Let the rectangular trait  $\mathcal{T}$  be defined by ranges  $(a_{11}, a_{12})$  and  $(a_{21}, a_{22})$  for attributes  $\mathcal{A}_1$  and  $\mathcal{A}_2$ , respectively. Let  $\text{Pr}(\mathcal{U} \in \mathcal{T})$  denote the probability of uncertain data at point  $P$  being in the interior of a trait, which we refer to as *interior probability*. The interior probability (i.e.,  $\text{Pr}(\mathcal{U} = (X, Y) \in \mathcal{T})$ ) can be computed in two steps: (1) the joint probability distribution of  $X$  and  $Y$  (i.e.,  $\text{pdf}_{X,Y}(x, y)$ ) is computed, and then (2) the joint probability distribution  $\text{pdf}_{X,Y}(x, y)$  is integrated over trait  $\mathcal{T}$ . Mathematically, the probability of point  $P$  being in the interior of a fiber surface can be expressed as the following double integral over a rectangular trait:

$$\text{Pr}(\mathcal{U} = (X, Y) \in \mathcal{T}) = \int_{a_{11}}^{a_{12}} \int_{a_{21}}^{a_{22}} \text{pdf}_{X,Y}(x, y) dx dy \quad (1)$$

Zheng and Sadlo [69] proposed a closed-form computation of the double integral in Equation 1 when  $\text{pdf}_{X,Y}(x, y)$  is Gaussian distributed. Sane et al. [52] proposed confidence-feature level sets for Gaussian distributed data and rectangular FSCP for bivariate fields. In their technique, they considered confidence intervals of uncertain ranges at a position  $P$  (i.e., random variables  $X$  and  $Y$ ) to generate respective confidence visualizations of fiber surfaces.

## 3 RELATED WORK

Our proposed methods mainly relate to two important research areas of scientific visualization: multivariate data visualization and uncertainty

visualization. Next, we discuss prior work in each of these areas that is relevant to fiber surface visualization.

### 3.1 Multivariate Data Visualization

Understanding correlations among variables is one of the main challenges in multivariate data analysis. Scatterplots [15] and parallel coordinate plots [23] are fundamental tools that facilitate effective exploration of attribute spaces of multivariate data. Several research efforts have focused on improving challenges associated with scatterplot and parallel-coordinate plot visualization. Continuous versions of scatterplots [5] and parallel coordinate plots [21] have been proposed to avoid sampling artifacts of their traditional discrete versions. Quadri and Rosen [46] used topological data analysis to identify and explore clusters in scatterplots. Sauber et al. [54] proposed multifield graphs to study correlations among variables and their correlation strength.

Data variables can be encoded into glyphs [63] to visualize a correlation in a physical domain but can suffer from occlusion and cluttering issues. Other effective alternatives for multivariate visualization include brushing of scatterplots or parallel coordinate plots and linking them with the spatial domain for exploration of correlations in a physical space. For example, in Fig. 1, each trait  $\mathcal{T}_i$  indicates brushing in the attribute space, and their respective visualization in physical space corresponds to a fiber surface. Hauser et al. [19] introduced angular brushes for exploration of parallel coordinate plots and scatterplots. Kniss et al. [33] proposed multidimensional transfer functions as a brushing technique for volume rendering of multidimensional data. Jänicke et al. [24] presented an approach for the projection of multivariate data to a 2D space (referred to as *attribute clouds*) followed by clustering and brushing for visualization.

We provide a brief overview of several variants and applications of fiber surfaces for visualizing multivariate data. Klacansky et al. [32] extended the fiber surface extraction technique to tetrahedral meshes and proposed a parallel computational framework for interactive exploration of fiber surfaces. Wu et al. [65] presented a direct volume rendering framework for visualization of fiber surfaces without explicit extraction of fiber surfaces. Sakurai et al. [51] extended the idea of flexible isosurfaces [10] to flexible fiber surfaces for the visualization of fiber surface components of interest without occlusion. Raith et al. [47] applied fiber surface extraction to tensor-field invariant spaces for the visualization of tensor fields. Tierny and Carr [59] derived fiber surfaces around the Jacobi sets of bivariate fields (a conceptual equivalent of critical points of univariate fields) for computing fiber surface topological descriptors known as *Reeb spaces* [11]. Our literature review covered a small subset of multivariate visualizations relevant to fiber surfaces. For a comprehensive overview of multivariate data analysis and visualization, we encourage readers to refer to the survey papers by Fuchs et al. [16] and Kehrer et al. [31].

### 3.2 Uncertainty Visualization

To date, most of the research in uncertainty visualization has analyzed noise propagation in univariate data and the associated visualization algorithms. Specifically, these algorithms, including level sets [1, 4, 45, 48, 64], direct volume rendering [2, 14, 35, 38, 50], and topology-based visualizations [3, 12, 17, 66, 68], have been extensively studied in the context of uncertain univariate data. A few studies have investigated the uncertainty in visualizations of vector-field [13, 18, 36, 40, 55] and tensor-field [26, 29, 56] data.

There have been a few developments in uncertainty visualization of multivariate data. Zheng and Sadlo [69] proposed a framework for visualizing continuous scatterplots and uncertain fibers when the uncertainty in bivariate data is modeled with Gaussian distributions. Hazarika et al. [20] characterized multivariate data distributions with copula-based models and proposed a sampling strategy for copula-based models to quantify the uncertainty of features such as level sets and vortices. Xie et al. [67] encoded data quality into scatterplots and parallel coordinate plots for multivariate data in the context of information visualization. Nagaraj et al. [39] studied the consistency and inconsistency of gradient fields derived across attributes of multivariate data. In this paper, we take a step toward uncertainty quantification for

multivariate data by investigating uncertainty in fiber visualizations of bivariate data for parametric and nonparametric noise distributions and arbitrary FSCP shapes.

## 4 FIBER UNCERTAINTY FOR INDEPENDENT PARAMETRIC NOISE MODELS

In this section, we discuss two approaches, the closed-form integration and Monte Carlo integration, for fiber uncertainty computations. Although we present a closed-form fiber uncertainty quantification framework for integrable kernels (uniform, Epanechnikov, and Gaussian), a more generic Monte Carlo approach can be utilized to estimate fiber uncertainties for integrable as well as nonintegrable kernels.

### 4.1 Closed-Form Integration

We address the challenge of uncertainty quantification of fibers, similar to Zheng and Sadlo [69], for arbitrary shapes of FSCP. Specifically, we present a closed-form framework for computing interior probabilities (i.e.,  $\Pr(\mathcal{U} = (X, Y) \in \mathcal{T})$ ) per grid vertex when data noise is modeled with the independent uniform, Epanechnikov, and Gaussian distributions, which are integrable. The techniques detailed in this section are then leveraged as building blocks of our derivations for nonparametric noise models in Sect. 5.1. When the FSCP shape is not rectangular, the double integral in Equation 1 cannot be utilized because it considers the ranges of two attributes. For FSCP with an arbitrary shape, we apply Green's theorem to integrate the joint probability distribution  $\text{pdf}_{X,Y}(x, y)$ . Our approach is depicted in Fig. 3.

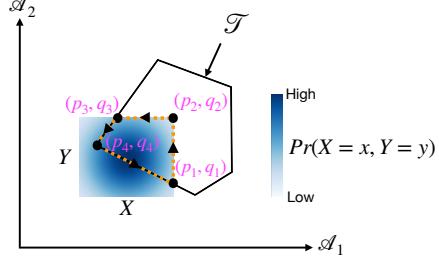


Fig. 3. Illustration of interior probability computation for a single point of a spatial domain when FSCP has an arbitrary shape.  $X$  and  $Y$  denote the uncertain data ranges of a point, with blue denoting the joint probability distribution over the uncertain range. The polygon with black boundaries denotes trait  $\mathcal{T}$ . The intersection of the two regions is indicated by dotted orange lines, and the coordinates of the intersection are indicated by pairs  $(p_i, q_i)$ . The probability of a point being in the interior of a fiber surface can be computed with Green's theorem by summing the line integrals of the probability density function (blue) along the edges of the intersection polygon in a counterclockwise direction (see arrow heads).

In Fig. 3, the blue region indicates the joint probability distribution of random variables  $X$  and  $Y$  (i.e.,  $\text{pdf}_{X,Y}(x, y)$ ). The polygon indicated by the orange dotted lines in Fig. 3 results from the intersection of the uncertain region (blue) with trait  $\mathcal{T}$ . Let  $(p_i, q_i)$  denote the coordinates of the intersection polygon. We compute the interior probability (i.e.,  $\Pr(\mathcal{U} = (X, Y) \in \mathcal{T})$ ) by integrating the joint probability distribution  $\text{pdf}_{X,Y}(x, y)$  over the intersection polygon (depicted by the orange dotted lines) using Green's theorem.

To use Green's theorem, we compute the new polynomials,  $L = (-1/2) \int \text{pdf}_{X,Y}(x, y) dy$  and  $M = (1/2) \int \text{pdf}_{X,Y}(x, y) dx$ . Assuming that  $X$  and  $Y$  are independent, the  $\text{pdf}_{X,Y}(x, y) = \text{pdf}_X(x) \cdot \text{pdf}_Y(y)$ . Thus, the polynomials  $L$  and  $M$  can be rewritten as follows:

$$\begin{aligned}
 L &= (-1/2) \int \text{pdf}_X(x) \cdot \text{pdf}_Y(y) dy \\
 &= (-1/2) \text{pdf}_X(x) \left[ \int \text{pdf}_Y(y) dy \right] \\
 M &= (1/2) \int \text{pdf}_X(x) \cdot \text{pdf}_Y(y) dx \\
 &= (1/2) \text{pdf}_Y(y) \left[ \int \text{pdf}_X(x) dx \right]. \tag{2}
 \end{aligned}$$

Table 1. Integration table

Kernel	$K(x)$	$\int K(x)dx$
Uniform	$\frac{1}{2}\mathbb{1}_{( x  \leq 1)}$	$\frac{1}{2}x_{( x  \leq 1)}$
Epanechnikov	$\frac{3}{4}(1 - x^2)_{( x  \leq 1)}$	$\frac{3}{4}(x - \frac{x^3}{3})_{( x  \leq 1)}$
Gaussian	$\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}_{(\mu=0, \sigma=1)}$	$\frac{1}{2}erf(\frac{x}{\sqrt{2}})$

The integral over a line segment with end coordinates  $(p_1, q_1)$  and  $(p_2, q_2)$ , denoted by  $I_{(p_1, q_1), (p_2, q_2)}$ , can be computed as follows:

$$I_{(p_1, q_1), (p_2, q_2)} = \int_{p_1}^{p_2} Ldx + \int_{q_1}^{q_2} Mdy. \quad (3)$$

Substituting the expressions for  $L$  and  $M$  presented in Equation 2 into Equation 3, we have the following:

$$\begin{aligned} I_{(p_1, q_1), (p_2, q_2)} &= \int_{x=p_1}^{x=p_2} (-1/2)\text{pdf}_X(x) \left[ \int \text{pdf}_Y(y)dy \right] dx + \\ &\quad \int_{y=q_1}^{y=q_2} (1/2)\text{pdf}_Y(y) \left[ \int \text{pdf}_X(x)dx \right] dy \\ &= \frac{-1}{2} \int_{x=p_1}^{x=p_2} \text{pdf}_X(x)dx \int \text{pdf}_Y(y)dy + \\ &\quad \frac{1}{2} \int_{y=q_1}^{y=q_2} \text{pdf}_Y(y)dy \int \text{pdf}_X(x)dx. \end{aligned} \quad (4)$$

For each vertex  $(p_i, q_i)$  of a polygon, the line integral (Equation 4) is computed for the line segment that connects coordinates  $(p_i, q_i)$  and  $(p_{(i\%N)+1}, q_{(i\%N)+1})$  in a counterclockwise direction (refer to the arrow heads of the polygon with dotted orange lines in Fig. 3), where  $N$  indicates the number of vertices of the intersection polygon. The expressions of integration of the uniform, Epanechnikov, and Gaussian distribution functions (computed with Wolfram alpha [22]) are provided in Table 1. Finally, the interior probability,  $\text{Pr}(\mathcal{U} = (X, Y) \in \mathcal{T})$ , can be computed using Green's theorem (i.e., by summing the line integrals [Equation 4] computed for each edge of the intersection polygon). Please refer to Sect. 1 of the supplementary material for a numerical integration experiment validating our derivation in Equation 4.

## 4.2 Monte Carlo Integration

In Sect. 4.1, we derived interior probabilities in closed form for parametric distributions with the uniform, Epanechnikov, and Gaussian kernels. Such analytical derivations are possible since a closed-form integration exists for each of the three kernels (see Table 1). For parametric distributions that are difficult to integrate or do not have a closed-form integration, a more generic Monte Carlo integration can be employed to obtain an approximate solution. In the Monte Carlo approach for the independent noise assumption,  $R$  samples are independently drawn from distributions  $\text{pdf}_X(x)$  and  $\text{pdf}_Y(y)$  at a vertex. If  $S$  samples lie in the interior of trait  $\mathcal{T}$ , we quantify or estimate the interior probability as  $\text{Pr}(\mathcal{U} \in \mathcal{T}) = S/R$ . This estimation technique is similar to the probabilistic marching cubes [43] technique, which estimates the spatial probability for level sets via Monte Carlo sampling of probability distributions. The accuracy of Monte Carlo solutions increases and converges with an increase in the number of samples.

## 5 FIBER UNCERTAINTY FOR NONPARAMETRIC NOISE MODELS

### 5.1 Independent Noise Assumption

We will first discuss a closed-form integration approach for fiber uncertainty quantification. For our closed-form approach, we utilize the analytical derivations of independent parametric noise models proposed in Sect. 4.1 as building blocks for deriving polynomial integration of nonparametric noise models. In nonparametric density models, the probability distribution of each uncertain variable per vertex can be estimated by deriving a histogram or by using KDE [41]. Let  $\{x_1, x_2, \dots, x_m\}$

denote  $m$  independently drawn samples from an unknown probability distribution  $\text{pdf}_X(x)$  of a random variable  $X$ . The  $\text{pdf}_X(x)$  can then be estimated from samples with KDE as follows:

$$\text{pdf}_X(x) = \frac{1}{m} \sum_{i=1}^{i=m} K_{h_x}(x - x_i). \quad (5)$$

The  $h_x$  in Equation 5 denotes a non-negative bandwidth of a kernel  $K$  associated with a sample.  $K_{h_x}(x - x_i)$  represents a scaled kernel, where  $K_{h_x}(x - x_i) = \frac{1}{h_x} K(\frac{x - x_i}{h_x})$ . The functions  $K(x)$  in Table 1 denote kernels with a unit bandwidth centered at  $x_i = 0$ . We estimate the bandwidth of a kernel from samples using Silverman's rule of thumb [57]. Similarly, for samples  $\{y_1, y_2, \dots, y_m\}$  independently drawn from an unknown probability distribution  $\text{pdf}_Y(y)$ , the probability distribution  $\text{pdf}_Y(y)$  for bandwidth  $h_y$  can be estimated as follows:

$$\text{pdf}_Y(y) = \frac{1}{m} \sum_{j=1}^{j=m} K_{h_y}(y - y_j). \quad (6)$$

Substituting Equation 5 and Equation 6 in Equation 4, the line integral for the independent nonparametric density estimation can be written as

$$\begin{aligned} I_{(p_1, q_1), (p_2, q_2)} &= \\ &\quad \frac{-1}{2} \int_{x=p_1}^{x=p_2} \frac{1}{m} \sum_{i=1}^{i=m} K_{h_x}(x - x_i)dx \int \frac{1}{m} \sum_{j=1}^{j=m} K_{h_y}(y - y_j)dy + \\ &\quad \frac{1}{2} \int_{y=q_1}^{y=q_2} \frac{1}{m} \sum_{j=1}^{j=m} K_{h_y}(y - y_j)dy \int \frac{1}{m} \sum_{i=1}^{i=m} K_{h_x}(x - x_i)dx. \end{aligned} \quad (7)$$

Rearranging the order of summations in Equation 7 gives us

$$\begin{aligned} I_{(p_1, q_1), (p_2, q_2)} &= \\ &\quad \frac{1}{m^2} \sum_{i=1}^{i=m} \sum_{j=1}^{j=m} \left[ \frac{-1}{2} \int_{x=p_1}^{x=p_2} K_{h_x}(x - x_i)dx \int K_{h_y}(y - y_j)dy + \right. \\ &\quad \left. \frac{1}{2} \int_{y=q_1}^{y=q_2} K_{h_y}(y - y_j)dy \int K_{h_x}(x - x_i)dx \right]. \end{aligned} \quad (8)$$

Equation 8 takes a form similar to the one for the parametric statistics presented in Equation 4, except that it loops through all pairs of kernels for random variables  $X$  and  $Y$ . Kernel  $K$  in Equation 8 can be replaced with the uniform, Epanechnikov, or Gaussian functions presented in Table 1.

Histograms could be considered as a special case of KDE with nonoverlapping uniform kernels associated with fixed-bin centers (pre-computed from noise samples and user-specified number of bins) and with a bandwidth equivalent to the bin width. Thus, for histograms, looping through all pairs of bins of histograms of random variables  $X$  and  $Y$  is required for computing the line integral in Equation 8. Finally, the interior probability (i.e.,  $\text{Pr}(\mathcal{U} = (X, Y) \in \mathcal{T})$ ) can be computed using Green's theorem by summing line integrals for all edges of the intersection polygon, as illustrated in Fig. 3.

When kernel  $K$  of a nonparametric distribution is difficult to integrate or nonintegrable, a Monte Carlo approach similar to the one for independent parametric distributions (Sect. 4.2) can be employed. In other words, for  $R$  samples independently drawn from nonparametric distributions  $\text{pdf}_X(x)$  and  $\text{pdf}_Y(y)$  with the base kernel  $K$ , the interior probability can be estimated as  $\text{Pr}(\mathcal{U} \in \mathcal{T}) = S/R$  if  $S$  samples lie in the interior of trait  $\mathcal{T}$ .

### 5.2 Correlated Noise Assumption

When random variables  $X$  and  $Y$  are correlated,  $\text{pdf}_{X,Y}(x, y)$  is not equal to the product  $\text{pdf}_X(x) \cdot \text{pdf}_Y(y)$ . In the correlated noise case, 2D histograms or bivariate KDE can be used to capture the correlation between variables  $X$  and  $Y$  and the multimodality of their probability distributions. Bivariate KDE extends the idea of univariate KDE in

Equation 5 to two variables [62]. An example of 2D histograms and bivariate KDE is provided in Sect. 2 of the supplementary material.

For a 2D histogram, the interior probability (i.e.,  $\Pr(\mathcal{U} = (X, Y) \in \mathcal{T})$ ) can be computed in closed form. Specifically, we loop through each 2D bin of a 2D histogram and quantify the amount of bin overlap with trait  $\mathcal{T}$  by taking into account the bin extent. We then scale the quantified overlap for a 2D bin with the weight of a bin computed from a 2D histogram. Our approach here is similar to the histogram approach for the independent nonparametric case (Sect. 5.1), except that the weight of each bin is not a product  $\text{pdf}_X(x) \cdot \text{pdf}_Y(y)$ , but the weight computed based on a 2D histogram.

In the case of a bivariate KDE, we propose a numerical integration approach similar to the one presented in Sect. 1 of the supplementary material. Specifically, we discretize the uncertain ranges of random variables  $X$  and  $Y$  with a uniform grid. We then estimate the density at each grid vertex using a bivariate KDE. We use the `gaussian_kde` function from the Python SciPy package [61] for bivariate KDE. We then perform a point-in-polygon test for user-specified FSCP at each grid vertex and sum the weights of all vertices that lie inside FSCP or trait  $\mathcal{T}$  to compute the interior probability (i.e.,  $\Pr(\mathcal{U} = (X, Y) \in \mathcal{T})$ ).

## 6 MEMORY AND COMPUTATIONAL COMPLEXITY

Here, we discuss the memory and computational complexity of our proposed parametric and nonparametric statistical models. Let the size of uncertain input data be  $D_1 \times D_2 \times D_3 \times V \times M$ , where  $D_i$  denotes the  $i^{\text{th}}$  dimension of a uniform grid that represents domain  $\mathcal{D}$ ,  $V$  indicates the number of variables ( $V = 2$  for bivariate data), and  $M$  is the number of ensemble members or noise samples that represent uncertainty in the data. For closed-form derivations with the independent parametric noise models in Sect. 4.1, we summarize  $M$  noise samples with two parameters per grid vertex. For example, we store the mean and width per vertex for a uniform distribution and the mean and standard deviation for a Gaussian distribution. Thus, for parametric noise models, the memory consumed is  $D_1 \times D_2 \times D_3 \times (V = 2) \times 2$ , which amounts to data reduction by a factor of  $M/2$ . To compute the interior probability volume for parametric noise models, one polynomial integration is performed per trait  $\mathcal{T} \subset \mathcal{A}$  per grid vertex if uncertain data at a vertex intersect trait  $\mathcal{T}$ . Thus, the worst-case computational complexity is proportional to the grid size (i.e.,  $D_1 \times D_2 \times D_3$ ).

For closed-form derivations with the independent nonparametric statistical approach (Sect. 5.1), the probability density per vertex can be estimated with KDE or histograms. For KDE, if a kernel is associated with each of the  $M$  noise samples per vertex and per variable, then the memory requirement is the same as the original data (i.e.,  $D_1 \times D_2 \times D_3 \times (V = 2) \times M$ ). If nonparametric density is estimated with a histogram with  $B$  bins per vertex and per variable, then memory consumption is reduced to  $D_1 \times D_2 \times D_3 \times (V = 2) \times B$ , which amounts to data reduction by a factor of  $M/B$ . To compute the interior probability volume, the number of computations per vertex increases quadratically with the number of kernels or histogram bins owing to the nested summation in Equation 8 (see the experiment in Fig. 6).

For the more generic Monte Carlo approach described in Sect. 4.2 and Sect. 5.1, the number of computations grows in proportion to the grid size and number of samples drawn per grid vertex (i.e.,  $D_1 \times D_2 \times D_3 \times S$ ). In the case of the correlated nonparametric statistical approach (Sect. 5.2), the space complexity is again equivalent to that of the independent nonparametric approach. For 2D histograms, the number of computations grows quadratically with the number of bins because we loop through each 2D bin. For bivariate KDE, we perform numerical integration by discretizing the uncertain data range at a vertex with a uniform grid of resolution,  $i \times i$ . Thus, the worst-case time complexity is proportional to  $D_1 \times D_2 \times D_3 \times i \times i$  per trait  $\mathcal{T}$  if the uncertain data range at each grid vertex intersects trait  $\mathcal{T}$ .

## 7 VISUALIZATION OF UNCERTAIN FIBERS

We visualize the positional uncertainty of fibers, i.e., the interior probability volumes derived using our proposed parametric (Sect. 4) and nonparametric (Sect. 5) statistics, via most probable fiber surface extraction, direct volume rendering, and probabilistic segmentation. Athawale et

al. proposed a vertex-based classification framework [4] for extracting the most probable isosurface from uncertain scalar fields. We extend their idea for extracting the most probable fiber surface from uncertain bivariate data. Specifically, we derive the interior probability  $\Pr(\mathcal{U} = (X, Y) \in \mathcal{T})$  per vertex, as described in Sect. 4 and Sect. 5. We then probabilistically predict the vertex sign as  $+$  if  $\Pr(\mathcal{U} \in \mathcal{T}) \geq 0.5$ ; otherwise, we predict the vertex sign to be  $-$ . The isosurface extracted from a grid with this sign classification indicates the most probable fiber surface topology.

We visualize the positional uncertainty of fibers via direct volume rendering of interior probability volumes derived using parametric and nonparametric statistics, similar to the visualizations by Zheng and Sadlo [69]. We render the positions with relatively high interior probability using higher opacity and the ones with relatively low interior probability using lower opacity. We visualize the probabilistic segmentation by thresholding computed interior probabilities. Specifically, for threshold  $t$ , we visualize the positions with  $\Pr(\mathcal{U} \in \mathcal{T}) \geq t$ . Such visualizations provide probabilistic insight into fiber positions.

## 8 RESULTS AND DISCUSSIONS

We validate our techniques and present their effectiveness via experiments on synthetic and simulation datasets.

### 8.1 Synthetic Data

#### 8.1.1 Independent Noise Models

We present the results of our proposed closed-form independent parametric (Sect. 4) and nonparametric (Sect. 5.1) statistical models for the synthetic data in Fig. 4. For our experiment, we sample the synthetic sphere and tangle [34] functions on a 3D grid with resolution  $64 \times 64 \times 64$ . Fig. 4a visualizes a 2D continuous scatterplot of sphere and tangle values plotted on the horizontal and vertical axes, respectively, for each grid vertex. The cyan polygon in Fig. 4a denotes FSCP or trait  $\mathcal{T}$ , and the corresponding fiber surface is visualized in Fig. 4b, which we treat as the reference.

We mix the synthetic data sampled on a grid with noise to generate an ensemble of 40 members. Specifically, we draw samples from a bimodal probability distribution, in which the mode with 80% cumulative probability density is centered around the ground truth, and the mode with 20% cumulative probability density is situated relatively far away from the ground truth. Thus, the noise samples from the mode with 20% cumulative probability density denote the outlier samples. We study the uncertainty in fiber positions arising from the ensemble.

Fig. 4c visualizes the fiber surface for the mean-field. The mean-field fiber surface is color-mapped with respect to its signed distance [6] from the ground truth fiber surface. Figs. 4d–4f visualize the results for the independent Gaussian noise assumption. Fig. 4d visualizes the interior probability volume (i.e.,  $\Pr(\mathcal{U} \in \mathcal{T})$ ) per grid vertex via direct volume rendering, in which yellow with higher opacity denotes positions with relatively high interior probabilities, and white with lower opacity denotes positions with relatively low interior probabilities. Fig. 4e visualizes the most probable fiber surface computed using the vertex-based classification (Sect. 7) color-mapped with the signed distance from the ground truth fiber surface. Fig. 4f visualizes the probabilistic segmentation (Sect. 7) extracted for isovalue 0.9, 0.75, and 0.6, and mapped to high, moderate, and low opacity, respectively. The probabilistic segmentation provides insight into the probability of fiber positions being in the interior of a fiber surface for different segments. Figs. 4g–4i visualize results similar to Figs. 4d–4f for the independent nonparametric KDE with the Gaussian base kernel.

The most probable fiber surface for the nonparametric noise assumption (Fig. 4h) is spatially closer to the ground truth fiber surface (Fig. 4b) when compared with the mean-field fiber surface (Fig. 4c) and the most probable fiber surface for the parametric assumption (Fig. 4e), as is evident by the color-mapped signed distance. The result implies a greater accuracy of interior probability computations per grid vertex with the nonparametric approach than with the mean-field or parametric approach. This result is expected because the nonparametric models can capture the bimodality of the underlying noise distribution more

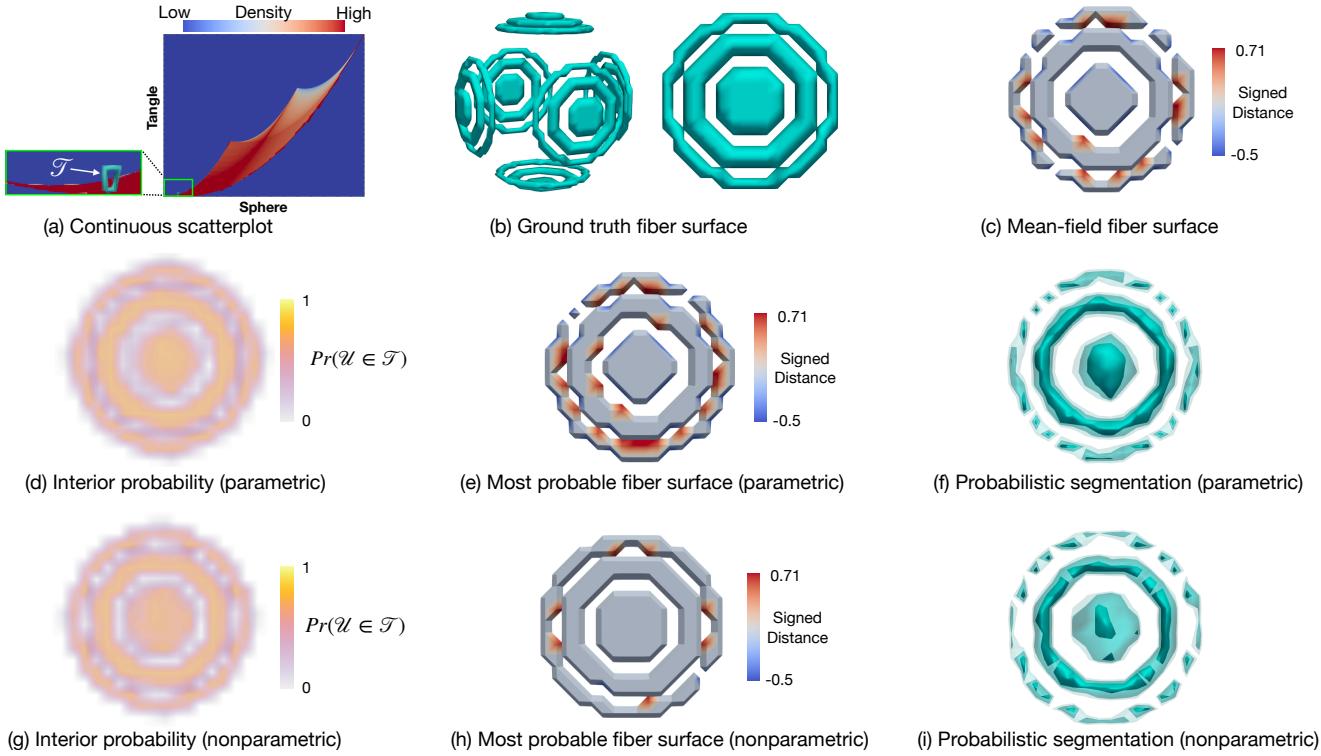


Fig. 4. Fiber visualizations for the synthetic tangle-sphere dataset. A continuous scatterplot for the dataset is visualized in image (a) with trait  $\mathcal{T}$  indicated by a cyan polygon. Image (b) visualizes the ground truth fiber surface that corresponds to trait  $\mathcal{T}$ . The ground truth dataset is mixed with random samples drawn from a bimodal noise distribution to generate an ensemble. Image (c) visualizes the mean-field fiber surface with a color-mapped signed distance from the ground truth fiber surface. The volume rendering of the interior probability volume, most probable fiber surface, and probabilistic segmentation are visualized for parametric (d)–(f) and nonparametric (g)–(i) statistical models. The volume rendering for nonparametric noise models visually represents the ground truth fiber positions more accurately than the volume rendering for parametric models (see Fig. 5 for quantitative analysis). Similarly, the most probable fiber surface for nonparametric models is spatially closer to the ground truth fiber surface than the surfaces for the mean-field and parametric models, as is evident by the color-mapped signed distance. The probabilistic segmentation provides insight into the positions with relatively high (opaque) or low (translucent) probability of fiber existence.

accurately than the parametric models, which results in nonparametric models that are less sensitive to outliers than the parametric models.

We further confirm the enhanced accuracy of nonparametric models compared to mean-field and parametric models using a quantitative error analysis, as shown in Fig. 5. Specifically, we compute and plot a Euclidean 2-norm of the difference between the interior probability volume (i.e.,  $Pr(\mathcal{U} \in \mathcal{T})$  per vertex) computed using different statistical techniques and the interior probability volume for the ground truth data, which we refer to as *interior probability error*. In the case of the ground truth data, the interior probability is 1 when data at a vertex lie inside trait  $\mathcal{T}$ ; otherwise, it is 0. Fig. 5a visualizes the results for parametric

noise models with uniform (red), Epanechnikov (green), and Gaussian (blue) assumptions. Fig. 5b visualizes the results for nonparametric KDE with uniform (red), Epanechnikov (green), and Gaussian (blue) base kernels. The dotted magenta line in Fig. 5a represents the error for the mean-field. In Fig. 5, we observe that the error computed for the nonparametric approach (Fig. 5b) is consistently lower than the error computed for the parametric density assumption and mean-field (Fig. 5a). The Gaussian kernel for the parametric and nonparametric statistics has the highest computational accuracy when compared with the uniform and Epanechnikov kernels.

We compare the results of the Monte Carlo sampling technique (Sect. 4.2 and Sect. 5.1) with our closed-form solutions in Fig. 5. In Fig. 5, the dashed lines denote the fixed error corresponding to our proposed closed-form solutions, and the solid curves plot the error as a function of the number of Monte Carlo samples. The error computed for the Monte Carlo solutions (solid curves) converges to the one computed for our proposed analytical solutions (dashed lines) as we increase the number of Monte Carlo samples. Such convergence confirms the correctness of our derivations (Equations 4 and 8) and implementations.

Fig. 6 visualizes the accuracy and time curves as a result of histogram rebinnning for the independent nonparametric noise assumption. At each grid vertex of the domain, we perform nonparametric density estimation by computing a histogram of ensemble members followed by the computation of interior probability. We then compute and plot the interior probability error (i.e., the Euclidean 2-norm of the difference between the interior probability volumes for the histogram noise model and the ground truth data). The number of bins determines the memory and computational requirements of the proposed nonparametric approach (Sect. 6) and impacts the accuracy. As observed in Fig. 6a, the

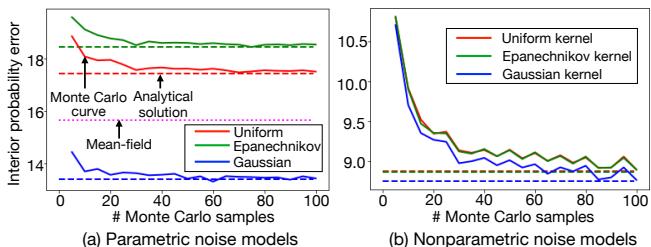


Fig. 5. Quantitative error analysis of interior probability computations for independent parametric and nonparametric noise models with respect to the ground truth interior probabilities. The solid lines plot the error as a function of the number of Monte Carlo samples. The magenta dotted line denotes the error for the mean-field, and the dashed lines denote the error for our proposed analytical solutions.

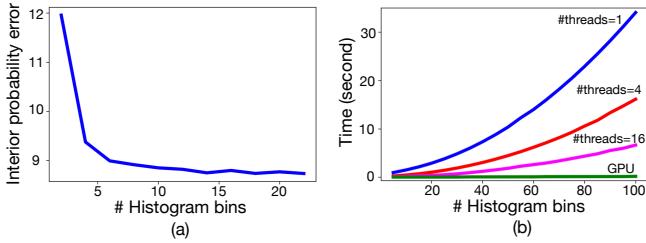


Fig. 6. The (a) accuracy vs. (b) computational cost curves for interior probability computation using the nonparametric histogram model. The computational cost grows quadratically, but the accuracy improves as the number of histogram bins increases. The parallel implementation with multiple openMP threads (red and magenta curves) and GPU (green curve) provides computational speed-up.

error reduces as the number of histogram bins increases. Note the sharp decline in errors for a relatively small number of bins. On the other hand, the computational complexity appears to grow quadratically, as we had anticipated. Since the interior probability at each grid vertex can be computed independently, the computations can be parallelized. In Fig. 6b, the red and magenta curves indicate the speed-up achieved with the OpenMP C++ parallel implementation with four and 16 threads, respectively. The GPU CUDA version of our code achieved 23X average speed-up with the NVIDIA V100 graphics card (the green curve in Fig. 6b) compared to the Power9 CPU with 16 openMP threads.

### 8.1.2 Correlated Noise Models

We study the correlated noise models (Sect. 5.2) via a synthetic experiment. The experimental settings are similar to the ones for the independent noise, except the bimodal noise now comprises two *correlated* Gaussians (e.g., Fig. 2a in the supplementary material) with the correlation specified by covariance matrices. A correlated Gaussian with an 80% probability concentration is situated close to the ground truth, and one with 20% probability concentration is situated relatively far away from the ground truth (denoting the outliers). The noise samples drawn from such bimodal distributions are mixed with the ground truth to generate an ensemble.

We observe that the ability of bivariate KDE or 2D histograms to reliably capture the correlation between random variables  $X$  and  $Y$  at each grid vertex is influenced by the number of ensemble members and by the grid resolution used for numerical integration of bivariate KDE. When the number of noise samples is relatively large (e.g., 10,000 samples for bivariate KDE estimation in Fig. 2c of the supplementary material), the correlation is captured reliably. For fewer samples, the correlation might not be captured well and might lead to relatively large errors. As shown in Fig. 7a, the interior probability error (blue curve) reduces with an increase in the number of ensemble members. The interior probabilities computed with the independent nonparametric noise assumption (the green dotted line in Fig. 7a) appear to be more accurate than the correlated noise assumption for 50 ensemble members.

The error for nonparametric models with the correlated or independent noise assumptions (Fig. 7a) is again smaller than the errors for the

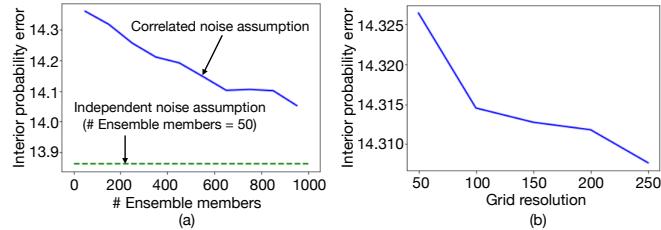


Fig. 7. Interior probability error (blue curves) for the correlated noise assumption plotted as a function of the (a) number of ensemble members and (b) grid resolution for numerical integration.

mean-field and parametric noise models owing to the higher robustness of nonparametric models to outliers. The mean-field for 50 ensemble members resulted in an interior probability error equal to 16.79. The parametric Gaussian noise assumption resulted in an interior probability error equal to 15.41. The error for the mean-field and parametric noise assumption does not fluctuate much with an increase in the number of ensemble members. Fig. 7b visualizes the error as a function of a grid resolution used for numerical integration in the case of bivariate KDE. The error reduces slowly with an increase in grid resolution.

## 8.2 Rectangular Vs. Arbitrary Shape of Polygonal Traits

The uncertainty visualization framework proposed by Zheng and Sadlo [69] is limited to a rectangular trait (Sect. 2.2). Our proposed framework extends their work to a polygonal trait with arbitrary shape. Fig. 8 demonstrates the advantage of an arbitrary polygon over rectangle for trait selection in the contexts of the original as well as uncertain data. Fig. 8a visualizes a continuous scatterplot of the electron density ( $\text{Rho}$ ) and reduced gradient ( $s$ ) attributes of the ethanediol molecule. The main separating axis or diagonal of the continuous scatterplot (with strong red hue in Fig. 8a) indicates regions with no chemical interactions. Thus, regions away from this axis are important for chemists in analyzing chemical interactions. The two traits  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are selected in the continuous scatterplot space to enclose a nondiagonal feature. As observed in Fig. 8a, the arbitrary shape of a polygon allows us to select a trait ( $\mathcal{T}_1$  in magenta) that is closer to a nondiagonal feature than the rectangular shape of a trait ( $\mathcal{T}_2$  in cyan).

The fiber surfaces corresponding to traits  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are visualized

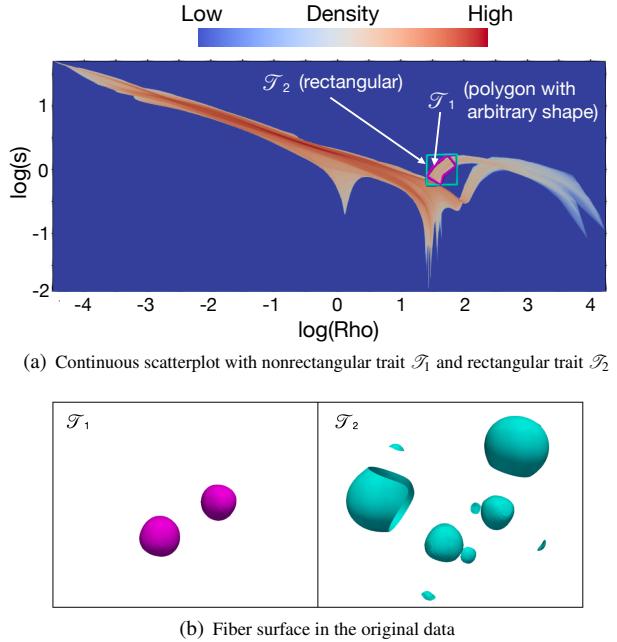


Fig. 8. Uncertainty visualization of the ethanediol dataset with a rectangular [69] vs. arbitrary (our contribution) shape of the polygonal trait. The arbitrarily shaped trait  $\mathcal{T}_1$  yields a more accurate extraction of carbon atom positions compared to the rectangular trait  $\mathcal{T}_2$  for both (b) the original data and (c) uncertain hixel data.

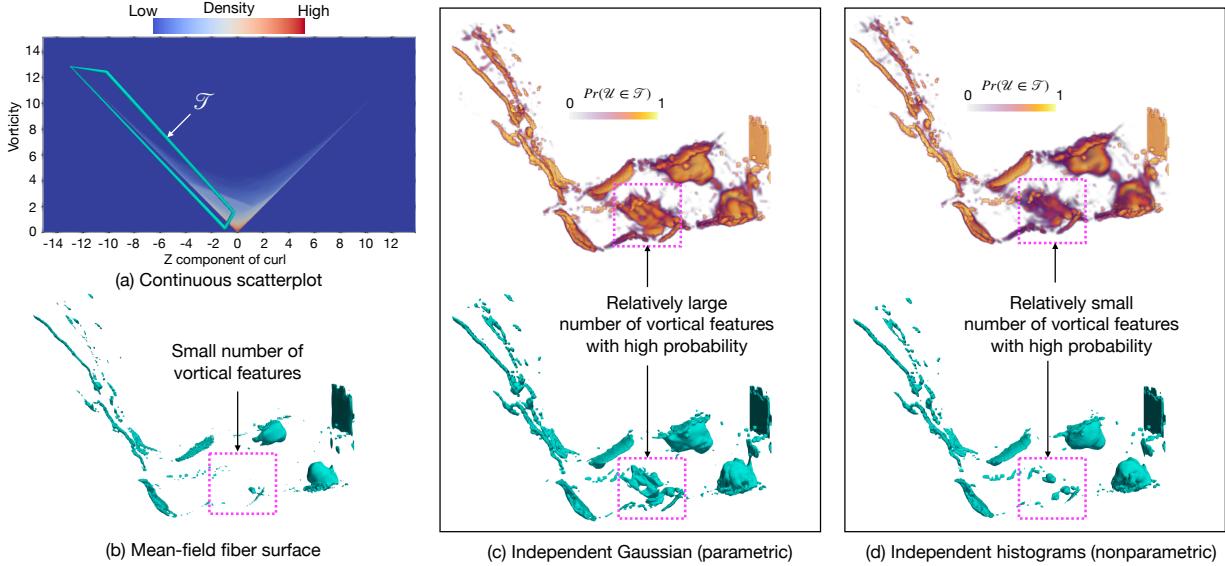


Fig. 9. Parametric vs. nonparametric noise models for uncertainty visualization of vortical features of the Red Sea ensemble dataset over the Gulf of Aden. Fiber positions are visualized for trait  $\mathcal{T}$  corresponding to anticyclonic (negative Z component of the curl of the velocity field) vortical features, as indicated by the cyan polygon in the continuous scatterplot shown in image (a). Image (b) visualizes the mean-field fiber surface. Images (c) and (d) visualize the results for the independent parametric (Gaussian) and independent nonparametric (histogram) noise models, respectively, with volume rendering of *interior probabilities* (top) and a *probabilistic segmentation* for isovalue 0.7 (bottom). Although the three statistical models in images (b)-(d) exhibit overall consistency in the positions of eddy features, the uncertainty of eddy features seems prominent in the region enclosed by the magenta box. For instance, the parametric noise model yields multiple high-probability vortex features inside the magenta box, whereas the nonparametric noise model yields a relatively low probability of vortex features inside the magenta box.

for the original data in Fig. 8b. The fiber surface for the nonrectangular trait  $\mathcal{T}_1$  (the left image in Fig. 8b) segments out carbon atoms more accurately than the fiber surface for the rectangular trait  $\mathcal{T}_2$  (the right image in Fig. 8b). Next, we reduce the original data with the Gaussian-distributed hixel representation [58]. Specifically, we partition the original data into blocks of size  $2 \times 2 \times 2$  and represent each block as a Gaussian distribution with mean and standard deviation. The fiber surfaces corresponding to traits  $\mathcal{T}_1$  and  $\mathcal{T}_2$  are visualized for the hixel representation in Fig. 8c. The carbon atoms are again segmented well for trait  $\mathcal{T}_1$  with the high probability (yellow) regions (the left image of Fig. 8c). The two newly formed blobs in the left image of Fig. 8c that do not exist in the left image of Fig. 8b are due to the uncertainty in data arising from the hixel representation, but they have a very low probability (purple) of existence. As observed in the right image of Fig. 8c, the rectangular trait  $\mathcal{T}_2$  results in an overestimation of features, i.e., a multiple high probability (yellow) regions, that do not quite correctly represent carbon atoms.

### 8.3 Analysis of Simulation Datasets with Parametric Vs. Nonparametric Models of Uncertainty

We demonstrate an application of our proposed parametric and nonparametric statistical techniques (Sect. 4 and Sect. 5) in the context of oceanology for the analysis of ocean eddies. The knowledge of eddy positions is important for oceanologists to understand the transport of energy and biogeochemical particles in oceans. Fig. 9 visualizes the results of our probabilistic analysis of fibers that represent vortical features of the Red Sea dataset [53]. The dataset was downloaded from the IEEE SciVis Contest 2020 website (<https://kaust-vislab.github.io/SciVis2020/>) and comprises ensembles of multiple variables pertinent to the oceanology simulated for 60 time steps. In this experiment, we visualize the positional uncertainty of fibers corresponding to eddy features of the Red Sea by analyzing 20 ensemble members of the velocity field for a time step of 40 over the Gulf of Aden sampled on a grid of resolution  $250 \times 200 \times 50$ .

Initially, we derive the vorticity field for each ensemble member by computing the curl of the velocity field. We then compute the vorticity magnitude and Z (vertical) component of a vorticity vector per grid

vertex for each member. Fig. 9a visualizes a continuous scatterplot with the mean vorticity magnitude plotted on the vertical axis and the mean Z component of vorticity vector plotted on the horizontal axis. Trait  $\mathcal{T}$  (cyan polygon) is selected to understand the fiber positions with relatively high vorticity and a negative value of the Z component of a vorticity vector which denote anticyclonic eddy positions.

Fig. 9b visualizes the mean-field fiber surface corresponding to trait  $\mathcal{T}$ . The magenta box in the mean-field visualization indicates a rare occurrence of vortical features. Fig. 9c visualizes the results for the independent Gaussian (parametric) noise model (the same assumption as in the paper by Zheng and Sadlo [69]), but for nonrectangular polygonal shape (our contributions). The volume rendering for the Gaussian noise model (top image in Fig. 9c) indicates the presence of a relatively large number of vortical features inside the region enclosed by the magenta box because there are multiple regions with high interior probability (mapped to yellow). The probabilistic segmentation of the interior probability volume with an isosurface for isovalue 0.7 (the bottom image in Fig. 9c) clearly shows the presence of multiple vortical features inside the region enclosed by the magenta box. Fig. 9d visualizes results similar to Fig. 9c, but for the independent histogram noise model with 10 bins. Both volume rendering and probabilistic segmentation for the nonparametric noise model indicate a relatively low probability of vortical features inside the region enclosed by the magenta box. Such inconsistency regarding the presence of vortical features inside the magenta box across the three statistical models indicates the need for further eddy analysis in the same region.

We parallelized the implementation of the independent nonparametric histogram models with OpenMP C++ and CUDA using an approach similar to the one for the histogram results in Fig. 6. The serial computation of interior probabilities required 83.84 seconds, whereas computations with four and 16 threads resulted in a significant speed-up with 22.21 and 6.755 seconds, respectively, of computational time. The GPU CUDA computations required only 0.203 seconds.

Fig. 10 visualizes the results of our proposed statistical techniques for the asteroid impact dataset [42]. For our experiment, we analyze time step 189 of a simulation, in which an asteroid 250 m in diameter bursts at an elevation of 5 km above sea level and impacts the sea at an

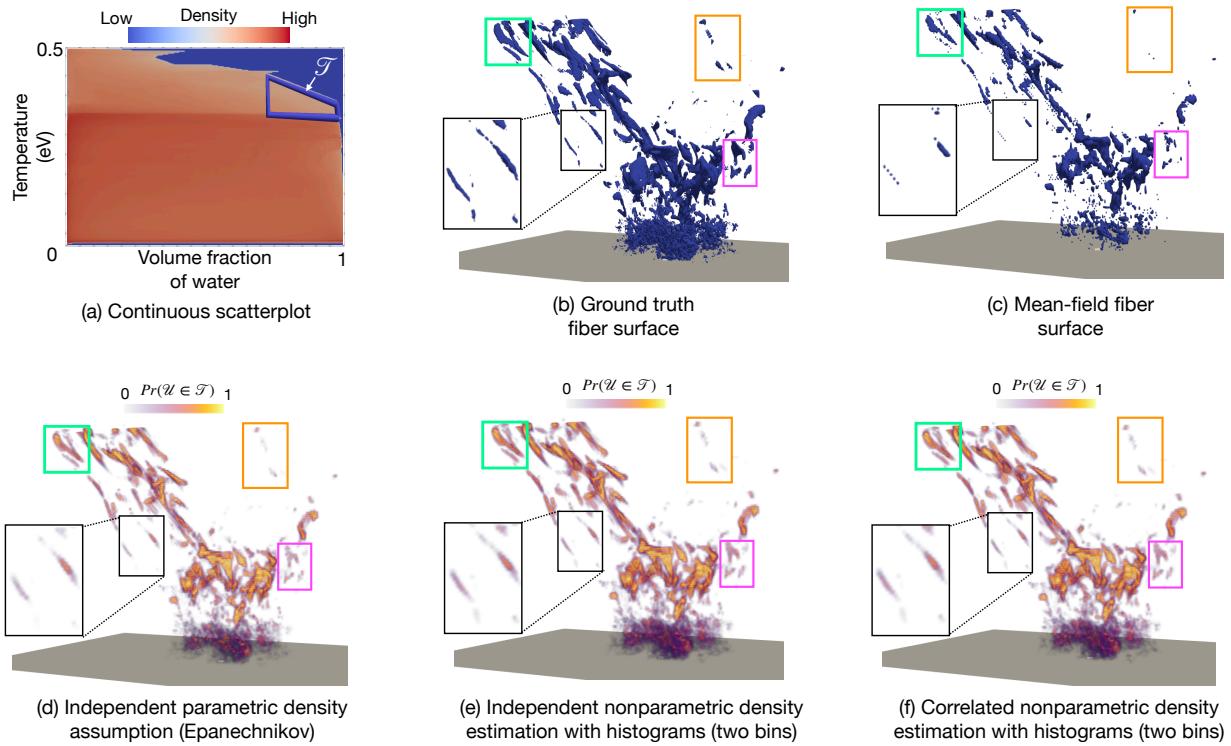


Fig. 10. Fiber visualizations for the deep-water asteroid impact simulation dataset comprising the temperature and volume fraction of water variables. Image (a) visualizes a continuous scatterplot of the two variables with trait  $\mathcal{T}$ , as indicated by a dark blue polygon. Image (b) visualizes the ground truth fiber surface that corresponds to trait  $\mathcal{T}$ . The data are reduced with the pixel representation, where we partition the data into blocks of size  $2 \times 2 \times 2$ , and each block is summarized with a probability distribution. The fiber positions are visualized for the pixel data using different statistical techniques in (c)–(f) for the trait  $\mathcal{T}$ . Both parametric and nonparametric noise models show enhanced feature recovery when compared to the mean-field, as illustrated by the colored boxes and inset zoom views, with image (b) treated as the reference.

angle of  $45^\circ$  (simulation series YB31 on the IEEE SciVis 2018 contest website <https://sciviscontest2018.org/>). We treat the dataset with resolution  $460 \times 280 \times 240$  as the reference. Fig. 10a visualizes a continuous scatterplot of the temperature and volume fraction of water variables for the reference dataset. Fig. 10b visualizes the fiber surface extracted from the reference volume for trait  $\mathcal{T}$ , as indicated by the dark blue polygon in Fig. 10a. Trait  $\mathcal{T}$  denotes the positions with a relatively high volume fraction of water and temperature. Next, we partition the data into  $2 \times 2 \times 2$  bricks and summarize each brick with a probability distribution similar to the pixel technique [58].

Fig. 10c visualizes the mean-field, which results in a significant loss of features caused by data variations in each brick. Figs. 10d–10f visualize the interior probabilities derived assuming the independent Epanechnikov (parametric) distribution (Sect. 4), independent histograms with two bins per variable (Sect. 5.1), and correlated histograms with two bins per variable (Sect. 5.2), respectively. Note that for the correlated 2D histograms in Fig. 10f, we compute interior probabilities in closed form, as described in Sect. 5.2. The probabilistic feature recovery with our proposed statistical frameworks is quite remarkable compared to the mean-field features, as illustrated by the boxes in Fig. 10, with Fig. 10b as the reference. The nonparametric statistics in Figs. 10e–10f show a slightly improved recovery of probabilistic features than the parametric statistics in Fig. 10d, as anticipated based on our synthetic data experiments (e.g., Fig. 5). A relatively small number of noise samples (here, eight per vertex) may not effectively capture noise correlation in the case of correlated nonparametric statistics (e.g., Fig. 7), thus yielding quite similar results in Figs. 10e–10f.

## 9 CONCLUSION AND FUTURE WORK

We extend the prior work by Zheng and Sadlo [69] to study the positional uncertainty of fibers for uncertain bivariate data. Specifically,

we present a closed-form statistical framework for spatial uncertainty quantification of fibers when data noise is characterized by independent parametric and nonparametric probability distributions. We perform our analysis for the uniform, Epanechnikov, and Gaussian kernels. Additionally, we present a numerical integration approach for uncertainty quantification of fibers when noise in bivariate data is correlated. We present our statistical frameworks for arbitrary shapes of FSCP as opposed to being limited to rectangular FSCP [52, 69]. We show that the nonparametric statistics improve the accuracy of uncertainty quantification and visualization when compared to the parametric and mean-field statistics via experiments on synthetic and simulation datasets.

In the future, we would like to extend our research to multivariate data with more than two variables. For bivariate data with correlated noise assumption, we would like to derive closed-form solutions (instead of numerical integration) similar to the independent noise assumption, if computable. Finally, we are interested in investigating the effect of uncertainty in multivariate data on the topology of fiber surfaces.

## ACKNOWLEDGMENTS

This work was partially supported by the Scientific Discovery through Advanced Computing (SciDAC) program in the U.S. Department of Energy, the Intel Graphics and Visualization Institutes of XELLENCE, the Intel OneAPI CoE, the NIH under award number R24 GM136986, the DOE under grant number DE-FE0031880, and the Utah Office of Energy Development. This research used resources of the Oak Ridge Leadership Computing Facility at the Oak Ridge National Laboratory, which is supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC05-00OR22725. We wish to thank Dr. Jieyang Chen at the Oak Ridge National Laboratory for helping us with the GPU code implementation and the reviewers of this article for their valuable feedback.

## REFERENCES

- [1] T. M. Athawale and C. R. Johnson. Probabilistic asymptotic decider for topological ambiguity resolution in level-set extraction for uncertain 2D data. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1163–1172, Jan. 2019. doi: 10.1109/TVCG.2018.2864505
- [2] T. M. Athawale, B. Ma, E. Sakaee, C. R. Johnson, and A. Entezari. Direct volume rendering with nonparametric models of uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1797–1807, Feb. 2021. doi: 10.1109/TVCG.2020.3030394
- [3] T. M. Athawale, D. Maljovec, L. Yan, C. R. Johnson, V. Pascucci, and B. Wang. Uncertainty visualization of 2D Morse complex ensembles using statistical summary maps. *IEEE Transactions on Visualization and Computer Graphics*, 28(4):1955–1966, Apr. 2022. doi: 10.1109/TVCG.2020.3022359
- [4] T. M. Athawale, E. Sakaee, and A. Entezari. Isosurface visualization of data with nonparametric models for uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2723–2732, Jan. 2016. doi: 10.1109/TVCG.2015.2467958
- [5] S. Bachthaler and D. Weiskopf. Continuous scatterplots. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1428–1435, Nov.-Dec. 2008. doi: 10.1109/TVCG.2008.119
- [6] J. Baerentzen and H. Aanaes. Signed distance computation using the angle weighted pseudonormal. *IEEE Transactions on Visualization and Computer Graphics*, 11(3):243–253, May-June 2005. doi: 10.1109/TVCG.2005.49
- [7] G. Bonneau, H. Hege, C. R. Johnson, M. Oliveira, K. Potter, P. Rheingans, and T. Schultz. Overview and state-of-the-art of uncertainty visualization. In M. Chen, H. Hagen, C. Hansen, C. R. Johnson, and A. Kauffman, eds., *Scientific Visualization: Uncertainty, Multifield, Biomedical, and Scalable Visualization*, pp. 3–27. Springer London, 2014. doi: 10.1007/978-1-4471-6497-5\_1
- [8] K. Brodlie, R. A. Osorio, and A. Lopes. A review of uncertainty in data visualization. In J. Dill, R. Earnshaw, D. Kasik, J. Vince, and P. C. Wong, eds., *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 81–109. Springer Verlag London, 2012.
- [9] H. Carr, Z. Geng, J. Tierny, A. Chattopadhyay, and A. Knoll. Fiber surfaces: Generalizing isosurfaces to bivariate data. *Computer Graphics Forum*, 34(3):241–250, July 2015. doi: 10.1111/cgf.12636
- [10] H. Carr, J. Snoeyink, and M. van de Panne. Flexible isosurfaces: Simplifying and displaying scalar topology using the contour tree. *Computational Geometry*, 43(1):42–58, Jan. 2010. Special Issue on the 14th Annual Fall Workshop. doi: 10.1016/j.comgeo.2006.05.009
- [11] H. Edelsbrunner, J. Harer, and A. K. Patel. Reeb spaces of piecewise linear mappings. In *Proceedings of the Twenty-Fourth Annual Symposium on Computational Geometry*, SCG '08, p. 242–250. Association for Computing Machinery, New York, NY, USA, June 2008. doi: 10.1145/1377676.1377720
- [12] G. Favelier, N. Faraj, B. Summa, and J. Tierny. Persistence atlas for critical point variability in ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1152–1162, Jan. 2019. doi: 10.1109/TVCG.2018.2864432
- [13] F. Ferstl, K. Bürger, and R. Westermann. Streamline variability plots for characterizing the uncertainty in vector field ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):767–776, Jan. 2016. doi: 10.1109/TVCG.2015.2467204
- [14] N. Fout and K.-L. Ma. Fuzzy volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2335–2344, Dec. 2012. doi: 10.1109/TVCG.2012.227
- [15] M. Friendly and D. Denis. The early origins and development of the scatterplot. *Journal of the History of the Behavioral Sciences*, 41(2):103–130, Spring 2005. doi: 10.1002/jhbs.20078
- [16] R. Fuchs and H. Hauser. Visualization of multi-variate scientific data. *Computer Graphics Forum*, 28(6):1670–1690, Sept. 2009. doi: 10.1111/j.1467-8659.2009.01429.x
- [17] D. Günther, J. Salmon, and J. Tierny. Mandatory critical points of 2D uncertain scalar fields. *Computer Graphics Forum*, 33(3):31–40, July 2014. doi: 10.1111/cgf.12359
- [18] H. Guo, W. He, T. Peterka, H.-W. Shen, S. M. Collis, and J. J. Helmus. Finite-time Lyapunov exponents and Lagrangian coherent structures in uncertain unsteady flows. *IEEE Transactions on Visualization and Computer Graphics*, 22(6):1672–1682, June 2016. doi: 10.1109/TVCG.2016.2534560
- [19] H. Hauser, F. Ledermann, and H. Doleisch. Angular brushing of extended parallel coordinates. In *IEEE Symposium on Information Visualization, 2002. INFOVIS 2002.*, pp. 127–130, Oct. 2002. doi: 10.1109/INFVIS.2002.1173157
- [20] S. Hazarika, A. Biswas, and H.-W. Shen. Uncertainty visualization using copula-based analysis in mixed distribution models. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):934–943, Jan. 2018. doi: 10.1109/TVCG.2017.2744099
- [21] J. Heinrich and D. Weiskopf. Continuous parallel coordinates. *IEEE Transactions on Visualization and Computer Graphics*, 15(6):1531–1538, Nov.-Dec. 2009. doi: 10.1109/TVCG.2009.131
- [22] W. R. Inc. Mathematica, Version 13.0.0. Champaign, IL, 2021.
- [23] A. Inselberg and B. Dimsdale. Parallel coordinates. In A. Klinger, ed., *Human-Machine Interactive Systems*, pp. 199–233. Springer US, Boston, MA, 1991. doi: 10.1007/978-1-4684-5883-1\_9
- [24] H. Jänicke, M. Böttger, and G. Scheuermann. Brushing of attribute clouds for the visualization of multivariate data. *IEEE Transactions on Visualization and Computer Graphics*, 14(6):1459–1466, Nov.-Dec. 2008. doi: 10.1109/TVCG.2008.116
- [25] J. Jankowai and I. Hotz. Feature level-sets: Generalizing iso-surfaces to multi-variate data. *IEEE Transactions on Visualization and Computer Graphics*, 26(2):1308–1319, Feb. 2020. doi: 10.1109/TVCG.2018.2867488
- [26] F. Jiao, J. M. Phillips, Y. Gur, and C. R. Johnson. Uncertainty visualization in HARDI based on ensembles of ODFs. In *2012 IEEE Pacific Visualization Symposium*, pp. 193–200, Feb.-Mar. 2012. doi: 10.1109/PacificVis.2012.6183591
- [27] C. R. Johnson and A. R. Sanderson. A next step: Visualizing errors and uncertainty. *IEEE Computer Graphics and Applications*, 23(5):6–10, Sept.-Oct. 2003. doi: 10.1109/MCG.2003.1231171
- [28] E. R. Johnson, S. Keinan, P. Mori-Sánchez, J. Contreras-García, A. J. Cohen, and W. Yang. Revealing noncovalent interactions. *Journal of the American Chemical Society*, 132(18):6498–6506, Apr. 2010. PMID: 20394428. doi: 10.1021/ja100936w
- [29] D. K. Jones. Determining and visualizing uncertainty in estimates of fiber orientation from diffusion tensor MRI. *Magnetic Resonance in Medicine*, 49(1):7–12, Dec. 2002. doi: 10.1002/mrm.10331
- [30] A. Kamal, P. Dhakal, A. Javaid, V. K. Devabhaktuni, D. Kaur, J. Zaientz, and R. Marinier. Recent advances and challenges in uncertainty visualization: a survey. *Journal of Visualization*, 24(5):861–890, May 2021. doi: 10.1007/s12650-021-00755-1
- [31] J. Kehrer and H. Hauser. Visualization and visual analysis of multifaceted scientific data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 19(3):495–513, Mar. 2013. doi: 10.1109/TVCG.2012.110
- [32] P. Klacansky, J. Tierny, H. Carr, and Z. Geng. Fast and exact fiber surfaces for tetrahedral meshes. *IEEE Transactions on Visualization and Computer Graphics*, 23(7):1782–1795, July 2017. doi: 10.1109/TVCG.2016.2570215
- [33] J. Kniss, G. Kindlmann, and C. Hansen. Multidimensional transfer functions for interactive volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 8(3):270–285, Jul.-Sept. 2002. doi: 10.1109/TVCG.2002.1021579
- [34] A. Knoll, Y. Hijazi, A. Kensler, M. Schott, C. Hansen, and H. Hagen. Fast ray tracing of arbitrary implicit surfaces with interval and affine arithmetic. *Computer Graphics Forum*, 28(1):26–40, Feb. 2009. doi: 10.1111/j.1467-8659.2008.01189.x
- [35] S. Liu, J. A. Levine, P.-T. Bremer, and V. Pascucci. Gaussian mixture model based volume visualization. In *IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 73–77, Oct. 2012. doi: 10.1109/LDAV.2012.6378978
- [36] S. Lodha, A. Pang, R. Sheehan, and C. Wittenbrink. Uflow: visualizing uncertainty in fluid flow. In *Proceedings of Seventh Annual IEEE Visualization '96*, pp. 249–254, Oct.-Nov. 1996. doi: 10.1109/VISUAL.1996.568116
- [37] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3D surface construction algorithm. *SIGGRAPH Computer Graphics*, 21(4):163–169, Aug. 1987. doi: 10.1145/37402.37422
- [38] C. Lundström, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty visualization in medical volume rendering using probabilistic animation. *IEEE Transactions on Visualization and Computer Graphics*, 13(6):1648–1655, Nov.-Dec. 2007. doi: 10.1109/TVCG.2007.70518
- [39] S. Nagaraj, V. Natarajan, and R. S. Nanjundiah. A gradient-based comparison measure for visual analysis of multifield data. In *Proceedings of the 13th Eurographics / IEEE - VGTC Conference on Visualization*, EuroVis'11, p. 1101–1110. The Eurographs Association & John Wiley & Sons, Ltd., Chichester, GBR, June 2011. doi: 10.1111/j.1467-8659.

2011.01959.x

- [40] M. Otto, T. Germer, and H. Theisel. Uncertain topology of 3D vector fields. In *2011 IEEE Pacific Visualization Symposium*, pp. 67–74, Mar. 2011. doi: 10.1109/PACIFICVIS.2011.5742374
- [41] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, Sept. 1962. doi: 10.1214/aoms/1177704472
- [42] J. M. Patchett, F. Samsel, K. Tsai, G. R. Gisler, D. H. Rogers, G. Abram, T. L. Turton, and L. Alamos. Visualization and analysis of threats from asteroid ocean impacts. Technical report, Nov. 2016.
- [43] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. *Computer Graphics Forum*, 30(3):931–940, June 2011. doi: 10.1111/j.1467-8659.2011.01942.x
- [44] K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In A. M. Dienstfrey and R. F. Boisvert, eds., *Uncertainty Quantification in Scientific Computing*, pp. 226–249. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. doi: 10.1007/978-3-642-32677-6\_15
- [45] K. Pöthkow and H.-C. Hege. Nonparametric models for uncertainty visualization. *Computer Graphics Forum*, 32(3pt2):131–140, July 2013. doi: 10.1111/cgf.12100
- [46] G. Quadri and P. Rosen. Modeling the influence of visual density on cluster perception in scatterplots using topology. *IEEE Transactions on Visualization & Computer Graphics*, 27(02):1829–1839, Feb. 2021. doi: 10.1109/TVCG.2020.3030365
- [47] F. Raith, C. Blecha, T. Nagel, F. Parisio, O. Kolditz, F. Günther, M. Stommel, and G. Scheuermann. Tensor field visualization using fiber surfaces of invariant space. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):1122–1131, Jan. 2019. doi: 10.1109/TVCG.2018.2864846
- [48] P. J. Rhodes, R. S. Laramee, R. D. Bergeron, and T. M. Sparr. Uncertainty Visualization Methods in Isosurface Rendering. In *Eurographics 2003 - Short Presentations*. Eurographics Association, 2003. doi: 10.2312/egs.20031054
- [49] O. Saeki. *Topology of singular fibers of differentiable maps*, vol. 1854 of *Lecture Notes in Mathematics*. Springer Heidelberg, Germany, 2004. doi: 10.1007/b100393
- [50] E. Sakaee and A. Entezari. A statistical direct volume rendering framework for visualization of uncertain data. *IEEE Transactions on Visualization and Computer Graphics*, 23(12):2509–2520, Dec. 2017. doi: 10.1109/TVCG.2016.2637333
- [51] D. Sakurai, K. Ono, H. Carr, J. Nonaka, and T. Kawanabe. *Flexible Fiber Surfaces: A Reeb-Free Approach*, pp. 187–201. Mathematics and Visualization. Springer Science and Business Media Deutschland GmbH, Germany, Dec. 2020. doi: 10.1007/978-3-030-43036-8\_12
- [52] S. Sane, T. M. Athawale, and C. R. Johnson. Visualization of Uncertain Multivariate Data via Feature Confidence Level-Sets. In M. Agus, C. Garth, and A. Kerren, eds., *Proc. EuroVis - Short Papers*. The Eurographics Association, 2021. doi: 10.2312/evs.20211053
- [53] S. Sanikommu, H. Toye, P. Zhan, S. Langodan, G. Krokos, O. Knio, and I. Hoteit. Impact of atmospheric and model physics perturbations on a high-resolution ensemble data assimilation system of the red sea. *Journal of Geophysical Research: Oceans*, 125(8):e2019JC015611, July 2020. doi: 10.1029/2019JC015611
- [54] N. Sauber, H. Theisel, and H.-p. Seidel. Multifield-graphs: An approach to visualizing correlations in multifield scalar data. *IEEE Transactions on Visualization and Computer Graphics*, 12(5):917–924, Sept.–Oct. 2006. doi: 10.1109/TVCG.2006.165
- [55] D. Schneider, J. Fuhrmann, W. Reich, and G. Scheuermann. A variance based FTLE-like method for unsteady uncertain vector fields. In R. Peikert, H. Hauser, H. Carr, and R. Fuchs, eds., *Topological Methods in Data Analysis and Visualization II: Theory, Algorithms, and Applications*, pp. 255–268. Springer Berlin Heidelberg, Berlin, Heidelberg, 2012. doi: 10.1007/978-3-642-23175-9\_17
- [56] F. Siddiqui, T. Höllt, and A. Vilanova. A progressive approach for uncertainty visualization in diffusion tensor imaging. *Computer Graphics Forum*, 40(3):411–422, June 2021. doi: 10.1111/cgf.14317
- [57] B. W. Silverman. *Density Estimation for Statistics and Data Analysis*. Taylor & Francis, New York, Oct. 1998. doi: 10.1201/9781315140919
- [58] D. Thompson, J. A. Levine, J. C. Bennett, P.-T. Bremer, A. Gyulassy, V. Pascucci, and P. P. Pébay. Analysis of large-scale scalar data using hixels. In *2011 IEEE Symposium on Large Data Analysis and Visualization*, pp. 23–30, Oct. 2011. doi: 10.1109/LDAV.2011.6092313
- [59] J. Tierny and H. Carr. Jacobi fiber surfaces for bivariate Reeb space computation. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):960–969, Jan. 2017. doi: 10.1109/TVCG.2016.2599017
- [60] J. Tierny, G. Favelier, J. A. Levine, C. Gueunet, and M. Michaux. The topology toolkit. *IEEE Transactions on Visualization and Computer Graphics*, 24(1):832 – 842, Jan. 2018. doi: 10.1109/TVCG.2017.2743938
- [61] P. e. a. Virtanen. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. *Nature Methods*, 17:261–272, Feb. 2020. doi: 10.1038/s41592-019-0686-2
- [62] M. P. Wand and M. C. Jones. Comparison of smoothing parameterizations in bivariate kernel density estimation. *Journal of the American Statistical Association*, 88(422):520–528, June 1993. doi: 10.1080/01621459.1993.10476303
- [63] M. O. Ward. *Multivariate Data Glyphs: Principles and Practice*, pp. 179–198. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008. doi: 10.1007/978-3-540-33037-0\_8
- [64] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2713–2722, Dec. 2013. doi: 10.1109/TVCG.2013.143
- [65] K. Wu, A. Knoll, B. J. Isaac, H. Carr, and V. Pascucci. Direct multifield volume ray casting of fiber surfaces. *IEEE Transactions on Visualization and Computer Graphics*, 23(1):941–949, Jan. 2017. doi: 10.1109/TVCG.2016.2599040
- [66] K. Wu and S. Zhang. A contour tree based visualization for exploring data with uncertainty. *International Journal for Uncertainty Quantification*, 3:203–223, 2012. doi: 10.1615/IntJ.UncertaintyQuantification.2012003956
- [67] Z. Xie, S. Huang, M. O. Ward, and E. A. Rundensteiner. Exploratory visualization of multivariate data with variable quality. In *2006 IEEE Symposium On Visual Analytics Science And Technology*, pp. 183–190, Oct.–Nov. 2006. doi: 10.1109/VAST.2006.261424
- [68] L. Yan, Y. Wang, E. Munch, E. Gasparovic, and B. Wang. A structural average of labeled merge trees for uncertainty visualization. *IEEE Transactions on Visualization and Computer Graphics*, 26(1):832–842, Jan. 2020. doi: 10.1109/TVCG.2019.2934242
- [69] B. Zheng and F. Sadlo. Uncertainty in continuous scatterplots, continuous parallel coordinates, and fibers. *IEEE Transactions on Visualization and Computer Graphics*, 27(2):1819–1828, Feb. 2021. doi: 10.1109/TVCG.2020.3030466
- [70] T. Zuk and S. Carpendale. Visualization of uncertainty and reasoning. In *Proceedings of the 8th international symposium on Smart Graphics*, pp. 164–177, June 2007. doi: 10.1007/978-3-540-73214-3\_15