Isosurface Visualization of Data with Nonparametric Models for Uncertainty

3 authors, including:

Tushar Athawale
University of Utah
4 PUBLICATIONS   **50** CITATIONS

Some of the authors of this publication are also working on these related projects:

Project   Marching Squares/Cubes Analysis for Uncertain Data View project

Isosurface Visualization of Data with Nonparametric Models for Uncertainty

Tushar Athawale, Elham Sakhaee, and Alireza Entezari, *Senior Member, IEEE*

Abstract—The problem of isosurface extraction in uncertain data is an important research problem and may be approached in two ways. One can extract statistics (e.g., mean) from uncertain data points and visualize the extracted field. Alternatively, data uncertainty, characterized by probability distributions, can be propagated through the isosurface extraction process. We analyze the impact of data uncertainty on topology and geometry extraction algorithms. A novel, edge-crossing probability based approach is proposed to predict underlying isosurface topology for uncertain data. We derive a probabilistic version of the midpoint decider that resolves ambiguities that arise in identifying topological configurations. Moreover, the probability density function characterizing positional uncertainty in isosurfaces is derived analytically for a broad class of nonparametric distributions. This analytic characterization can be used for efficient closed-form computation of the expected value and variation in geometry. Our experiments show the computational advantages of our analytic approach over Monte-Carlo sampling for characterizing positional uncertainty. We also show the advantage of modeling underlying error densities in a nonparametric statistical framework as opposed to a parametric statistical framework through our experiments on ensemble datasets and uncertain scalar fields.

Index Terms—Uncertainty quantification, linear interpolation, isosurface extraction, marching cubes, nonparametric statistics

✦

1 Introduction

Visualization techniques have become indispensable for exploration and analysis of data produced by simulations or acquired in observational studies. One of the key challenges in visualization research is the integration of uncertainty, inherent in acquisition or simulation processes, into the visualization pipeline [44, 13, 3]. Addressing this challenge is of paramount importance for building reliable visualization systems and can lead to a greater penetration of visualization tools in diverse application domains [23].

Data uncertainty is often attributed to inaccuracies in the acquisition or in the limited and incomplete measurements available to the computational simulation process. In observational studies, uncertainties are introduced at the acquisition stage, and one can characterize this uncertainty, based on the physical models describing the sensing process by a probability distribution at each data point. Moreover, sampling (discretization) and quantization errors further contribute to data uncertainty and lead to non-trivial transformations of probability distributions. Similarly, in simulation science the uncertainty associated with model parameters leads to variability in the computed solution. A common approach is to run an *ensemble simulation* that seeks to capture the diversity of solutions, as the input variables are sampled in the uncertainty range. Based on the diversity of the solutions, one can make inference with statistical analysis of the ensemble members [42, 41].

On the other hand, with the ever-growing size of data sets produced from simulations or acquired (e.g., biomedical applications), the data is usually processed, sometimes reduced, and often compressed before analysis and visualization. These post-processing operators also transform these probability distributions that describe the uncertainty at data points. More importantly, data processing in the visualization algorithms introduce further transformations to these distributions making the quantification of uncertainty as the data propagates through the analysis pipeline particularly challenging.

In this paper, we study the interaction of data uncertainty with isosurface visualization and characterize the transformations introduced by isosurface extraction to the uncertainty of the input data. From a high level point of view, the issue of data uncertainty in isosurface visualization can be dealt with in two ways. One can extract a statistic (e.g., mean, median) from the probability distribution at each data point – effectively denoising the data – and visualize the resulting field. On the other hand, one can propagate the probability distributions through the cell topology and geometry extraction stages of isosurface visualization. We analyze the cell configuration and the geometry extraction steps of the marching cubes algorithm in a (nonparametric) statistical framework. The former is cast as a classification problem which is solved via optimization, and the latter is solved in closed-form to quantify positional uncertainties in isosurfaces. This approach can be used for the analysis of uncertainty throughout and after the extraction and for quantifying the uncertainty specific to the visualized isosurface. We contrast and discuss the advantages of the propagation approach to the static (e.g., mean) field visualization.

As the data propagates through various stages of processing, analysis and visualization pipeline, these probability distributions are transformed by various data processing operators. To enable the analysis of uncertainty propagation, we employ nonparametric distributions (e.g., histograms, empirical distributions, kernel density methods) for modeling uncertainty since they provide the flexibility needed to characterize these transformations. In nonparametric statistics, one does not assume that the structure of the density function is fixed unlike parametric statistics, and the complexity of the model grows in size to accommodate the complexity of the task at hand. Parametric models significantly restrict characterization of the realistic distributions, whereas nonparametric models allow us to capture a broader set of distributions for modeling uncertainty at each data point and analyze its propagation in applications. The flexibility of nonparametric densities has recently been advocated for uncertainty visualization [33].

In ensemble simulation datasets, the visual analysis of variability of an isosurface through simultaneous visualization of all ensemble members can be challenging. Our method provides a probabilistic isosurfacing approach that can be used for the characterization of the variations of the isosurface after the extraction from the ensemble. In our approach, we represent the data uncertainty at each grid point by a nonparametric distribution that is estimated from ensemble members (e.g., histograms in the simplest case). This approach also can be used for compact representation of data distributions when the number of ensemble members is large and data reduction is needed (e.g., re-binning histograms/kernels).

Specifically, our contributions include:

- We propose a technique for predicting isosurface topology by casting the marching cubes configuration cases (i.e., vertex signs), in a probabilistic framework, as a classification prob-

Tushar Athawale, Elham Sakhaee and Alireza Entezari are with the *Department of Computer and Information Science and Engineering at the University of Florida.*

E-mails: athawale@cise.ufl.edu, esakhaee@cise.ufl.edu, entezari@cise.ufl.edu

lem. Our approach leverages *vertex-crossing probabilities* and *edge-crossing probabilities* [34] for determination of cell configuration. The proposed probabilistic approach provides more accurate prediction of the isosurface topology than the non-probabilistic approach e.g., statistical mean in uncertain data.

- We quantify and visualize the spatial uncertainties in an isosurface extracted from uncertain data when the underlying uncertainty is modeled in a nonparametric statistical framework. In this setting, Monte Carlo sampling is the simplest but expensive approach to estimate *the distribution of inverse interpolation* for characterizing isosurface geometry. However, we provide, in closed-form, the distribution of inverse interpolation characterizing the isosurface level-crossing positions. This closed-form solution can be used for efficiently reconstructing the expected isosurface, and its variance can illustrate the positional uncertainty in the isosurface. Our closed-form derivation generalizes uniform distribution studied in [1] to general kernels such as a triangle, Epanechnikov, and other common kernels used in non-parametric density estimation. The flexibility offered by the non-parametric models can be used to provide more accurate isosurface reconstruction than the one corresponding to the parametric models.

- When a grid cell is classified to be crossed by isosurface, and cell configuration happens to be one of the ambiguous cases [17], we devise a *probabilistic midpoint decider* for ambiguity resolution. While the existing techniques resolve the topological ambiguities when data is *not* uncertain, the *probabilistic midpoint decider* generalizes the ambiguity resolution methods to resolve the ambiguities when data is uncertain (e.g., noisy or ensemble). Uncertain midpoint decider derives the midpoint distribution in closed-form for nonparametric density functions using splines.

- Non-local means (NLM) is a common technique used in image processing that forms a non-local neighborhood of a data point in a noisy dataset, based on their (patch) similarity. The mean computed over this non-local neighborhood is used to denoise that data point [5]. Since our framework can be used for propagating the probability distributions through isosurface extraction, we demonstrate that the *distribution of non-local neighbors* (e.g., histogram) – as opposed to just the mean – could provide a more accurate reconstruction of the isosurface compared to the isosurface extracted from the NLM-denoised dataset.

2 Related Work

Uncertainty quantification and visualization of the quantified uncertainties are considered among top challenges in visualization [44, 13, 12], and there is a growing body of research for developing effective means for visualizing uncertainty [3]. Uncertainty in isosurfaces can be visualized by mapping color and opacity proportional to spatial uncertainty [37]. Osorio and Brodlie devised techniques [24] for visualization of the possible spatial locations for uncertain 2-D isocontours. A point-based displacement method was proposed by Grigoryan and Rheingans [10, 11] to interactively visualize uncertainty in the isosurface. Various animation techniques such as visual vibrations and probabilistic animation can be used to represent uncertainty in the data [4, 20]. Wittenbrink et al. proposed a non-overloading visualization technique [43, 28] (as opposed to overloading techniques such as pseudocoloring, opacity mapping) called glyphs to encode and visualize uncertainty in the vector field data. The visualization techniques for scalar and vector field data were extended in [21, 19] for visualizing spatial distribution datasets, in which at each grid location data distribution is known.

Box plots are often used to encode various statistical quantities such as mean, median, maximum, minimum etc. [22]. Whitaker et al. devised contour boxplots [42] to visualize statistical quantities related to the ensemble of the isocontours extracted from the ensemble simulation datasets. Pfaffelmoser and Westermann studied structural

variability [31] of the height fields by visualizing the global correlation structures. Otto et al. visualized global uncertainty in 2-D vector fields [25] by computing probability density functions of particle flows. The authors further extended this technique to uncertain 3-D vector fields [26] with some performance improvements. Schlegel et al. [40] demonstrated the suitability of Gaussian process regression for interpolating uncertain data, given that a suitable covariance model exists (whose parameters are estimated from the data). Thompson et al. proposed a fuzzy isosurface visualization technique [41] to visualize the possible isosurface locations in an uncertain scalar field. In this technique, the researchers sliced the histogram of an ensemble of values corresponding to each cell and performed direct volume rendering of the likelihood that cell will be crossed by the isosurface. Lee and Shen [16] devised an algorithm for efficient storage of integral histograms and fast query of local histograms at arbitrary locations within the data.

Pöthkow and Hege developed a volume rendering approach for visualizing positional uncertainty in an extracted isosurface [32]. In their approach, given an isosurface, the probability of it crossing each cell of the grid is computed when datapoints are modeled as independent normal distributions. These cell crossing probabilities are estimated using a Monte Carlo approach and are visualized using a volume rendering approach. Pöthkow et al. further extended this work to take into account the correlated Gaussian random fields [35]. Pfaffelmoser et al. integrated computations of isosurface-crossing probabilities into the front-to-back ray casting approach for volume rendering by considering distance dependent pairwise correlations among the samples along a ray [30]. For the interactive visualization of the level-crossing uncertainties, expensive Monte Carlo computations were replaced by fast techniques such as maximum edge crossing probability and linked-pairs [34]. Motivated by this work, Athawale and Entezari [1] devised an efficient approach for closed-form computation of the level-crossing probabilities when the underlying uncertainties are modeled by uniform distributions (in a parametric framework). This approach eliminates the expensive Monte Carlo computations which need to be carried out per cell with closed-form formulas that are calculated a priori [1].

Empirical distribution, histograms, kernel density estimation, and weighted kernel density estimation are some of the nonparametric statistical approaches that can be used for characterizing data uncertainty. The density estimate obtained from empirical distribution is discontinuous, since this technique estimates the probability mass only for the observed sample values. Continuous estimate of the underlying density can be obtained with a histograming approach, where kernels are non-overlapping, and the choice of kernels is limited to the uniform kernel.

Kernel-based nonparametric density estimation approaches (i.e., Parzen [29] and Rosenblatt [38]) determine the set of the possible density estimates that asymptotically approximate the underlying distributions. This nonparametric approach relaxes two restrictions of histograms, i.e., smoother kernels such as a triangle, cubic, or Epanechnikov may be chosen, and they can overlap. Kernel density estimation is, therefore, considered as a generalization of histograms. Each of the kernels may be assigned equal or different weights based on the sample location. Smooth estimates of the underlying density may be obtained by a suitable choice of kernel and optimal bandwidth (i.e., scaling of kernel).

Pöthkow and Hege extended uncertainty analysis for the isosurfaces to nonparametric uncertainty models [33]. The authors approximated the level-crossing probabilities again using the Monte Carlo sampling. One of the contributions of this paper is the generalization of the analytic formulation of the level-crossing probability (i.e., the distribution of inverse linear interpolation) from the uniform density model to the nonparametric densities with commonly used kernels. Our approach for quantifying spatial uncertainty in isosurfaces is applicable when nonparametric densities are estimated with compactly-supported kernels such as a triangle, Epanechnikov, cubic, and so on. As the choice of kernel bandwidth is key to kernel density estimation, we detail our approach for bandwidth selection for density estimation in

the Appendix. For clarity of presentation, we present our approach for closed-form characterization of level-crossing probability of an isosurface with the triangle kernel for nonparametric models. This approach is directly applicable to other compactly-supported kernels (e.g., quadratic, cubic, Epanechnikov).

3 Marching Uncertain Cubes

The marching cubes algorithm (MC) [18, 3] provides a triangulation of an isosurface from a scalar field defined on a Cartesian grid and consists of two steps. The first step involves classifying grid vertices as above or below, depending on the scalar field's value at each grid point and its relation to the isovalue c. Vertex signs are then used to construct the cell configuration, where the topology of the isosurface inside each grid cell is determined. The second step determines isosurface geometry by computing the crossing locations on the grid edges that belong to a cell that has been identified as crossed in the first step. Level-crossing locations are computed using the inverse linear interpolation formula in order to reconstruct isosurface that is consistent with trilinear interpolation. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ denote adjacent vertices in a cell that have been *classified* with opposite signs during the cell configuration step, and x and y be the data values at $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively. The isosurface corresponding to the isovalue c has a vertex $\mathbf{v}_c$ whose location can be computed using the *inverse linear interpolation* formula as follows:

$$\mathbf{v}_c = (1-z)\mathbf{v}_1 + z\mathbf{v}_2, \quad \text{where} \quad z = \frac{c-x}{y-x}. \quad (1)$$

Here, the ratio z determines the location of level-crossing between $\mathbf{v}_1$ and $\mathbf{v}_2$.

When there is uncertainty in the values assumed at each grid point, the classification of vertices (i.e., determination of sign) as well as the inverse linear interpolation steps need to be revisited in a probabilistic view. In our work, we assume an independent noise model for propagating the data uncertainty into the steps of classical MC algorithm. In the following section, we present our classification approaches for identifying cell topology, and in Subsection 3.3, we discuss the probabilistic extension of ambiguity resolution methods when a cell is identified as an ambiguous case [17]. Subsection 3.2 presents our closed-form characterization of level-crossing probability for geometry extraction.

3.1 Cell Topology from Classification of Vertices

We first describe the problem setting. Let $\mathbf{v}_1, \dots \mathbf{v}_N$ denote the vertices in a Cartesian grid, and $f(\mathbf{v}_n)$ be the scalar value attained at vertex $\mathbf{v}_n$ for $1 \leq n \leq N$. When the scalar values are known exactly, as in the classical MC algorithm, the classification of vertices is carried out based on the isovalue c: vertex $\mathbf{v}_n$ is classified as $+1$ when $f(\mathbf{v}_n) > c$ and -1 otherwise. When there is uncertainty present in the data, the common approach is to model $f(\mathbf{v}_n)$ with a random variable, X_n, whose probability distribution, $\text{pdf}_{X_n}(x)$, reflects the uncertainty at the grid point $\mathbf{v}_n$. We want to find a scheme that best represents the underlying vertex classification for uncertain data.

In a given problem setting, one can classify the grid vertices based on the most probable sign: vertex $\mathbf{v}_n$ classified as $+1$ when $Pr(X_n > c) = \int_c^\infty \text{pdf}_{X_n}(x)dx > 0.5$ and -1 otherwise. We refer to this method as *vertex-based classification*. One widely used approach for estimating vertex signs is taking the statistical mean of uncertain data. Classification based on the statistical mean can be poor if the sign corresponding to statistical mean and the most probable sign do not agree. The behavior is usually observed when noise samples are far away from the underlying value, as demonstrated in Section 5 (e.g., Fig. 7).

Since the vertex-based classification approach processes each vertex independently, it might lead to inaccurate vertex classification when the vertex in question is almost equally likely to be on either side of the isovalue (e.g., $Pr(X > c) \approx 0.5$). We might avoid this problem by classifying the vertices in a *collaborative* fashion. We propose a vertex classification approach that decides vertex signs by processing vertices in a collaborative fashion, and we also show an experiment illustrating the benefit of a proposed method over a vertex-based classification method (Fig. 9).
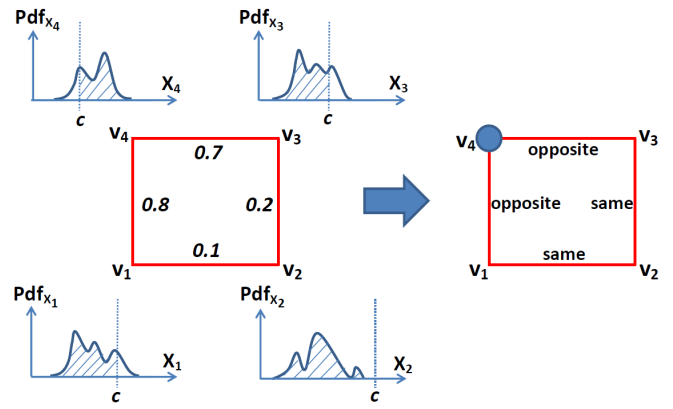

Fig. 1: Edge-based vertex classification approach for deciding cell topology. Random variable X_i with probability density Pdf_{X_i} represents data uncertainty at cell corner v_i. Numbers on edges represent exemplar level-crossing probabilities (LCPs) for isovalue c. When LCP is relatively high, edge vertices are likely to be classified with opposite signs, e.g., edges v_1v_4 and v_3v_4. When LCP is relatively low, edge vertices are likely to be classified with same signs, e.g., edges v_1v_2 and v_2v_3. Marked vertices (with dark blue circle) represent signs opposite to unmarked vertices.

Specifically, one can focus on the probability of a pair of vertices being on the opposite sides of the isovalue in a grid. As a vertex is being examined against other vertices for being in the same or opposite side of the isovalue, one can classify it as ± 1 so that there is a global fitness optimality. The level-crossing probability (LCP) concept, introduced by Pöthkow and Hege [34], assigns a probability of the isosurface crossing an edge between two adjacent vertices in a grid. Conceptually, if a vertex is deemed opposite to all of its neighbors because the level-crossing probabilities (LCPs) on its edges are high, it should be classified as opposite to its neighbors (if there is consensus among them). The two vertices sharing an edge with low LCP are likely to be classified with the same sign (i.e., both $+1$ or both -1), whereas two vertices sharing a high LCP edge are likely to be classified as opposite signs. Fig. 1 depicts the *edge-based classification* approach for uncertain data.

The basic idea behind this edge-based classification is: given the likelihood of an isosurface crossing the edges of a Cartesian grid, what is the optimal assignment of ± 1 to its vertices that best fits these crossing probabilities? Although crossing probabilities are computed from vertex distributions, it is the collaborative classification of vertices (a vertex with respect to all of its neighbors) that differentiates this method from the vertex-based classification. To build a global notion of fitness for this classification, we build an objective function that penalizes same-sign classification of vertices sharing a high LCP edge and favors opposite-sign classification of vertices sharing a high LCP edge. Let s_i and s_j denote the signs assigned to $\mathbf{v}_i$ and $\mathbf{v}_j$ vertices, respectively, and $w_{i,j} = 1 - Pr(X_i > c)Pr(X_j > c) - Pr(X_i < c)Pr(X_j < c)$ be the LCP [34] on the edge between $\mathbf{v}_i$ and $\mathbf{v}_j$. Then $s_i w_{i,j} s_j > 0$ has a positive penalty for same-sign classification and a negative cost for opposite-sign classification. Let $\mathbf{s} = [s_1, \dots, s_N]$ denote the signs for all grid vertices, and $\mathbf{W} = [w_{i,j}]$ with $1 \leq i, j \leq N$ denote the LCP matrix for all grid edges. Then the optimal classification is determined by:

$$\mathbf{s}^* = \arg\min_{s_n = \pm 1} \mathbf{s}^T \mathbf{W} \mathbf{s}. \quad (2)$$

The cost for the combinatorial approach for solving this binary quadratic assignment problem [39] is prohibitive (even though $\mathbf{W}$ is very sparse and banded), and for a practical solution one can examine a relaxed version of this objective function. One approximate solution can be obtained by relaxing the $s_n = \pm 1$ constraint to the quadratic constraint $\mathbf{s}^T \mathbf{s} = 1$. The solution (via the method of Lagrange multipli-
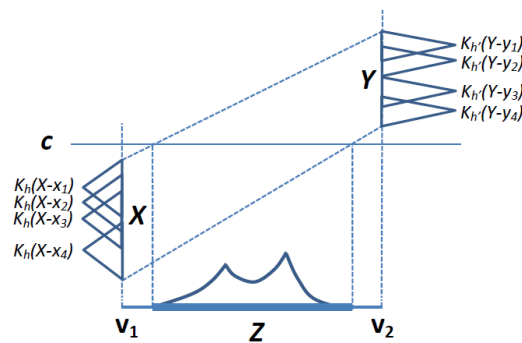
Fig. 2: A ratio density in nonparametric statistics. Random variables X and Y represent uncertain data values at the grid vertices $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, having nonparametric distributions with the triangle function as the basis kernel. $K_h(X - x_i)$ denotes a triangle function with bandwidth h centered at $X = x_i$, and $K_{h'}(Y - y_j)$ denotes a triangle function with bandwidth h' centered at $Y = y_j$. Distribution of the ratio random variable, Z, is illustrated for the isovalue of c.

Fig. 3: The joint density of the *dependent* random variables Z_1 and Z_2 when kernel-based densities for the random variables X and Y consist of one triangle each. X and Y take values in the range $[x_1 - h, x_1 + h]$ and $[y_1 - h', y_1 + h']$, respectively, and c denotes the isovalue. A, B, C, and D represent the midpoints of the edges PQ, QR, RS, and SP, respectively, and E is at the center of the parallelogram. The joint density function of Z_1 and Z_2 is a piecewise quadratic function consisting of four polynomials highlighted with different colors with their supports being the subparallelograms $PAED$, $AQBE$, $EBRC$, and $DECS$.

ers for example) to this relaxed problem is the eigenvector of $\mathbf{W}$ corresponding to its largest (negative) eigenvalue. For large datasets, eigenvector computation is still costly, and one can resort to the Nyström method [9] for efficient computation. In Section 5, we demonstrate the effectiveness of the approximate solution to an edge-based classification problem compared to a vertex-based classification.

3.2 Ratio Distribution for Geometry Extraction

Our earlier work [1] characterized the ratio density, in closed-form, for uniform parametric densities. We now derive ratio density analytically for kernel-based density estimators. Our experiments in Section 5 (e.g., Fig. 8) show that nonparametric models provide more accurate characterization of the error densities than the parametric models, and therefore, providing more accurate reconstruction results.

Problem setting: Let X and Y be the random variables representing the uncertainty in the data values at adjacent grid vertices $\mathbf{v}_1$ and $\mathbf{v}_2$, respectively, and x_i $(i = 1 \ldots n)$ and y_j $(j = 1 \ldots m)$ be the independent, identically distributed (IID) samples available from the distributions of X and Y. The kernel-based density estimators for the random variables X and Y are given by the functions $\text{pdf}_X(x) = \frac{1}{n} \sum_{i=1}^{n} K_h(x - x_i)$ and $\text{pdf}_Y(y) = \frac{1}{m} \sum_{j=1}^{m} K_{h'}(y - y_j)$, respectively, where $K(\cdot)$ denotes kernel function, and h and h' are the bandwidth parameters [14]. Given that $\mathbf{v}_1$ and $\mathbf{v}_2$ are classified as opposite signs, the uncertainty in the location of the level-crossing $\mathbf{v}_c$ on the edge between them is given by the ratio random variable $Z = \frac{c - X}{Y - X}$ as a result of the inverse interpolation in Eq. (1). We now explain the approach to derive the density function, in closed-form, of the level-crossing random variable Z. Fig. 2 depicts the problem setting.

3.2.1 Nonparametric Density With Single Kernel

Initially, we describe a three-step process for analytically deriving the ratio distribution for a simple case, where pdf_X and pdf_Y consist of a single triangle kernel each, i.e., when $n = 1$ and $m = 1$. In the first step, we establish the joint distribution of the *dependent* random variables $Z_1 = c - X$ and $Z_2 = Y - X$. Since $n = 1$ and $m = 1$ in this case, the random variables X and Y assume the values in the intervals $[x_1 - h, x_1 + h]$ and $[y_1 - h', y_1 + h']$, respectively. Hence, the distribution function of the random variable Z_1 is a triangle function consisting of two linear polynomials over the range $[c - x_1 - h, c - x_1 + h]$. The density function of the random variable Z_2 is also a triangle function composed of 2 linear polynomials over the range $[y_1 - h' - x, y_1 + h' - x]$, where the random variable X takes the value x. The support of the joint density function of Z_1 and Z_2 is a parallelogram [1].

When the random variables X and Y are independent, their joint distribution is given by a product of their densities, pdf_X and pdf_Y. In

case of single triangle kernels, it is given by $K_h(X - x_1) \cdot K_{h'}(Y - y_1)$. Since $Pr(Z_1, Z_2) = Pr(c - X, Y - X)$, the joint distribution of Z_1 and Z_2 is a reflected and sheared version of $Pr(X, Y)$. In the case of a single triangle, $Pr(Z_1, Z_2)$ is simply a reflected and sheared version of the product $K_h(X - x_1) \cdot K_{h'}(Y - y_1)$. As $K_h(X - x_1) \cdot K_{h'}(Y - y_1)$ is a piecewise polynomial function when the kernel is a triangle function, $Pr(Z_1, Z_2)$ is also a piecewise polynomial function consisting of four quadratic polynomials with their supports being the subparallelograms shown in Fig. 3. In contrast, the joint is a constant function for uniform distributions, as discussed in [1].

The second step consists of formulating the cumulative density function of a random variable Z. The value of the ratio random variable Z can be represented by slope of the line $Z_1 = zZ_2$. The expression for the cumulative density function, $\text{cdf}_Z(z)$, can be obtained by integrating the joint density of Z_1 and Z_2 as we sweep the line $Z_1 = zZ_2$ from slope $z = 0$ to $z = 1$:

$$
\begin{aligned}
\text{cdf}_Z(z) &= Pr(0 \le Z \le z) \\
&= Pr(0 \le Z_1 \le zZ_2) + Pr(zZ_2 \le Z_1 < 0).
\end{aligned}
\tag{3}
$$

This sweeping process amounts to the integration of the joint as discussed in [1]; however, unlike the simple case of constant functions in [1], the integration of the joint involves integrating high-order polynomials for kernel functions, such as triangle or Epanechnikov. To perform this integration, we use *Green's theorem*, which relates the integral over a domain of the joint to its line integrals. The process of computing the $\text{cdf}_Z(z)$ is illustrated in Fig. 4. We integrate the polynomials over the subparallelograms using Green's theorem.

Lastly, the density function $\text{pdf}_Z(z)$ for the ratio random variable Z is obtained by differentiating the cumulative density function $\text{cdf}_Z(z)$ with respect to z. Our approach for the analytic characterization of the ratio random variable can be expanded for other kernels of higher degree with compact support, such as Epanechnikov, quartic, tricube, etc. When using kernels of even degree, e.g. uniform or Epanechnikov, the joint density function of the Z_1 and Z_2 is a single polynomial over its entire support. For the kernels of odd degree, e.g. triangle or tricube, the resulting joint distribution function is a piecewise polynomial. In either case, the polynomials can be integrated using Green's theorem.

3.2.2 Nonparametric Density With Multiple Kernels

When $n > 1$ or $m > 1$, we follow an approach similar to one described in Subsection 3.2.1 for the closed form characterization of the ratio distribution. Since $Pr(Z_1, Z_2) = Pr(c - X, Y - X)$, the joint distribution
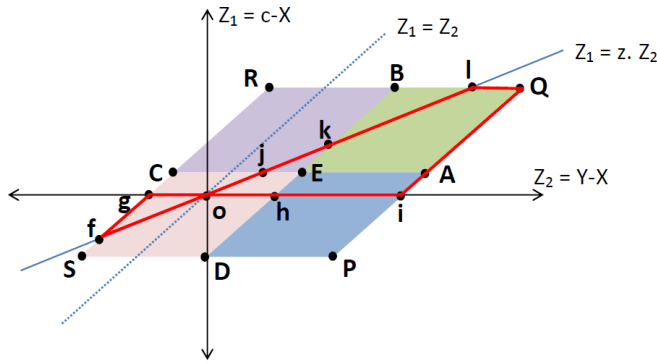
Fig. 4: The process for computing the cumulative distribution of Z, $\mathrm{cdf}_Z(z)$, as the line $Z_1 = z \cdot Z_2$ is swept from $z = 0$ (Z_2 axis) to $z = 1$ (dotted line). The $\mathrm{cdf}_Z(z)$ is determined by integrating a piecewise polynomial function of the joint density (denoted by subparallelograms with different colors) of Z_1 and Z_2 swept by the line $Z_1 = z \cdot Z_2$ (polynomials falling in the region enclosed by the red border). The polynomials over the polygons ogf, $ohej$, $hiAE$, $AQlkE$, and jEk can be integrated using the Green's theorem.

of random variables Z_1 and Z_2 is, again, the reflected and sheared version of $Pr(X, Y)$. Thus, in case of multiple kernels, as shown in Fig. 5, $Pr(Z_1, Z_2)$ is the reflected and sheared version of the following product:

$$Pr(X = x, Y = y) = \frac{1}{nm} \sum_{i=1}^{n} \sum_{j=1}^{m} K_h(x - x_i) K_{h'}(y - y_j). \quad (4)$$

The analysis for each term, $K_h(X - x_i) \cdot K_{h'}(Y - y_j)$, of the joint in Eq. (4) is analogous to the simple case of a single kernel, as explained in Subsection 3.2.1. $Pr(Z_1, Z_2)$ is, therefore, the superposition of joint distribution for each pair of kernels from pdf_X and pdf_Y. In the case of a triangle kernel, the density of the joint distribution of Z_1 and Z_2 is the sum of the piecewise polynomial functions corresponding to each pair of kernels from the nonparametric densities, pdf_X and pdf_Y, scaled by $\frac{1}{nm}$. The sweeping/integration process is similarly a superposition of every pair of kernels from the two distributions. The process for deriving the cumulative density function, $\mathrm{cdf}_Z(z)$, is shown in Fig. 5 when $n = 2$ and $m = 2$, from which pdf_Z is obtained via analytic differentiation. This process is general for any compactly-supported kernel, such as a triangle, Epanechnikov, and cubic. We provide an in-depth analysis of the complexity of the proposed model for computing ratio distribution in the Appendix.
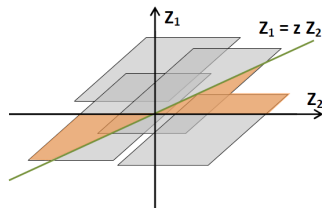
Fig. 5: The joint distribution of Z_1 and Z_2 when nonparametric distributions for X and Y consist of 2 triangle kernels each. Each parallelogram represents the piecewise quadratic polynomial function consisting of 4 polynomials over 4 subparallelograms (not explicitly shown as in Fig. 3) for every pair of kernels. The cumulative distribution of Z, $\mathrm{cdf}_Z(z)$, is determined by integrating a piecewise polynomial function for each of the parallelograms swept by the line $Z_1 = z \cdot Z_2$ (highlighted in orange) using the Green's theorem.
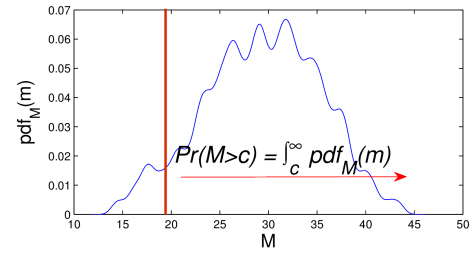

Fig. 6: Example of uncertain midpoint decider. M denotes the midpoint random variable, $\mathrm{pdf}_M(m)$ is the probability density of the random variable M, and c represents the isovalue. For the given example, $Pr(M > c) > Pr(M < c)$, i.e., a topological decision is performed assuming the underlying configuration for the midpoint is positive.

3.3 Uncertain Midpoint Decider

When the predicted topology corresponds to ambiguous cases in the marching cubes algorithm, the midpoint decider averages values at the cell corners to make a topological decision (see Section 17.4.1 in [17]). However, the uncertainty in the data values at the cell corners introduces uncertainty in the average value associated with a cell. We, therefore, devise an uncertain midpoint decider to resolve topological ambiguities probabilistically for uncertain data. The fact that density of sum of random variables is given by convolving their density functions, allows us to analytically characterize the midpoint distribution when the data uncertainty at the cell vertices is modeled with nonparametric densities. As uniform and triangle kernels represent the first and second order B-splines (piecewise polynomial basis functions), this enables us to derive a closed-form density function for the uncertain midpoint in terms of higher order box splines (generalization of B-splines) [7]. This method can be applied to other kernels if we determine the convolution of kernels analytically.

We notice that $\mathrm{pdf}_{\frac{1}{a}X}(x) = \mathrm{pdf}_X(ax)$ for any $a > 0$; therefore, given L independent random variables X_ℓ as $\ell = 1..L$, each distributed as per pdf_{X_ℓ}, distribution of the midpoint $M = \frac{1}{L} \sum_{\ell=1}^{L} X_\ell$ can be formulated as:

$$\mathrm{pdf}_M = \mathrm{pdf}_{\frac{x_1}{L}} * \cdots * \mathrm{pdf}_{\frac{x_L}{L}}, \quad (5)$$

where $*$ denotes convolution.

Again, consider two random variables X and Y, corresponding to grid vertices $\mathbf{v}_1$ and $\mathbf{v}_2$, where pdf_X and pdf_Y consist of single kernels each, with arbitrary bandwidths h and h', respectively. If X and Y are uniformly distributed and $h = h'$, the convolution of the two box functions is a hat (triangle) function. When $h \neq h'$ the convolution is a general box spline. Similarly, when the kernels are triangle functions, the convolution is, in general, a cubic box spline. We compute convolution of L kernels, in closed form, using the finite differencing framework introduced in [8]. This framework enables us to obtain the high-order box spline kernel as the result of convolving kernels with arbitrary bandwidths. The density function of the midpoint turns out to be superposition of the box spline kernels shifted at sample points which are mean of the samples of random variables X_ℓ. The most probable midpoint configuration is determined by slicing the midpoint density function at the isovalue and integrating the density falling in each of the slices. A topological decision is made based on the most probable midpoint configuration. Fig. 6 presents an example of uncertain midpoint decider. As the figure suggests, unlike Gaussian distribution, the derived function can model the multi-modality of a real density functions.

4 From Non-Local Means to Non-Local Density Estimation

While in ensemble simulation datasets samples of the distribution at each data point are available from the ensemble, the uncertainty at each data point can sometimes be estimated from the data itself. For instance, in many denoising algorithms (e.g., Wiener filter), local neighbors of a data point are considered as samples of the distribution at
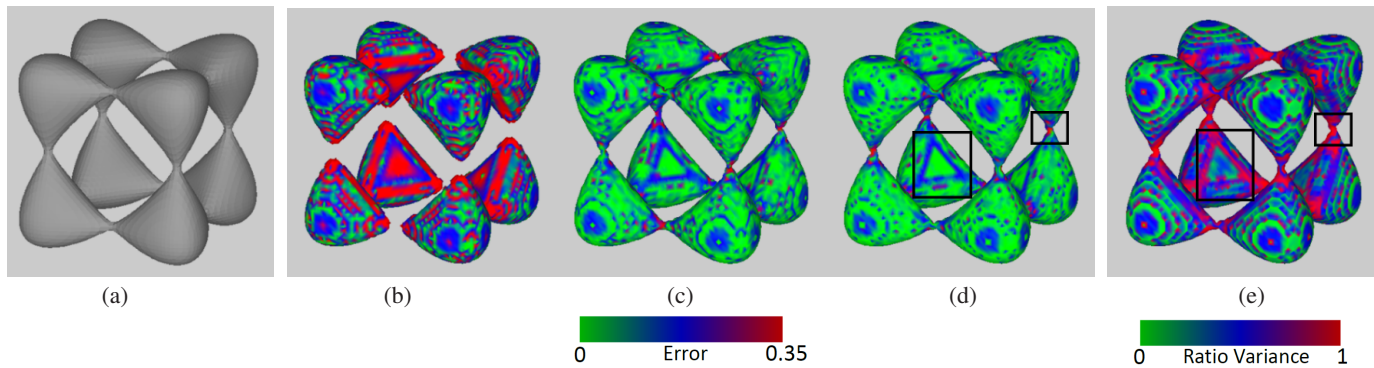
Fig. 7: A comparison of vertex-classification techniques for the isosurface visualization ($c = -0.59$) of tangle function. (a) The groundtruth visualization. Color mapped error visualization relative to the groundtruth (map shown by error colorbar) for the isosurface in noisy data corresponding to (b) statistical mean, (c) vertex-based classification, (d) edge-based classification. (e) Color mapped spatial uncertainties in the isosurface (map shown by ratio variance colorbar) for the result with edge-based classification. The advantage of the probabilistic vertex classification techniques over statistical mean technique is clearly evident from the color mapped error visualization. From the similar patterns in subfigures (d) and (e) indicated by the boxes, it can be observed that the regions of relatively high ratio variance are likely to have relatively high error in the expected isosurface computation relative to the groundtruth.

that data point. Moreover, in probabilistic marching cubes [32, 35] the uncertainty is modeled as a Gaussian distribution whose covariance is estimated from a local neighborhood (via PCA).

Non-local filtering methods have been proven to be superior to local methods for denoising applications [36, 2]. For instance, the non-local means (NLM) method [5] identifies a non-local neighborhood for each data point based on the similarity of patches around the data point and the members of the non-local neighborhood. The denoising is then usually carried out by averaging the data point with its non-local neighborhood. Since our framework allows the marching cubes algorithm for propagation of uncertainty through the isosurface extraction, we leverage the non-local neighborhood for estimating the distribution at each data point. The non-local density estimation approach for isosurface visualization shows subtle reconstruction improvements over the NLM technique, as shown in Section 5 (e.g., Fig. 11). The distance between two vertices $\mathbf{v}_i$ and $\mathbf{v}_j$ is defined based on the similarity between the patches around them:

$$\text{dist}(\mathbf{v}_i, \mathbf{v}_j) = \|\mathcal{N}_k(\mathbf{v}_i) - \mathcal{N}_k(\mathbf{v}_j)\|. \tag{6}$$

Here the notion of distance between the neighborhoods, $\|\mathcal{N}_k(\mathbf{v}_i) - \mathcal{N}_k(\mathbf{v}_j)\|$, is a (weighted) sum of square differences between the corresponding patches, of size k, centered on $\mathbf{v}_i$ and $\mathbf{v}_j$ respectively.

We leverage the non-local matching idea, that is now widely used in denoising algorithms, for kernel density estimation. The density estimation for random variable X_i modeling uncertainty in voxel $\mathbf{v}_i$ is obtained by:

$$\text{pdf}_{X_i} = \frac{1}{\tau} \sum_{j=1}^{N} f(\mathbf{v}_j) K\big(\text{dist}(\mathbf{v}_i, \mathbf{v}_j)\big), \tag{7}$$

where τ is a normalizing constant. When K is the uniform kernel this provides a non-local generalization of the histograming approach that forms a distribution from the values of all voxels in the data whose surrounding patch is similar to that of the voxel $\mathbf{v}_i$. To avoid searching the entire dataset for finding non-local neighbors, often NLM implementations restrict the search to a window around each voxel for this neighborhood analysis [5]. More smooth density estimation can be obtained by choosing K from smooth kernels such as the triangle, Parzen, or Epanechnikov. Moreover, based on the non-local neighbors and the choice of kernel, we can perform bandwidth estimation as discussed in the Appendix.

5 Experiments

We first provide a brief overview of the experiment setup. In our first set of experiments, we use ensemble datasets for isosurface visualization, whereas our second set of experiments consists of visualizing

spatial uncertainties in the isosurface extracted from a single uncertain scalar field. In ensemble datasets, error samples are acquired through multiple simulations or by injecting the noise in the dataset for testing the performance of different techniques. For a single uncertain scalar field, samples within the user specified window size are considered as the error samples at each grid point.

The underlying error densities at each grid point are estimated from the error samples using parametric or nonparametric models. For a local statistics, each base kernel associated with error sample is assigned equal weight. For non-local statistics, each base kernel associated with error sample is assigned a weight based on neighborhood similarity as explained in Section 4.

We quantify and visualize uncertainty in isosurfaces using nonparametric density estimates at grid vertices. Isosurface topology is determined using techniques described in Subsection 3.1 and Subsection 3.3. The geometric uncertainty is then characterized in closed form by computing ratio distribution on each grid edge using the techniques described in Subsection 3.2 and Section 4. Analytic characterization of the ratio distribution can be used for closed-form computation of the expected value and variance on each grid edge. While former quantity can be used to reconstruct a stable isosurface, the latter can be used to visualize spatial uncertainty in an expected isosurface on the grid edges.

Our first experiment on a synthetic dataset consists of visualizing the spatial uncertainty in the level-set extracted from the tangle function [15] for the isovalue of $c = -0.59$. In this experiment, we show the advantage of vertex-based classification and edge-based classification techniques for determining isosurface topology over vertex classification based on the statistical mean of data. The ensemble representing uncertainty in data is obtained by injecting the noise from shifted triangle kernels in the tangle dataset. Five noise samples are generated per kernel from 10 shifted triangle kernels and are injected into the dataset to create an ensemble of 50 realizations. Eight kernels are placed closer to the underlying value, whereas two kernels are placed far away.

Fig. 7 visualizes the tangle function for nonparametric noise model. At each grid vertex, nonparametric density is estimated with the triangle function as base kernel and plugin rule for bandwidth estimation, as explained in the Appendix. Fig. 7a visualizes the result of linear interpolation in the sample data, which we consider to be the ground truth. Fig. 7b shows the expected isosurface extracted from the mean of noisy data. Fig. 7c and Fig. 7d visualize the expected surface extracted for the same noisy data when isosurface topology corresponds to vertex-based classification and edge-based classification techniques, respectively. Fig. 7e visualizes the color mapped ratio distribution variance for the isosurface with edge-based classification.

The difference for various vertex-classification techniques is visually noticeable, particularly in the areas where blobs of the isosurface connect each other. For statistical mean technique, the noise samples far away from the underlying value considerably shift the mean of the data from the underlying value. This causes isosurface to break. However, the probabilistic techniques for vertex classification take into account noise distribution and preserve the connection.

We support our visual evidence by performing quantitative analysis. The Metro program [6] provides a functionality to compute the geometric difference between two meshes, and the program has been integrated into Meshlab software in the Hausdorff distance sampling filter. Isosurface for each vertex-classification technique is sampled at the mesh vertices using Metro program. The samples with low, moderate, and high error with respect to the groundtruth are mapped to green, blue, and red, respectively. We also compute the root mean squared error (RMS) for each of the vertex-classification techniques. The RMS for statistical mean, vertex-based classification, and edge-based classification results are 0.342, 0.111, and 0.110, respectively. RMS and error visualization clearly indicate the advantage of the latter two techniques over the former one.

Ratio distribution variance visualization can be useful, especially in real dataset visualization, in identifying parts of the isosurface that are very sensitive to noise, and are therefore, likely to have expectation relatively far away from the groundtruth. High ratio variance implies high sensitivity of isosurface to noise and vice versa. In Fig. 7(d) and Fig. 7(e), it can be observed that the areas of low sensitivity to noise (green) have low error relative to the groundtruth; however, the areas of high sensitivity to noise (red/blue) can have high or low error relative to the groundtruth. From similar patterns in subfigures Fig. 7(d) and Fig. 7(e) indicated in the boxes, it can be observed that regions of relatively high ratio variance are likely to have relatively high error in expected isosurface computation relative to the groundtruth.

Our second experiment on a synthetic dataset consists of visualization of the spatial uncertainty in the level-set extracted from the teardrop function [15] for the isovalue of $c = -0.002$. In this experiment, we show the advantage of nonparametric models over parametric models for characterizing error densities. The ensemble representing uncertain data is obtained by a same procedure as for the tangle dataset.

Fig. 8(a) visualizes the groundtruth. Fig. 8(b) and Fig. 8(c) visualize expected isosurface in the noisy dataset assuming parametric and nonparametric densities, respectively. Parametric noise model assumes uniform error density. The mean and width of the uniform density at each grid location is estimated from the ensemble members. In nonparametric density estimation, uniform kernel is associated with each noise sample, and the plugin rule is used for bandwidth estimation. Fig. 8(d) visualizes color mapped ratio density variance for the result corresponding to nonparametric models. Topology for the results in the noisy dataset is decided with edge-based classification technique.

Visually, the isosurface in nonparametric statistics looks closer to the groundtruth than the one in parametric statistics. In the case of uniform parametric noise model, outlier samples at each grid location result in poor estimation of the parameters of uniform density, i.e., the mean and the width. This leads to more geometry in the extracted isosurface. In the case of nonparametric density estimation, outliers do not significantly affect the shape of the estimated density.

We, again, support our observation by performing a quantitative analysis similar to the experiment for the tangle dataset. The RMS for parametric and nonparametric density models are 0.626 and 0.096, respectively. RMS and color mapped error relative to the groundtruth clearly show the advantage of nonparametric models over parametric models for characterizing the underlying densities. Ratio distribution variance visualization in Fig. 8d, as in the case of tangle function, can be useful in identifying parts of the isosurface that are very sensitive to noise, and therefore, likely to have expectation relatively far away from the groundtruth.

Fig. 9 visualizes isocontour ($c = 0.008$) extracted from a mixture of 2-d Gaussian probability density functions. The density function is sampled on a high resolution (600×600) grid. The hixel data [41]
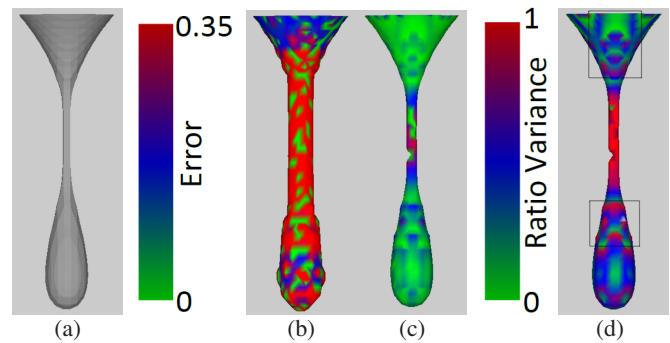

Fig. 8: Parametric versus nonparametric noise density models for the isosurface visualization ($c = -0.002$) of teardrop function. (a) The groundtruth visualization. Color mapped error visualization relative to the groundtruth (map shown by error colorbar) for the isosurface extracted in noisy data assuming (b) a parametric statistical framework, (c) a nonparametric statistical framework. (d) Color mapped spatial uncertainties (map shown by ratio variance colorbar) in the isosurface for a nonparametric framework. Color mapped errors relative to the groundtruth clearly show the advantage of nonparametric models over parametric models for characterizing the error densities. From similar patterns in subfigures (c) and (d) indicated by boxes, it can be observed that regions of relatively high ratio variance are likely to have relatively high error in the isosurface with respect to the groundtruth.
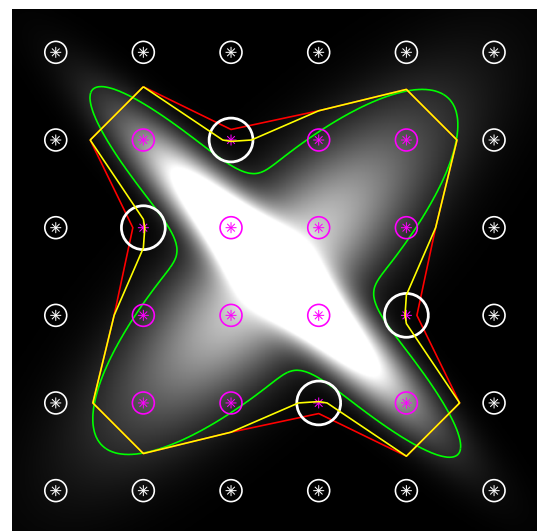

Fig. 9: Vertex-based classification versus edge-based classification. Image shows the result of isocontour extraction ($c = 0.008$) from a mixture of 2 Gaussian density functions. Isocontour in a high resolution data is shown in green. Isocontours corresponding to vertex-based classification and edge-based classification are extracted from the hixel dataset [41] corresponding to high resolution data. While the former is shown in red, the latter is shown in yellow. Stars correspond to vertex-based classification, whereas circles correspond to edge-based classification. White represents positive vertices ($+1$), whereas magenta represents negative vertices (-1) for both vertex-classification techniques. Places where colors of stars and circles do not match (indicated in large circles) represent grid locations where vertex-based classification and edge-based classification do not agree.

is generated from a high resolution grid with block size of 100×100 and histogram bin count of 10. Histogram derived at each location

(a) Classification based on mean field

(b) Vertex-based classification method

(c) Edge-based classification method

(d) Mean   (e) Vertex-based   (f) Edge-based
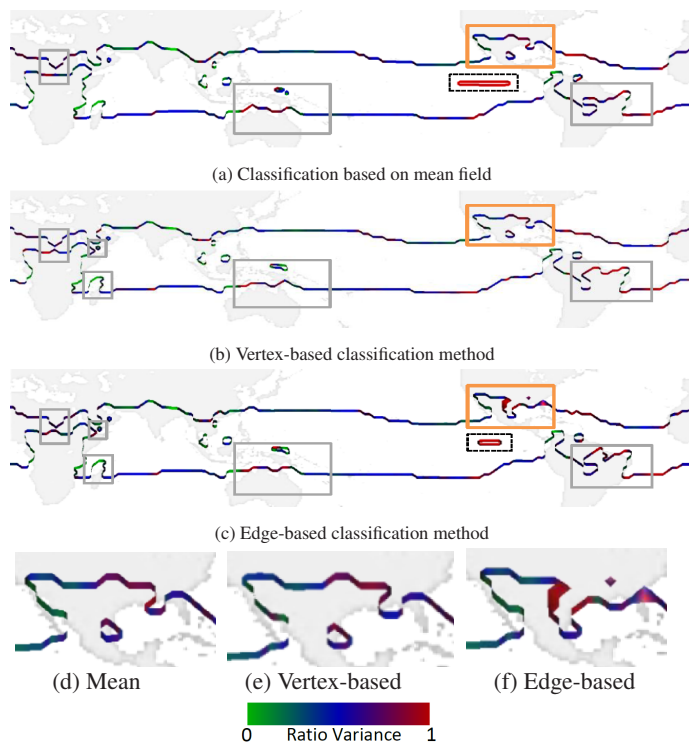
0   Ratio Variance   1

Fig. 10: A Comparison of vertex-classification techniques for the isosurface visualization ($c = 24\,^\circ\mathrm{C}$) of an ensemble representing uncertain temperature field [27]. Boxes mark differences in isosurface topology. Subfigures (d), (e), and (f) provide zoomed-in views of the regions enclosed by the orange boxes in subfigure (a), (b), and (c), respectively. Edge-based classification in (f) extracts the isocontour along a coastline of the Gulf of Mexico unlike the other vertex classification techniques. Positional uncertainties in expected isosurface are color mapped (quantiles formed after histogram equalization).

in hixel data (6×6) is used to characterize data uncertainty. Fig. 9 shows results for vertex-based classification and edge-based classification. Classification for the two techniques disagree when isocontour in a low resolution grid is close to grid vertices, i.e., when classification probabilities for vertices are close to 0.5 in the case of vertex-based classification. Also, edge-based classification approach pulls isocontour (yellow) closer to isocontour (green) in a high resolution dataset when compared to isocontour (red) based on vertex-based classification.

In an experiment on a real ensemble dataset, we quantify and visualize spatial uncertainty in an expected isosurface extracted from uncertain temperature field for the isovalue of $c = 24\,^\circ\mathrm{C}$, as shown in Fig. 10. We analyse the impact of various vertex classification methods on isosurface topology. The climate simulations representing uncertainty in 2-meter termperature field are obtained from the DEMETER project [27]. The ensemble of simulations consists of 63 realizations acquired through 7 different models for day 90 of the $2000-02$ hindcast. The data is provided on a regular grid of size 144×72 pixels with 2.5 degrees resolution along the longitude and latitude. This experiment setting is similar to the one in section 7 of [33].

Fig. 10a, Fig. 10b, and Fig. 10c show the effect of vertex-classification techniques on isosurface topology. Nonparametric density estimation for mentioned subfigures is performed with Epanechnikov base kernel and plugin rule for bandwidth estimation. The vertex-based classification method in Fig. 10b shows significant differences in topology when compared to vertex classification based on statistical mean in Fig. 10a. Gray boxes mark places where isosurface topology is affected, e.g., merged rings of isosurfaces in Fig. 10b that are separate in Fig. 10a, and newly formed isosurface rings in Fig. 10b

that are absent in Fig. 10a. Isosurface topology in Fig. 10b is different from Fig. 10a mainly in the areas of relatively high spatial uncertainty (red colored regions) in Fig. 10a. We have already shown the advantage of vertex-based classification method over a classification based on statistical mean of data for tangle function. Edge-based classification in Fig. 10c also captures topological changes comparable to vertex-based classification in Fig. 10b, e.g., merged and newly formed isosurfaces.

The dotted and the orange boxes indicate places where edge-based classification shows topological differences for corresponding vertex-based classification. As we can observe from the zoomed-in views of the regions marked by the orange boxes in Fig. 10, the edge-based classification extracts the isocontour along a coastline of the Gulf of Mexico unlike the other classification methods. It can be observed that the cell-crossing probabilities are relatively high along the mentioned coastline region for the same dataset visualization in [33]. We also observe that the isothermal region follows the coastline at many places in Fig. 10 e.g., the coastline of the Gulf of California.

A large number of cells in the area of the dotted box in Fig. 10a have very low cell-crossing probabilities for nonparametric models, as visualized in [33]. The red color of isosurface enclosed by the dotted box in Fig. 10a indicates high spatial uncertainty of isosurface. Vertex-based classification method in Fig. 10b classifies all vertices with same sign, thus, eliminating the entire ring of isosurface present in Fig. 10a. However, the edge-based classification method in Fig. 10c takes into account relative edge crossing probabilities and therefore classifies vertices for edges with relatively high edge-crossing probabilities with opposite signs. Thus, the edge-based classification method avoids complete elimination of isosurfaces with high spatial uncertainty from visualization.

Now, we perform the uncertainty analysis on isosurface extracted from a single uncertain scalar field. We compare non-local statistics with local statistics for isosurface visualization of the bonsai tree dataset ($c = 84$). Fig. 11a visualizes the result of the linear interpolation in the locally computed mean field. All the bonsai tree results corresponding to probabilistic techniques visualize expected isosurface with topology corresponding to the vertex-based classification method. Fig. 11b visualizes an isosurface assuming the underlying noise model to be a parametric uniform density [1]. The mean and the width of the uniform density at each grid point is estimated from the local neighbors. The isosurface visualization for Fig. 11c is obtained by performing nonparametric density estimation on local neighbors at each grid location. The nonparametric density estimation is performed with the uniform base kernel and a plugin rule for bandwidth estimation. Each base kernel for density estimation on local neighbors is equally weighed.

For the experiments with non-local statistics, the voxels in the window of size $5 \times 5 \times 5$ around each voxel are considered as error samples, and the neighborhood similarity window size is set to $3 \times 3 \times 3$. Fig. 11d shows the result of linear interpolation in data denoised using non-local means (NLM) algorithm [5]. Fig. 11e visualizes the expected isosurface by performing nonparametric density estimation on non-local neighbors, as explained in Section 4. The weights for uniform base kernel associated with each of the error samples are derived using NLM algorithm [5]. The bandwidth estimation is performed using rule of thumb approach with normal density as a reference. We justify using the rule of thumb approach for weighted kernel density estimation in Appendix. Fig. 11f visualizes the relative spatial uncertainties for the isosurface in Fig. 11e by colormapping the variance of ratio density function after histogram equalization. In parts of the isosurface with relatively high ratio density variance, it is very likely that error in expected computation is relatively high with respect to the groundtruth.

The isosurface reconstruction in non-local statistics recovers most of the bonsai tree branches without much loss of detail. However, isosurface visualization in local statistics results in a significant loss of detail. Advantage of non-local techniques over local techniques for level-set extraction is evident in the areas marked by the black boxes. Among the probabilistic techniques for the isosurface extraction, non-
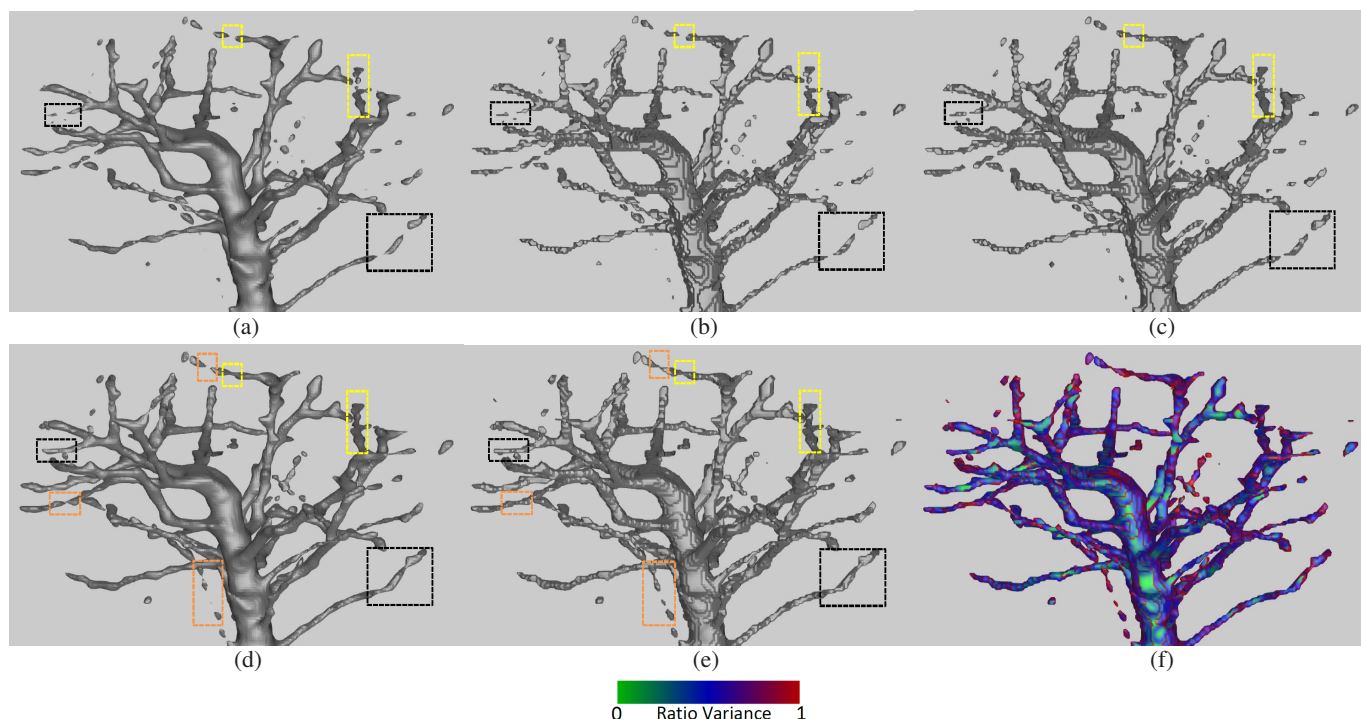
Fig. 11: Non-local techniques (the bottom row) versus local techniques (the top row) for the isosurface visualization ($c = 84$) of the bonsai tree dataset. Isosurface visualization corresponding to (a) locally computed mean of data, (b) locally estimated uniform density, (c) locally estimated nonparametric density, (d) data denoised using a non-local means (NLM) technique [5], (e) nonparametric density estimation on non-local neighbors. (f) Color mapped spatial uncertainties in the isosurface in subfigure (e) (quantiles after histogram equalization). Boxes with the same color indicate the reconstruction performance for each of the techniques. Black boxes mark a few areas where local techniques break the bonsai tree branches, whereas the branches are recovered in non-local techniques. Similarly, the nonparametric models recover more branches than the parametric models as shown by regions marked by yellow boxes. Moreover, the orange boxes indicate the places where nonparametric density estimation on non-local neighbors shows subtle reconstruction improvements over the NLM technique.

parametric statistics recovers few branches that are broken in parametric statistics, as indicated in yellow boxes. The isosurface corresponding to nonparametric density estimation on non-local neighbors recovers branches at a few places where they are not recovered for the NLM denoised dataset. Few such regions are enclosed in orange boxes.

Fig. 12 visualizes results similar to the bonsai tree experiment. The fuel dataset is visualized for the isovalue of $c = 97.74$, assuming data to be uncertain. Nonparamatric density in local and non-local techniques is estimated with an Epanechnikov base kernal. While the former uses the plugin rule, the latter uses the rule of thumb for bandwidth estimation. For non-local techniques, search window size and neighborhood similarity window size are set to $5 \times 5 \times 5$ and $3 \times 3 \times 3$, respectively. The non-local techniques recover more data features than the local techniques, especially in the top portion of the fuel dataset. Among the probabilistic techniques for recovering isosurface, the non-local techniques provide smoother reconstruction when compared to the local techniques.

6 Conclusion & Future Work

We study the problem of isosurface extraction in an uncertain scalar field when the data uncertainty is characterized in the general framework of nonparametric statistics.

We devise a probabilistic techniques for handling uncertainty in cell configurations for isosurface topology determination. We propose vertex-based classification and edge-based classification methods to classify vertex signs. While the vertex-based classification approach works independently on each grid vertex, the edge-based classification approach works in collaborative fashion by taking into account spatial neighbors. In the case of cell configuration corresponding to ambiguous topology, we develop a probabilistic midpoint decider to resolve

the topological ambiguities. We derive, in closed form, probability density at cell midpoint when data uncertainty is characterized with parametric and nonparametric models.

In order to study geometric uncertainty in level-crossing locations of an isosurface, we analyze interaction between linear interpolation and data uncertainty characterized with parametric and nonparametric densities. We also propose a scheme for accurately characterizing underlying noise densities by leveraging non-local statistics. We show the advantage of non-local statistics approach for characterizing data uncertainty over locally estimated parametric and nonparametric densities. We plan to analytically characterize the spatial uncertainties in the isosurfaces by taking into account correlations in the random field.

7 Acknowledgements

References

[1] T. Athawale and A. Entezari. Uncertainty quantification in linear interpolation for isosurface extraction. *IEEE Trans. on Vis. and Comp. Graph.*, 19(12):2723–2732, 2013.

[2] S. P. Awate and R. T. Whitaker. Unsupervised, information-theoretic, adaptive image filtering for image restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28(3):364–376, 2006.

[3] K. Brodlie, R. A. Osorio, and A. Lopes. A review of uncertainty in data visualization. *Expanding the Frontiers of Visual Analytics and Visualization*, pages 81–109, 2012.
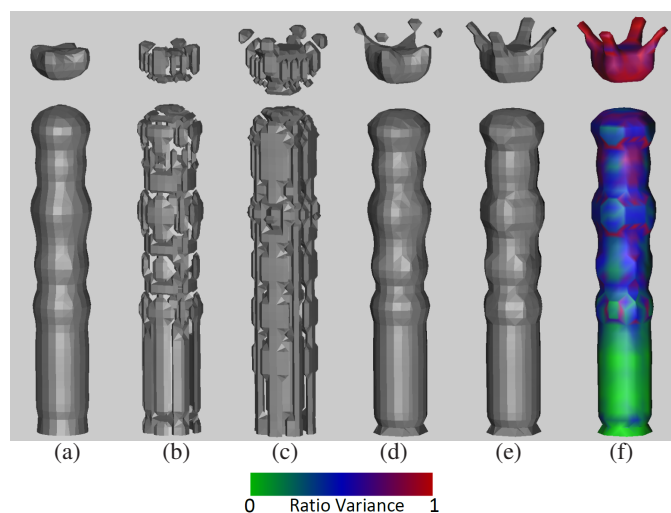
Fig. 12: Non-local techniques (last 3 subfigures) versus local techniques (first 3 subfigures) for the isosurface visualization ($c = 97.74$) of the fuel dataset. Isosurface visualization corresponding to (a) locally computed mean of data, (b) locally estimated uniform density, (c) locally estimated nonparametric density, (d) data denoised using a non-local means (NLM) technique [5], (e) nonparametric density estimation on non-local neighbors. (f) Color mapped spatial uncertainties in the isosurface for subfigure (e). The non-local techniques recover more data features than the local techniques, especially in the top portion of the fuel dataset. Among the probabilistic techniques for recovering isosurface, non-local techniques provide smoother reconstruction when compared to local techniques.

[4] R. Brown. Animated visual vibrations as an uncertainty visualisation technique. In *GRAPHITE '04: Proc. of the 2nd Int. Conf. on Comp. Graphics and Interactive Techniques*, pages 84–89, 2004.

[5] A. Buades, B. Coll, and J. M. Morel. A non-local algorithm for image denoising. *Comp. Vision and Pattern Recognition, CVPR 2005. IEEE Comp. Society Conf. on*, 2:60–65, 2005.

[6] P. Cignoni, C. Rocchini, and R. Scopigno. Metro: Measuring error on simplified surfaces. *Comp. Graphics Forum*, 17:167–174, 1998.

[7] C. De Boor, K. Höllig, and S. D. Riemenschneider. *Box Splines*, volume 98 of *Applied Mathematical Sciences*. Springer-Verlag, 1993.

[8] A. Entezari, M. Nilchian, and M. Unser. A box spline calculus for the discretization of computed tomography reconstruction problems. *Med. Img., IEEE Trans. on*, 31(8):1532–1541, 2012.

[9] C. Fowlkes, S. Belongie, F. Chung, and J. Malik. Spectral grouping using the nystrom method. *Pattern Analysis and Machine Intelligence, IEEE Trans. on*, 26(2):214–225, 2004.

[10] G. Grigoryan and P. Rheingans. Probabilistic surfaces: Point based primitives to show surface uncertainty. In *IEEE Vis., 2002. (VIS 2002).*, pages 147–153, 2002.

[11] G. Grigoryan and P. Rheingans. Point-based probabilistic surfaces to show surface uncertainty. *Vis. and Comp. Graphics, IEEE Trans. on*, 10(5):564–573, 2004.

[12] C. Johnson. Top scientific visualization research problems. *Comp. Graphics and Applications, IEEE*, 24(4):13–17, July-Aug. 2004.

[13] C. Johnson and A. Sanderson. A next step: Visualizing errors and uncertainty. *Comp. Graphics and App., IEEE*, 23(5):6–10, Sept.-Oct. 2003.

[14] M. C. Jones, J. S. Marron, and S. J. Sheather. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association*, 91:401–407, 1996.

[15] 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. In *Comp. Graphics Forum*, volume 28, pages 26–40, 2009.

[16] T. Y. Lee and H. W. Shen. Efficient local statistical analysis via integral histograms with discrete wavelet transform. *IEEE Trans. on Vis. and Comp. Graphics*, 19(12):2693–2702, Dec. 2013.

[17] A. Lopes and K. Brodlie. Interactive approaches to contouring and isosurfacing for geovisualization. *Exploring Geovisualization*, 2005.

[18] W. E. Lorensen and H. E. Cline. Marching cubes: A high resolution 3d surface construction algorithm. 1987.

[19] A. L. Love, A. Pang, and D. L. Kao. Visualizing spatial multivalue data. *IEEE Comp. Graphics and Applications*, 25(3):69–79, 2005.

[20] C. Lundstrom, P. Ljung, A. Persson, and A. Ynnerman. Uncertainty visualization in medical volume rendering using probabilistic animation. *Vis. and Comp. Graphics, IEEE Trans. on*, 13(6):1648–1655, 2007.

[21] A. Luo, D. Kao, and A. Pang. Visualizing spatial distribution data sets. *Proc. Symp. Data Visualization 2003 (VISSYM 03)*, pages 29–38, 2003.

[22] R. McGill, J. W. Tukey, and W. A. Larsen. Variations of box plots. *The American Statistician*, 32(1):12–16, February 1978.

[23] T. Munzner, C. Johnson, R. Moorhead, H. Pfister, P. Rheingans, and T. S. Yoo. NIH-NSF visualization research challenges report summary. *Comp. Graphics and Applications, IEEE*, 26(2):20–24, 2006.

[24] R. A. Osorio and K. Brodlie. Contouring with uncertainty. *Theory and Practice of Comp. Graphics 2008.*, pages 59–66, 2008.

[25] M. Otto, T. Germer, H. C. Hege, and H. Theisel. Uncertain 2d vector field topology. In *Comp. Graph. Forum*, volume 29, pages 347–356, 2010.

[26] M. Otto, T. Germer, and H. Theisel. Uncertain topology of 3d vector fields. In *IEEE Pacific Vis. Symp. (PacificVis)*, pages 67–74, 2011.

[27] T. Palmer, A. Alessandri, U. Andersen, P. Cantelaube, M. Davey, P. Délécluse, E. Díez, F. J. Doblas-Reyes, H. Feddersen, R. Graham, S. Gualdi, J.-F. Guérémy, R. Hagedorn, M. Hoshen, N. Keenlyside, M. Latif, A. Lazar, E. Maisonnave, V. Marletto, A. P. Morse, B. Orfila, P. Rogel, J.-M. Terres, and M. C. Thomson. Development of a european multi-model ensemble system for seasonal to interannual prediction (demeter). Technical report, ECMWF, 2004.

[28] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Comp.*, 13(8):370–390, 1997.

[29] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.

[30] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Comp. Graphics Forum*, 30(3):951–960, 2011.

[31] T. Pfaffelmoser and R. Westermann. Visualization of global correlation structures in uncertain 2d scalar fields. *Comp. Graphics Forum*, 31(3):1025–1034, 2012.

[32] K. Pöthkow and H. C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *Vis. and Comp. Graphics, IEEE Trans. on*, 17(10):1393–1406, 2011.

[33] K. Pöthkow and H. C. Hege. Nonparametric models for uncertainty visualization. *Comp. Graphics Forum*, 32(3.2):131–140, 2013.

[34] K. Pöthkow, C. Petz, and H. C. Hege. Approximate level-crossing probabilities for interactive visualization of uncertain isocontours. *Int. Journal for Uncertainty Quantification*, 3(2), 2012.

[35] K. Pöthkow, B. Weber, and H. C. Hege. Probabilistic marching cubes. In *Comp. Graphics Forum*, volume 30, pages 931–940, 2011.

[36] I. Ram, M. Elad, and I. Cohen. Image denoising using nl-means via smooth patch ordering. In *Acoustics, Speech and Signal Processing (ICASSP), IEEE Int. Conf. on*, pages 1350–1354. IEEE, 2013.

[37] P. J. Rhodes, R. S. Laramee, R. D. Bergeron, T. M. Sparr, et al. Uncertainty visualization methods in isosurface rendering. In *Eurographics*, volume 2003, pages 83–88, 2003.

[38] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.

[39] S. Sahni and T. Gonzalez. P-complete approximation problems. *J. ACM*, 23(3):555–565, 1976.

[40] S. Schlegel, N. Korn, and G. Scheuermann. On the interpolation of data with normally distributed uncertainty for visualization. *Vis. and Comp. Graphics, IEEE Trans. on*, 18(12):2305–2314, 2012.

[41] 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 *Large Data Analysis and Visualization (LDAV), 2011 IEEE Symp. on*, pages 23–30, 2011.

[42] R. T. Whitaker, M. Mirzargar, and R. M. Kirby. Contour boxplots: A method for characterizing uncertainty in feature sets from simulation ensembles. *IEEE T. Vis. and Comp. Graph.*, 19(12):2713–2722, 2013.

[43] C. M. Wittenbrink, A. T. Pang, and S. K. Lodha. Glyphs for visualizing uncertainty in vector fields. *Vis. and Comp. Graphics, IEEE Trans. on*, 2(3):266–279, 1996.

[44] P. C. Wong, H.-W. Shen, C. R. Johnson, C. Chen, and R. B. Ross. The top 10 challenges in extreme-scale visual analytics. *Comp. Graphics and Applications, IEEE*, 32(4):63–67, 2012.