

Direct Volume Rendering with Nonparametric Models of Uncertainty

Tushar Athawale, Bo Ma, Elham Sakhaei, Chris R. Johnson, *Fellow, IEEE*,
and Alireza Entezari, *Senior Member, IEEE*

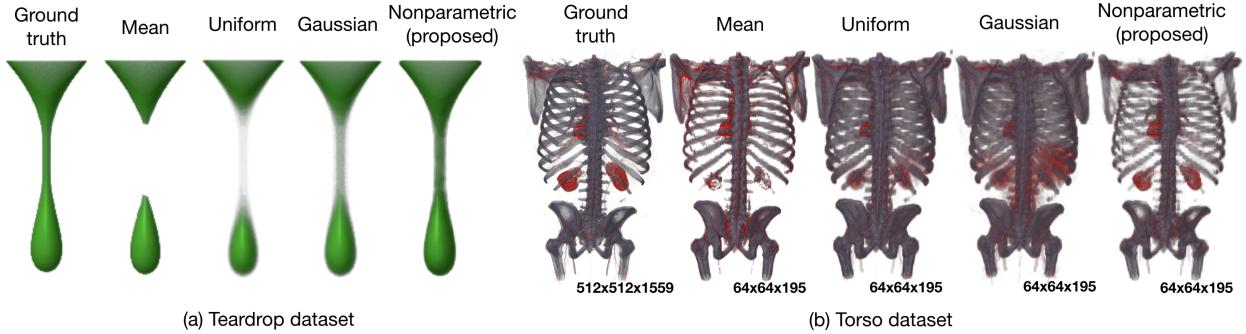


Fig. 1. Nonparametric models of uncertainty improve the quality of reconstruction and classification within an uncertainty-aware direct volume rendering framework. (a) Improvements in topology of an isosurface in the teardrop dataset ($64 \times 64 \times 64$) with uncertainty due to sampling and quantization. (b) Improvements in classification (i.e., bones in gray and kidneys in red) of the torso dataset with uncertainty due to downsampling.

Abstract—We present a nonparametric statistical framework for the quantification, analysis, and propagation of data uncertainty in direct volume rendering (DVR). The state-of-the-art statistical DVR framework allows for preserving the transfer function (TF) of the ground truth function when visualizing uncertain data; however, the existing framework is restricted to parametric models of uncertainty. In this paper, we address the limitations of the existing DVR framework by extending the DVR framework for nonparametric distributions. We exploit the *quantile interpolation* technique to derive probability distributions representing uncertainty in viewing-ray sample intensities in closed form, which allows for accurate and efficient computation. We evaluate our proposed nonparametric statistical models through qualitative and quantitative comparisons with the mean-field and parametric statistical models, such as uniform and Gaussian, as well as Gaussian mixtures. In addition, we present an extension of the state-of-the-art rendering parametric framework to 2D TFs for improved DVR classifications. We show the applicability of our uncertainty quantification framework to ensemble, downsampled, and bivariate versions of scalar field datasets.

Index Terms—Volumes, uncertainty, nonparametric, 2D transfer function

1 INTRODUCTION

As visualization techniques continue to facilitate the exploration of scientific simulations and biomedical datasets, analysis of data uncertainties, inherent in all forms of acquisition, modeling, and representation, has emerged as an important research area. Uncertainties present in data, such as those intrinsic to acquisition or modeling (e.g., sampling, quantization), as well as those introduced within data processing (e.g., filtering/downsampling), adversely impact the reliability of visualizations. To facilitate reliable visualization in the presence of uncertainty, several studies have advocated redesigning visualization algorithms to treat data as probability distributions to account for various types of uncertainty [25, 57]. Quantifying the impact of uncertainty in the computational process and its propagation throughout the visualization pipeline poses several mathematical as well as algorithmic challenges.

Visualization of uncertain data is an active field of research, including several advances in innovative ways for the visual depiction of uncertainty [4]. In contrast, analysis and propagation of uncertainty in the various stages of the rendering pipeline and quantifying their impact on transforming the uncertainty remain challenging tasks. Recent studies [6, 14, 30] have considered sources of uncertainty within the visual analytics process and analyzed the contribution of each stage to the uncertainty associated with the volume data.

In our work, we study the propagation of data uncertainty through the stages of the direct volume rendering (DVR) pipeline. DVR is a fundamental visualization technique for gaining insights into volumetric datasets. A transfer function (TF) plays a central role in DVR, as it translates scalar or multifield data to optical properties, such as color and opacity. The visual mappings produced by TFs help users understand interesting features or patterns in the dataset. Such a process of feature identification through TF space exploration is referred to as *classification* [17].

The classification task in DVR can be challenging when the data have uncertainty. The DVR of uncertain data by reusing the TF design for the original function can lead to poor classification results [58]. A simple workaround would be to generate a new TF when data have uncertainty; but the design of TFs is known to be a time-consuming and laborious task, especially in the case of multidimensional TFs. Thus, developing new techniques that seek to improve the quality of visualizations of uncertain data while reusing the TF design for the original function is more desirable. An expressive rendering of uncertain data must

• T. Athawale and C. R. Johnson are with Scientific Computing & Imaging (SCI) Institute, University of Utah. E-mail: {tushar.athawale, crj}@sci.utah.edu,
• B. Ma, E. Sakhaei, and A. Entezari are with the Department of CISE at the University of Florida, Gainesville, FL, 32611. E-mail: {bbo, esakhaei, and entezari}@cise.ufl.edu

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. Digital Object Identifier: xx.xxxx/TVCG.201x.xxxxxxx

preserve all visible features in the rendering of the ground truth data and also indicate the uncertainties engendered by the various stages of the rendering pipeline.

A recently developed statistical framework by Sakhaei *et al.* [51] introduced a novel approach for DVR that addressed the issue of preserving TF designs of the original function for visualizations of uncertain data. In their approach, data uncertainty is integrated against 1D TF in the reconstruction stage of the DVR pipeline. Their framework opens up new directions for the exploration of uncertain data, because it allows for uncertainty propagation and aggregation within the reconstruction and the traditional classification stages. In the framework proposed by Sakhaei *et al.* [51], the data input into the DVR process are considered as a field of random variables described by parametric probability density functions (PDFs). Liu *et al.* [31] proposed a DVR framework for visualization of uncertain 3D data when PDFs are modeled using Gaussian mixture models (GMM). Their framework used an expensive Monte Carlo (MC) sampling approach for uncertainty estimation of interpolated samples of a DVR raycaster.

Inspired by contributions on DVR for parametric- [51] and GMM- [31] based uncertainty, we propose a closed-form DVR framework for nonparametric density models. Recently, noise modeling using nonparametric distributions has been advocated over parametric distributions for taking into account the skewness or multimodality of distributions, and hence improving the precision of uncertainty visualizations [3, 47]. Although the extension to nonparametric distributions for DVR has been discussed in previous work [51], such an extension is challenging, especially from the computational cost point of view, and no recipe for implementation or empirical results has been provided.

In our work, we present an efficient quantile interpolation technique for DVR of uncertain data, where uncertainty is characterized using nonparametric distributions. Read [49] first introduced the 1D quantile interpolation technique for the interpolation of histograms. Hollister and Pang [22] leveraged the quantile interpolation technique for bilinear interpolation of nonparametric distributions characterizing uncertain vector fields. We present quantile interpolation for trilinear interpolation of nonparametric distributions, and we successfully integrate interpolated distributions with a DVR framework for the visualization of uncertain data.

Although the-state-of-the-art spline-based technique [51] explores 1D classification of uncertain scalar fields for DVR, the classification of uncertain scalar volumes with multidimensional transfer functions (TFs) and the visualization of multifield data remain challenging tasks. Specifically, the intensity-gradient magnitude (2D) TF has proved valuable due to its effectiveness in isolating complex boundaries with overlapping materials [27]. Unlike the previous work that dealt with uncertainties in the data and gradient field for data separately, we leverage the simultaneous estimation of uncertainty in both the scalar and gradient fields. Specifically, we apply a spline-based statistical framework to intensity-gradient magnitude 2D TFs and study its ramifications in visualizing bivariate datasets.

1.1 Contributions

We generalize the recently developed spline-based statistical framework [51] to nonparametric statistical models and 2D TFs for visualization of uncertain data. Specifically, we propose the following methods for the visualization of uncertain data:

- Given an uncertain scalar field, represented as a field of PDFs, we analytically derive the interpolation of the PDFs of the intensities for any arbitrary sample point along the viewing rays for DVR. Each grid point is modeled using a nonparametric PDF in contrast to a parametric one. A previous study [51] considered nonparametric models only as a possible venue for investigation within the framework for uncertainty visualization, but did not explore nonparametric models or their potential advantages. In our work, we fill this gap by proposing the use of the quantile interpolation technique for nonparametric statistics. Specifically, we present an analytic formulation of the quantile interpolation technique for trilinear interpolation of nonparametric PDFs. Our closed-form

formulation permits efficient integration of nonparametric statistics with a DVR framework. We demonstrate the effectiveness of our proposed nonparametric models through qualitative and quantitative comparisons with mean-field and parametric models.

- The quantile interpolation technique presents an example of order statistics, where quantiles are ordered using a cumulative density function for a random variable. We, thus, take advantage of order statistics to investigate uncertainty in ensemble data by devising a tool called the quartile view.
- Similar to intensities, we analytically derive a formulation for the interpolation of PDFs of the gradient magnitudes for samples along the viewing rays. The PDFs of intensities and gradient magnitudes are then integrated against 2D TF (gradient magnitude vs. intensity). Improved classification of uncertain scalar fields using this approach signifies the importance of the simultaneous handling of uncertainty in data and its gradient field.
- We demonstrate an application of our proposed DVR framework for the visualization of ensemble and downsampled data. We also present an application of the reconstruction of PDFs for DVR of bivariate data.

The paper is organized as follows: In Sec. 2, we review the prior work on uncertainty visualization and multidimensional TFs. In Sec. 3.1, we briefly revisit the state-of-the-art theory [51] on the interpolation of uncertain scalar fields and linear interpolation of histograms in 1D using the quantile interpolation technique [49]. We then present an extension of the quantile interpolation technique for trilinear interpolation of nonparametric PDFs and its integration into a DVR framework in Sec. 3.2 and Sec. 3.3, respectively. In Sec. 3.4, we describe our quartile view technique. In Sec. 4.1 and Sec. 4.2, we describe a spline-based model for the interpolation of uncertain gradient fields and propose integration of interpolated intensity and gradient magnitude distributions against 2D TFs for visualizations. In Sec. 5, we demonstrate experimental results for the visualization of uncertain data using reconstructed uncertain scalar fields and uncertain gradient fields. Finally, we conclude our work and discuss possible future work in Sec. 6.

2 RELATED WORK

2.1 Uncertainty Visualization

Uncertainty visualization has been recognized as one of the top challenges in the visualization community due to its significance in decision-making [24, 25, 57, 59]. Specifically, uncertainty visualization has been shown to be important in avoiding misleading interpretations regarding the underlying data. Whereas classical visualization approaches consider uncertainty associated with the volume data [41, 48], recent works account for aggregated uncertainty due to rendering algorithms [6, 14, 51]. Brodlie *et al.* [4] discuss the impact of propagating uncertainty in the data to uncertainty in the final image. They define the propagation problem as determining the PDF of the output entities from the PDF of the input entities, or discuss that often a MC sampling method is required to obtain the PDF of the output. In this paper, we derive PDFs analytically. Correa *et al.* [6] describe uncertainty propagation and aggregation for data transformations, such as regression, principal component analysis, and k-means clustering.

Statistical uncertainty analysis of topological features of data, such as level sets and critical points, has drawn increasing attention in the study of data uncertainty. Contour [56], curve [37], and surface [15] box-plots extend the concept of functional-depth ranking [32] for deriving quantiles that represent the spatial variability of ensembles of isocontours, arbitrary curves, and 2D images, respectively. The level-crossing probability method of Pöthkow and Hege [46, 47] and uncertainty-aware marching cubes algorithm proposed by Athawale *et al.* [1–3] demonstrated the benefits of nonparametric statistical noise modeling over parametric modeling for deriving positional uncertainty in level sets. Hixels [55] summarized information of a brick of volume as a histogram for visualizing fuzzy isosurfaces. Suter *et al.* [53] exploited

shape similarities based on Hausdorff distance for extracting isosurfaces from 3D scalar fields. Günther *et al.* [16] and Favelier *et al.* [12] devised statistical approaches for characterizing spatial variations in critical points of uncertain data. Otto *et al.* performed statistical gradient field analysis for visualizing topological variations of uncertain 2D [39] and 3D [40] vector fields. Recently, He *et al.* [18] devised a nonparametric method, known as surface density estimation (SDE), for analyzing spatial inconsistency in level sets.

In the context of DVR, a considerable body of literature has analyzed uncertainty propagation in the DVR pipeline and the impact of errors on final visualizations. Fout and Ma [14] discussed the contribution of each stage of DVR (quantization, reconstruction/filtering, classification, shading, and integration) to the uncertainty associated with volume data. Kniss *et al.* [28] presented rendering based on probabilistic classification that allows the user to interactively explore the uncertainty and the information computed during fuzzy segmentation. Pfaffelmoser *et al.* [44] assumed Gaussian-distributed data uncertainty for visualizing geometric uncertainty in isosurfaces extracted using DVR. Etian *et al.* [10, 11] proposed a novel strategy for verifying the correctness of DVR implementations by analyzing the correlation between discretization errors caused by sampling along viewing rays and the rendering quality. Kronander *et al.* [30] evaluated the effects of the propagation of numerical errors, caused by finite precision of data representation and processing, on the volume rendered images. Djurcicov *et al.* [8] employed features such as speckle, texture, or noise to represent uncertainty in the volume rendering process. In medical volume rendering, probabilistic animation has been used to visualize uncertainty [33].

The focus of our work is on advocating nonparametric noise modeling over parametric noise modeling for preserving the TF design of the original function when performing DVR of uncertain scalar fields. For DVR of uncertain data, we adopt a probabilistic view of TFs since it allows incorporating uncertainty in both classification and visual parameter mapping. A probabilistic view is proposed by Drebin *et al.* [9], where the application of TFs involves two steps: (1) map each sample to a set of material probabilities, and assign each material an RGBA color, and (2) compute the color for each sample as a weighted average of material colors based on material probabilities. Please refer to Sec. 3.3 for more details regarding the probabilistic view.

2.2 Multidimensional Transfer Functions

Traditionally, a TF classifies voxels to optical properties, such as colors and opacity, according to a 1D function of the scalar values. The function can be designed either manually, which is a tedious task, or automatically based on attributes of the underlying volume data [45]. Since 1D TF classification has limited power in exploring and classifying the embedded features in the data, subsequent studies have considered TFs with multiple dimensions. Multidimensional TFs, as proposed by Kniss *et al.* [26], have been proven superior to traditional 1D transfer functions due to their ability to isolate complex materials with overlapping intensities. In particular, the gradient magnitude and second-order derivatives are commonly used as additional properties to expand the TF domain [19, 34, 54]. In this work, we demonstrate the benefits of incorporating gradient magnitude uncertainty, computed analytically within the reconstruction stage into a 2D TF, where a 2D TF is characterized by intensity and gradient magnitude. Broadly speaking, we advocate the extension of a methodology involving the integration of 1D TFs against data uncertainty [51] to 2D TFs for the improved efficiency and reliability of DVR classifications.

3 INTERPOLATION OF UNCERTAIN SCALAR FIELD FOR DVR

3.1 Mathematical Model and the State of the Art

3.1.1 Interpolation of Parametric Distributions

We state the mathematical model for our methods and briefly revisit the state of the art in interpolation of uncertain data when the uncertainty is modeled as probability distributions. Given 3D discrete uncertain scalar data $f(\mathbf{v}_i)$, the uncertainty can be modeled at each voxel \mathbf{v}_i by a random variable X_i . We assume random variables modeling uncertainty at voxels to be independent. The reconstruction of the random field

at an arbitrary position, \mathbf{v} , results in a random variable X , which is a linear combination of random variables at positions \mathbf{v}_i 's: $X = \sum_i w_i X_i$, with weights $w_i = \varphi(\mathbf{v} - \mathbf{v}_i)$, where $\varphi : \mathbb{R}^3 \rightarrow \mathbb{R}$ is the basis function that determines the weights for the neighboring voxels contributing to the interpolated sample \mathbf{v} . When the data are sampled on a Cartesian grid, φ is commonly chosen as a tensor product-based trilinear B-spline. The goal is to analytically obtain the probability distribution of X at any arbitrary sampling point \mathbf{v} . The probability distribution of a linear combination of independent random variables is the convolution of their individual distributions [20]. Therefore, the PDF of X can be derived from the convolution of the PDFs of X_i 's, i.e., $\text{pdf}_X(x) = \text{pdf}_{w_1 X_1}(x) * \dots * \text{pdf}_{w_K X_K}(x)$, where pdf_X represents the uncertainty at the interpolated point, and $\text{pdf}_{w_i X_i}$ denotes the scaled distribution of a random variable X_i : $\text{pdf}_{w_i X_i}(x) = \frac{1}{w_i} \text{pdf}_{X_i}(x/w_i)$, for $1 \leq i \leq K$.

A common choice for modeling uncertainty is the Gaussian distribution. When PDFs of X_i 's are represented as Gaussian distributions, the PDF of the interpolated point X is also a Gaussian distribution whose mean and variance are linearly transformed from the means and variances of X_i 's [13]. Liu *et al.* [31] modeled PDFs of X_i using Gaussian mixture models (GMM). Their framework comprised fitting a GMM to uncertain data $f(\mathbf{v}_i)$ using expectation maximization [7], and involved MC sampling of estimated GMM for the PDF estimation at the interpolation point \mathbf{v} . Sakhaei *et al.* [51] modeled PDFs of X_i 's using compactly supported box splines, e.g., uniform distributions, and demonstrated the benefits of a box spline framework over a Gaussian assumption for DVR of uncertain data.

3.1.2 Quantile Interpolation of Histograms in 1D

Even though the box-spline method allows for interpolation of nonparametric distributions in a closed form [51] (described again in the supplementary material for this paper), it is computationally expensive because of its exponential nature. The exponential time complexity of the convolution-based nonparametric approach is a challenge to handle in the DVR reconstruction stage even when the system has high-performance hardware. In our contribution, we take advantage of the linear time complexity quantile interpolation technique for interpolation of nonparametric distributions. Read [49] proposed a quantile interpolation method for 1D interpolation of histograms. To summarize their approach, a histogram or nonparametric density characterizing the PDF at each vertex \mathbf{v}_i is broken into a fixed number of quantiles, and the respective quantiles at each vertex are interpolated to compute a probability distribution at the interpolated point \mathbf{v} . Thus, the computational complexity of computing the PDF at an interpolated position is *linearly* proportional to the number of quantiles.

The use of quantile interpolation in the context of uncertainty visualization was first advocated by Hollister and Pang [21]. The quantile interpolation method was shown to have two desirable qualities in the context of uncertainty visualization. First, quantile interpolation preserves the modality of probability distributions at cell vertices \mathbf{v}_i . Second, the variance of interpolation data is thresholded from below by variances of probability distributions at grid vertices. Moreover, quantile interpolation of histograms has been shown to better capture the shape of the interpolated distribution than the interpolation of parametric or GMM distributions. The quantile interpolation allows for a closed-form solution of the PDF at the interpolated position, and hence allows for efficient and accurate PDF computations.

Let $\text{pdf}_{X_1}(x)$ and $\text{pdf}_{X_2}(x)$ be the continuous probability distributions for random variables X_1 and X_2 at 1D cell vertices \mathbf{v}_1 and \mathbf{v}_2 , respectively. The probability distributions $\text{pdf}_{X_1}(x)$ and $\text{pdf}_{X_2}(x)$ can be estimated from noise samples using histograms or kernel density estimation (KDE) [42, 50]. Suppose the distributions $\text{pdf}_{X_1}(x)$ and $\text{pdf}_{X_2}(x)$ are broken into q quantiles each, where $qval$ denotes the quantile value. For example, setting $qval = 0.5$, $qval = 0.25$, and $qval = 0.125$ results in median ($q=2$), quartiles ($q=4$), and octiles ($q=8$), respectively.

Let $Q_{i1}, Q_{i2}, \dots, Q_{iq}$ denote q quantiles ordered by cumulative density function (CDF) with widths $w_{i1}, w_{i2}, \dots, w_{iq}$, respectively, for a random variable X_i associated with grid vertex \mathbf{v}_i when the quantile value is set to $qval$. The quantile representation for the PDF of each random variable X_i is a piecewise constant function in which each quantile

(piece) j assumes a constant probability density, i.e., $\Pr(Q_{ij}) = \frac{q_{val}}{w_{ij}}$. We, therefore, present $\text{pdf}_{X_i}(x)$ as a tuple $\{\Pr(Q_{i1}), \dots, \Pr(Q_{iq})\}$ in its quantile representation, in which each entry of a tuple denotes the probability density over its respective quantile. Let α indicate the spatial distance parameter between 1D cell vertices v_1 and v_2 . Let $\{Q_1, \dots, Q_q\}$ denote the quantile representation for a linearly interpolated random variable $X = \alpha X_2 + (1 - \alpha)X_1$ when the quantile value is q_{val} . Based on the previous work in [49] and [22], the probability density for the j 'th quantile $\Pr(Q_j)$ of the interpolated random variable can be computed as follows:

$$\begin{aligned}\Pr(Q_j) &= \frac{\Pr(Q_{1j})\Pr(Q_{2j})}{(1 - \alpha)\Pr(Q_{2j}) + \alpha\Pr(Q_{1j})} \\ &= \frac{q_{val}}{\alpha w_{2j} + (1 - \alpha)w_1}\end{aligned}\quad (1)$$

where $\alpha \in [0, 1]$ and $j \in \{1, 2, \dots, q\}$. As can be seen from Eq. 1, the quantile width for the j 'th quantile (quantile value = q_{val}) of a linearly interpolated random variable X is essentially a linear interpolation of j 'th quantile widths of random variables at grid vertices. Since the arithmetic operations in Eq. 1 are applied quantile-wise for each quantile $j \in \{1, 2, \dots, q\}$, the computational complexity of the quantile interpolation is linearly proportional to the number of quantiles q . As the number of quantiles q approaches infinity, the PDF at an interpolated position converges to a closed-form continuous probability distribution.

3.2 Quantile Interpolation of Nonparametric Distributions in 3D

Recently, Hollister and Pang [22] presented a closed-form solution for a bilinear interpolation of histograms/nonparametric distributions on a 2D grid using the quantile interpolation technique. We extend the derivation for quantile interpolation of histograms or nonparametric distributions to a 3D case using a similar approach, as proposed in [22]. Let α , β , and γ denote the spatial distance parameters for vertices v_i of a 3D grid cell along three dimensions. Deriving quantile interpolation in the interior of a 3D cell comprises three steps: first, an interpolated PDF along the cell edges can be computed by applying Eq. 1 to PDFs at grid vertices with parameter α . Second, the interpolated PDFs in the interior of cell faces can be computed by again applying Eq. 1 to interpolated PDFs computed in step one with parameter β . Finally, the interpolated PDF in the interior of 3D cell can be computed by applying Eq. 1 to interpolated PDFs computed in step two with parameter γ . For brevity, we represent the probability $\Pr(Q_{ij})$ as \Pr_i for the j 'th quantile of random variable X_i . Then the formula for the interpolated PDF of the j 'th quantile, i.e., $\Pr(Q_j)$, in 3D is as follows:

$$\Pr(Q_j) = \frac{\Pr_1 \Pr_2 \Pr_3 \Pr_4 \Pr_5 \Pr_6 \Pr_7 \Pr_8}{t_1 t_2 t_3 t_4 t_5 t_6 t_7} \quad (2)$$

where: $t_1 = \alpha \Pr_1 + (1 - \alpha) \Pr_2$, $t_2 = \alpha \Pr_3 + (1 - \alpha) \Pr_4$, $t_3 = \alpha \Pr_5 + (1 - \alpha) \Pr_6$, $t_4 = \alpha \Pr_7 + (1 - \alpha) \Pr_8$, $t_5 = \beta \Pr_1 \Pr_2 / t_1 + (1 - \beta) \Pr_3 \Pr_4 / t_2$, $t_6 = \beta \Pr_5 \Pr_6 / t_3 + (1 - \beta) \Pr_7 \Pr_8 / t_4$, $t_7 = \gamma \Pr_1 \Pr_2 \Pr_3 \Pr_4 / (t_1 t_2 t_5) + (1 - \gamma) \Pr_5 \Pr_6 \Pr_7 \Pr_8 / (t_3 t_4 t_6)$, $\alpha \in [0, 1]$, $\beta \in [0, 1]$, and $\gamma \in [0, 1]$.

Similar to Eq. 1, Eq. 2 simplifies to $\Pr(Q_j) = \frac{q_{val}}{w_j}$, where w_j denotes the width of the j 'th quantile of an interpolated random variable X . The width w_j is essentially a trilinear interpolation of widths of j 'th quantiles for PDFs of random variables X_1, \dots, X_8 at 3D cell vertices, evaluated with interpolation parameters α , β , and γ . We compute the formula in Eq. 2 using the MATLAB Symbolic Math Toolbox. We validate Eq. 2 through an experiment on synthetic data. For our experiment, we define histograms representing continuous probability distributions at eight vertices of a 3D cell. We randomly draw 2×10^6 samples from each histogram and perform KDE on the samples to estimate continuous PDFs, $\text{pdf}_{X_1}(x) \dots \text{pdf}_{X_8}(x)$, from the samples. A fixed number of

quantiles are then computed for estimated PDFs at each of the eight vertices with the quantile value $q_{val} = 0.001$. Setting the quantile value to 0.001 results in $q = 1000$ quantiles at each cell vertex. The PDF at the interpolated position is then computed in closed form using Eq. 2. Fig. 2(a) and (f) visualize the PDF computed at an interpolated position using MC sampling and our analytic derivation, respectively. Having the same shape for both distributions confirms the correctness of Eq. 2.

We compare an interpolated PDF shape computed using quantile interpolation with two GMM models, i.e., the GMM (ordered) [21] and the GMM (MC) [31]. The GMM (ordered) model is a variant of the GMM (MC) model in that Gaussians in each GMM are ranked based on sorted Gaussian means. The Gaussians at each grid vertex with the same rank are then interpolated. The PDF interpolation proposed in [31] resorts to MC sampling because of its exponential computational complexity. Fig. 2(b) shows the result of the interpolated PDF using the GMM (ordered) model, which emphasizes the peaks of the interpolated PDF seen in Fig. 2(a) and is smoother. Fig. 2(c) shows the result for the GMM (MC) model, in which a lack of ordering among Gaussians produces a PDF shape significantly different from the quantile interpolation or GMM (ordered) models. We assume four Gaussians for the GMM results since they require storage comparable to the quantile interpolation with eight quantiles (for more details, please see Sec. 5).

Fig. 2(d) and (e) visualize quantile interpolation PDF results when the PDF at each grid vertex is summarized using coarse quantile representation. The interpolated PDF with four quantiles in Fig. 2(d) appears similar to a uniform distribution without providing any useful insight regarding which data values have a higher probability of occurrence, whereas the interpolated PDF in Fig. 2(e) captures the shape of the interpolated PDF reasonably well for an eight-quantile PDF representation. Fig. 2(g-i) show results similar to those in Fig. 2(d-f), respectively, except that the number of noise samples at the grid vertices is reduced from 2^6 to 2^4 . For the reduced number of noise samples, the quantile interpolation results can fluctuate significantly for a relatively high number of quantiles, e.g., Fig. 2(f), because of poor KDE caused by reduced sample size.

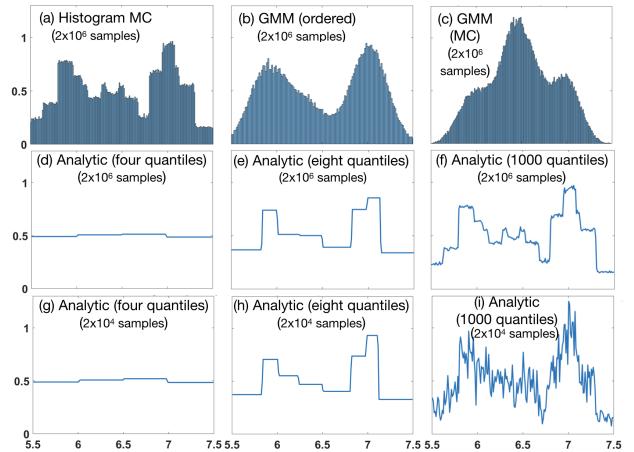


Fig. 2. A probability distribution at a 3D interpolated position computed using MC sampling (image (a)) vs. our analytic formula (image (f)) for quantile interpolation (Eq. 2). Images (b) and (c) visualize interpolated PDFs for GMM noise models. Images (d) and (e) visualize interpolated PDFs for coarse quantile representations. Images (g-i) show plots similar to images (d-f), respectively, for reduced sample size.

3.3 DVR with Quantile Interpolation

We describe a three-step approach for DVR of uncertain data when data uncertainty is characterized by histograms or nonparametric distributions. In the first step, we preprocess uncertain data. At each grid position, we estimate histogram or nonparametric density from noise samples. We then partition the continuous distribution $\text{pdf}_{X_i}(x)$ at each grid position v_i into q quantiles based on a user-set quantile value

$qval$ and compute a quantile representation of $\text{pdf}_{X_i}(x)$. We denote quantile representation of $\text{pdf}_{X_i}(x)$ as a tuple $\{\Pr(Q_{i1}), \dots, \Pr(Q_{iq})\}$ (see Sec. 3.1.2 for details). The quantile representations for PDFs are provided as inputs to the DVR framework.

In the second step, for each sample along a viewing ray of the DVR framework, we look up quantile-based PDFs at the neighboring eight vertices in a fragment shader. We apply Eq. 2 to neighboring vertex quantile densities and compute, in closed form, a quantile representation $\{\Pr(Q_1), \dots, \Pr(Q_q)\}$ for a continuous distribution $\text{pdf}_X(x)$ of the interpolated random variable X . In the third step, the opacity and color for an interpolated random variable X can be computed by applying the uncertainty integration framework proposed by Sakhaei and Entezari [51] as follows:

$$E(\text{TF}(X)) = \int \text{TF}(x) \text{pdf}_X(x) dx \quad (3)$$

where $\text{TF}(x)$ is the color and opacity sampled at intensity x in the TF domain, and $E(\text{TF}(X))$ represents the expected value of the classified color and opacity. The piecewise constant quantile representation for interpolated $\text{pdf}_X(x)$, i.e., a tuple $t = \{\Pr(Q_1) = \frac{qval}{w_1}, \dots, \Pr(Q_q) = \frac{qval}{w_q}\}$, may be integrated with a TF in different ways to produce meaningful DVR visualizations. We propose two integration schemes, namely *quantile range* and *quantile mean*.

Quantile range technique: In this method, we integrate all intensities contained in a quantile with a TF. Suppose $[a, b]$ denote the domain of interpolated random variable X . By substituting $\text{pdf}_X(x)$ as a quantile-based piecewise constant density function represented by the tuple t in Eq. 3, we get the following formula for the expected fragment color:

$$\begin{aligned} E(\text{TF}(X)) &= \int_a^{a+w_1} \text{TF}(x) \Pr(Q_1) dx + \dots + \int_{b-w_q}^b \text{TF}(x) \Pr(Q_q) dx \\ &= \frac{qval}{w_1} \int_a^{a+w_1} \text{TF}(x) dx + \dots + \frac{qval}{w_q} \int_{b-w_q}^b \text{TF}(x) dx \end{aligned} \quad (4)$$

In summary, the expected fragment color can be computed by averaging the TFs for each quantile and performing a weighted sum of average colors computed for each quantile, where the weight is equal to $qval$.

Quantile mean technique: In this method, we integrate quantile mean intensities with the TF. Let m_1, \dots, m_q denote the quantile means for quantiles $Q_1 \dots Q_q$, respectively, of interpolated random variable X . Then the discrete probability density for the mean of j 'th quantile is $\Pr(x = m_j) = \frac{p_j}{p}$, in which $p_j = \frac{qval}{w_j}$, and p is a normalization constant, i.e., $\frac{qval}{w_1} + \dots + \frac{qval}{w_q}$. The discretized version of Eq. 3 for the quantile mean technique results in the following formula for the expected fragment color:

$$E(\text{TF}(X)) = \sum_{j=1}^q \text{TF}(x = m_j) \Pr(x = m_j) \quad (5)$$

Memory and Time Complexity: We comment on the memory and time complexity of our approach for DVR with nonparametric statistics. The preprocessing step is computationally expensive. It involves computing KDE at each grid vertex from noise samples [42, 50] and reducing continuous distributions to their quantile representations. Let $L \times M \times N$ denote the grid size, and q denote the number of quantiles at each grid vertex. The memory consumed by input data for our nonparametric DVR framework is, therefore, $L \times M \times N \times (q+1)$, in which $q+1$ represents the number of boundaries defining quantiles at each grid vertex. Thus, the memory requirement for our nonparametric framework increases linearly with the number of quantiles. The fragment shader for our DVR framework computes the distribution at viewing-ray samples through quantile interpolation of neighboring quantile representations of PDFs with q quantiles (Eq. 2). The computational complexity of the quantile interpolation is linearly proportional

to the number of quantiles because of the quantile-wise arithmetic of probability distributions in Eq. 2. Hence, the computational complexity of our DVR framework increases linearly with the number of quantiles.

3.4 Quartile View

The quantile interpolation is an example of order statistics, where quantiles are ranked based on CDF for a random variable. Specifically, the j 'th quantile (ranked by CDF) of input probability distribution $\text{pdf}_{X_i}(x)$ corresponds to the j 'th quantile of a distribution computed using quantile interpolation (see Eq. 2). We take advantage of the order statistics exhibited by quantile interpolation to derive a boxplot-like view for the input data, which we refer to as the quartile view. Specifically, we visualize populations corresponding to three quantiles, i.e., first 25% (lower quartile), middle 50%, and last 25% (upper quartile), in the quartile view. We compute the expected color for each of the three populations with our quantile range method (Sec. 3.3). The middle 50% population represents a visualization corresponding to the interquartile range (IQR) for the input data, which is considered as a robust range in order statistics. Also, the quartile view can help users understand uncertainty in input data by examining commonalities and differences among visualizations for the lower, middle 50%, and upper quartiles. Our results in Fig. 5 and Fig. 7 illustrate the quartile views.

4 2D TF FOR CLASSIFICATION OF UNCERTAIN DATA

4.1 Interpolation of Uncertain Gradient Field

The reconstruction of the gradient field from the scalar field is an essential process in volume rendering [36]. The gradient information can be employed in both multidimensional TFs (for material classification) [9] and various local illumination techniques [17]. The gradient is mathematically the first-order derivative of the scalar field $f(\mathbf{v})$, $\nabla f(\mathbf{v}) = (\frac{\partial f(\mathbf{v})}{\partial x}, \frac{\partial f(\mathbf{v})}{\partial y}, \frac{\partial f(\mathbf{v})}{\partial z})^T$, and points in the direction of the steepest ascent. The reconstruction of the gradient field at an arbitrary position \mathbf{v} in an uncertain scalar field results in a trivariate random vector $\mathbf{Y} = (Y_x, Y_y, Y_z)^T$, where Y_x, Y_y , and Y_z are random variables denoting the uncertainties of a partial derivative along directions x, y , and z , respectively. The uncertain gradient \mathbf{Y} can be obtained from linear combinations of random variables at neighboring voxels \mathbf{v}_i 's of \mathbf{v} :

$$\begin{bmatrix} Y_x \\ Y_y \\ Y_z \end{bmatrix} = \sum_i \begin{bmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{bmatrix} X_i \text{ with weights } \begin{bmatrix} \alpha_i \\ \beta_i \\ \gamma_i \end{bmatrix} = \rho(\mathbf{v} - \mathbf{v}_i) \quad (6)$$

where $\rho : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ is the basis function that determines the amount of the contribution of each neighboring random variable X_i to gradient \mathbf{Y} at \mathbf{v} . The basis function ρ can be represented as a derivative reconstruction filter obtained by convolving a continuous interpolation filter with a digital derivative filter [38]. In the context of volume rendering, the most common choice is the combination of linear interpolation and central difference. For voxels that fall out of the support of ρ , zero weights are assigned. As a result, only M neighbors of X are involved in the gradient estimation, i.e., $1 \leq i \leq M$.

As shown in Eq. 6, the uncertain gradient estimation at arbitrary point \mathbf{v} is also obtained by linear combinations of random variables X_i 's of the uncertain scalar field. Therefore, the distribution of the gradient at \mathbf{v} can also be derived analytically using the spline framework. For example, when data uncertainties are modeled as (scaled) uniform distributions, the distribution of each partial derivative is also a box spline whose direction vectors are the interpolation weights. In this setup, the joint distribution of the random vector \mathbf{Y} , i.e., the uncertain gradient, is analytically derived as a trivariate box spline whose directions are obtained by the weights in Eq. 6, i.e., the i -th direction vector is $[\alpha_i, \beta_i, \gamma_i]^T$.

4.2 2D TF Integration with Interpolated Gradient Field

Unlike the gradient, it is challenging to analytically derive the PDF of the gradient magnitude given that random variables Y_x, Y_y , and Y_z

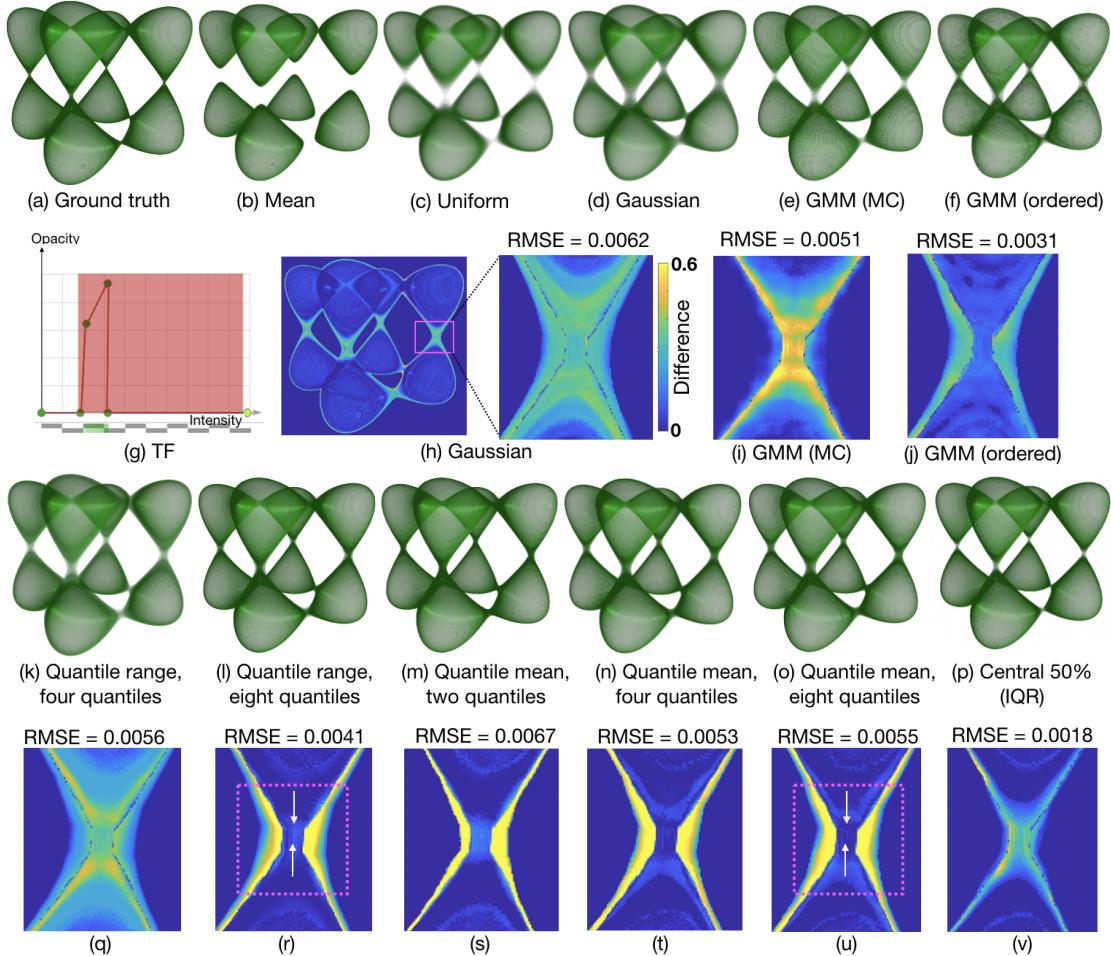


Fig. 3. Nonparametric (third and fourth rows) vs. parametric statistics (first and second rows) for visualizations of the tangle function: The ground truth volume visualized in (a) is mixed with noise to generate an ensemble representing uncertain data. The DVR visualizations for various statistical models in the first and third row are rendered with the same TF (image (g)). The second and the fourth rows quantify and visualize the differences with respect to the ground truth for their corresponding DVR images in the first and third rows, respectively. In images (r) and (u), the white arrows illustrate the positions of high reconstruction accuracy and the dotted pink boxes enclose the positions that illustrate error bands.

are correlated, even when these random variables are modeled as parametric normal distributions [35]. Therefore, the directional derivative along the mean gradient direction has been commonly used as an approximation in uncertainty visualization [43]. Mathematically, the gradient magnitude $\|\mathbf{Y}\| \approx \frac{\mu_{\mathbf{Y}}}{\|\mu_{\mathbf{Y}}\|} \mathbf{Y}$, where $\mu_{\mathbf{Y}}$ is the mean gradient, which can be obtained by substituting X_i 's with their mean values in (6). Since the directional derivative is also linear with respect to the original scalar field, we can formulate the interpolation of intensity X and gradient magnitude $\|\mathbf{Y}\|$ as linear combinations of M neighboring voxels of \mathbf{v} , i.e., $X = \sum_{i=1}^M w_i X_i$ and $\|\mathbf{Y}\| = \sum_{i=1}^M u_i X_i$. The joint PDF of the bivariate random variable $\mathbf{Z} = [X, \|\mathbf{Y}\|]^T$ is exactly described by a bivariate box spline with parameters $[(w_1, u_1)^T, \dots, (w_M, u_M)^T]$. Similar to (3), we compute the color at position \mathbf{v} by integrating the 2D TF with the joint PDF of intensity and gradient magnitude \mathbf{Z} , i.e., $E(\text{TF}(\mathbf{Z})) = \int \int \text{TF}(x, y) \text{pdf}_{\mathbf{Z}}(x, y) dx dy$.

5 RESULTS AND DISCUSSIONS

We demonstrate the effectiveness of our nonparametric noise modeling (Sec. 3.3) for DVR on two kinds of synthetic and real datasets: ensembles of volumetric datasets (Fig. 1a, Fig. 3, and Fig. 6) and downsampled versions of high-resolution datasets (Fig. 1b, Fig. 8, and Fig. 9). Next, we demonstrate improved DVR classifications with 2D TFs (Fig. 10, Fig. 11), where underlying uncertain data are represented as probability distributions (Sec. 4) as opposed to mean statistics.

All volume renderings are performed on a machine with Nvidia GPU Quadro P6000, with 24 GB memory. We integrate the fragment shaders for our statistical frameworks into the Voreen volume rendering engine (<http://voreen.uni-muenster.de>) for DVR of uncertain data.

5.1 Statistical Rendering: Nonparametric vs. mean/parametric statistics

Ensemble Datasets: In Fig. 3, we perform qualitative and quantitative assessment of DVR reconstruction accuracy for different noise models on a synthetic tangle [29] dataset. Fig. 3(a) visualizes the ground truth volume for a fixed TF design (Fig. 3(g)). The same TF is used for all visualizations in Fig. 3. The ground truth volume is mixed with noise to generate an ensemble of 50 members representing uncertain data. Specifically, we inject noise samples randomly drawn from a bimodal probability distribution, in which the mode with 80% probability concentration is centered around the ground truth, and the mode with 20% probability concentration (representing outliers) is centered far away from the ground truth. The injected noise, thus, has a shape similar to the one-tailed asymmetric distribution.

Fig. 3(b) and Fig. 3(c) visualize the results for the mean-field and uniform noise models, respectively. The presence of the outliers in the noise samples shifts the sample mean at each grid vertex substantially, hence breaking the regions connecting the blobs of the tangle function. The reconstruction with the uniform model still shows improved topological recovery compared to the mean-field. The visualizations using

Gaussian, GMM (MC), and GMM (ordered) in Fig. 3(d-f), respectively, and our proposed nonparametric models in Fig. 3(k-p) show further reconstruction improvements.

In the second and fourth rows of Fig. 3, we perform quantitative analysis of the reconstruction accuracy for Gaussian, GMM (MC), GMM (ordered), and nonparametric statistical models. Specifically, at each pixel, we compute an absolute difference between the mean of the RGB values for a DVR image specific to a noise model and the DVR of the ground truth (Fig. 3(a)). We then visualize computed differences using a blue-yellow diverging color map, in which yellow and blue indicate relatively high and low difference regions, respectively. In Fig. 3, we also report the root mean squared error (RMSE) for each noise model. Note that the difference images for the mean and uniform noise models are not shown in Fig. 3 because of their relatively high RMSE values, i.e., 0.0245 and 0.0200, respectively.

We assume four Gaussians for modeling uncertainty with GMMs since they consume a memory comparable to the quantile interpolation with eight quantiles (for details, see the next paragraph). In the case of the GMM (MC) model, we draw MC samples from a GMM until no significant variations are observed in the visualizations. The same process is followed for all the datasets. In Fig. 3(e), we drew 200 MC samples per grid vertex. The difference images for the GMM (ordered) (Fig. 3(j)) and our nonparametric approach (Fig. 3(r-u)) show relatively high reconstruction accuracy in the interior of the dataset, as indicated by the white arrows in Fig. 3(r) and (u), when compared to the difference images for the Gaussian (Fig. 3(h)) and GMM (MC) (Fig. 3(i)) noise models. Moreover, the error pattern is systematic and refined in the case of our quantile range and quantile mean techniques with eight quantiles. Specifically, the error bands are observed (enclosed within the pink dotted boxes in Fig. 3(r) and (u)) with a relatively high error near the boundary between the dataset and background with the error gradually decreasing as we move farther away from the dataset. The RMSE for our nonparametric approach with eight quantiles is smaller than the ones for mean, uniform, and Gaussian statistics. Also, increasing the number of quantiles for our quantile interpolation techniques decreases the RMSE for tangle visualizations when the number of quantiles is increased from two to four, e.g., Fig. 3(q-r). Fig. 3(v) visualizes a difference image for the expected visualization in Fig. 3(p) corresponding to the middle 50% quantile (IQR), ordered by CDF, for the random variable at each grid vertex. The RMSE for the IQR visualization is the lowest among all visualizations in Fig. 3.

The uniform and Gaussian noise assumptions for the visualizations in Fig. 3(c) and (d), respectively, require storing the mean and width/variance of uncertain values, and hence consume twice the memory needed for the mean-field approach in Fig. 3(b). The visualizations in Fig. 3(e) and (f) with four Gaussians per GMM consume 12 times the memory needed for the mean-field (see [31]), as each Gaussian requires storing its mean, variance, and weight. Our proposed nonparametric models consume memory linearly proportional to the number of quantiles (Sec. 3.3). For example, the quantile representations for the results shown in Fig. 3(l) and (o) consume nine times the memory needed for the mean-field. The IQR visualization Fig. 3(p) interpolates only a single quantile (middle 50%), and, thus, requires only twice the memory needed for the mean-field.

In Fig. 4(a), we analyze the effects of the number of quantiles in our nonparametric framework and the number of ensemble members on the reconstruction accuracy of the tangle dataset. As observed in Fig. 2, analytically computed quantile interpolation can be quite unreliable when the average number of samples per quantile is relatively low (Fig. 2(i)). We observe a similar trend in Fig. 4(a) for the tangle dataset. Analysis with a small sample size, e.g., five ensemble members, results in poor KDE at grid vertices, hence resulting in a RMSE for the four/eight quantile case greater than the one for the two quantile case. In contrast, the RMSE for the four/eight quantile case becomes smaller than the RMSE for the two quantile case and becomes more stable, as the number of ensemble members increases to 50.

Fig. 4(b) visualizes the relative RMSE and frame rate for various noise models. The mean-field is rendered at a frame rate of 10 fps. We measured a rendering performance of the visualizations of the tangle

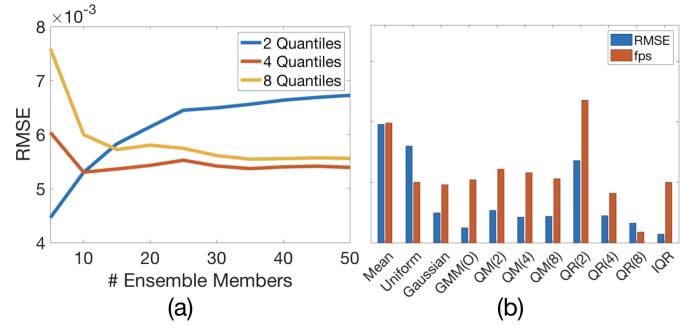


Fig. 4. (a) Effects of number of quantiles and sample size on reconstruction accuracy. (b) The relative RMSE and frame rate (fps) for different noise models. In image (b), GMM (O), QM, and QR denote GMM (ordered), quantile mean, and quantile range noise models, respectively.

dataset ($64 \times 64 \times 64$) using Nvidia’s Frameview tool. Our proposed quantile mean method achieves frame rates comparable to parametric noise models for eight quantile representations. Note that the frame rate is not shown for the GMM (MC) model in Fig. 4(b), as the GMM (MC) noise model [31] uses screen space integration of MC samples for producing static images. Fig. 5 visualizes a quartile view (Sec. 3.4) for the tangle dataset. The magenta boxes in Fig. 5 highlight the positions that exhibit reconstruction variability across the three quartile populations, hence indicating data uncertainty in those regions.

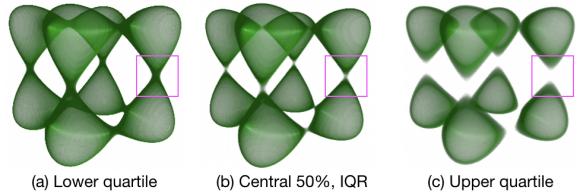


Fig. 5. The quartile view for the uncertain tangle dataset. The pink boxes mark positions that exhibit reconstruction variations across three populations.

Fig. 1(a) visualizes the synthetic teardrop function [29] for an experiment similar to the tangle function. For the teardrop dataset, we again analyze an ensemble of 50 members. In Fig. 1(a), the reconstruction in the case of the nonparametric density assumption is superior compared to the parametric density assumptions with ground truth as a reference.

We perform an experiment similar to the ones for the tangle and teardrop functions on a real dataset. We analyze the Red Sea eddy simulation ensemble comprising 20 members made available at the IEEE SciVis Contest 2020 (<https://kaust-vislab.github.io/SciVis2020/>). Each member of the ensemble dataset is generated based on the MIT ocean general circulation model (MITgcm) and the Data Research Testbed (DART) [23] with varying initial conditions. The ensembles are sampled for 60 time steps on a grid with resolution $500 \times 500 \times 50$ to represent a time-varying 3D flow [52].

Fig. 6 shows the DVR of the uncertain velocity magnitude field for an ensemble (time step = 40) over a portion of the domain using the mean-field, parametric, and our nonparametric statistical frameworks. Fig. 6(a) visualizes a velocity vector field using arrow glyphs colored by magnitude for a single ensemble member. High-velocity magnitude generally is observed on a vortex rim. Fig. 6(b) visualizes the result for the mean-field. We set a TF (Fig. 6(c)) for the mean-field visualization, such that relatively high-, moderate-, and low-velocity magnitudes are assigned red, blue, and yellow, respectively. The opacities are set to recover eddy-like structures from the dataset. The same TF is then used for all visualization in Fig. 6. The results for all noise models in Fig. 6 look significantly different from the mean-field visualization. Fig. 7 visualizes a quartile view for the Red Sea dataset. The dotted black boxes in Fig. 7 highlight the positions where the variability or

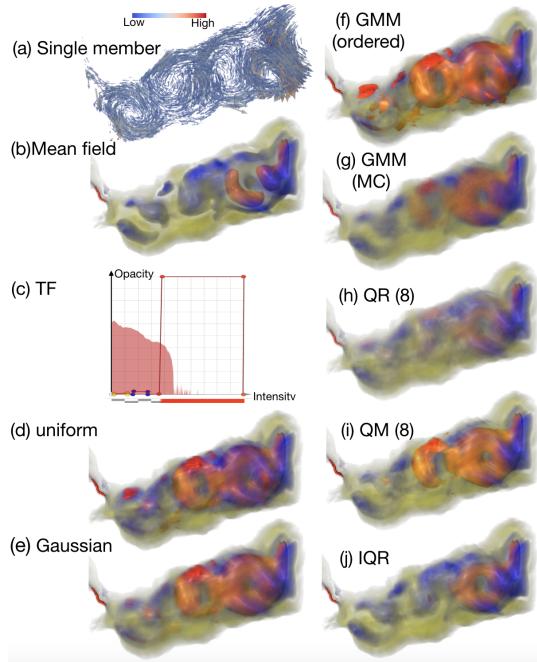


Fig. 6. Visualizations of an uncertain velocity magnitude field for the Red Sea eddy simulations: (a) arrow glyph visualization of a velocity vector field for a single member colored by magnitude, (b, d-j) DVR using various statistical models with the TF shown in image (c). The red, blue, and yellow in the TF indicate relatively high-, moderate-, and low-velocity magnitudes. QR(8) and QM(8) denote our proposed quantile range and mean techniques, respectively, with eight quantiles.

uncertainty of eddy presence is prominent across populations belonging to the lower, central 50%, and upper quartiles. In contrast, the solid black boxes highlight the positions where eddy presence is consistently observed across the three populations.

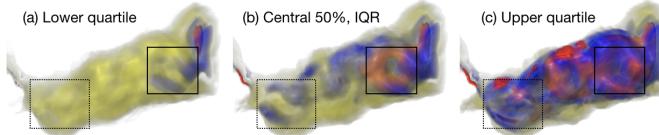


Fig. 7. The quartile view for the Red Sea eddy dataset. The solid and dotted boxes indicate the positions with relatively high and low confidence, respectively, regarding the eddy presence.

Downsampling Datasets: We perform an experiment on a nested spheres dataset similar to the experiment described in Section 4.2 of the previous work on GMM-based DVR [31]. In Fig. 8, we compare the visualizations for various noise models. For our experiment, we sample the nested spheres function on a high-resolution $512 \times 512 \times 512$ grid. In the nested spheres function, spheres attain nonzero values on their surface, and the empty spaces between the spheres are assigned zero intensity. We treat the visualization of a high-resolution volume as a reference image (Fig. 8(a)). We set TF (Fig. 8(b)) for the ground truth such that four spheres are assigned distinct colors. The histogram in Fig. 8(b) shows four intensity peaks corresponding to four spheres. The high-resolution nested spheres dataset is partitioned into $4 \times 4 \times 4$ bricks, and the mean intensity for each brick is stored at a grid vertex of the downsampled version. The mean-field is visualized in Fig. 8(c). For distribution-based analysis, we store a probability distribution characterizing noise per brick (similar to the hixel idea proposed by Thompson *et al.* [55]) and apply our DVR framework for visualization of the distribution data. Fig. 8(d-i) visualize the results for

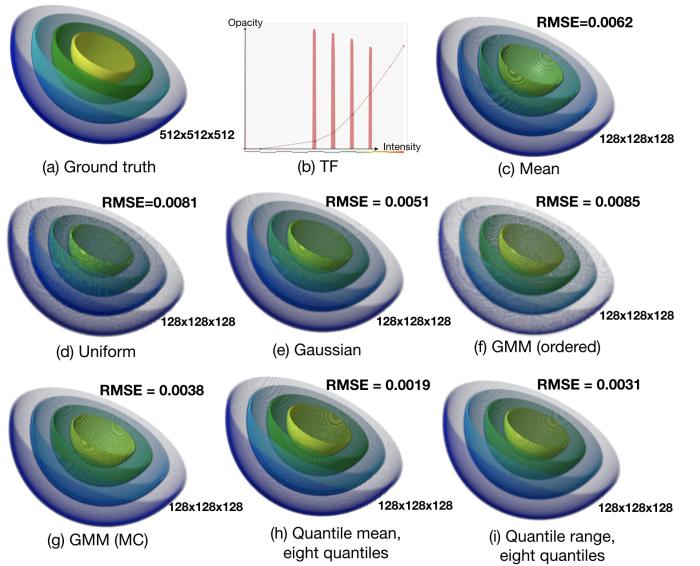


Fig. 8. Visualizations of the nested spheres function: All results are rendered with the same TF shown in image (b).

various noise models, in which our quantile mean technique exhibits the lowest RMSE. Our closed-form nonparametric models enable us to efficiently generate visualizations without needing to perform MC sampling or screen space integration as in [31].

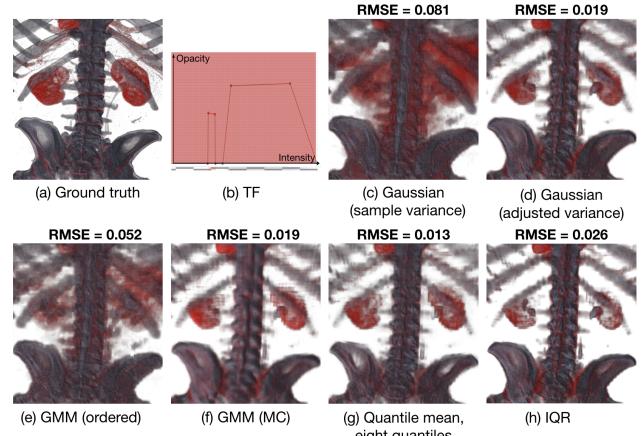


Fig. 9. Zoomed-in views for the Osirix OBELIX dataset visualizations in Fig. 1: All results are rendered with the same TF (image (b)). The ground truth in (a) with resolution $512 \times 512 \times 1559$ is reduced to $64 \times 64 \times 195$, and the reduced data with uncertainty are visualized by applying various noise models in (c)-(h).

Fig. 1(b) visualizes results similar to Fig. 8 on the Osirix OBELIX dataset and classifies bones (gray) and kidneys (red). In Fig. 1(b), the mean-field loses a significant amount of information near the kidneys. The visualizations using the uniform and Gaussian noise models result in thicker bone rendering and fail to localize the right-kidney classification. For the zoomed-in visualizations of the Osirix OBELIX dataset in Fig. 9, the RMSE for our quantile mean technique is the smallest among all noise models.

Parameter Sensitivity: We briefly discuss the parameters of statistical models that can potentially influence the quality of visualizations, and hence, RMSE. In the case of our proposed nonparametric models, the reliability of quantile interpolation results is, admittedly, sensitive to the sample size with respect to the number of quantiles (see Fig. 2 and

Fig. 4(a)), which in turn depends on the complexity of underlying data distributions. We ensure a sufficient sampling density for the dataset at hand, e.g., 50 for the tangle dataset and 20 for the Red Sea eddy simulations, with an empirical approach. Specifically, we keep track of how much visualizations vary with increasing the sample size. We stop when the increase in sample size does not change or affect the quality of visualizations (more details in the supplementary material). We follow the same empirical approach for ensuring sufficient sampling density for all nonparametric statistical visualizations.

In the case of the parametric models, the RMSE is again sensitive to the choice of the parameter values. For example, for the uniform and Gaussian noise models, we estimate the mean and width/variance from the noise samples. Adjusting the Gaussian variance estimated from the noise samples can, however, significantly improve the classification result, as we have demonstrated for the Osirix OBELIX dataset in Fig. 9(c) and (d). In the case of our GMM visualizations, we use four Gaussians per mixture since they consume a memory comparable to the quantile interpolation with eight quantiles. Still, the estimation of the parameters, such as the number of Gaussians, regularization value, covariance matrix estimation, and number of MC samples, can be further improved to enhance the quality of the GMM results. Note that the GMMs demand a memory consumption equal to three times the number of Gaussians in a GMM (for storing mean, variance, and weight per Gaussian). In summary, the study regarding the choice of the noise model in the context of DVR and analytical identification of optimal parameter values for data with arbitrary complexity of noise distributions is nontrivial, and we plan to research it in the future.

5.2 Application: 2D TFs and Visualizing Bivariate Data

We demonstrate our statistical rendering with 2D TFs using the classic tooth dataset in Fig. 10. Fig. 10(a) shows the ground truth rendering (top) with a 2D TF (bottom), where the tooth is classified as four materials: dentine (yellow), tooth-holding material (blue), enamel (red), and root and dentine boundaries (green). Fig. 10(b) shows the standard DVR of the bivariate field characterized by mean intensities and the mean gradient magnitudes for data at reduced resolutions. Fig. 10(c) shows our statistical rendering using the interpolated distributions of uncertain intensities and gradient magnitudes at the same reduced resolutions as Fig. 10(b). When comparing Fig. 10(b) and Fig. 10(c), our methods show improved recovery of the features in the ground truth, whereas the mean-field visualizations result in poor classifications.

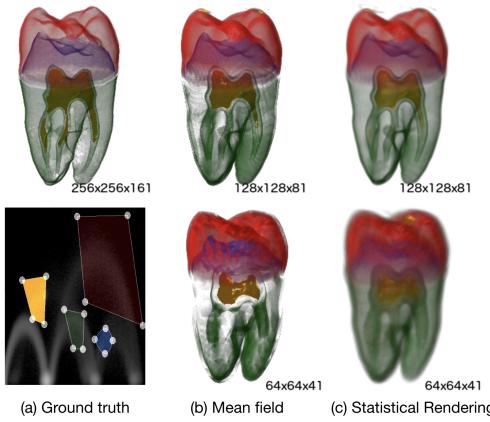


Fig. 10. Volume rendering of the tooth dataset with a 2D TF.

Thompson *et al.* [55] proposed a technique to visualize *fuzzy isosurfaces* of a scalar field in reduced data. The multifield equivalents of isosurfaces are fiber surfaces [5]. In the bivariate case, a fiber surface is a contour defined by a curve that is composed of a number of points (i.e., fibers) in the 2D range space. In the case of reduced bivariate data, our proposed uncertainty-aware 2D TF integration scheme can be applied for the visualization of *fuzzy fiber surfaces*. Specifically, we can model the uncertainties at grid points of the reduced data as the tensor

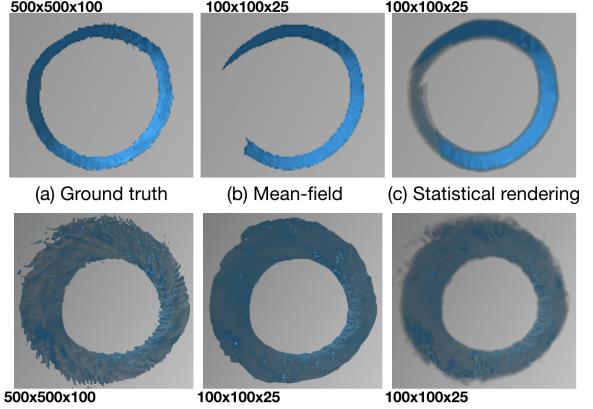


Fig. 11. Visualizations of fiber surfaces with $600 < \text{Pressure} < 800$ and Water Vapor = 0.011 in the top row and $1500 < \text{Pressure} < 1800$ and Water Vapor = 0.028 in the bottom row.

product of two univariate box splines defined by any two uncertain fields. Then, the 2D TF reconstruction scheme introduced in Sec. 4.2 can be directly applied to the spline-modeled bivariate uncertain field.

We conducted experiments on a multifield dataset, *Isabel*, a simulation of hurricane Isabel from the West Atlantic region in 2003. The data dimensions are $500 \times 500 \times 100$ with 48 time steps. Fig. 11(a) shows the volume rendering of a fiber surface defined in a bivariate field: Pressure and Water Vapor of *Isabel* at time step 18. The fiber surface is identified by a line in the 2D range space with $600 < \text{Pressure} < 800$ and Water Vapor = 0.011 (the top row of Fig. 11). When the data are represented as the mean statistics with a resolution of $100 \times 100 \times 25$, as shown in Fig. 11(b), volume rendering of the reduced fiber surface resulted in a disconnected surface. Fig. 11(c) shows our statistical rendering of the fiber surface, which preserves the topology of the original surface with "fuzzy" depiction of less certain areas. The bottom row in Fig. 11 shows DVR results for another fiber surface similar to the top row results for $1500 < \text{Pressure} < 1800$ and Water Vapor = 0.028.

6 CONCLUSION AND FUTURE WORK

We expand the spline-based parametric DVR framework [51] to more flexible nonparametric statistics for visualization of an uncertain scalar field. We leverage the quantile interpolation technique [21, 49] for efficient integration of nonparametric PDFs with a DVR framework. We evaluate our proposed nonparametric statistical models by presenting qualitative and quantitative comparisons with respect to mean-field and parametric statistical models. We show that the time and memory complexity of our nonparametric DVR framework increases linearly with the number of quantiles used in quantile interpolation. We demonstrate the application of a previous study [51] to 2D TFs for improved DVR classification compared to the 2D TF classification with mean statistics.

Our approach has a few limitations that we plan to address in subsequent work. In this work, we employ quantile interpolation with even quantile values. We would like to study an efficient use of uneven quantile values for improving DVR classification accuracy, which we briefly discuss in the supplementary material. Next, we would like to investigate methods for efficient implementation of an exponentially complex box-spline framework [51] (see also the supplementary material) and study their effectiveness in DVR for uncertain data. We plan to generalize our uncertainty-aware 2D TF framework (Sec. 4) to TFs with a variable number of dimensions using nonparametric statistics. Finally, we plan to study the nonparametric models for dependent random fields for further improvements in DVR reconstruction accuracy.

ACKNOWLEDGMENTS

This work was supported in part by the NSF grant IIS-1617101; the NIH grants P41 GM103545-18 and R24 GM136986; the DOE grant DE-FE0031880; and the Intel Graphics and Visualization Institutes of XeLLENCE.

REFERENCES

- [1] T. Athawale and A. Entezari. Uncertainty quantification in linear interpolation for isosurface extraction. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2723–2732, Oct. 2013. doi: 10.1109/TVCG.2013.208
- [2] T. 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
- [3] T. Athawale, E. Sakhaei, and A. Entezari. Isosurface visualization of data with nonparametric models for uncertainty. *IEEE Transactions on Visualization and Computer Graphics*, 22(1):777–786, Jan. 2016. doi: 10.1109/TVCG.2015.2467958
- [4] K. Brodlie, R. Allendes Osorio, and A. Lopes. A review of uncertainty in data visualization. *Expanding the Frontiers of Visual Analytics and Visualization*, pp. 81–109, 2012. doi: 10.1007/978-1-4471-2804-5_6
- [5] 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
- [6] C. Correa, Y. H. Chan, and K. L. Ma. A framework for uncertainty-aware visual analytics. In *Proceedings of 2009 IEEE Symposium on Visual Analytics Science and Technology (VAST 2009)*, pp. 51–58. IEEE, October 2009. doi: 10.1109/VAST.2009.5332611
- [7] A. Dempster, N. Laird, and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.
- [8] S. Djurcicov, K. Kim, P. Lermusiaux, and A. Pang. Visualizing scalar volumetric data with uncertainty. *Computers and Graphics*, 26(2):239–248, April 2002. doi: 10.1016/S0097-8493(02)00055-9
- [9] R. A. Drebin, L. Carpenter, and P. Hanrahan. Volume rendering. In *Proceedings of the 15th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH '88*, pp. 65–74. ACM, New York, NY, USA, June 1988. doi: 10.1145/54852.378484
- [10] T. Etienne, D. Jönsson, T. Ropinski, C. Scheidegger, J. L. Comba, L. G. Nonato, R. M. Kirby, A. Ynnerman, and C. T. Silva. Verifying volume rendering using discretization error analysis. *IEEE transactions on Visualization and Computer Graphics*, 20(1):140–154, Jan 2014. doi: 10.1109/TVCG.2013.90
- [11] T. Etienne, R. Kirby, and C. Silva. *An Introduction to Verification of Visualization Techniques*. Morgan & Claypool Publishers, 2015. doi: 10.2200/S00679ED1V01Y201511CGR022
- [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, September 2018. doi: 10.1109/TVCG.2018.2864432
- [13] W. Feller. *An introduction to probability theory and its applications: volume I*, vol. 3. John Wiley & Sons New York, Jan 1968.
- [14] N. Fout and K. L. Ma. Fuzzy volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, 18(12):2335–2344, Oct 2012. doi: 10.1109/TVCG.2012.227
- [15] M. G. Genton, C. R. Johnson, K. Potter, G. Stenchikov, and Y. Sun. Surface boxplots. *Stat Journal*, 3(1):1–11, 2014. doi: 10.1002/sta4.39
- [16] 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
- [17] M. Hadwiger, J. M. Kniss, C. Rezk-salama, D. Weiskopf, and K. Engel. *Real-time Volume Graphics*. A. K. Peters, Ltd., Natick, MA, USA, 2006.
- [18] W. He, H. Guo, H. W. Shen, and T. Peterka. eFESTA: Ensemble feature exploration with surface density estimates. *IEEE Transactions on Visualization and Computer Graphics*, 26(4):1716–1731, Nov 2018. doi: 10.1109/TVCG.2018.2879866
- [19] F. V. Higuera, N. Sauber, B. Tomandl, C. Nimsky, G. Greiner, and P. Hasstreiter. Automatic adjustment of bidimensional transfer functions for direct volume visualization of intracranial aneurysms. In *Proc. SPIE 5367, Medical Imaging 2004: Visualization, Image-Guided Procedures, and Display*, pp. 275–284. International Society for Optics and Photonics, May 2004. doi: 10.1117/12.535534
- [20] R. V. Hogg, A. T. Craig, and J. W. McKean. *Introduction to mathematical statistics*, 6th ed. Pearson, 2004.
- [21] B. Hollister and A. Pang. Interpolation of non-Gaussian probability distributions for ensemble visualization. Technical report, Jack Baskin School of Engineering, Oct 2013.
- [22] B. Hollister and A. Pang. Bivariate quantile interpolation for ensemble derived probability density estimates. *International Journal for Uncertainty Quantification*, 5(2):123–137, 2015. doi: 10.1615/Int.J.UncertaintyQuantification.2015011789
- [23] I. Hoteit, T. Hoar, G. Gopalakrishnan, J. Anderson, N. Collins, B. Cornuelle, A. Khl, and P. Heimbach. A MITgcm/DART ocean prediction and analysis system with application to the Gulf of Mexico. *Dynamics of Atmospheres and Oceans*, 63:1–23, September 2013. doi: 10.1016/j.dynatmoce.2013.03.002
- [24] C. R. Johnson. Top scientific visualization research problems. *IEEE Computer Graphics and Applications: Visualization Viewpoints*, 24(4):13–17, July/August 2004. doi: 10.1109/MCG.2004.20
- [25] C. R. Johnson and A. R. Sanderson. A next step: Visualizing errors and uncertainty. *Computer Graphics and Applications, IEEE*, 23(5):6–10, Sept.–Oct. 2003. doi: 10.1109/MCG.2003.1231171
- [26] J. Kniss, G. Kindlmann, and C. Hansen. Interactive volume rendering using multi-dimensional transfer functions and direct manipulation widgets. In *Proceedings of the conference on Visualization'01*, pp. 255–262. IEEE Computer Society, Oct 2001. doi: 10.1109/VISUAL.2001.964519
- [27] 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, Nov 2002. doi: 10.1109/TVCG.2002.1021579
- [28] J. M. Kniss, R. V. Uitert, A. Stephens, G.-S. L., T. Tasdizen, and C. Hansen. Statistically quantitative volume visualization. In *Proceedings of IEEE Visualization 2005*, pp. 287–294, Oct 2005. doi: 10.1109/VISUAL.2005.1532807
- [29] 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. 28(1):26–40, Feb 2009. doi: 10.1111/j.1467-8659.2008.01189.x
- [30] J. Kronander, J. Unger, T. Möller, and A. Ynnerman. Estimation and modeling of actual numerical errors in volume rendering. In *Computer Graphics Forum*, vol. 29, pp. 893–902. Wiley Online Library, 2010.
- [31] 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)*. IEEE, Oct 2012. doi: 10.1109/LDAV.2012.6378978
- [32] S. López-Pintado and J. Romo. On the concept of depth for functional data. *Journal of the American Statistical Association*, 104(486):718–734, June 2009. doi: 10.1198/jasa.2009.0108
- [33] C. Lundstrom, 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 2007. doi: 10.1109/TVCG.2007.70518
- [34] B. Ma and A. Entezari. Volumetric feature-based classification and visibility analysis for transfer function design. *IEEE Transactions on Visualization and Computer Graphics*, 24(12):3253–3267, 2018.
- [35] A. Mathai and S. Provost. *Quadratic Forms in Random Variables*. Statistics: A Series of Textbooks and Monographs. Taylor & Francis, 1992.
- [36] Z. Mihajlovic, Z. Comput., L. Budin, and N. Quid. Reconstruction of gradient in volume rendering. *IEEE Transactions on Visualization and Computer Graphics*, Dec 2003. doi: 10.1109/ICIT.2003.1290301
- [37] M. Mirzargar, R. Whitaker, and R. M. Kirby. Curve boxplot: Generalization of boxplot for ensembles of curves. *IEEE Transactions on Visualization and Computer Graphics*, 20(12):2654–63, Nov 2014. doi: 10.1109/TVCG.2014.2346455
- [38] T. Möller, R. Machiraju, K. Mueller, and R. Yagel. A comparison of normal estimation schemes. In *Proceedings of the IEEE Conference on Visualization 1997*, pp. 19–26. IEEE, Oct. 1997. doi: 10.1109/VISUAL.1997.663848
- [39] M. Otto, T. Germer, H.-C. Hege, and H. Theisel. Uncertain 2D vector field topology. 29(2):347–356, June 2010. doi: 10.1111/j.1467-8659.2009.01604.x
- [40] M. Otto, T. Germer, and H. Theisel. Uncertain topology of 3D vector fields. In *2011 IEEE Pacific Visualization Symposium*, pp. 67–74. IEEE, 2011. doi: 10.1109/PACIFICVIS.2011.5742374
- [41] A. T. Pang, C. M. Wittenbrink, and S. K. Lodha. Approaches to uncertainty visualization. *The Visual Computer*, 13(8):370–390, Nov 1997. doi: 10.1007/s003710050111
- [42] E. Parzen. On estimation of a probability density function and mode. *The Annals of Mathematical Statistics*, 33(3):1065–1076, 1962.
- [43] T. Pfaffelmoser, M. Mihai, and R. Westermann. Visualizing the variability of gradients in uncertain 2D scalar fields. *IEEE Transactions on Visual-*

- ization and Computer Graphics*, 19(11):1948–1961, June 2013. doi: 10.1109/TVCG.2013.92
- [44] T. Pfaffelmoser, M. Reitinger, and R. Westermann. Visualizing the positional and geometrical variability of isosurfaces in uncertain scalar fields. *Computer Graphics Forum*, 30(3):951–960, June 2011. doi: 10.1111/j.1467-8659.2011.01944.x
- [45] H. Pfister, B. Lorensen, C. Bajaj, G. Kindlmann, W. Schroeder, L. S. Avila, K. Raghu, R. Machiraju, and J. Lee. The transfer function bake-off. *IEEE Computer Graphics and Applications*, 21(3):16–22, May 2001. doi: 10.1109/38.920623
- [46] K. Pöthkow and H. C. Hege. Positional uncertainty of isocontours: Condition analysis and probabilistic measures. *IEEE Transactions on Visualization and Computer Graphics*, 17(10):1393–1406, Nov 2010. doi: 10.1109/TVCG.2010.247
- [47] K. Pöthkow and H.-C. Hege. Nonparametric models for uncertainty visualization. *Computer Graphics Forum*, 32(3.2):131–140, July 2013. doi: 10.1111/cgf.12100
- [48] K. Potter, P. Rosen, and C. R. Johnson. From quantification to visualization: A taxonomy of uncertainty visualization approaches. In *Uncertainty Quantification in Scientific Computing*, pp. 226–249. Springer, 2012. doi: 10.1007/978-3-642-32677-6_15
- [49] A. L. Read. Linear interpolation of histograms. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, 425(1-2):357–360, April 1999. doi: 10.1016/S0168-9002(98)01347-3
- [50] M. Rosenblatt. Remarks on some nonparametric estimates of a density function. *The Annals of Mathematical Statistics*, 27(3):832–837, 1956.
- [51] E. Sakhaei 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
- [52] S. Sivareddy, H. Toyé, 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*, 2020.
- [53] S. K. Suter, B. Ma, and A. Entezari. Visual analysis of 3d data by isovalue clustering. In *International Symposium on Visual Computing*, pp. 313–322. Springer, 2014. doi: 10.1007/978-3-319-14249-4_30
- [54] N. Svakhine, D. S. Ebert, and D. Stredney. Illustration motifs for effective medical volume illustration. *IEEE Computer Graphics and Applications*, 25(3):31–39, June 2005. doi: 10.1109/MCG.2005.60
- [55] D. Thompson, J. Levine, J. Bennett, P. Bremer, A. Gyulassy, V. Pascucci, and P. Pebay. Analysis of large-scale scalar data using hixels. In *2011 IEEE Symposium on Large Data Analysis and Visualization (LDAV)*, pp. 23–30, 2011. doi: 10.1109/LDAV.2011.6092313
- [56] R. Whitaker, M. Mirzargar, and R. 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, Oct. 2013. doi: 10.1109/TVCG.2013.143
- [57] P. C. Wong, H. W. Shen, C. R. Johnson, C. Chen, and R. B. Ross. The top 10 challenges in extreme-scale visual analytics. *IEEE Computer Graphics and Applications*, 32(4):63–67, Aug 2012. doi: 10.1109/MCG.2012.87
- [58] H. Younesy, T. Möller, and H. Carr. Improving the quality of multi-resolution volume rendering. In *EUROVIS'06: Proceedings of the Eighth Joint Eurographics / IEEE VGTC conference on Visualization*, pp. 251–258. Citeseer, May 2006.
- [59] T. Zuk and S. Carpendale. Theoretical analysis of uncertainty visualizations. In *Proceedings Volume 6060, Visualization and Data Analysis 2006*, vol. 6060, pp. 66–79, Jan 2006. doi: 10.1117/12.643631