Data-Driven Computation of Probabilistic Marching Cubes for Efficient Visualization of Level-Set Uncertainty

Tushar M. Athawale, Zhe Wang, Chris R. Johnson, and David Pugmire

Abstract
Uncertainty visualization is an important emerging research area. Being able to visualize data uncertainty can help scientists improve trust in analysis and decision-making. However, visualizing uncertainty can add computational overhead, which can hinder the efficiency of analysis. In this paper, we propose novel data-driven techniques to reduce the computational requirements of the probabilistic marching cubes (PMC) algorithm. PMC is an uncertainty visualization technique that studies how uncertainty in data affects level-set positions. However, the algorithm relies on expensive Monte Carlo (MC) sampling for the multivariate Gaussian uncertainty model because no closed-form solution exists for the integration of multivariate Gaussian. In this work, we propose the eigenvalue decomposition and adaptive probability model techniques that reduce the amount of MC sampling in the original PMC algorithm and hence speed up the computations. Our proposed methods produce results that show negligible differences compared with the original PMC algorithm demonstrated through metrics, including root mean squared error, maximum error, and difference images. We demonstrate the performance and accuracy evaluations of our data-driven methods through experiments on synthetic and real datasets.

CCS Concepts
• Human-centered computing → Scientific visualization; • Mathematics of computing → Probabilistic algorithms; Sequential Monte Carlo methods; Multivariate statistics;

1. Introduction
Uncertainty visualization is one of the top research challenges [JS03, Joh04] that must be addressed to ensure reliability of scientific analysis and decision-making. Although there have been multiple advances in uncertainty visualization for scalar [ASE16, PWB+09, PWH11, FFST19, LLBP12, AMS21], vector [OGT11, FBW16, GHP+16], and tensor data [Kim04, Jon02], they bear an additional cost of computing uncertainty in visualization. Such computational overhead can limit our ability to integrate uncertainty analysis techniques into production visualization tools (e.g., ParaView [ABM+01], VisIt [Chi11], VTK-m [Mea16]) for use by scientists. Researching novel techniques that speed up computation of uncertainty in scientific visualizations is therefore critical to make uncertainty usable in visualization tools. In this paper, we propose novel data-driven techniques to increase the performance of positional uncertainty computation of marching cubes level-sets [LC87].

Probabilistic marching cubes (PMC) [PWH11], devised by Pöthkow et al., is the first probabilistic method to understand how uncertainty in data is propagated into level-set positions. In particular, the authors modeled uncertainty in data as probability distributions to derive the probability distribution of level-sets. In their work, it was shown that uncertainty computations with the multivariate Gaussian distribution assumption are more accurate than independent uniform/Gaussian distribution models because they can capture the correlation among dimensions. However, the higher accuracy of the multivariate Gaussian model comes at the cost of computational overhead resulting from the use of expensive Monte Carlo (MC) sampling. MC sampling is required because multivariate Gaussian cannot be integrated in closed form. For example, as presented in the PMC paper, the computation of level-set uncertainty takes 194 minutes for a data grid of size 432 × 219 × 68 with eight MC samples per grid cell. Such high computational times cannot be accommodated in visualization tools.

A few recent advances have investigated the fast computation of PMC. Han et al. [HAPJ22] proposed a novel deep learning network for fast prediction of level-set uncertainty for a multivariate Gaussian noise model. This approach, however, is limited to time-varying ensembles of 2D datasets and involves significant training time specific to a dataset before efficiently predicting level-set uncertainty. Wang et al. [WAM+23] recently developed a parallel filter FunMC² that can run on multicore CPUs and GPUs for fast
computations of level-set uncertainty. The filter, however, still resorts to traditional MC sampling per compute thread to account for data correlation and compute uncertainty. In this work, we propose new data-driven optimizations that can improve the speed of level-set uncertainty computations on a single-core CPU. Our methods produce results that have negligible differences compared with the original PMC algorithm. Furthermore, we integrate our techniques with the FunMC filter to achieve further speedup on multicore devices.

Contributions: We propose two novel data-driven techniques for the efficient computation of level-set uncertainty.

- In our first technique, we extract low-rank structures of a multivariate Gaussian distribution to reduce the amount of MC sampling, thereby speeding up computations. Specifically, we employ the eigendecomposition to extract low-rank structures that carry important information followed by their sampling.
- In our second technique, we leverage the Pearson’s correlation coefficient \([\text{Kir}08]\) to estimate the level of data correlation and adaptively choose a probabilistic model. Such adaptive model decision reduces the amount of MC sampling and subsequently the time of uncertainty computation.
- We evaluate the performance vs. accuracy of our techniques for uncertain 2D and 3D datasets. We observed the maximum speedups up to 3.38x on synthetic dataset and 2.30x on real datasets with negligible differences compared with the original PMC [PWH11] results.

2. Background

2.1. Marching Cubes

Marching cubes [LC87] is a workhorse algorithm used in many visualization tools for depiction of level-sets of 2D or 3D scalar functions. If \( f : \mathbb{R}^3 \rightarrow \mathbb{R} \) is a scalar function, then the level-set or isosurface \( \mathcal{L} \) of a scalar field \( f \) for the isovalue \( k \) corresponds to a preimage of function value \( k \). Mathematically, \( \mathcal{L}(k) \equiv \{ P : f(P) = k \} \), where \( P \) denotes the domain positions. Assuming that \( f \) is sampled on a 3D uniform grid, the marching cubes extracts a triangular mesh corresponding to an isosurface with isovalue \( k \). Consider a single grid cell with fixed data values at eight cell vertices denoted by \( d = (d_0, \ldots, d_7)^T \). If \( d_i \geq k \) (or \( \leq k \)) \( \forall i \), then the level-set is inferred not to cross the cell; otherwise, the level-crossing takes place (i.e., the level-set crosses a grid cell). If level-crossing takes place, then the level-set topology within a cell is determined based on the marching cubes topology cases [LC87].

2.2. Probabilistic Marching Cubes

Uncertain grid vertex data create ambiguity regarding the level-crossing event. The PMC addresses such ambiguity by estimating the level-crossing probability (LCP) for a grid cell. In particular, LCP represents the probability of the level-set to cross a grid cell for isovalue \( k \). Although LCP computed with the multivariate Gaussian model is more accurate than the independent uniform/Gaussian model because of its ability to capture data correlation [PWH11], there is no closed-form solution for the multivariate Gaussian model. Thus, MC sampling is the only practical solution in the case of multivariate Gaussian to compute LCP. Let \( D = (D_0, \ldots, D_7)^T \) denote a random variable representing uncertain data at the 3D cell vertices. The random variable is assumed to have a multivariate Gaussian distribution with the sample mean \( \overline{\mu} = (\bar{\mu}_0, \ldots, \bar{\mu}_7)^T \) and sample covariance matrix \( \Sigma = \mathbb{E}[(D - \overline{\mu})(D - \overline{\mu})^T] \). If \( S \) MC samples are then drawn from the distribution \( \mathcal{N}(\overline{\mu}, \Sigma) \) and if the level-set with isovalue \( k \) crosses a cell for \( C \) number of samples then the LCP is computed as \( \frac{S}{C} \).

3. Methods

In this paper, we propose efficient data-driven solutions to optimize the PMC performance. Data-driven solutions (e.g., data histograms) have been proposed in several prior works [BDL+21, ZJW21, ZWZ+19] for feature preservation and optimization. Below, we describe our eigenvalue decomposition and adaptive probability model techniques for performance optimization of the PMC algorithm [PWH11] through a reduced level of MC sampling.

3.1. Eigenvalue Decomposition for Sampling

The original PMC technique (Sec. 2.2) samples 8D multivariate distribution \( \mathcal{N}(\overline{\mu}, \Sigma) \) space for computing the level-set uncertainty (i.e., LCP) in 3D datasets. We leverage a low-dimensional equivalent of a multivariate space for reducing the amount of MC sampling and subsequently enhancing the efficiency of LCP computation. Specifically, we utilize the eigenvalue decomposition to extract low-dimensional structures and perform sampling.

In an \( n \)-dimensional grid cell, by diagonalizing the covariance matrix \( \Sigma \) into \( \Gamma \Lambda \Gamma^T \) using the eigenvalue decomposition, uncertain data \( D \sim \mathcal{N}(\overline{\mu}, \Sigma) \) can be rewritten as \( D \sim \overline{\mu} + \Gamma \Lambda^{\frac{1}{2}} Z \), where \( \overline{\mu} \in \mathbb{R}^n \) is the sample mean, \( \Lambda \in \mathbb{R}^{n \times n} \) denotes eigenvectors, and \( Z \sim \mathcal{N}(0, I_n) \) represents the data with standard normal distribution. \( I_n \) denotes the \( n \)-dimensional identity matrix. The matrix form of multivariate Gaussian data can be expressed as follows:

\[
\begin{bmatrix}
D_1 \\
\vdots \\
D_n
\end{bmatrix} = \begin{bmatrix}
\mu_1 \\
\vdots \\
\mu_n
\end{bmatrix} + \begin{bmatrix}
\tilde{V}_{11} & \cdots & \tilde{V}_{1n} \\
\vdots & \ddots & \vdots \\
\tilde{V}_{n1} & \cdots & \tilde{V}_{nn}
\end{bmatrix} \begin{bmatrix}
\lambda_1^{\frac{1}{2}} \\
\vdots \\
\lambda_n^{\frac{1}{2}}
\end{bmatrix} \begin{bmatrix}
Z_1 \\
\vdots \\
Z_n
\end{bmatrix} \tag{1}
\]

In the above form, each column of \( \Gamma \) denotes an eigenvector. The eigenvectors are sorted by their eigenvalues such that \( \lambda_1 > \lambda_2 > \cdots > \lambda_n \), indicated as a diagonal matrix. Thus, a random variable \( D_j \) denoting uncertain data at the \( j \)th vertex can be represented as follows:

\[
D_j = \mu_j + \sum_{i=1}^{m} \tilde{V}_{ij} \lambda_i^{0.5} Z_i \tag{2}
\]

The key idea of this study is to use the significant eigenvalues in Equation 2 to reduce the amount of MC sampling and enhance the speed of uncertainty computations. For a given \( n \)-dimensional cell, if only \( m \) dimensions have significant eigenvalues, then we can rewrite Equation 2 with the following approximate equation because multiplication with small eigenvalues corresponds to a minor contribution to the sum in Equation 2:

\[
D_j \approx \mu_j + \sum_{i=1}^{m} \tilde{V}_{ij} \lambda_i^{0.5} Z_i \tag{3}
\]

The original PMC algorithm involves drawing \( n \) MC samples of a
random variable $Z$ (Equation 2), whereas the proposed eigenvalue-based approach (Equation 3) requires drawing $m$ MC samples per cell. If $m \ll n$, then Equation 3 can provide a significant speed-up while producing solution close to the original PMC. For the guidance and results for significant eigenvalue decisions, please refer to the supplementary material. Note that choosing only nonzero eigenvalues provides the same solution as the original PMC method except for the errors from random number generation.

### 3.2. Adaptive Probability Model

The multivariate Gaussian model ([PWH11] and Sec. 3.1) can lead to unnecessary computations if the underlying data are independent or have a weak correlation. We therefore propose an adaptive strategy that selects between the multivariate Gaussian and independent models for higher efficiency of uncertainty computations. In our adaptive strategy, we compute the Pearson’s correlation coefficient [Kir08] to test the level of data independence within a grid cell. In particular, for an $n$-dimensional grid cell, we compute the Pearson’s correlation coefficient for each pair of dimensions followed by the maximum among computed correlation values, $\rho_{\text{max}}$. If $\rho_{\text{max}}$ attains smaller values, then we predict data to be independent and use the closed-form independent Gaussian model [PH11, ASI21]; otherwise, we use our proposed eigenvalue-based model (Sec. 3.1) for LCP computation. The utilization of the closed-form independent model avoids MC sampling, thereby providing a speedup. We note that computing pairwise correlation can quickly become expensive with an increase in $n$. For example, for 2D data, the number of pairwise correlations per grid cell is 6, but for 3D data, it is 28. Thus, we compute the $\rho_{\text{max}}$ array offline, which is independent of the isovalue. This precomputed $\rho_{\text{max}}$ array is then used for adaptive model decisions.

### 4. Results

First, we demonstrate efficient computation of level-set uncertainty (i.e., LCP) through experiments on a synthetic Ackley function [Ack87] using our proposed techniques (Sec. 3.1 and 3.2). For our experiments, we generate two ensemble datasets by adding independent and correlated noise to the Ackley function. For the independent noise model, we randomly generate samples $r \sim [-0.1, 0.1]$ and add them to the Ackley function. For the correlated noise model, we draw a random number $r \sim [0.9, 1.1]$ and generate an ensemble member by multiplying the Ackley function by $r$.

Figure 1a visualizes the results of the proposed eigenvalue decomposition technique (Sec. 3.1) applied to the correlated ensemble dataset and their comparison with the original PMC [PWH11] results. For both eigenvalue-based and PMC techniques, 2,000 MC samples are drawn per grid cell. As expected for the correlated ensemble, only one eigenvalue ($\lambda_1$) is dominant, and the maximum Pearson’s correlation coefficient ($\rho_{\text{max}}$) is strong (yellow regions in the correlation map). The $\lambda_1$ value in Figure 1a denotes the sum of the top eigenvalues across all grid cells. All other eigenvalues are calculated in a similar fashion for all results in the paper. For computing LCP, we utilize the nonzero eigenvalues (Equation 3) per grid cell, which provides a speedup of $2.60 \times$. The root mean squared error (RMSE) and maximum error (Max error) between the results of our eigenvalue-based method and the original PMC are 0.007 and 0.048, respectively. These errors are about the same as the RMSE and Max_error between the two runs of MC sampling in the original PMC. Thus, the errors are within the tolerance of random number generation, which confirms that our eigenvalue-based method produces the same solution as the original PMC. The speedup of $2.60 \times$ is observed because only one dimension (with the largest eigenvalue $\lambda_1$) is sampled for most grid cells as opposed to MC sampling of all four dimensions in the original PMC.

Figure 1b visualizes the results for the adaptive probability model technique (Sec. 3.2) applied to the independent ensemble dataset. The data independence test is performed using the $\rho_{\text{max}}$ test. The precomputation time for the $\rho_{\text{max}}$ is 0.56 seconds on a serial i7 Intel processor, which is fast because of the low-resolution 2D grid. As observed in Figure 1b, $\rho_{\text{max}}$ attains smaller values for each grid cell because of the independence among members. For
the results shown, if $\rho_{\text{max}} \leq 0.2$, then we predict that the data are independent. We chose this threshold 0.2 based on a few previous studies [Ako18, SBS18], but we show variation in speed and accuracy with a change in the threshold in the supplementary material. If the independence test is satisfied, then we use the fast closed-form model for LCP computation [PH11, ASJ21]; otherwise, we utilize our eigenvalue-based approach with MC sampling of dimensions with nonzero eigenvalues (Sec. 3.1). We observe a speedup of $3.38 \times$ with negligible differences (i.e., RMSE, maximum error, and difference visualization) with respect to the original PMC. This speedup is attributed to the adaptive usage of the fast independent model that eliminates the need for sampling of a distribution.

Next, we present the results for real datasets. Figure 2a–e visualizes the results for the velocity magnitude fields of the wind ensemble [Vit17] with 15 members each with a $68 \times 68$ grid resolution. The proposed eigenvalue decomposition approach with nonzero eigenvalues (Figure 2b) provides a $2.30 \times$ speedup compared with the original PMC (Figure 2a), and minor differences (RMSE, maximum error, and difference visualization) in the LCP computation arise from the random number generation, similar to the Ackley dataset. The postanalysis of the ensemble recovers a significant correlation among members, as evident through one dominant eigenvalue ($\lambda_1$) and strong $\rho_{\text{max}}$ Values in Figure 2d–e. The presence of only one dominant eigenvalue significantly reduces the amount of MC sampling (with $m = 1$ for Equation 3 in most cases) compared with the original PMC. The adaptive method in Figure 2c with an independence prediction based on $\rho_{\text{max}} < 0.2$ provides about the same speedup as the eigenvalue-based method because the fast independent model is rarely used due to a weak independence in data. The precomputation of a $\rho_{\text{max}}$ is fast (0.96 seconds) because of the low-resolution 2D grid. We present further analysis regarding the sensitivity of results to thresholds for significant eigenvalues and $\rho_{\text{max}}$ in the supplementary material.

Figure 2f–j depicts the results for the beetle dataset [GGK05]. Specifically, we partition the original data with spatial resolution $832 \times 832 \times 494$ into blocks of size $4 \times 4 \times 4$ and summarize the block uncertainty by using a multivariate Gaussian [TLB*11]. Because this dataset is large, we utilize the FunMC2 filter implementation with the VTK-m library [WAM*23, Mea16] for accelerating the performance. In particular, we run the original PMC and our proposed methods on Oak Ridge National Laboratory’s Frontier supercomputer, which is equipped with AMD GPUs [Fro]. The proposed eigenvalue-based approach with nonzero eigenvalues took 16.72 seconds (Figure 2g), whereas the original PMC (Figure 2f) took 27.18 seconds for 500 MC samples. A $1.61 \times$ speedup is observed because the existence of four dominant eigenvalues ($\lambda_1, \ldots, \lambda_4$) reduces the amount of MC sampling (with $m = 4$ for Equation 3 in most cases) compared with the original PMC that samples all eight dimensions. There are again minor differences (RMSE, maximum error, and difference visualization) in the LCP computation with respect to the original PMC that arise from the random number generation, similar to the Ackley and wind datasets.

The adaptive method in Figure 2h provides about the same speedup as the eigenvalue-based method because the data exhibit weak independence (similar to the wind dataset). The precomputation time for the correlation map is 94.6 minutes on a serial i7 Intel processor because of the 28 correlation computations per
References


