

**A Multi-Phase Framework for Vector Field Sampling Using FTLE, Context, and Gradient
Features**

by

Meghanto Majumder

A thesis accepted and approved in partial fulfillment of the
requirements for the degree of

Master of Science
in Computer Science

Thesis Committee:

Hank Childs, Chair

Brittany Erickson, Member

Jee Choi, Member

University of Oregon

Spring 2025

© 2025 Meghanto Majumder

All Rights Reserved

THESIS ABSTRACT

Meghanto Majumder

Master of Science in Computer Science

Title: A Multi-Phase Framework for Vector Field Sampling Using FTLE, Context, and Gradient Features

Computational simulations on high-end supercomputers frequently cannot save high-resolution data to disk for post-hoc analysis, due to recent trends where overall compute power has accelerated much more quickly than I/O bandwidth. Data-driven intelligent sampling addresses this challenge, by helping to capture and analyze features for post-hoc analysis without requiring prior knowledge about data features. Recent work has demonstrated that data-driven sampling approaches can significantly outperform random sampling for large-scale scalar field visualization and analysis tasks. However, vector field sampling for such scientific simulations presents fundamental challenges in preserving both local accuracy and global flow structure.

We introduce a novel multi-phase sampling framework that combines finite-time Lyapunov exponent analysis, structured spatial context preservation, and gradient-based feature detection. Our approach outperforms random sampling and existing data-driven scalar sampling techniques by significantly improving streamline and pointwise angle accuracy, as demonstrated on six diverse scientific data sets at sampling rates of 1% to 5%.

This thesis includes previously unpublished coauthored material.

Table of Contents

1	Introduction	8
2	Related Work	11
2.1	Sampling-Based Data Analysis and Visualization	11
2.2	Large-Scale Data Reduction and Visualization	12
2.3	Subsampling Techniques for Vector Fields	13
3	Methodology	14
3.1	Fundamental Considerations in Vector Field Sampling	14
3.2	Multi-Stage Analysis Framework	15
3.2.1	Feature Identification Through FTLE Analysis	16
3.2.2	Structured Neighborhood Sampling	16
3.2.3	Local Structure Characterization Through Gradient Analysis	17
3.2.4	Integrated Multi-Phase Sampling Algorithm	17
3.2.5	Computational Characteristics and Scalability	19
4	Experimental Overview	21
4.1	Data Set Selection	21
4.2	Methods Compared	22
4.3	Evaluation Protocol	23
4.4	Sampling Configuration	24

4.5	Comparative Analysis	24
5	Results	25
5.1	Limitations of Scalar Field Techniques	25
5.2	Multi-Phase Sampling Performance	26
5.3	Comparative Analysis of Methods	27
5.4	Sampling Rate Effects	28
5.5	Visual Analysis of Streamline Reconstruction	30
6	Discussion and Parameter Selection	34
6.1	FTLE Integration Time Selection	34
6.2	Histogram Configuration	35
6.3	Budget Allocation Optimization	35
6.4	Alternative Approaches Explored	36
7	Conclusion	38

List of Tables

5.1	Relative Angular Accuracy Improvement Over Random Sampling	32
5.2	Relative Streamline Accuracy Improvement Over Random Sampling	33

List of Figures

4.1	Original streamlines for the six datasets used in the study	22
5.1	Performance evaluation of scalar field sampling adaptations	27
5.2	Performance comparison of sampling methods across sampling rates	29
5.3	Comparison of streamline visualizations for Asteroid Impact dataset	31

Chapter 1

Introduction

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

The exponential growth in computational power has revolutionized scientific discovery through large-scale simulations. Modern supercomputers generate petabyte-scale data sets with unprecedented spatiotemporal resolution, offering insights into complex phenomena across disciplines. However, this computational advancement has created a critical challenge: while processing power continues to scale, I/O capabilities lag significantly behind. This disparity has rendered traditional post hoc workflows, where simulation data is saved to permanent storage for later analysis, increasingly impractical.

Recent work has demonstrated that data-driven sampling approaches can significantly outperform random sampling for large-scale scalar field visualization [4]. These methods, which combine value-based importance with local smoothness metrics, have shown promise for in situ data reduction. However, extending these techniques to vector fields—ubiquitous in scientific simulations

from fluid dynamics to astrophysics—presents several fundamental challenges.

First, while scalar fields have clear value-based importance metrics that can identify rare or important values, the extension to vector fields is non-trivial. Traditional scalar importance metrics, such as sampling a vector field by assigning importance based on vector magnitude or direction rarity may fail to capture the complex spatial and temporal relationships inherent in vector fields. Vector fields contain critical directional information that cannot be captured by magnitude alone, requiring simultaneous preservation of both local directional changes and global flow behavior.

Second, the extension of smoothness-based sampling to vector fields requires careful consideration. In scalar fields, the gradient magnitude provides a clear measure of local variation. However, vector field gradients are represented by 3×3 Jacobian matrices, making it challenging to distill this rich directional information into meaningful sampling criteria.

Third, the evaluation of sampling quality for vector fields demands different approaches than those used for scalar fields. Traditional metrics like Signal-to-Noise Ratio (SNR) or correlation coefficients fail to capture important vector field characteristics and flow behavior preservation.

Unlike existing approaches that treat vector fields as collections of independent scalar components, our method specifically addresses the interdependent nature of vector components and their spatial relationships. Our approach addresses these challenges through a novel sampling strategy that combines:

1. Finite-Time Lyapunov Exponent (FTLE)-based feature identification and sampling to capture regions of significant flow separation
2. Context-aware neighborhood sampling around high FTLE regions
3. Gradient-based sampling using vector field Jacobians

We evaluate this approach across six scientific data sets, including asteroid impacts, atmospheric phenomena, and fluid dynamics simulations. Using streamline reconstruction accuracy and vector angle preservation as our primary metrics, we demonstrate consistent improvement over random sampling and adaptations of existing scalar field techniques. Our evaluations demonstrate consistent improvements in both local and global accuracy metrics, with relative improvements

in angular accuracy and streamline distance ranging from 1.02 to 1.39 across diverse scientific datasets. This approach enables scientists to reduce vector field storage requirements while maintaining the ability to accurately reconstruct critical flow features for post-hoc analysis.

Chapter 2

Related Work

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

2.1 Sampling-Based Data Analysis and Visualization

The challenges posed by large-scale data sets have led the visualization community to develop various data sampling techniques aimed at reducing the data size for timely analysis and visualization. Park et al. proposed a visualization-aware sampling technique optimized for specific visualization techniques such as scatter plots and map plots [19]. Nguyen and Song addressed simple random sampling limitations by proposing a centrality clustering-based data sampling scheme [16]. Woodring et al. introduced a stratified random sampling-based scheme for summarizing cosmology simulations, enabling interactive post-hoc visualization[33]. Wei et al. extended this approach by incorporating bitmap indexing and information-theoretic measures, creating in situ compressed subsampled data sets [30]. Information theoretic measures such as information entropy have also

been leveraged by researchers for effective data subsampling. Building on information theory, Biswas et al. presented a scheme for in situ sampling of large-scale data sets, preserving important features through probability-based importance assignment [3]. Rapp et al. contributed a sampling approach for scattered data sets, that identified a subset of points in order to preserve the blue noise properties of the original data [20]. Biswas et al. extended their work on in situ sampling by incorporating local smoothness of the data during the importance assignment phase [4]. Our proposed methods extend the work of Biswas et al., working on regular grid data sets and intending to capture the feature regions of vector fields at very low sampling rates.

2.2 Large-Scale Data Reduction and Visualization

With the ever-increasing size of simulation data, scientists have explored techniques for data reduction to facilitate interactive analysis and visualization. In situ visualization infrastructure integrated into existing frameworks enables direct visualization during simulation execution[9, 15, 27, 32]. However, certain data analysis and visualization tasks require post hoc visualization owing to time and resource constraints. Dutta et al. addressed this by developing in situ to post-hoc-capable flexible data summarization techniques, employing Gaussian mixture model-based data representations[6, 7]. Wang et al. extended data reduction efforts by incorporating spatial distributions for accurate data recovery[28, 29]. In situ generated histograms, as utilized by Ye et al., use an accelerated post hoc query-based visual analysis[34]. Cinema and Tikhonova’s image-based approaches represent emerging techniques for efficient post hoc visualization [2, 26]. Lakshminarasimhan et al.’s ISABELA focused on in situ sort-and-B-spline error-bounded lossy abatement for scientific data [13]. This work, in contrast to data-modeling-based approaches, emphasizes the reduction of data sets by preserving a small subset of representative data samples. The subsampled data serve as a proxy for full-resolution raw data, facilitating the direct performance of various visualization tasks without additional post hoc processing.

2.3 Subsampling Techniques for Vector Fields

There have been many vector field analysis and compression techniques, and they can be considered based on their methodologies and applications. There have been approaches that explore the effectiveness of Lagrangian representations, as demonstrated by Sane et al., who introduced an approach for in situ reduction in cosmology and seismology simulations [21]. Another approach focuses on radial basis function (RBF) approximation, as seen in Smolik et al. and Smolik et al., demonstrating the suitability of RBF for preserving critical points and handling large scattered vector fields [22, 23]. The third category involves topological simplification, with papers by Weinkauff et al., Lodha et al., and Theisel et al. addressing the extraction of higher-order critical points, hierarchical refinement for 2D vector fields, and the combination of topological simplification with topology-preserving compression [14, 25, 31]. Key papers have concentrated on compression techniques, including subsampling-based methods by Agranovsky et al., and detail-preserving strategies for smoke-based flow visualization by Yuan et al. [1, 35]. Finally, papers like Koch et al. and Dey et al. delve into novel vector field approximation techniques using linear neighborhoods and Delaunay triangulations, respectively [5, 12]. In contrast, our study aims to develop and evaluate data-driven importance-based sampling methods specifically designed for vector fields.

Chapter 3

Methodology

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

3.1 Fundamental Considerations in Vector Field Sampling

The sampling of vector fields presents distinct methodological challenges that require a more sophisticated approach than scalar field sampling. While scalar field sampling techniques could be applied for vector fields using metrics such as vector direction or magnitude distributions, such simplifications risk losing crucial information about how vectors interact and evolve across space and time. In vector fields, the significance of any given vector cannot be determined by its components alone - the same vector can be part of vastly different flow behaviors depending on its surrounding context. This spatial dependency manifests primarily in two ways: through long-term trajectory behavior that reveals how flow regions evolve and separate over time, and through the rate at which vector directions change across space. The long-term behavior itself cannot be fully understood

without capturing the local context around significant regions, since flow evolution depends on the interaction between neighboring vectors. Each of these aspects captures different characteristics of the underlying field structure, and missing any one of them could result in an incomplete representation of the field’s behavior. Moreover, features evident through one type of analysis might be entirely invisible through another, making any single approach insufficient for comprehensive field understanding. This inherent multi-faceted nature of vector field features, combined with the context-dependent significance of individual vectors, demands a sampling strategy that can:

1. Identify regions of significant dynamic behavior
2. Preserve the spatial context necessary for understanding flow features
3. Capture areas of rapid local variation in the field

These requirements - identifying dynamically significant regions, preserving spatial context, and capturing rapid local variation - naturally lead to a multi-stage sampling framework that addresses each aspect systematically.

3.2 Multi-Stage Analysis Framework

Our framework addresses the multi-faceted nature of vector field analysis through an integrated, multi-stage approach. Rather than treating feature identification, context preservation, and local structure analysis as separate concerns, we develop a methodology that considers these components in relation to each other. This approach combines FTLE-based feature analysis with structured spatial sampling to investigate both local and global field characteristics. This method is then combined with the smoothness-based importance sampling outlined by Biswas et al [4]. Each stage of the framework builds upon information gathered in previous stages, aiming to satisfy the requirements outlined above.

3.2.1 Feature Identification Through FTLE Analysis

We extend the value-based importance sampling framework of Biswas et al. [4] to the vector field context through FTLE computation. For a normalized vector field $\mathbf{v}(\mathbf{x})$, we compute the FTLE field $\sigma(\mathbf{x})$ through analysis of the flow map gradient tensor:

$$\sigma(\mathbf{x}) = \frac{1}{T} \ln \sqrt{\lambda_{\max}(\mathbf{M}(\mathbf{x}))}$$

where $\mathbf{M}(\mathbf{x})$ represents the Cauchy-Green deformation tensor computed over integration time T .

The FTLE field provides a scalar measure that quantifies trajectory separation, enabling the application of value-based importance sampling. We achieve approximately uniform coverage across the range of observed FTLE values, ensuring representation of diverse separation behaviors. This approach maintains the statistical rigor of the original method while adapting it to capture vector field dynamics.

3.2.2 Structured Neighborhood Sampling

To preserve the spatial context when sampling vector field features, we develop a structured neighborhood sampling strategy. For each point $\mathbf{p}$ identified through FTLE-based sampling, we select additional samples along the six cardinal directions $(\pm x, \pm y, \pm z)$ at a fixed radius r :

$$\mathbf{p}_i = \mathbf{p} + r\mathbf{e}_i$$

where $\mathbf{e}_i$ represents the unit vectors in the cardinal directions. In cases where proposed neighborhood points coincide with previously selected samples, we employ random selection from remaining unsampled points in the vicinity, maintaining the sampling budget allocated. This structured approach using cardinal directions provides consistent spatial context while avoiding the computational overhead and potential bias of more complex neighborhood sampling patterns.

3.2.3 Local Structure Characterization Through Gradient Analysis

To identify regions of significant directional variation in the vector field, we analyze the local gradient structure. The Jacobian matrix $\mathbf{J}(\mathbf{x})$ of the vector field provides a measure of local directional change along the three axes. We compute its Frobenius norm to obtain a scalar measure of local structure:

$$L(\mathbf{x}) = \|\mathbf{J}(\mathbf{x})\|_F = \sqrt{\sum_{i,j} |J_{ij}|^2}$$

This measure quantifies the magnitude of directional change in the vector field, with higher values indicating regions of lower smoothness. Our sampling strategy prioritizes regions where $L(\mathbf{x})$ is large, making it more likely to select regions where the vector field exhibits rapid directional variation. This approach ensures detailed representation of field features that manifest as discontinuities or sharp transitions, while allowing sparser sampling in regions of relative uniformity.

The combination of this gradient-based analysis with our FTLE-based feature identification creates a complementary sampling strategy. While FTLE analysis captures longer-term trajectory behavior and separation dynamics, the gradient-based approach ensures representation of immediate local structure. By incorporating both measures into our sampling framework, we address the multi-scale nature of vector field features - from instantaneous directional changes to evolving flow patterns.

3.2.4 Integrated Multi-Phase Sampling Algorithm

The algorithm for our multi-phase sampling framework implements a hierarchical selection strategy that addresses the different aspects of vector field representation. Algorithm 1 formalizes this process in a way that maintains the strict sampling budgets while integrating the previously described theoretical components.

The sampling budget allocation, governed by coefficients α , β , and γ where $\alpha + \beta + \gamma =$

Algorithm 1 Multi-Phase Vector Field Sampling

Require: Vector field $\mathcal{V}$, total points N , sampling ratio η , coefficients α, β, γ , neighbor radius r

Ensure: Set of sampled points $\mathcal{S}$

```
1:  $\mathcal{S} \leftarrow \emptyset$  ▷ Initialize empty sample set
2:  $M_1 \leftarrow N\eta\alpha$  ▷ Points for FTLE sampling
3:  $M_2 \leftarrow N\eta\beta$  ▷ Points for neighborhood sampling
4:  $M_3 \leftarrow N\eta\gamma$  ▷ Points for gradient sampling
5: // Phase 1: FTLE-based Feature Sampling
6:  $\sigma \leftarrow \text{COMPUTEFTLE}(\mathcal{V})$ 
7:  $H_1 \leftarrow \text{ValueBasedImportanceSampling}(\sigma, M_1)$ 
8:  $\mathcal{S}_1 \leftarrow \text{PROBABILISTICSELECT}(H_1, M_1)$ 
9:  $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{S}_1$ 
10: // Phase 2: Structured Neighborhood Sampling
11:  $H_\sigma \leftarrow \text{COMPUTEHILOGRAM}(\sigma)$ 
12:  $\text{ftle}_{\text{threshold}} \leftarrow \text{FINDTHRESHOLD}(H_\sigma, M_2/6)$ 
13: for  $\mathbf{p} \in \mathcal{S}_1$  do
14:   if  $\sigma(\mathbf{p}) \geq \text{ftle}_{\text{threshold}}$  then
15:     for  $\mathbf{d} \in \{\pm\mathbf{x}, \pm\mathbf{y}, \pm\mathbf{z}\}$  do
16:        $\mathbf{q} \leftarrow \mathbf{p} + r\mathbf{d}$ 
17:       if  $\mathbf{q} \notin \mathcal{S}$  then
18:          $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathbf{q}\}$ 
19:       else
20:          $\mathbf{q}' \leftarrow \text{RANDOMUNPICKEDNEIGHBOR}(\mathbf{p}, r)$ 
21:          $\mathcal{S} \leftarrow \mathcal{S} \cup \{\mathbf{q}'\}$ 
22:       end if
23:     end for
24:   end if
25: end for
26: // Phase 3: Gradient-based Sampling
27:  $\mathbf{J} \leftarrow \text{COMPUTEJACOBIAN}(\mathcal{V})$ 
28:  $L \leftarrow \text{FROBENIUSNORM}(\mathbf{J})$ 
29:  $\Omega_{\text{remain}} \leftarrow \{\mathbf{x} \in \Omega : \mathbf{x} \notin \mathcal{S}\}$  ▷ Unpicked points
30:  $H_2 \leftarrow \text{SMOOTHNESSBASEDIMPORTANCESAMPLING}(L|_{\Omega_{\text{remain}}}, M_3)$ 
31:  $\mathcal{S}_3 \leftarrow \text{PROBABILISTICSELECT}(H_2, M_3)$ 
32:  $\mathcal{S} \leftarrow \mathcal{S} \cup \mathcal{S}_3$ 
33: return  $\mathcal{S}$ 
```

1, establishes precise bounds on the contribution of each sampling criterion. The FTLE-based importance sampling operates on the initial budget of $M_1 = N\eta\alpha$ points, where the importance function construction inherently prioritizes regions of significant flow separation. `ValueBasedImportanceSampling` refers to the algorithm described in Biswas et al [4], and uses the FTLE

field as the input.

For the next phase, the algorithm dynamically computes $\text{ftle}_{\text{threshold}}$ based on the statistical distribution of FTLE values, ensuring that the neighborhood sampling budget $M_2 = N\eta\beta$ is never exceeded. The threshold FTLE value is the cutoff point, above which the neighbors will be selected.

The smoothness-based sampling phase restricts the sampling domain to Ω_{remain} , the complement of previously selected points. This constraint ensures that the final $M_3 = N\eta\gamma$ points effectively capture additional structural information not already represented in the feature-based or neighborhood samples. `SmoothnessBasedImportanceSampling` is similar in nature to the Smoothness-based importance sampling in Biswas et al [4], however the smoothness field is calculated as described above.

This algorithmic formulation provides a framework for implementing the theoretical components described earlier while maintaining strict control over the number of sampled points through explicit budget allocation. The resulting sampling set $\mathcal{S}$ thus satisfies the three components of vector field sampling identified earlier.

Our implementation employs empirically determined coefficients for budget allocation, with $\alpha = 0.1$ devoted to FTLE-based feature sampling, $\beta = 0.4$ for structured neighborhood sampling, and $\gamma = 0.5$ for smoothness-based sampling, and $r = 5 \times \text{grid cell size}$. The FTLE computation integrates trajectories over 20 time steps, providing sufficient temporal extent to identify meaningful flow separation while maintaining computational efficiency suitable for in situ deployment. These parameters were determined through extensive experimentation across our test data sets, optimizing the balance between feature preservation and computational overhead.

3.2.5 Computational Characteristics and Scalability

Our framework’s computational profile is characterized by linear-time operations and inherent parallelizability. Each phase - FTLE computation, neighborhood selection, and gradient analysis - scales linearly with input size and can be executed independently across spatial domains leading to

high parallelizability. The FTLE computation, while potentially intensive, is computed once using a low number of steps (20) and reused across sampling phases. The histogram-based importance functions require $O(n)$ operations for construction and $O(1)$ for lookups. Neighborhood sampling introduces minimal overhead, and can also be parallelized. Hence, the approaches considered are all suitable for in-situ deployment, where leveraging the computational power is crucial.

Chapter 4

Experimental Overview

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

Our evaluation methodology employs rigorous analysis of both streamline accuracy and vector field reconstruction quality across diverse scientific data sets. We assess performance through carefully designed comparative metrics while ensuring reproducibility through multiple trial runs.

4.1 Data Set Selection

We validate our approach across six scientific data sets representing different domains and flow characteristics:

1. Asteroid Impact: Water impact simulation, velocity field (100×100×100) [11]
2. Hurricane Isabel: Atmospheric simulation, velocity field (250×250×50) [17]
3. CM1 Tornado: Atmospheric simulation, vorticity field (176×222×100) [18]

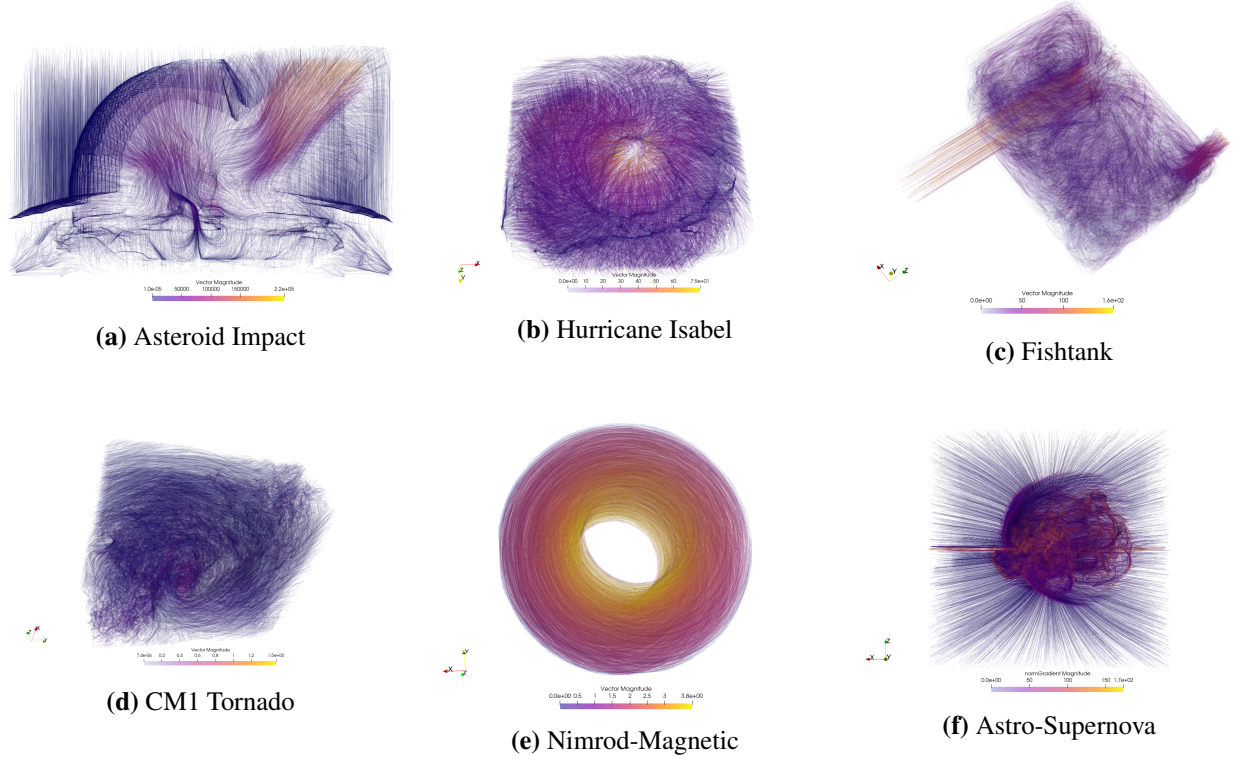


Figure 4.1: Original streamlines for the six datasets used in the study: (a) Asteroid Impact, (b) Hurricane Isabel, (c) Fishtank, (d) CM1 Tornado, (e) Nimrod-Magnetic, and (f) Astro-Supernova.

4. Astro-Supernova: Astrophysical simulation, velocity field ($128 \times 128 \times 128$) [8]
5. Nimrod-Magnetic: Plasma physics simulation, magnetic field ($150 \times 150 \times 150$) [24]
6. Fishtank: Fluid dynamics simulation, velocity field ($100 \times 100 \times 100$)

[10]

Fig. 4.1 shows a visualization of streamlines generated from these data sets.

4.2 Methods Compared

We evaluate our multi-phase approach against several baseline and component methods:

1. Random Sampling: Traditional uniform random sampling serving as the primary baseline.
2. Component Methods:

1. Value-based FTLE Sampling: Uses only FTLE field values for importance-based sampling
 2. Smoothness-based Sampling: Uses only Frobenius norm of the Jacobian for sampling
 3. Value-based FTLE + Neighborhood Sampling: Picks 20% of samples using Value-based FTLE sampling, and the rest using structured neighborhood sampling from high FTLE values
 4. Value-based FTLE + Smoothness Sampling:
Equal weighting between FTLE and smoothness-based sampling
3. Our Multi-phase Method:
- 10% budget for FTLE-based feature sampling
 - 40% budget for structured neighborhood sampling for high FTLE values
 - 50% budget for smoothness-based sampling

By testing each component independently (FTLE Only, Smoothness Only) and in strategic combinations (FTLE + Context, FTLE + Smoothness), we can isolate the impact of each sampling strategy.

4.3 Evaluation Protocol

For each data set, we employ two primary evaluation strategies:

1. Streamline-Based Analysis:
 - Generate 10,000 uniformly seeded streamlines
 - Trace each streamline for 200 integration steps
 - Compare streamlines between original and reconstructed fields using average pairwise distances
 - Compute relative improvement over random sampling baseline

2. Vector Field Reconstruction:

- Create Delaunay triangulation from sampled points
- Perform linear interpolation within tetrahedra
- Calculate angular differences between original and reconstructed vectors
- Express results as relative improvement over random sampling

4.4 Sampling Configuration

We evaluate our method using three sampling fractions: - 0.01 (1% of original points) - 0.03 (3% of original points) - 0.05 (5% of original points)

To ensure robust and reproducible results, each configuration (method-dataset-sampling fraction combination) is evaluated using 5 different random seeds. This approach mitigates potential bias from specific random number sequences and enables statistical analysis of performance variability.

4.5 Comparative Analysis

Our primary baseline for comparison is random sampling at equivalent sampling fractions. For each metric (streamline distance and angular difference), we compute the relative improvement:

$$\text{Relative Distance} = \frac{\text{Random Distance}}{\text{Method Distance}}$$

This relative measure provides a clear indication of our method’s effectiveness compared to the baseline approach while normalizing for data set specific characteristics. Higher values for relative score indicates that the method accrues lower error when compared to random sampling.

Chapter 5

Results

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

Our experimental evaluation revealed two key findings: first, that traditional scalar field sampling techniques are fundamentally inadequate for vector field preservation, and second, that our multi-phase approach achieves consistent performance improvements across diverse data sets and sampling rates.

5.1 Limitations of Scalar Field Techniques

Our experimental evaluation of scalar field sampling adaptations revealed consistent underperformance across multiple metrics. We systematically evaluated three scalar field sampling adaptations for vector fields:

1. **Vector Magnitude-Driven Sampling:** Applying the Biswas et al. [4] multi-criteria impor-

tance sampling to vector magnitude

2. **Vector Direction Data-Driven Sampling:** Utilizing directional components as the basis for importance sampling
3. **Per-Component Data-Driven Sampling:** Implementing sampling independently to each vector component

Figure 5.1 presents the quantitative performance assessment of these methods across sampling rates of 1%, 3%, and 5%. Notably, all three methodologies consistently underperformed relative to random sampling across both metrics, with performance deteriorating further at higher sampling densities where improvement would typically be anticipated.

The empirical evidence demonstrates that traditional scalar sampling paradigms, despite their established efficacy for scalar field representation, fail to capture the complex structural and spatial interdependencies inherent in vector field data. These quantitative findings provided the foundational motivation for developing our multi-criteria sampling framework, which explicitly addresses the multifaceted characteristics of vector fields.

5.2 Multi-Phase Sampling Performance

Our multi-phase framework demonstrates statistically significant improvements over random sampling across all tested configurations. Table 5.1 and Table 5.2 show the relative improvements in angular accuracy and streamline distance metrics respectively, which is further supported by Figure 5.2. At 5% sampling rate, our method achieves average relative improvements of 1.16 for angular accuracy and 1.12 for streamline distance across all data sets, with the strongest improvements observed in the Astro dataset for both metrics. This consistency is particularly noteworthy given the diverse characteristics of our test data sets.

The stability of our method’s performance is evidenced by its compact interquartile ranges in both figures, especially when compared to single-criterion approaches. This suggests that combin-

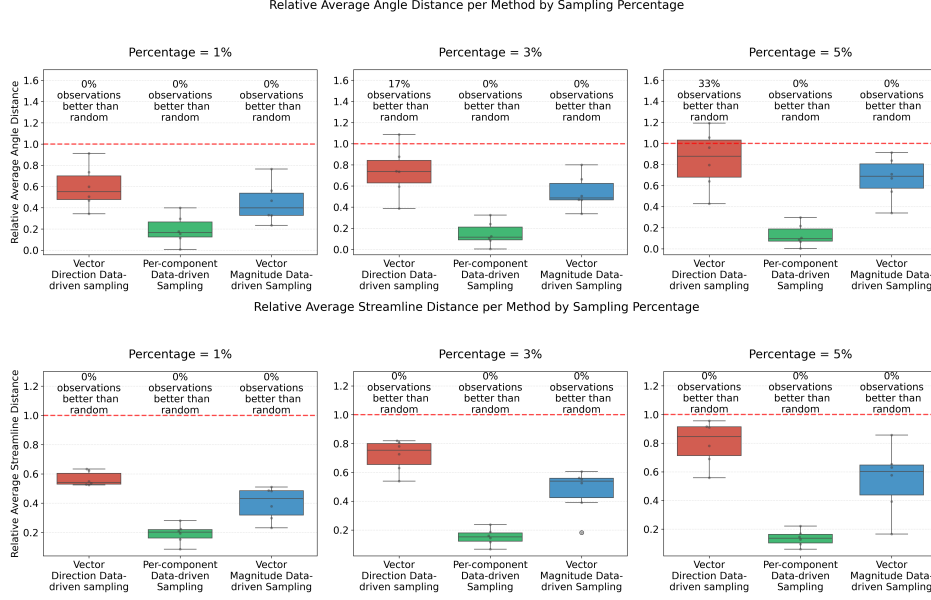


Figure 5.1: Performance evaluation of scalar field sampling adaptations across three sampling densities (1%, 3%, and 5%). The metrics from top to bottom include: (a) Average Angular Distance and (b) Average Streamline Distance, both normalized relative to random sampling baseline (dashed red line at 1.0). The quantitative analysis demonstrates consistent underperformance of all scalar-based methodologies, with 0% success rate in achieving performance parity with random sampling across most experimental configurations.

ing FTLE analysis with structured neighborhood sampling and smoothness criteria produces more reliable results than any individual approach alone.

5.3 Comparative Analysis of Methods

Each sampling strategy exhibits distinct characteristics:

1. FTLE + Context demonstrates strong performance comparable to our multi-phase method, particularly at higher sampling rates. However, its wider performance variance, especially visible in the angle distance metric at 5% sampling, indicates greater sensitivity to data set characteristics.
2. FTLE + Smoothness shows the highest peak performance in some cases but exhibits the widest variance. Its performance improves markedly with increased sampling rates, suggesting that this approach requires higher sampling densities to reliably capture field structure.

3. FTLE-only sampling shows the most variable performance, with only 33% of cases exceeding random sampling at 1% sampling rate. This underscores the limitations of single-criterion approaches for vector field sampling.

5.4 Sampling Rate Effects

Figure 5.2 reveals a clear correlation between sampling rate and method effectiveness. At 1% sampling, improvements over random sampling are modest across all methods, with median relative improvements rarely exceeding 1.1. As sampling rates increase to 3% and 5%, the advantages of structured sampling become more pronounced, particularly for our multi-phase method and FTLE + Context approaches. This trend suggests a minimum sampling density threshold necessary for effectively capturing vector field structure. Below this threshold (approximately 3% in our experiments), the distinction between methods becomes less significant, indicating that very sparse sampling may fundamentally limit the effectiveness of sophisticated selection criteria. These results provide strong evidence that our multi-phase sampling framework successfully balances the competing demands of reliability and effectiveness in vector field sampling. While other methods may achieve higher peak performance in specific cases, our approach consistently outperforms random sampling across all data sets and sampling rates, making it particularly suitable for general-purpose vector field sampling applications.

Performance comparison of sampling methods across sampling rates (1%, 3%, and 5%) using two key metrics. (a) Average Angular Distance (AAD) relative to random sampling, measuring local vector field accuracy preservation. (b) Average Streamline Distance relative to random sampling, quantifying preservation of global flow behavior. For both metrics, the y-axis shows relative improvement over random sampling (dashed red line at 1.0), with values above 1.0 indicating superior performance. Box plots represent the distribution across six scientific data sets and five random seeds, with whiskers extending to the most extreme non-outlier values. The percentage below each method indicates how often it outperformed random sampling across all data sets. We

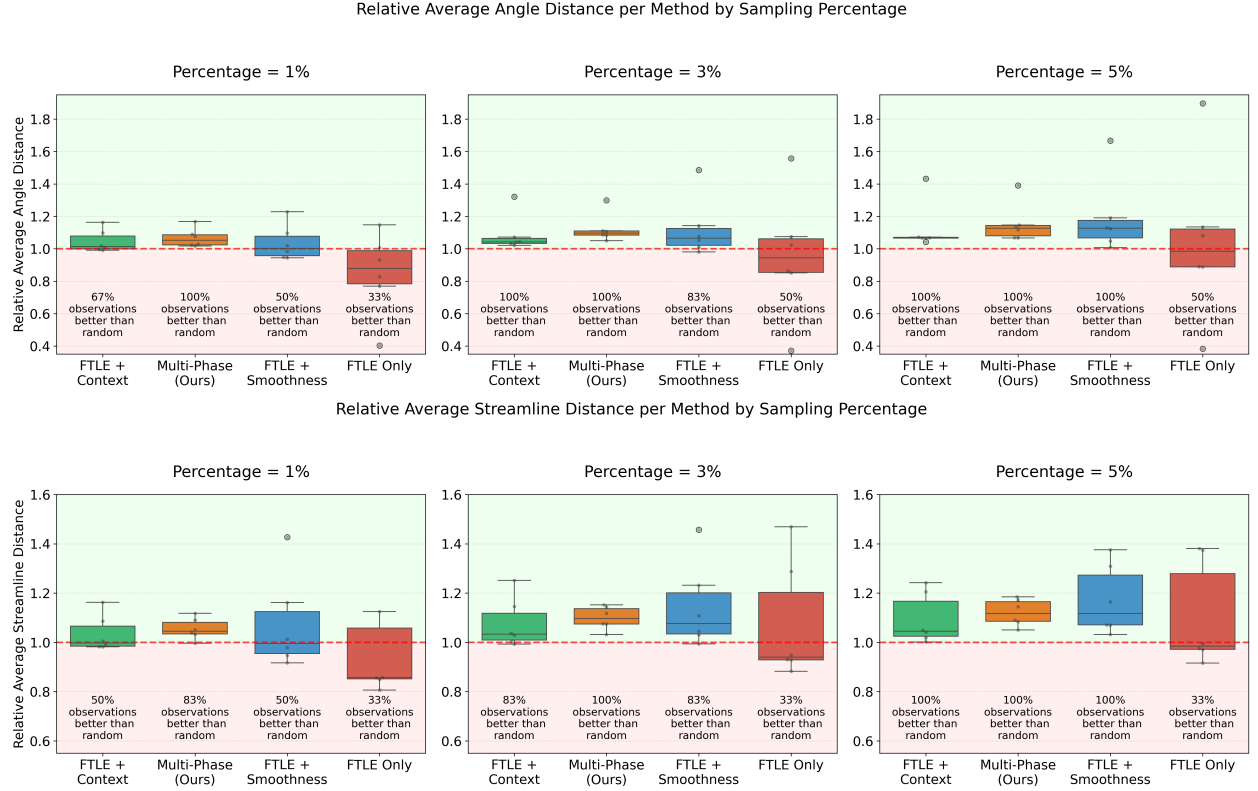


Figure 5.2: Performance comparison of sampling methods across sampling rates (1%, 3%, and 5%) using two key metrics. The metrics from top to bottom include: (a) Average Angular Distance (AAD) relative to random sampling, measuring local vector field accuracy preservation (b) Average Streamline Distance relative to random sampling, quantifying preservation of global flow behavior. For both metrics, values above 1.0 (dashed red line) indicate superior performance compared to random sampling. Box plots show distribution across six scientific datasets and five random seeds. Our multi-phase method (second from left) demonstrates consistent improvement over random sampling.

do not display smoothness only in this figure as it is not competitive with the other methods. Our multi-phase method (second from left) demonstrates both consistent performance and systematic improvement over random sampling, achieving 100% success rate in all cases but one. While FTLE + Smoothness shows higher peak performance in specific cases (visible as outliers), the multi-phase approach provides more reliable performance across diverse vector field types. The increasing separation from the random baseline at higher sampling rates suggests that structured sampling approaches become more effective as sampling density increases. Similarly, the variance of the results per method also decrease with increase in sampling percentage, leading to more reliable behavior with more samples selected.

5.5 Visual Analysis of Streamline Reconstruction

The qualitative comparison of streamline reconstructions from the Asteroid Impact data set (Figure 5.3) provides compelling visual evidence of our method’s effectiveness. The original vector field exhibits several distinctive flow features, most notably a pronounced curved “shell” structure characterized by abrupt directional changes in the flow field. This feature represents a challenging test case for sampling methods, as it combines both smooth continuous flow and sharp transitional regions.

Random sampling captures the general flow structure but introduces noticeable artifacts, particularly along the shell boundary where the flow direction changes rapidly. The reconstructed streamlines maintain the overall shell shape and flow patterns, though with some visible noise and perturbations along critical transition regions. While FTLE analysis effectively identifies regions of flow separation, using these values alone for sampling leads to significant degradation in streamline reconstruction quality. The resulting visualization exhibits more pronounced discontinuities and artifacts, particularly evident in the distorted shell structure and irregular streamline paths. This visual assessment aligns with our quantitative findings in Table 5.2, where FTLE-only sampling shows consistently lower performance compared to random sampling across multiple data

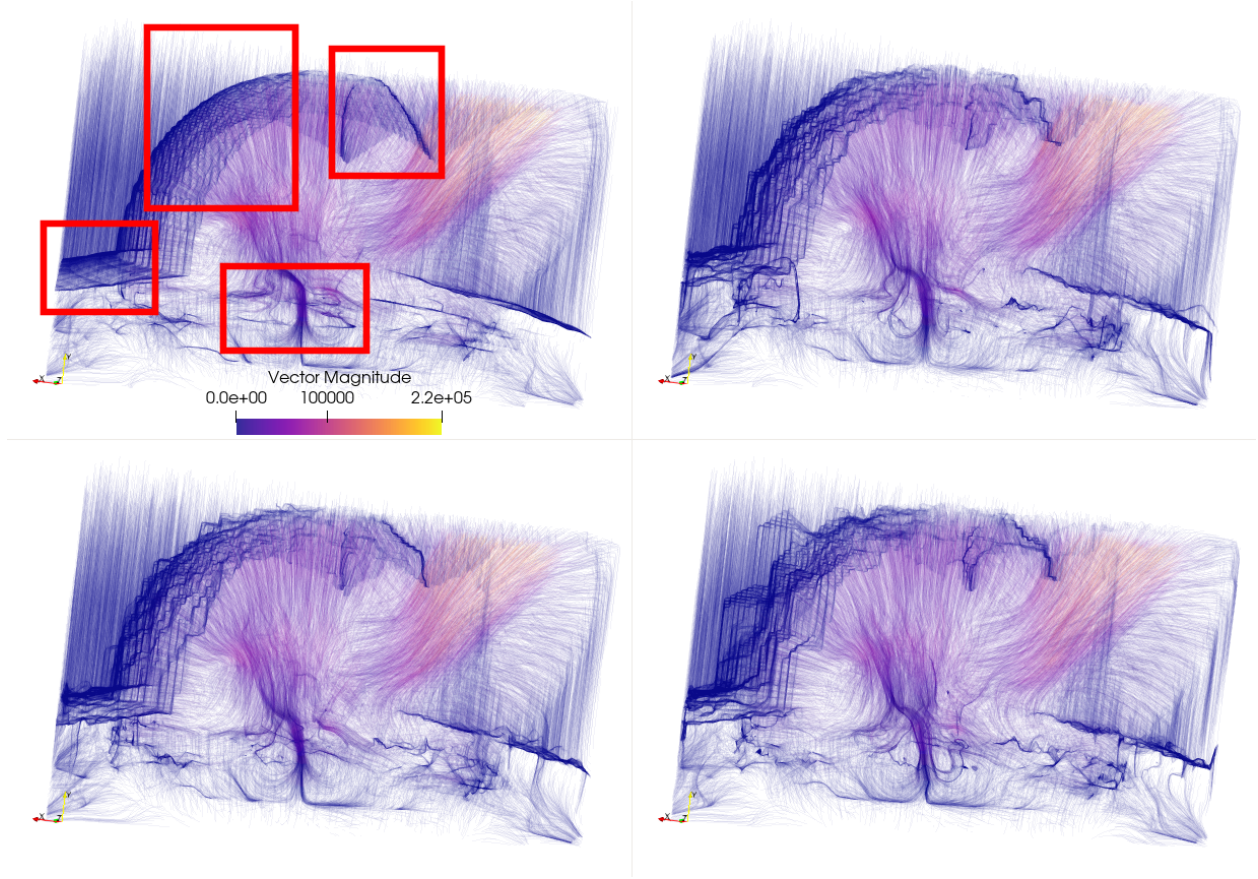


Figure 5.3: A comparison of streamline visualizations across sampling methods for the Asteroid Impact dataset. Clockwise from top left: (a) Original streamlines with annotated 'shell' feature, (b) streamlines from random sampling at 5%, (c) streamlines from FTLE-only sampling at 5%, and (d) streamlines from our multi-phase method at 5%. Note the distinctive 'shell' feature in the original data set, characterized by an abrupt directional change in the vector field. The regions of interest are annotated using red rectangles. While random sampling and FTLE-only sampling exhibit different sampling artifacts, our multi-phase approach better preserves this key structural feature with lesser distortion.

sets.

Our multi-phase approach achieves improved preservation of both the shell structure and the overall flow field compared to random sampling. The streamlines show smoother transitions across the shell boundary while maintaining the characteristic curved structure, as well as maintaining the secondary shell features that are annotated in the figure. The visual results align with our quantitative findings, particularly the improved Average Streamline Distance (ASD) metrics reported in Table 5.2.

Table 5.1: Relative Angular Accuracy Improvement Over Random Sampling

Dataset	Method	1%	3%	5%
		Relative Avg. Angle Distance		
Asteroid Impact	Multi-Phase	1.03	1.09	1.15
	FTLE + Context	1.00	1.04	1.07
	FTLE + Smooth	0.95	1.05	1.12
	FTLE Only	0.93	1.02	1.08
	Smoothness	0.29	0.31	0.33
Hurricane Isabel	Multi-Phase	1.02	1.05	1.07
	FTLE + Context	0.99	1.02	1.04
	FTLE + Smooth	0.94	0.98	1.01
	FTLE Only	0.83	0.86	0.89
	Smoothness	0.18	0.18	0.20
CM1 Tornado	Multi-Phase	1.03	1.09	1.12
	FTLE + Context	1.01	1.04	1.07
	FTLE + Smooth	0.98	1.08	1.13
	FTLE Only	0.93	1.02	1.08
	Smoothness	0.34	0.33	0.35
Astro	Multi-Phase	1.17	1.30	1.39
	FTLE + Context	1.16	1.32	1.43
	FTLE + Smooth	1.23	1.49	1.67
	FTLE Only	1.15	1.56	1.90
	Smoothness	0.18	0.26	0.28
Nimrod	Multi-Phase	1.09	1.08	1.07
	FTLE + Context	1.10	1.07	1.07
	FTLE + Smooth	1.02	1.01	1.05
	FTLE Only	0.40	0.37	0.38
	Smoothness	0.06	0.06	0.22
Fishtank	Multi-Phase	1.08	1.11	1.14
	FTLE + Context	1.02	1.03	1.06
	FTLE + Smooth	1.10	1.14	1.19
	FTLE Only	1.01	1.07	1.13
	Smoothness	0.61	0.78	0.90

Values greater than 1.0 indicate better performance than random sampling. Multi-Phase refers to our proposed method.

Table 5.2: Relative Streamline Accuracy Improvement Over Random Sampling

Dataset	Method	1%	3%	5%
		Rel. Avg. Streamline Distance		
Asteroid Impact	Multi-Phase	1.05	1.12	1.14
	FTLE + Context	1.00	1.04	1.04
	FTLE + Smooth	1.01	1.11	1.16
	FTLE Only	0.86	0.93	0.97
	Smoothness	0.35	0.37	0.40
Hurricane Isabel	Multi-Phase	1.03	1.07	1.09
	FTLE + Context	0.99	1.03	1.05
	FTLE + Smooth	0.98	1.04	1.07
	FTLE Only	0.85	0.93	0.97
	Smoothness	0.21	0.20	0.22
CM1 Tornado	Multi-Phase	1.00	1.03	1.05
	FTLE + Context	0.98	1.00	1.02
	FTLE + Smooth	0.92	0.99	1.03
	FTLE Only	0.86	0.93	0.97
	Smoothness	0.31	0.28	0.28
Astro	Multi-Phase	1.12	1.14	1.19
	FTLE + Context	1.09	1.15	1.21
	FTLE + Smooth	1.16	1.23	1.31
	FTLE Only	1.12	1.29	1.38
	Smoothness	0.23	0.29	0.33
Nimrod	Multi-Phase	1.09	1.15	1.17
	FTLE + Context	1.16	1.25	1.24
	FTLE + Smooth	1.43	1.46	1.38
	FTLE Only	1.61	1.47	1.37
	Smoothness	0.22	0.82	1.34
Fishtank	Multi-Phase	1.06	1.08	1.09
	FTLE + Context	0.99	1.00	1.01
	FTLE + Smooth	1.04	1.06	1.08
	FTLE Only	0.93	0.97	1.00
	Smoothness	0.55	0.66	0.72

Values greater than 1.0 indicate better performance than random sampling. Multi-Phase refers to our proposed method.

Chapter 6

Discussion and Parameter Selection

This chapter is based on a coauthored paper. I was the lead author and primary researcher, responsible for implementing all software, running all experiments, and writing the text. Ayan Biswas contributed ideas regarding initial project direction and guidance on specific techniques. Hank Childs served as my advisor and provided writing feedback as well as assistance in analyzing results and discussing adaptations to approach.

Through extensive experimentation, we identified several key parameters and design choices that significantly impact sampling effectiveness. This section details our exploration process and empirical findings that led to the final framework configuration.

6.1 FTLE Integration Time Selection

The selection of FTLE integration time proved crucial for effective feature identification. Our initial experiments explored integration periods ranging from 5 to 50 time steps, revealing that longer integration times did not necessarily improve sampling quality. We found that 10-20 integration steps provided optimal feature detection, while longer periods often introduced spurious features arising from accumulated numerical errors. The relationship between integration time and sam-

pling effectiveness was not monotonic - performance typically peaked around 20 time steps and declined thereafter, suggesting that excessive temporal information can actually degrade sampling quality.

6.2 Histogram Configuration

The binning strategy for both FTLE and smoothness-based importance sampling required careful consideration of the trade-off between resolution and statistical stability. We evaluated bin counts ranging from 4 to 128, conducting systematic tests across our data set collection. While finer binning (64+ bins) theoretically offered more precise importance discrimination, it also ends up oversampling on high-value regions, leading to poor sampling quality. We found that 32 bins consistently provided the best balance between granularity and robustness. This finding held across different sampling rates (1%, 3%, and 5%) and data sets, suggesting that this represents a fundamental sweet spot in the resolution-stability trade-off for vector field sampling.

6.3 Budget Allocation Optimization

Perhaps the most crucial aspect of our framework is the budget allocation between different sampling criteria. For the FTLE-based sampling with context preservation (FTLE + Context), we conducted a systematic sweep of allocation ratios from 10/90 through 60/40. The 20/80 split emerged as consistently optimal, suggesting that a relatively small number of feature-based seeds combined with extensive context preservation captures flow structure most effectively.

Similarly, for the combined FTLE and smoothness-based sampling, we explored various weightings through the same range. The 50/50 split showed the best performance, indicating equal importance of feature identification and local structure preservation. This allocation was not a clear winner however, as in some data sets, where FTLE-based sampling alone provided the best results, adding any percentage of smoothness-based sampling led to worse accuracy, becoming worse as the ratio tended towards the 50/50 split. However, the 50/50 split was the most consistent across all

six data sets, while those with a smaller weight for smoothness-based sampling performed worse than random on some data sets. This finding naturally led to our final three-phase allocation of 10:40:50 - essentially replacing the FTLE-based sampling with structured context preservation around FTLE-identified features.

A key finding from our analysis emerges from examining the relationship between vector field statistical distributions and sampling method performance. The data sets where FTLE-based sampling demonstrated superior performance (Astro and Nimrod) share a distinctive statistical property - their FTLE values and Jacobian Frobenius norms are concentrated within remarkably narrow ranges compared to the other data sets, with most values clustered tightly around their respective medians. This concentrated distribution suggests these fields have highly localized regions where flow separation and directional changes deviate significantly from the background, making feature identification particularly effective through FTLE analysis alone. In contrast, data sets with wider value distributions benefit more from our multi-phase approach, which can better handle varying scales of feature importance. This finding points to opportunities for a research direction investigating how data set characteristics might guide sampling approach optimization, without needing a human in the loop.

6.4 Alternative Approaches Explored

We investigated several alternative technical approaches that, while ultimately not included in our final framework, provided valuable insights. Log-scaled binning schemes were tested but showed no consistent improvement over linear binning. We explored alternative smoothness metrics including determinant-based measures and vorticity calculations, but found the Frobenius norm of the Jacobian provided the most reliable indication of local structure complexity. These explorations revealed that simpler, more robust approaches often outperformed more sophisticated techniques in practice.

Our systematic exploration of these parameters and alternatives provides a strong empirical

foundation for our framework’s design choices. While different scientific domains might benefit from parameter tuning, our findings suggest that these values represent robust defaults suitable for a wide range of vector field sampling applications.

Chapter 7

Conclusion

This research introduces a novel multi-phase sampling framework for vector fields that effectively preserves both local and global flow characteristics at sampling rates as low as 1%. Through comprehensive evaluation across six diverse scientific data sets, we have demonstrated that our approach consistently outperforms random sampling and single-criterion methods in preserving vector field structure and flow behavior. The experimental results validate our fundamental hypothesis that effective vector field sampling requires the simultaneous consideration of multiple complementary characteristics.

The key insight driving our method’s success lies in its unified treatment of both local and global vector field properties. By combining FTLE-based feature identification with structured neighborhood preservation and smoothness-based sampling, our framework captures the full spectrum of vector field behavior - from instantaneous directional changes to longer-term trajectory evolution. This comprehensive approach enables reliable preservation of complex flow features while maintaining computational efficiency suitable for in situ deployment. The structured neighborhood sampling strategy, in particular, proves crucial for maintaining spatial context around regions of significant flow separation, addressing a fundamental limitation of previous sampling approaches.

Our experimental results demonstrate that the multi-phase framework achieves consistent improvement over random sampling across all sampling fractions and data sets, with relative improvements in angular accuracy ranging from 1.02 to 1.39 and streamline distance metrics showing similar gains. At 5% sampling rate, our method achieves an average relative improvement of 1.16 for angular accuracy across all datasets, with particularly strong performance on the Astro dataset (1.39 $\times$) and the Fishtank dataset (1.14 $\times$). This consistency stands in marked contrast to single-criterion methods, which exhibit strong data set dependence and often fail to generalize across different flow types. The framework’s robust performance across diverse scientific domains underscores its potential as a general-purpose data-agnostic solution for in situ vector field data reduction through sampling.

Future research directions could extend this framework to address the unique challenges of time-varying vector fields, where temporal coherence introduces additional complexity to the sampling problem. The interaction between spatial and temporal features may require novel approaches to capture evolving flow structures while maintaining computational efficiency. Time-varying fields present particular challenges for FTLE computation and neighborhood preservation, as the temporal evolution of features may necessitate adaptive sampling strategies that can respond to changing flow characteristics. Other directions to consider would be to adaptively choose parameters based on data set statistical properties such as nature of distribution of values, dynamic range of values and more.

The empirical evidence presented in this work establishes clear guidelines for practical deployment of vector field sampling in production scientific workflows. The consistent performance improvements across diverse data sets, coupled with the framework’s inherent parallelizability and linear scaling characteristics, make it immediately applicable to current large-scale simulation environments. While our current evaluation focuses on demonstrating the effectiveness of the sampling strategy, future work should include explicit parallel implementation and evaluation on truly large-scale datasets to fully validate the framework’s potential for in situ deployment. The systematic analysis of parameter sensitivity and practical implementation considerations provides

a robust foundation for deployment across different scientific domains, while the identified paths for extension suggest rich opportunities for future development.

References

- [1] Alexy Agranovsky, David Camp, Kenneth I. Joy, and Hank Childs. Subsampling-based compression and flow visualization. *Electronic Imaging*, 2015. doi: 10.1117/12.2083251.
- [2] James Ahrens, Sébastien Jourdain, Patrick O’Leary, John Patchett, David H Rogers, and Mark Petersen. An image-based approach to extreme scale in situ visualization and analysis. In *SC’14: Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis*, pages 424–434. IEEE, 2014.
- [3] Ayan Biswas, Soumya Dutta, Jesus Pulido, and James Ahrens. In situ data-driven adaptive sampling for large-scale simulation data summarization. In *Proceedings of the Workshop on In Situ Infrastructures for Enabling Extreme-Scale Analysis and Visualization*, pages 13–18, 2018.
- [4] Ayan Biswas, Soumya Dutta, Earl Lawrence, John Patchett, Jon C. Calhoun, and James Ahrens. Probabilistic data-driven sampling via multi-criteria importance analysis. *IEEE Transactions on Visualization and Computer Graphics*, 27(12):4439–4454, 2021. doi: 10.1109/TVCG.2020.3006426.
- [5] Tamal K. Dey, Joshua A. Levine, and Rephael Wenger. A Delaunay simplification algorithm for vector fields. In *Pacific Graphics*, 2007. doi: 10.1109/pg.2007.34.
- [6] Soumya Dutta, Chun-Ming Chen, Gregory Heinlein, Han-Wei Shen, and Jen-Ping Chen. In situ distribution guided analysis and visualization of transonic jet engine simulations. *IEEE transactions on visualization and computer graphics*, 23(1):811–820, 2016.

- [7] Soumya Dutta, Jonathan Woodring, Han-Wei Shen, Jen-Ping Chen, and James Ahrens. Homogeneity guided probabilistic data summaries for analysis and visualization of large-scale data sets. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 111–120. IEEE, 2017.
- [8] Eirik Endeve, Christian Y Cardall, Reuben D Budiardja, and Anthony Mezzacappa. Generation of magnetic fields by the stationary accretion shock instability. *The Astrophysical Journal*, 713(2):1219, 2010.
- [9] Nathan Fabian, Kenneth Moreland, David Thompson, Andrew C. Bauer, Pat Marion, Berk Gevecik, Michel Rasquin, and Kenneth E. Jansen. The paraview coprocessing library: A scalable, general purpose in situ visualization library. In *2011 IEEE Symposium on Large Data Analysis and Visualization*, pages 89–96, 2011. doi: 10.1109/LDAV.2011.6092322.
- [10] Paul Fischer, James Lottes, David Pointer, and Andrew Siegel. Petascale algorithms for reactor hydrodynamics. In *Journal of Physics: Conference Series*, volume 125, page 012076. IOP Publishing, 2008.
- [11] Raphael Imahorn, Irene Baeza Rojo, and Tobias Günther. Visualization and analysis of deep water asteroid impacts. In *2018 IEEE Scientific Visualization Conference (SciVis)*, pages 85–96. IEEE, 2018.
- [12] Stefan Koch, Jens Kasten, Alexander Wiebel, Gerik Scheuermann, and Mario Hlawitschka. 2D vector field approximation using linear neighborhoods. *The Visual Computer*, 2016. doi: 10.1007/s00371-015-1140-9.
- [13] Sriram Lakshminarasimhan, Neil Shah, Stéphane Ethier, Scott Klasky, Robert Latham, Robert B. Ross, and Nagiza F. Samatova. Compressing the incompressible with ISABELA: in-situ reduction of spatio-temporal data. In *Euro-Par (1)*, volume 6852 of *Lecture Notes in Computer Science*, pages 366–379. Springer, 2011.

- [14] Suresh K. Lodha, N.M. Faaland, and J.C. Renteria. Topology preserving top-down compression of 2D vector fields using bintree and triangular quadrees. *IEEE Transactions on Visualization and Computer Graphics*, 2003. doi: 10.1109/tvcg.2003.1260738.
- [15] Jay F Lofstead, Scott Klasky, Karsten Schwan, Norbert Podhorszki, and Chen Jin. Flexible IO and integration for scientific codes through the adaptable IO system (ADIOS). In *Proceedings of the 6th international workshop on Challenges of large applications in distributed environments*, New York, NY, USA, June 2008. ACM.
- [16] Tam Thanh Nguyen and Insu Song. Centrality clustering-based sampling for big data visualization. In *2016 International Joint Conference on Neural Networks (IJCNN)*, pages 1911–1917. IEEE, 2016.
- [17] David S Nolan, Morris A Bender, Timothy P Marchok, Stephen T Garner, and Christopher L Kerr. Simulations of hurricane isabel (2003) in the wrf, gfdl, and zetac models. In *Preprints, 25th Conference on Hurricanes and Tropical Meteorology*, 2004.
- [18] Leigh Orf, Robert Wilhelmson, and Louis Wicker. Visualization of a simulated long-track ef5 tornado embedded within a supercell thunderstorm. *Parallel Computing*, 55:28–34, 2016.
- [19] Yongjoo Park, Michael Cafarella, and Barzan Mozafari. Visualization-aware sampling for very large databases. In *2016 IEEE 32nd international conference on data engineering (icde)*, pages 755–766. IEEE, 2016.
- [20] Tobias Rapp, Christoph Peters, and Carsten Dachsbacher. Void-and-cluster sampling of large scattered data and trajectories. *IEEE transactions on visualization and computer graphics*, 26(1):780–789, 2019.
- [21] Sudhanshu Sane, Chris R. Johnson, and Hank Childs. Investigating in situ reduction via lagrangian representations for cosmology and seismology applications. In *International Conference on Computational Science*, 2021. doi: 10.1007/978-3-030-77961-0_36.

- [22] Michal Smolik and Vaclav Skala. Efficient simple large scattered 3D vector fields radial basis functions approximation using space subdivision. In *International Conference on Computational Science and Its Applications*, 2019. doi: 10.1007/978-3-030-24289-3_25.
- [23] Michal Smolik and Vaclav Skala. Radial basis function and multi-level 2D vector field approximation. *Mathematics and Computers in Simulation*, 2020. doi: 10.1016/j.matcom.2020.10.009.
- [24] Carl R Sovinec, AH Glasser, TA Gianakon, DC Barnes, RA Nebel, SE Kruger, DD Schnack, SJ Plimpton, A Tarditi, MS Chu, et al. Nonlinear magnetohydrodynamics simulation using high-order finite elements. *Journal of Computational Physics*, 195(1):355–386, 2004.
- [25] Holger Theisel, Christian Rössl, and Hans-Peter Seidel. Combining topological simplification and topology preserving compression for 2D vector fields. In *Pacific Conference on Computer Graphics and Applications*, 2003. doi: 10.1109/pccga.2003.1238287.
- [26] Anna Tikhonova, Carlos D Correa, and Kwan-Liu Ma. Explorable images for visualizing volume data. *PacificVis*, 10(177-184):4, 2010.
- [27] Venkatram Vishwanath, Mark Hereld, and Michael E. Papka. Toward simulation-time data analysis and i/o acceleration on leadership-class systems. In *2011 IEEE Symposium on Large Data Analysis and Visualization*, pages 9–14, 2011. doi: 10.1109/LDAV.2011.6092178.
- [28] Ko-Chih Wang, Kewei Lu, Tzu-Hsuan Wei, Naeem Shareef, and Han-Wei Shen. Statistical visualization and analysis of large data using a value-based spatial distribution. In *2017 IEEE Pacific Visualization Symposium (PacificVis)*, pages 161–170. IEEE, 2017.
- [29] Ko-Chih Wang, Naeem Shareef, and Han-Wei Shen. Image and distribution based volume rendering for large data sets. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, pages 26–35. IEEE, 2018.

- [30] Tzu-Hsuan Wei, Soumya Dutta, and Han-Wei Shen. Information guided data sampling and recovery using bitmap indexing. In *2018 IEEE Pacific Visualization Symposium (PacificVis)*, pages 56–65. IEEE, 2018.
- [31] Tino Weinkauff, Holger Theisel, Kuangyu Shi, H.-C. Hege, and Hans-Peter Seidel. Extracting higher order critical points and topological simplification of 3D vector fields. In *IEEE Visualization*, 2005. doi: 10.1109/vis.2005.35.
- [32] Brad Whitlock, Jean M. Favre, and Jeremy S. Meredith. Parallel in situ coupling of simulation with a fully featured visualization system. In Torsten W. Kuhlen, Renato Pajarola, and Kun Zhou, editors, *11th Eurographics Symposium on Parallel Graphics and Visualization, EGPGV@Eurographics 2011, Llandudno, UK, April 10-11, 2011*, pages 101–109. Eurographics Association, 2011. doi: 10.2312/EGPGV/EGPGV11/101-109. URL <https://doi.org/10.2312/EGPGV/EGPGV11/101-109>.
- [33] Jonathan Woodring, James Ahrens, J Figg, Joanne Wendelberger, Salman Habib, and Katrin Heitmann. In-situ sampling of a large-scale particle simulation for interactive visualization and analysis. In *Computer Graphics Forum*, volume 30, pages 1151–1160. Wiley Online Library, 2011.
- [34] Yucong Chris Ye, Tyson Neuroth, Franz Sauer, Kwan-Liu Ma, Giulio Borghesi, Aditya Kon-duri, Hemanth Kolla, and Jacqueline Chen. In situ generated probability distribution functions for interactive post hoc visualization and analysis. In *2016 IEEE 6th Symposium on Large Data Analysis and Visualization (LDAV)*, pages 65–74. IEEE, 2016.
- [35] Zhi Yuan, Ye Zhao, Fan Chen, Sean Reber, Cheng-Chang Lu, and Yang Chen. Detail-preserving compression for smoke-based flow visualization. *Journal of Visualization*, 2019. doi: 10.1007/s12650-018-0526-y.