Statistical Rendering for Visualization of Red Sea Eddy Simulation Data

Tushar M. Athawale, Alireza Entezari, Bei Wang, and Chris R. Johnson

Abstract

Analyzing the effects of ocean eddies is important in oceanology for gaining insights into the transport of energy and biogeochemical particles. We present an application of statistical visualization techniques for the analysis of the Red Sea eddy simulation ensemble. Specifically, we demonstrate the applications of *statistical volume rendering*, *statistical summary maps*, and *statistical level sets* to velocity magnitude fields derived from the ensemble for the study of eddy positions. In statistical volume rendering, we model per-voxel data uncertainty using noise densities and study the propagation of uncertainty into the volume rendering pipeline. In the statistical summary maps, we characterize the uncertainty of gradient flow destinations to understand the structural variations of Morse complexes across the ensemble. In statistical level sets, we study the effects on isocontour positions for the Gaussian-distributed uncertainty. We demonstrate the utility of our statistical visualizations for the analysis of eddy positions and their spatial uncertainty, as well as for the exploration of the correlation between temperature and velocity magnitude fields and the study of water surface elevation data.

Index Terms: Human-centered computing—Visualization—Visualization application domains—Scientific visualization; Mathematics of computing—Probability and statistics—Probabilistic algorithms

1 Introduction

In this work, we visually explore the Red Sea eddy simulation dataset using statistical rendering techniques, as part of the IEEE SciVis Contest 2020[1]. The randomness in data acquisition is captured via an ensemble of simulations, in which each ensemble member is generated based on the MIT ocean general circulation model (MITgcm) and the Data Assimilation Research Testbed (DART) [5] with varying initial conditions. Each ensemble is sampled on a domain with a grid of resolution $500 \times 500 \times 50$, and ensemble members are sampled from 60 time steps to represent a time-varying 3D flow [9]. Fig. 1 illustrates the variability of velocity vector fields and eddy positions for the three ensemble members sampled at time step $t = 40$ over a subset of the domain.

Although the uncertainty inherent in physical systems may be represented through multiple simulations/sensors, the large size of ensemble simulations, e.g., 1.5 TB for the Red Sea dataset, can become a bottleneck for data transmission and visualization. As surveyed in the recent work by Wang *et al.* [10], visualization of statistical summaries for ensembles can be effective in understanding commonalities and differences in the features observed across the ensemble members. Moreover, statistical summarizations can help alleviate the storage burden through reduced representations of data. We employ three statistical summarization techniques for the visualization of the Red Sea eddy simulation dataset, namely statistical volume rendering [8], statistical summary maps [2], and statistical level sets [7]. We now demonstrate the results of the application of these techniques to the Red Sea eddy simulation ensemble.

[1] https: //kaust-vislab.github.io/SciVis2020/

2 Statistical 3D Volume Rendering

We employ statistical volume rendering frameworks [1, 6, 8] for 3D visualizations of the Red Sea eddy simulation ensemble. In these frameworks, per-voxel uncertainty is characterized using a probability distribution, which is estimated from the ensemble members. The probability distributions are then propagated through a direct volume rendering pipeline to derive the likely (expected) visualizations for the ensemble. In Fig. 2, we derive the expected visualizations for the uniform [8], Gaussian, Gaussian mixture [6], and nonparametric [1] noise models. The visualizations are derived for the velocity magnitude ensemble with 20 members over the domain 40°E-50°E and 10°N-20°N for the time step $t = 40$. The transfer function (TF) shown in Fig. 2b maps the regions with relatively high-, moderate-, and low-velocity magnitudes to red, blue, and yellow, respectively. The same TF is used for all statistical renderings in Fig. 2.

The expected visualizations derived using the distribution-based statistical models (Fig. 2 (c-h)) exhibit improved reconstructions of the eddy features compared to the mean-field (Fig. 2a) due to their relatively high resilience to the outlier members [1]. In Fig. 2c, the three potential eddy positions are annotated as e_1, e_2, and e_3. As can be inferred from the statistical renderings, the eddy e_1 can be observed across all distribution models, thus indicating a high likelihood of its presence. The eddy e_2 is clearly seen in the uniform, Gaussian, and Gaussian mixture (ordered) models. The eddy e_3 exhibits relatively high uncertainty regarding its presence, as no noise model shows a clear vortical structure in the same region.

Note that the reduced data representations with probability distributions enable us to perform statistical volume rendering at interactive frame rates. For example, the parametric noise models consume only twice the amount of memory needed for the mean-field statistical approach since they store mean and width/variance per voxel. The memory requirements for the Gaussian mixture and quantile interpolation frameworks increase linearly with the number of Gaussians and quantiles, respectively. Please refer to the supplementary video for the data interaction demo.

Fig. 3a visualizes a box-plot-like view for the velocity magnitude ensemble. Specifically, we derive the lower quartile (lower 25%), middle quartile (central 50%), and upper quartile (upper 25%) at each voxel of the dataset and visualize each quartile with the uniform statistics. The quartile view [1] gives us insight into variations in features across the three populations. Fig. 3b analyzes the effects of sample size on nonparametric statistical renderings. The dotted boxes in Fig. 3b illustrate the features with relatively high sensitivity to underlying data. Fig. 4 depicts how visualizations evolve for the time steps $t = 36, \cdots, 40$ for the mean and parametric statistics.

We also take advantage of 2D TFs in the context of statistical volume rendering [1] to visually explore the correlation between the temperature and velocity magnitude fields. Fig. 5a visualizes the regions with relatively low-velocity magnitudes, which are observed near the vortex cores, vortex rims, and ocean surface. We further segment these regions into the areas with relatively high temperatures in red and relatively low temperatures in yellow (see the TF shown in Fig. 5b). A relatively high temperature is observed near the vortex features, whereas a relatively low temperature is observed near the ocean surface. Fig. 6 depicts how visualizations evolve for the time steps $t = 35, \cdots, 40$. Fig. 7 segments the regions with relatively high- and low-velocity magnitudes for a relatively high temperature range. Velocity magnitudes appear to increase as we move toward the vortex cores.

3 Statistical Summary Maps of Morse Complexes

Morse and Morse-Smale complexes are topological descriptors that provide abstract representations of the gradient flow behavior of scalar fields. We study the variability of Morse complexes for the Red Sea ensemble members using the probabilistic maps proposed by Athawale et al. [2] to gain insight into the positional variability of expected vortex structures. For our analysis, we use an ensemble of 10 members, in which each member corresponds to a 2D slice perpendicular to the z-axis ($z = 1$) for time step $t = 40$. We again analyze the eddies over the domain 40°E-50°E and 10°N-20°N.

Each ensemble member represents a velocity vector field, and Morse complexes are computed from the negation of velocity magnitudes to focus on the local maxima. For the ensemble members illustrated in Figs. 8a-c, their corresponding simplified Morse complexes [3] contain 2-cells that highlight vortical features of ocean eddies (white boxes). The mean-field Morse complex in Fig. 8d, however, does not give any insight into the variabilities of these features across the ensemble. We, thus, analyze the variability of vortical features via computation of probabilistic maps as follows:

Persistence simplification. Initially, we employ persistence simplification [4] to eliminate potentially noisy features, guided by persistence graphs and spaghetti plots in Fig. 9. In particular, at the selected simplification scale (dotted red line) in Fig. 9a, 5 of 10 (50%) members agree on the number of maxima (11) after simplification. The spaghetti plots of the Morse complexes in Fig. 9b display the high variability of simulations even at a simplified scale.

Labeling. Next, we leverage the labeling strategies proposed in [2] for deriving associations among local maxima of simplified ensemble members. In Fig. 10, we compare the three labeling strategies, Morse mapping, k-means clustering with $k = 11$, and nearest mandatory maxima. Note that the mandatory maxima lose significant topological information because of large variations across the ensemble; hence, we derive the probabilistic maps for the labeling with k-means clustering and Morse mapping.

Probabilistic map. In Fig. 11, we visualize the probabilistic map using color blending [2], which highlights the positional uncertainty of 2-cell boundaries invisible to the mean-field of Fig. 8d. The expected 2-cell boundaries (black contours) extract the expected eddy positions for the ensemble. Figs. 12a-c visualize our entropy-based exploration of the probabilistic map for lower entropy thresholds of 1.5, 1.25, and 1, respectively. Figs. 12d-f carve out regions in the domain, where the ensemble agrees in their gradient destinations for at least 80%, 70%, and 60% members, respectively. Thus, the shared features denoting the eddy structures across the ensemble are discoverable in Figs. 12d-f. Fig. 13 depicts interactive probability distribution queries, in which the gradient flows originating at the query selections $0-3$ have the highest probability of terminating at the local maxima with green, yellow, gray, and pink labels, respectively.

In Fig. 14-16, we visualize results similar to those in Fig. 9-13 for the water surface elevation (η) field. The quasi-circular patterns with the η values lower or higher than the surroundings in the dataset indicate the eddy presence [11]. Our expected Morse complex and agreement-thresholded visualizations in Fig. 15 and Fig. 16, respectively, help us recover the shared quasi-circular features for the ensemble, thus indicating the potential eddy positions.

4 Statistical Level Sets

In Fig. 17, we study the expected isocontour positions by applying probabilistic marching squares (PMS) [7] to the velocity magnitude ensemble assuming the Gaussian-distributed uncertainty. In PMS, the level-crossing probability is estimated for each grid cell by sampling the underlying probability distributions and visualized via color mapping. Fig. 17b visualizes the level-crossing probabilities for a level set with low-velocity magnitude, which effectively highlights the potential vortex core positions when compared to the corresponding spaghetti plots visualized in Fig. 17a.

5 Implementation

The statistical volume renderings are performed on a machine with Nvidia GPU Quadro P6000, with 24 GB memory by integrating the fragment shaders into the Voreen volume rendering engine (`http://voreen.uni-muenster.de`). In the case of statistical summary maps, we extend the Python code for topological data analysis available at `https://pypi.org/project/topopy/`. We implement statistical level sets in MATLAB.

6 Conclusion

We demonstrate the applications of statistical visualization techniques coupled with volume rendering [1,6,8], Morse complexes [2], and level sets [7] to the Red Sea eddy simulation ensemble for effectively extracting the likely (expected) eddy positions and their variability. The distribution-based data representation in the case of statistical volume rendering allows for the exploration of the large-scale Red Sea dataset in 3D at interactive frame rates. Additionally, the distribution-based statistics show increased robustness to noise compared to the mean statistics and allow for an uncertainty integration with visualizations using statistical rendering techniques.

Acknowledgments

This work was supported in part by the NIH grants P41 GM103545-18 and R24 GM136986; the DOE grant DE-FE0031880; the Intel Graphics and Visualization Institutes of XeLLENCE; and the NSF grants IIS-1617101, IIS-1910733.

References

[1] T. M. Athawale, B. Ma, E. Sakhaee, C. R. Johnson, and A. Entezari. Direct volume rendering with nonparametric models of uncertainty. *IEEE Transactions on Visualization and Computer Graphics, to appear*, 2021. arXiv:2008.13576.

[2] T. M. Athawale, D. Maljovec, L. Yan, C. R. Johnson, V. Pascucci, and B. Wang. Uncertainty visualization of 2D Morse complex ensembles using statistical summary maps. *IEEE Transactions on Visualization and Computer Graphics, to appear*, 2020. doi: 10.1109/TVCG.2020.3022359

[3] H. Edelsbrunner, J. Harer, and A. Zomorodian. Hierarchical Morse complexes for piecewise linear 2-manifolds. In *Proceedings of the 17th Annual Symposium on Computational Geometry*, pp. 70–79, 2001.

[4] H. Edelsbrunner, D. Letscher, and A. J. Zomorodian. Topological persistence and simplification. *Discrete and Computational Geometry*, 28:511–533, 2002.

[5] I. Hoteit, T. Hoar, G. Gopalakrishnan, N. Collins, J. Anderson, B. Cornuelle, A. Köhld, and P. Heimbach. A MITgcm/DART ocean analysis and prediction system with application to the Gulf of Mexico. *Dynamics of Atmospheres and Oceans*, 63:1–23, 2013.

[6] S. Liu, J. Levine, P.-T. Bremer, and V. Pascucci. Gaussian mixture model based volume visualization. In *Proceedings of the IEEE Large-Scale Data Analysis and Visualization Symposium*, pp. 73–77, 2012.

[7] K. Pöthkow, B. Weber, and H.-C. Hege. Probabilistic marching cubes. *Computer Graphics Forum*, 30(3):931–940, 2011.

[8] E. Sakhaee 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, 2017.

[9] S. Sivareddy, H. Toye, P. Zhan, S. Langodan, G. Krokos, O. Knio, and I. Hoteit. Impact of atmospheric and model physics perturbations on a high-resolution ensemble data assimilation system of the Red Sea. *Journal of Geophysical Research: Oceans*, 2020.

[10] J. Wang, S. Hazarika, C. Li, and H.-W. Shen. Visualization and visual analysis of ensemble data: A survey. *IEEE Transactions on Visualization and Computer Graphics*, 25(9), 2018.

[11] P. Zhan, A. Subramanian, F. Yao, and I. Hoteit. Eddies in the red sea: A statistical and dynamical study. *Journal of Geophysical Research*, 119(6):3909–3925, 2014.

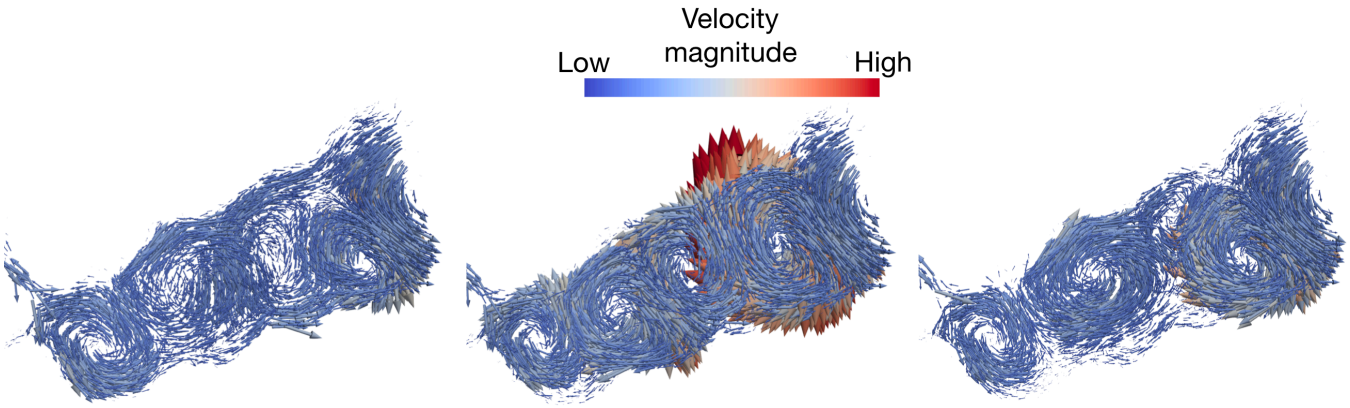

Figure 1. Arrow glyph visualizations for the velocity vector fields of three ensemble members at time step $t = 40$ colored by velocity magnitude.

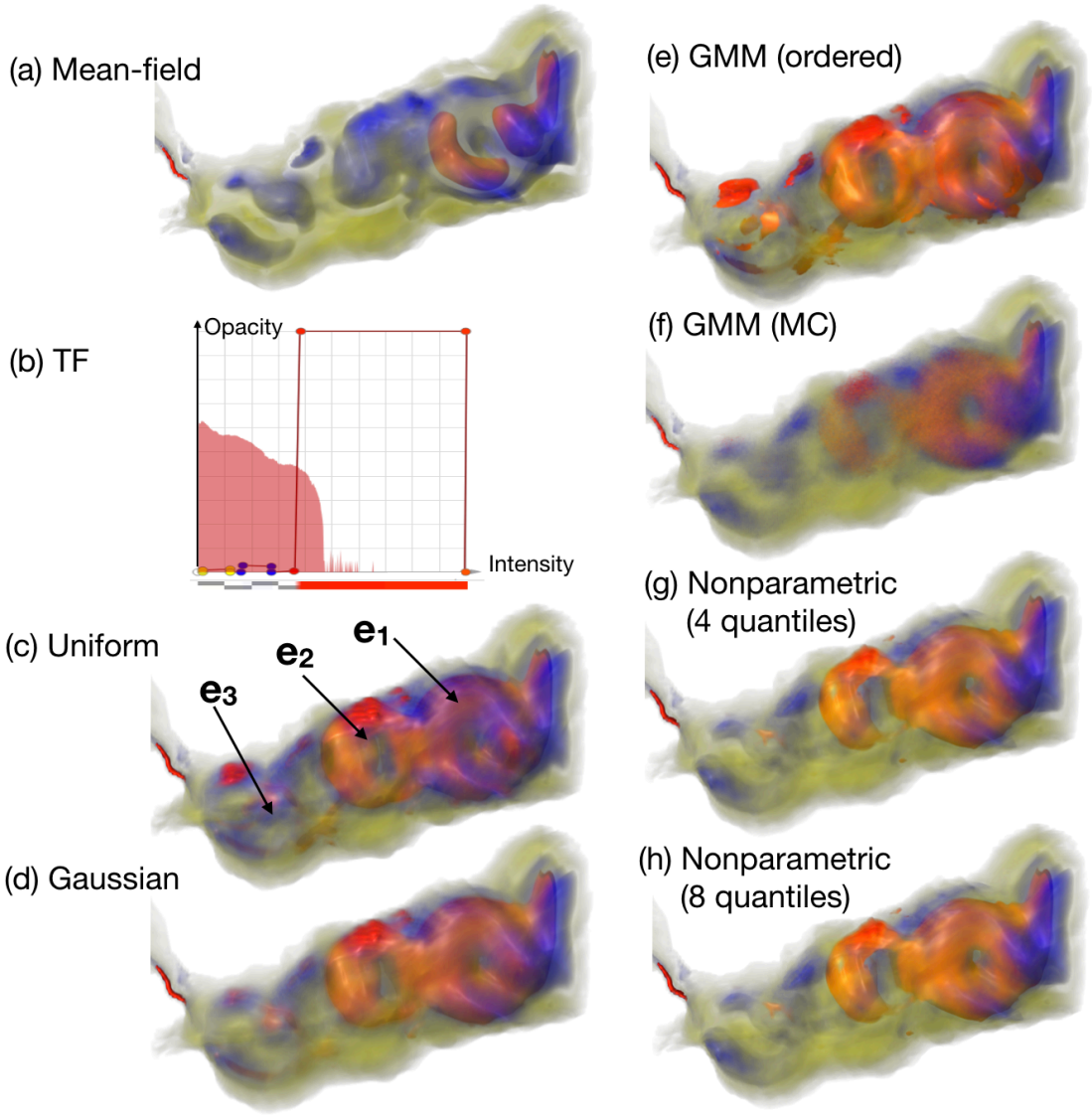

Figure 2. Statistical volume rendering of the velocity magnitude fields derived from the ensemble: (a) mean statistics, (c-d) parametric noise models, (e) Gaussian mixture models with ordered Gaussian means, (f) Gaussian mixture models with Monte Carlo (MC) sampling, (g-h) nonparametric density models with quantile representation. The red, blue, and yellow in the transfer function (b) indicate relatively high-, moderate-, and low-velocity magnitudes. e_1, e_2, and e_3 in (c) denote the potential eddy positions observed across different noise models.

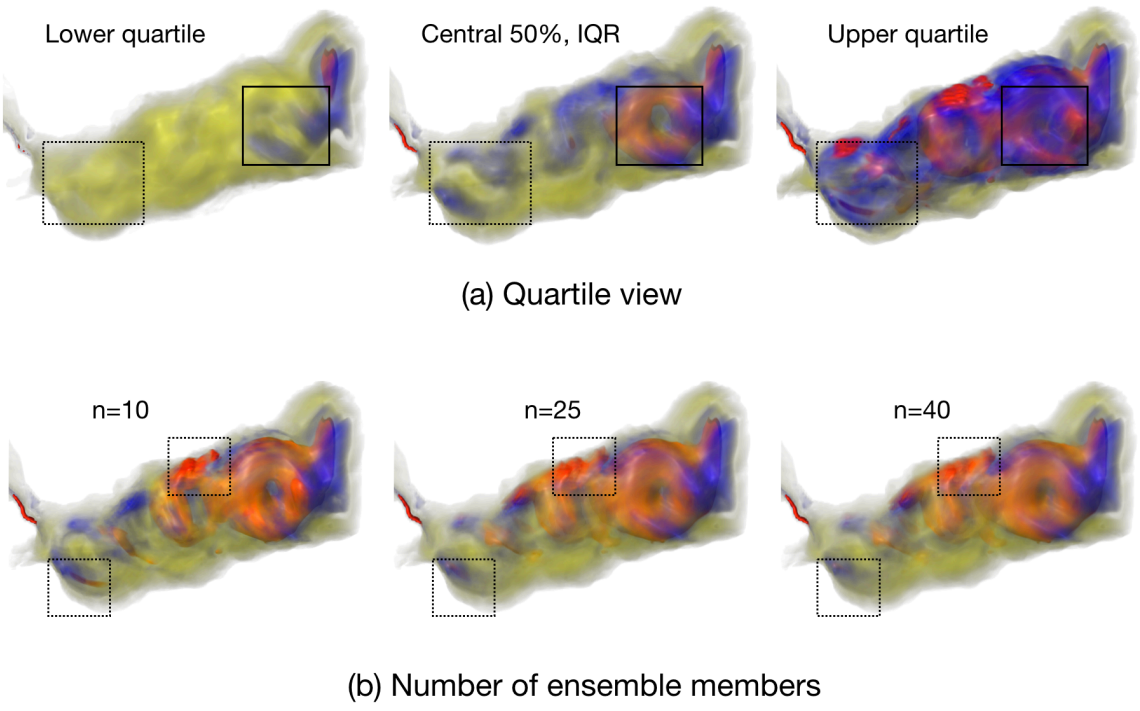

Figure 3. (a) The quartile view for uncertainty analysis of the presence of an eddy in the Red Sea dataset. The solid boxes enclose the positions that indicate the high likelihood of the presence of an eddy, whereas the dotted boxes mark positions with substantial uncertainty in the presence of an eddy. (b) Effect of sample size (n) on visualizations. The dotted boxes illustrate positions with variability in reconstruction.

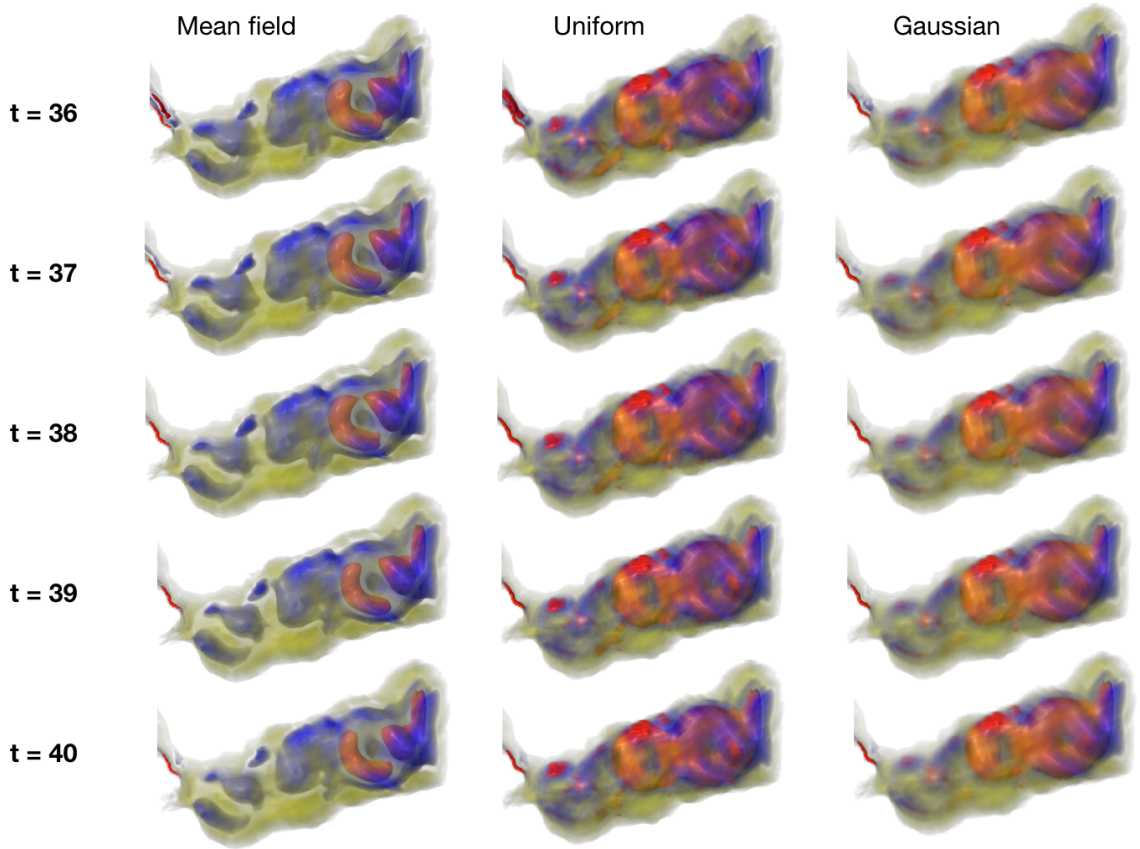

Figure 4. Visualizations of the uncertain velocity magnitude field for a series of time steps. The mean-field visualizations exhibit more fluctuations in reconstructions compared to the uniform and Gaussian noise models.

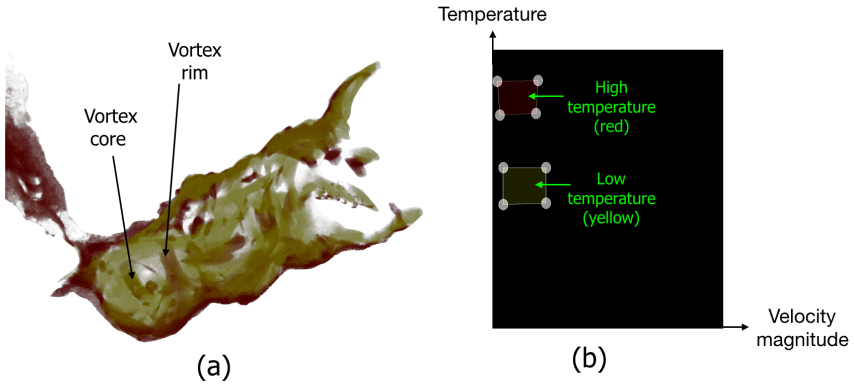

Figure 5. Statistical rendering with 2D TF to visually explore the correlation between the velocity magnitude and temperature fields. High-temperature and low-temperature regions are mapped to red and yellow, respectively.

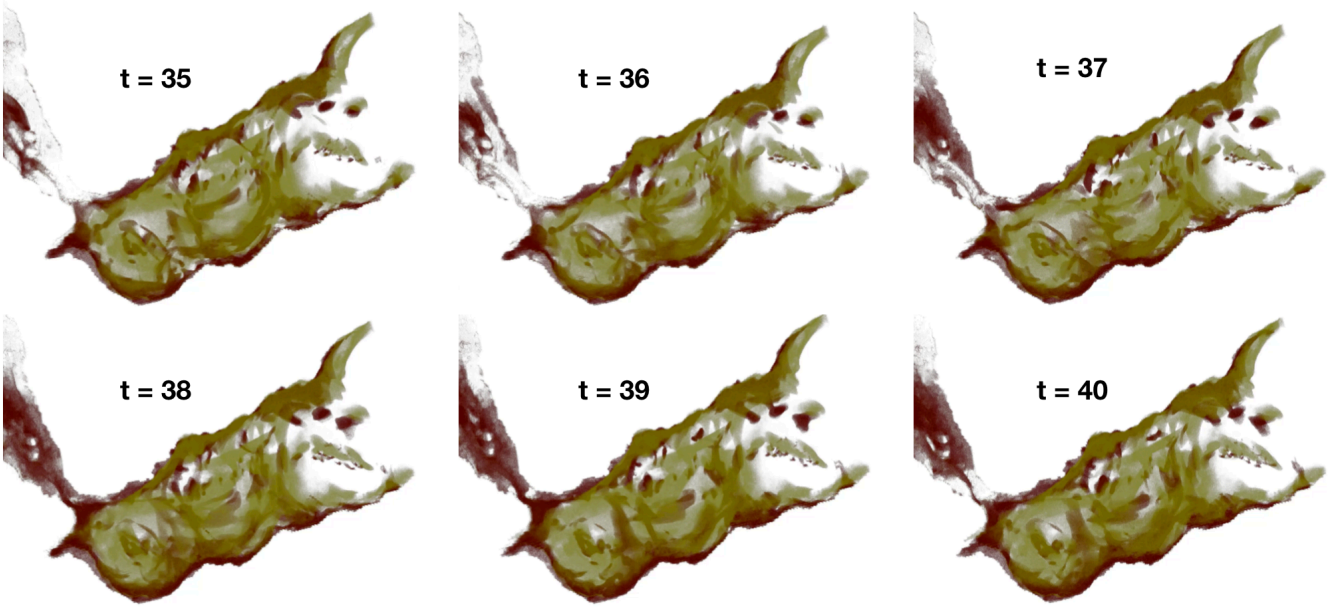

Figure 6. Visualizations of the correlation between the velocity magnitude and temperature fields for a series of time steps using the 2D TF shown in Fig. 5b.

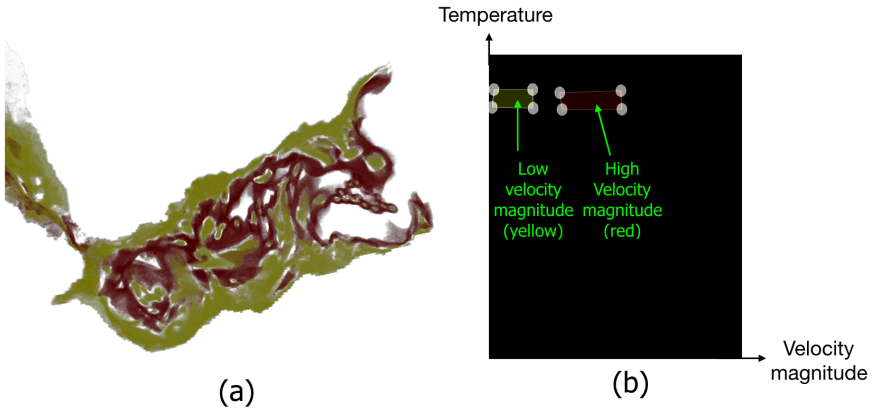

Figure 7. Statistical rendering with 2D TF to visually explore the correlation between velocity magnitude and temperature fields. High-velocity and low-velocity magnitude regions are mapped to red and yellow, respectively.

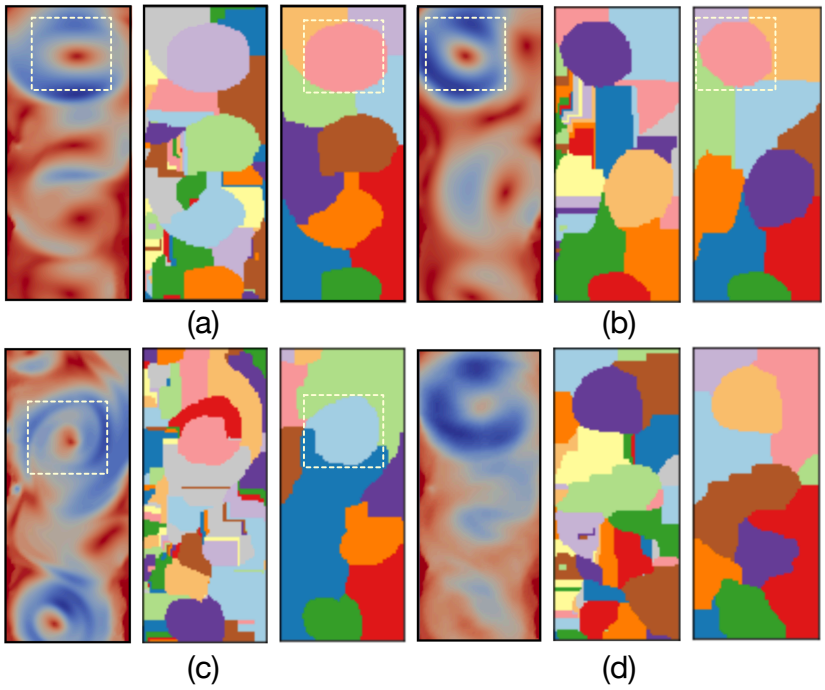

Figure 8. (a-c) Three ensemble members together with (d) the mean-field. Each subfigure visualizes, from left to right, the negated velocity magnitude field (red means low- and blue means high-velocity magnitude), and its corresponding Morse complexes before and after persistence simplification. The dotted white boxes illustrate the vortex features of each member.

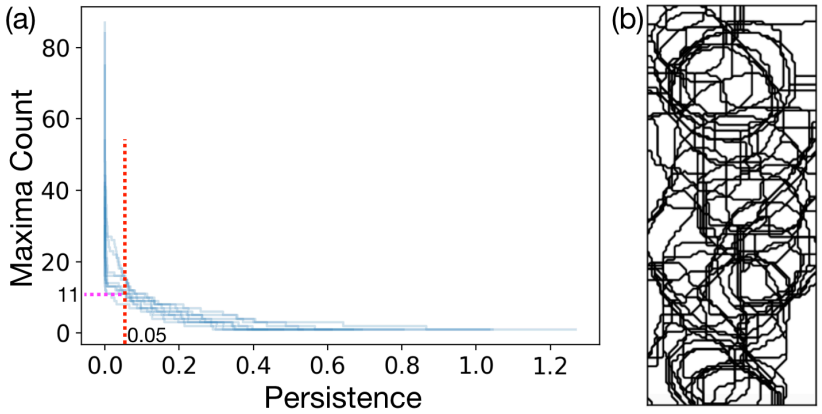

Figure 9. Persistence simplification. (a) Persistence graphs. (b) Spaghetti plots of the simplified Morse complexes.

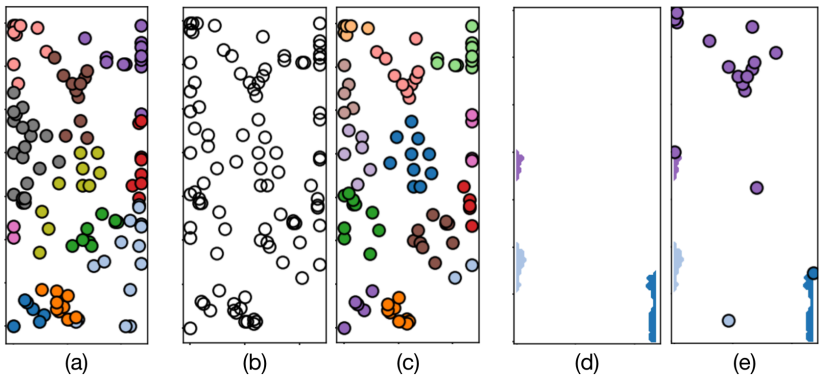

Figure 10. (a) Labeling with Morse mapping. (b-c) Labeling with k-means clustering. (d) Mandatory maxima are shown as colored regions. (e) Labeling with nearest mandatory maxima.

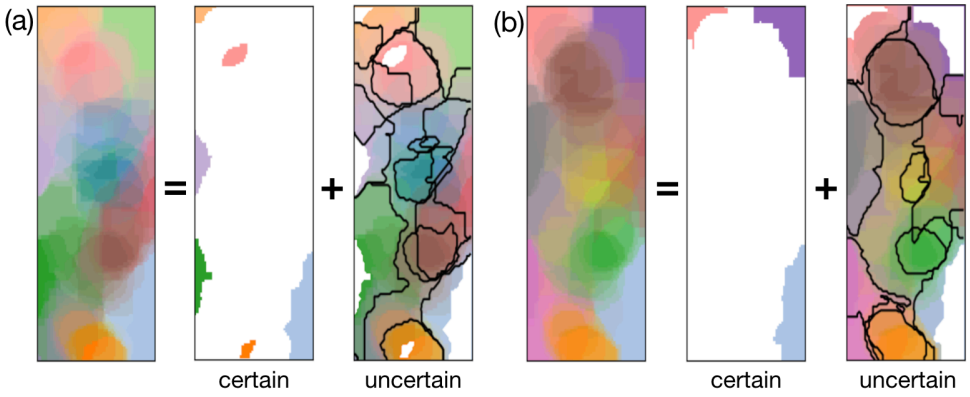

Figure 11. The probabilistic map is visualized based on (a) k-means clustering and (b) Morse mapping strategies.

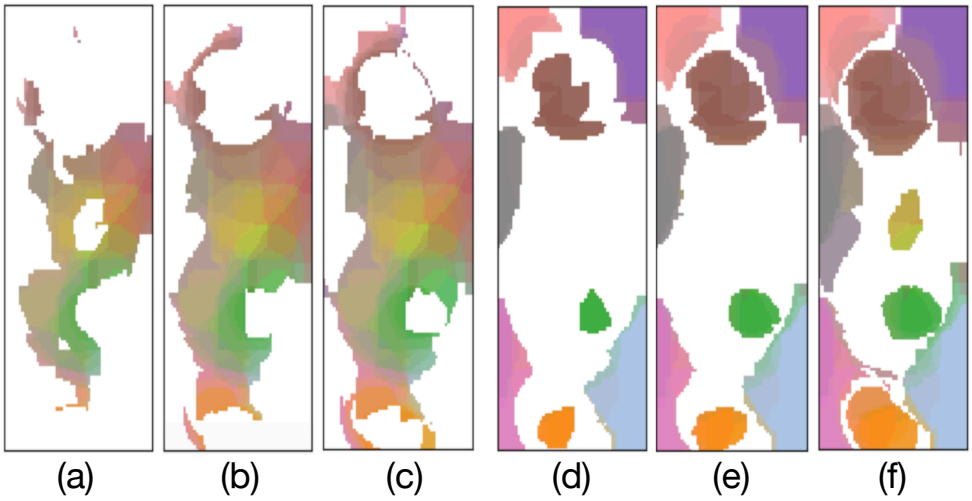

Figure 12. (a-c) Entropy-based exploration of uncertain regions representing entropy greater than or equal to 1.5 in (a), 1.25 in (b), and 1 in (c), respectively; (d-f) visualizations of the regions that agree in their gradient destinations for at least 80% members in (d), 70% members in (e), and 60% members in (f), respectively.

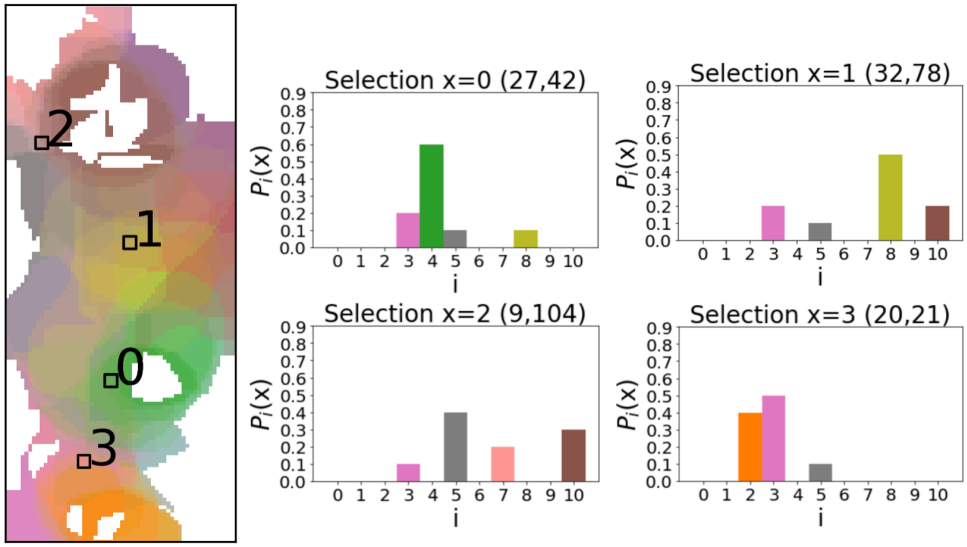

Figure 13. The probabilistic map for the positions with entropy greater than or equal to 0.8. The query selections $0-3$ have the highest probability of flowing to the maxima with green, yellow, gray, and pink labels, respectively.

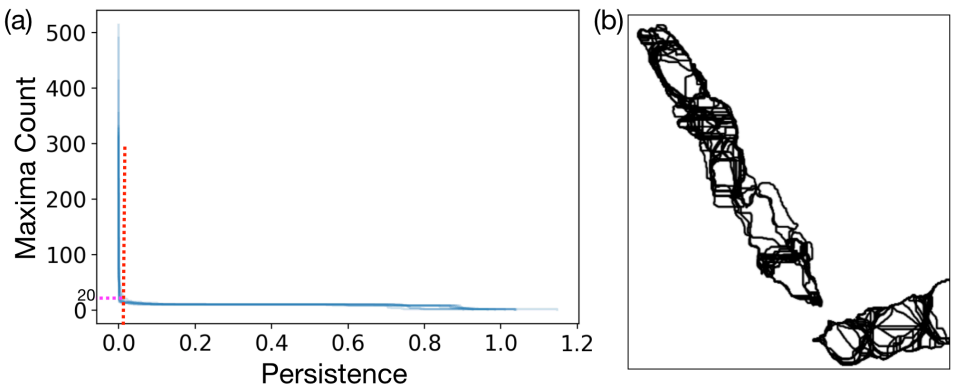

Figure 14. Persistence simplification. (a) Persistence graphs. (b) Spaghetti plots of the simplified Morse complexes.

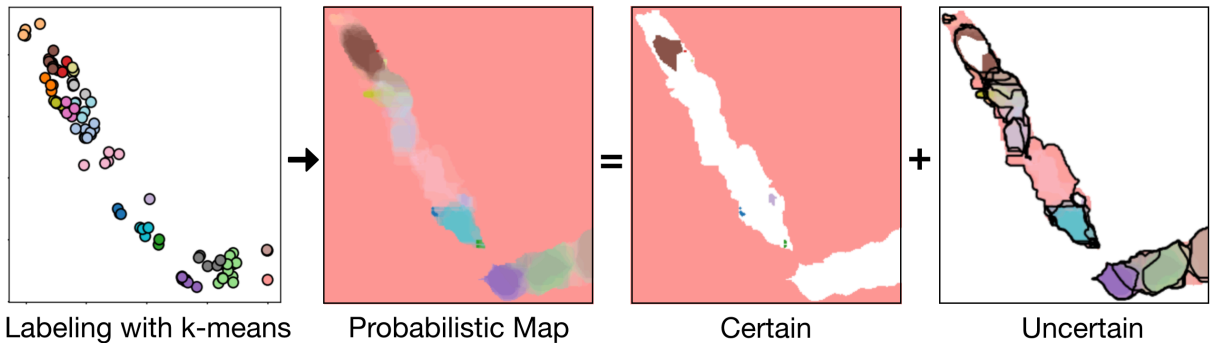

Figure 15. The probabilistic map visualized as a combination of certain and uncertain regions based on a labeling with k-means clustering. The black contours in the uncertain regions represent the expected Morse complex boundaries.

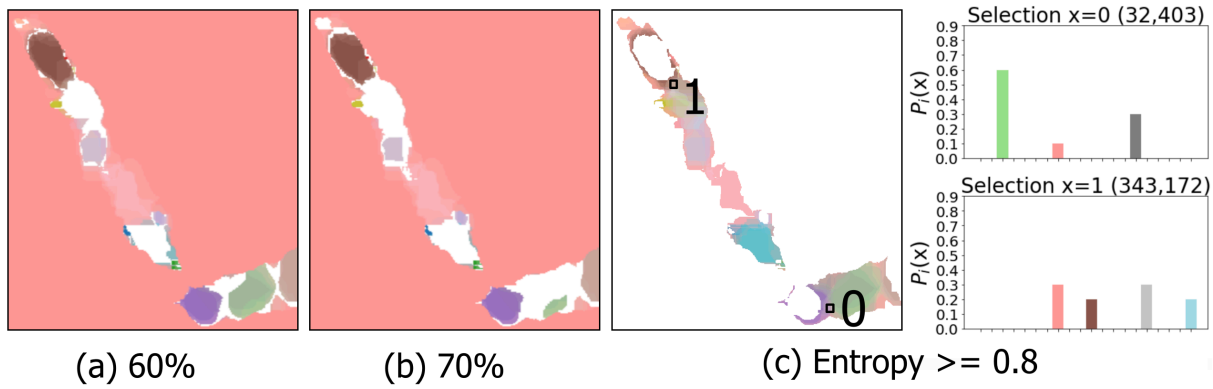

Figure 16. Exploration of a probabilistic map. Visualizations of the regions that agree in their gradient destinations for at least 60% members in (a) and 70% members in (b). Visualization of the regions for the lower entropy threshold of 0.8 along with the interactive probability queries in (c).

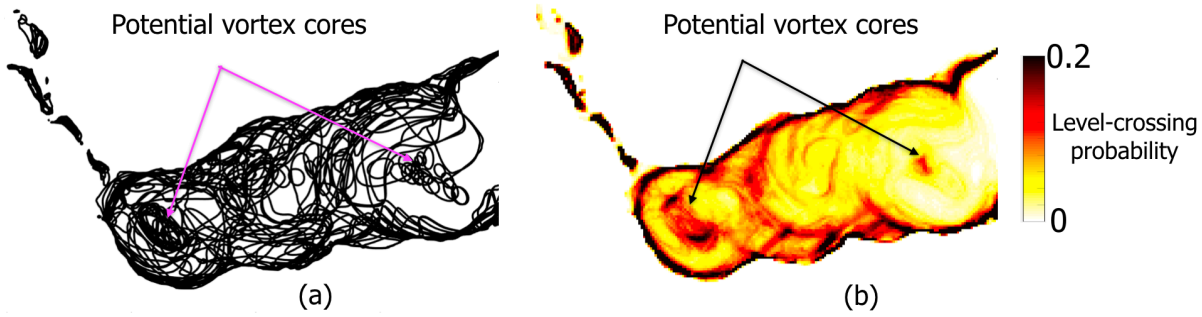

Figure 17. Uncertain level-set visualizations for the isovalue 0.2. (a) Spaghetti plots, (b) probabilistic marching squares.