Uncertainty Visualization of Critical Points of Scalar Fields for Parametric and Nonparametric Probabilistic Models

Tushar M. Athawale
Oak Ridge National Laboratory

Zhe Wang
Oak Ridge National Laboratory

David Pugmire
Oak Ridge National Laboratory

Kenneth Moreland
Oak Ridge National Laboratory

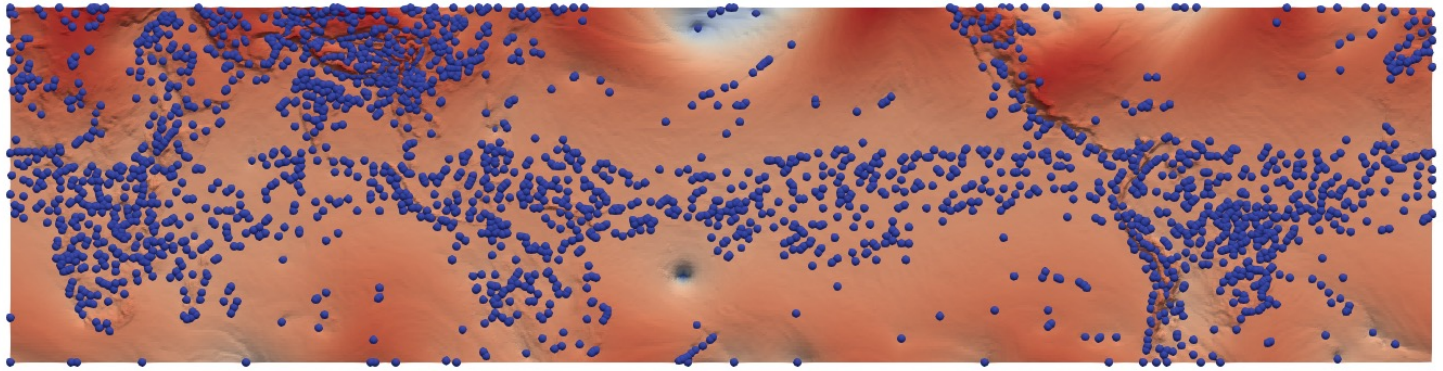
Qian Gong
Oak Ridge National Laboratory

Scott Klasky
Oak Ridge National Laboratory

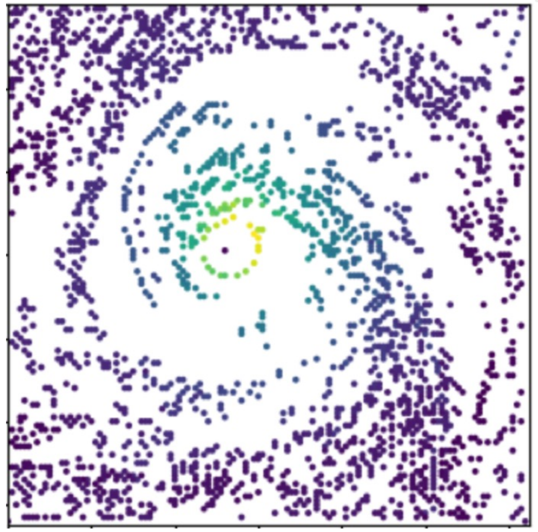
Chris R. Johnson
Scientific Computing & Imaging Institute, University of Utah

Paul Rosen
Scientific Computing & Imaging Institute, University of Utah

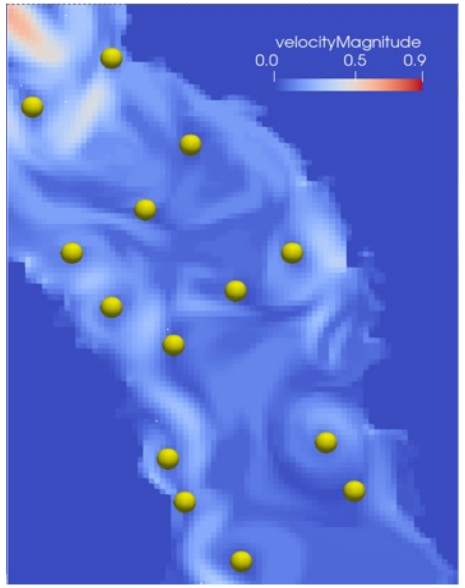
Critical Point Visualization



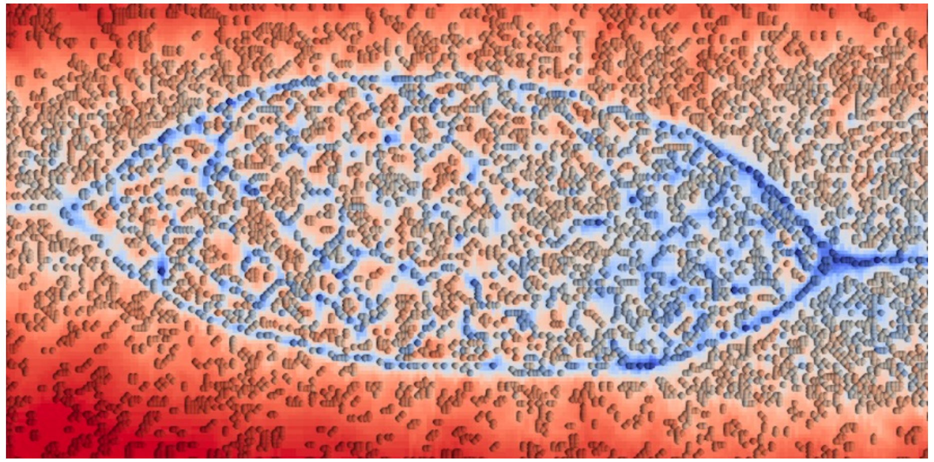
Climatology



Hurricane Tracking



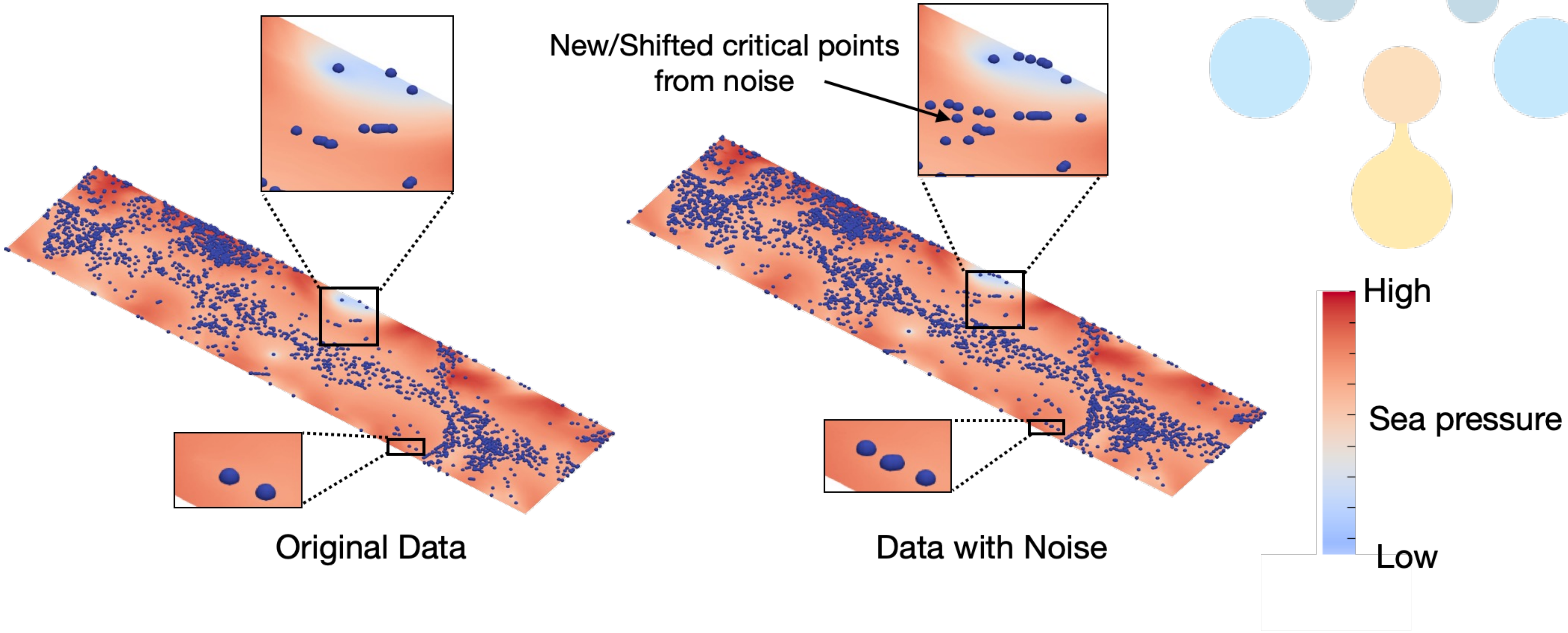
Oceanology



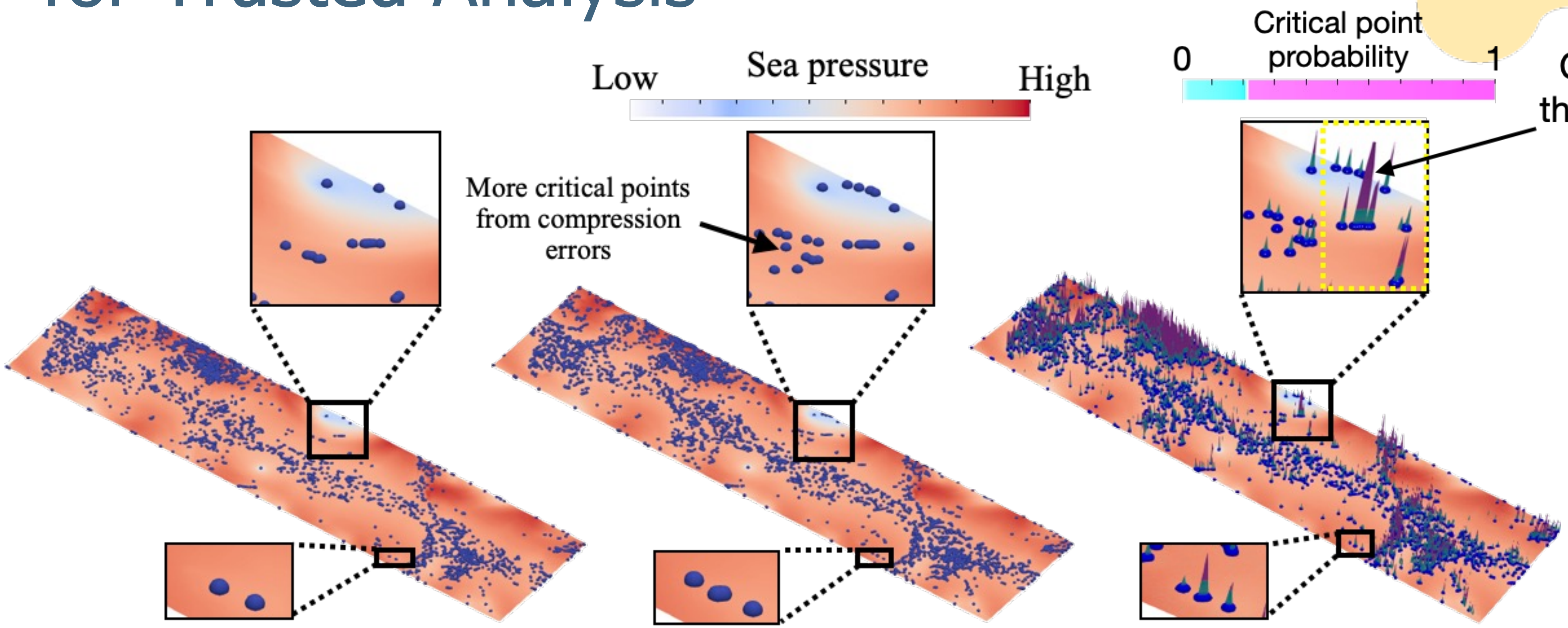
Biology

Critical point visualization with the Topology Toolkit (TTK)
[Tierny et al., 2018]

Noise in Data Creates Uncertainty in Critical Point Positions



Uncertainty Visualization of Critical Points for Trusted Analysis

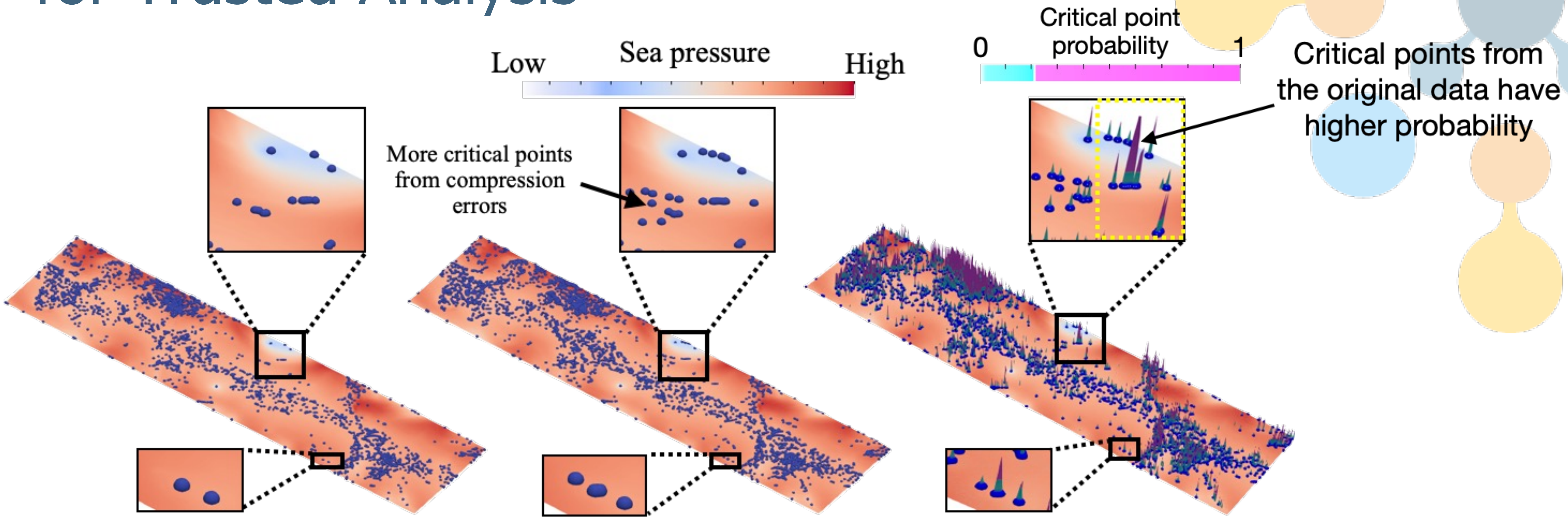


(a) Original critical points

(b) No uncertainty visualization

(c) With Uncertainty visualization

Uncertainty Visualization of Critical Points for Trusted Analysis



(a) Original critical points

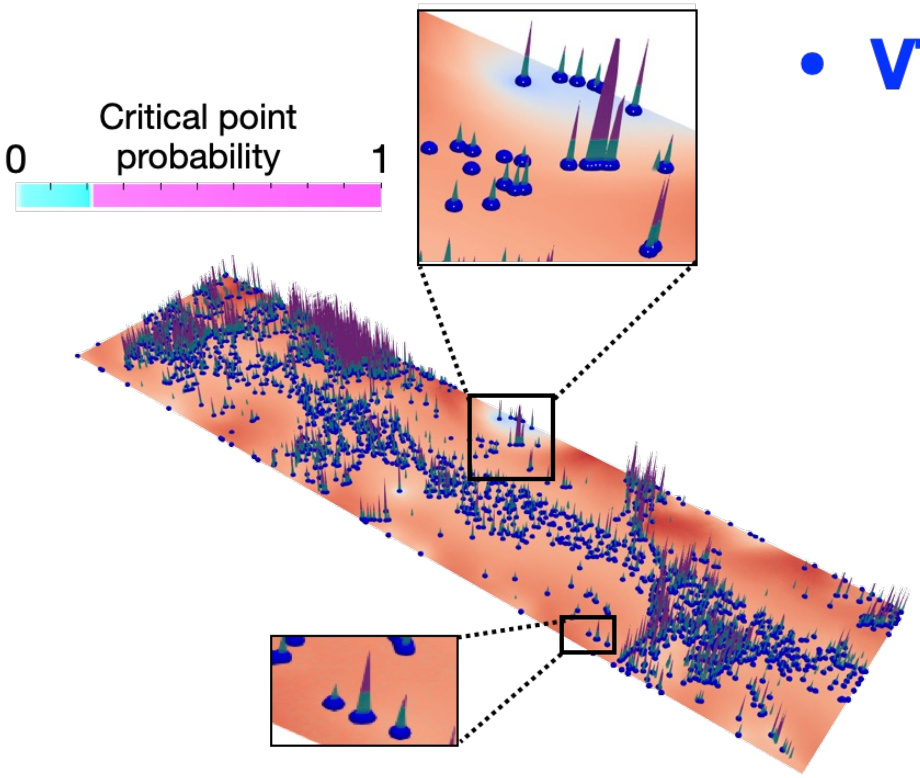
(b) No uncertainty visualization

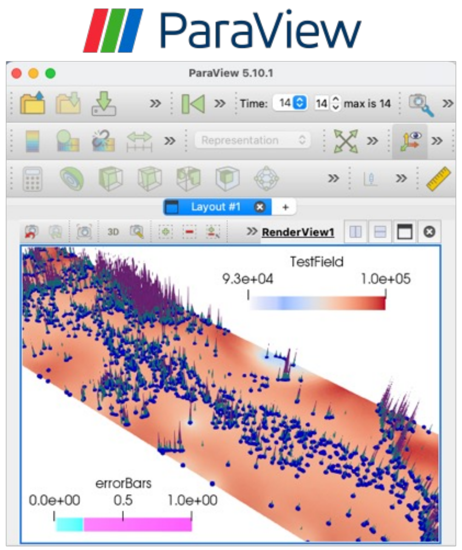
(c) With Uncertainty visualization

State of the art [Petz et al., 2012, Liebmann and Scheuermann, 2016, Vietinghoff et al., 2022]:
(1) resorts to computationally expensive **Monte Carlo** sampling
(2) restricted to **parametric** (e.g., uniform, Gaussian) noise models

Our Contribution

- **Closed-form solutions** (Upto 411× speed-up compared to Monte Carlo)
- **Nonparametric models (e.g., histograms)**
- **VTK-m [Moreland et al., 2016] GPU acceleration**





Makes integration with visualization tools practical!

(Upto 1646× faster and more accurate than conventional Monte Carlo sampling)

Uncertainty Visualization: Other Related Work

- Confidence intervals and likelihood of critical points
  [Mihai and Westermann, 2014, Günther et al., 2014, Vietinghoff et al., 2022]

- Uncertainty visualization algorithms
  Scalar fields (Isosurfaces [Pöthkow et al. 2013, Athawale et al. 2016], Direct Volume Rendering [Liu et al., 2012], Contour trees [Wu et al. 2012, Yan et al., 2020], Persistence diagrams [Vidal et al., 2020]), **Vector fields** (Streamlines [Ferstl et al., 2016], Finite-time Lyapunov exponents [Guo et al., 2016]), **Tensor fields** (diffusion tensor [Siddiqui et al., 2021] and HARDI [Jiao et al., 2012] imaging)

- Distribution models of uncertainty
  Independent uniform/Gaussian [Athawale et al., 2013, Günther et al., 2014], Correlated Gaussian [Pöthkow et al. 2013, Petz et al., 2016], Nonparametric [Pöthkow et al. 2013, Liu et al., 2012, Athawale et al. 2020]

- Acceleration of uncertainty computation
  ML for uncertainty [Han et al., 2022], FunMC^2: GPU acceleration [Wang et al., 2023], Hierarchical data structures [Li et al., 2024]

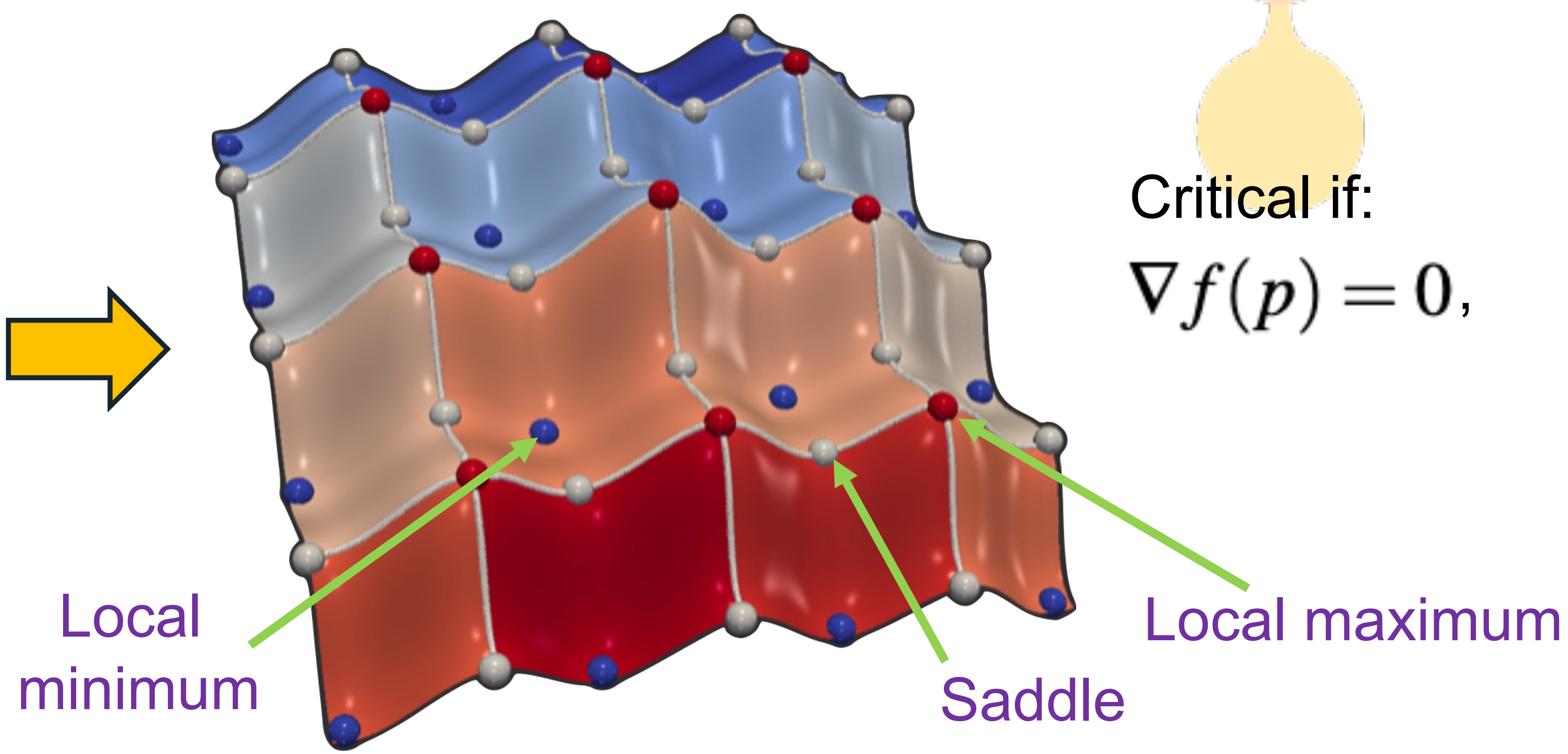
- Rendering of uncertainty
  Colormapping [Rhodes et al., 2003], Elevation maps [Petz et al., 2012], Glyphs [Wittenbrink et al., 1996]

Background and Problem Statement

Critical points in deterministic data

(This work only considers uniform-grided data)

2	3	2	3	2	3	2
3	4	3	5	3	4	3
2	3	2	3	2	3	2
4	5	4	6	4	5	4
2	3	2	3	2	3	2
3	4	3	5	3	4	3
2	3	2	3	2	3	2



Heightmap

Critical if:

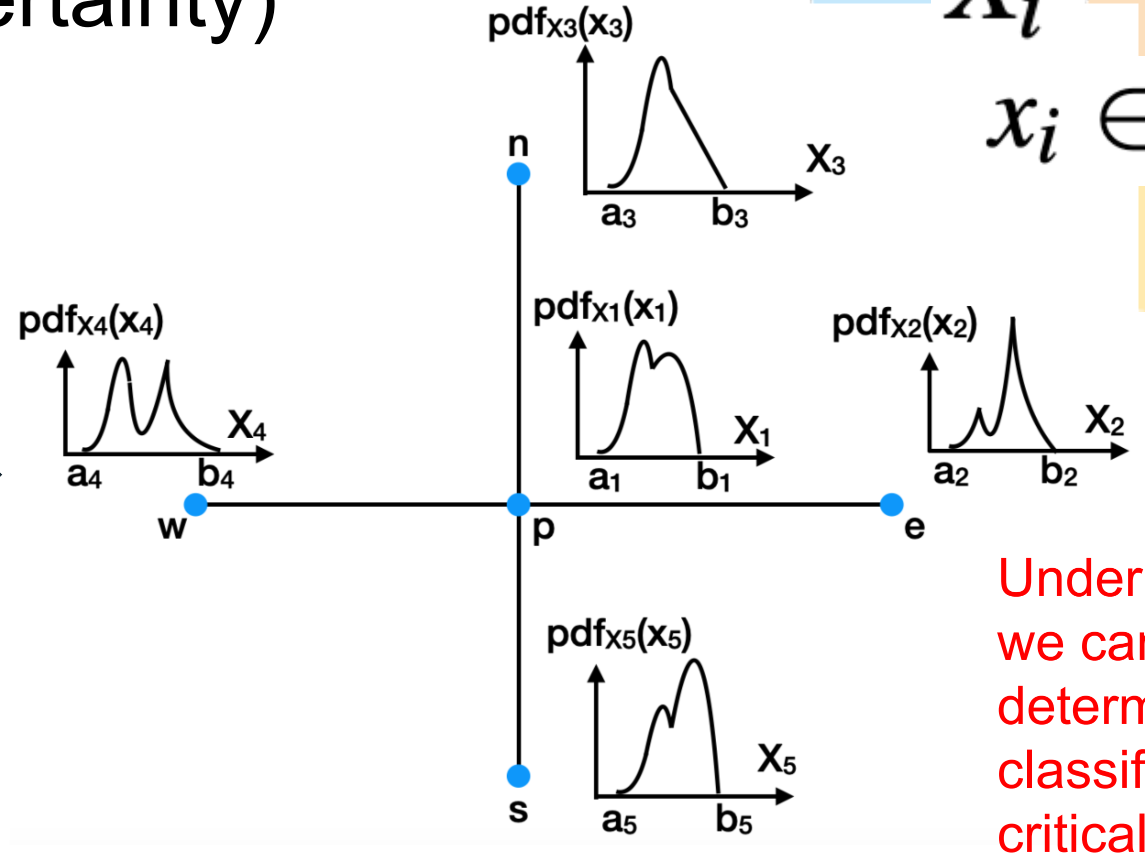
$$\nabla f(p) = 0,$$

Background and Problem Statement

Critical points in <u>uncertain</u> data?
(Most real data have uncertainty)

2 ± 0.1	3 ± 0.4	2 ± 0.1	3 ± 0.2	2 ± 0.1	3 ± 0.4	2 ± 0.1
3 ± 0.2	4 ± 0.3	3 ± 0.1	5 ± 0.2	3 ± 0.4	4 ± 0.3	3 ± 0.2
2 ± 0.1	3 ± 0.2	2 ± 0.3	3 ± 0.1	2 ± 0.2	3 ± 0.5	2 ± 0.1
4 ± 0.2	5 ± 0.4	4 ± 0.1	6 ± 0.2	4 ± 0.3	5 ± 0.2	4 ± 0.1
2 ± 0.2	3 ± 0.3	2 ± 0.1	3 ± 0.5	2 ± 0.1	3 ± 0.2	2 ± 0.4
3 ± 0.3	4 ± 0.1	3 ± 0.4	5 ± 0.1	3 ± 0.3	4 ± 0.1	3 ± 0.2
2 ± 0.3	3 ± 0.4	2 ± 0.5	3 ± 0.1	2 ± 0.1	3 ± 0.4	2 ± 0.2

$$X_i \sim \mathrm{Pdf}_{X_i}(x_i)$$
$$x_i \in [a_i, b_i]$$



Under uncertainty, we cannot deterministically classify if a point is critical!

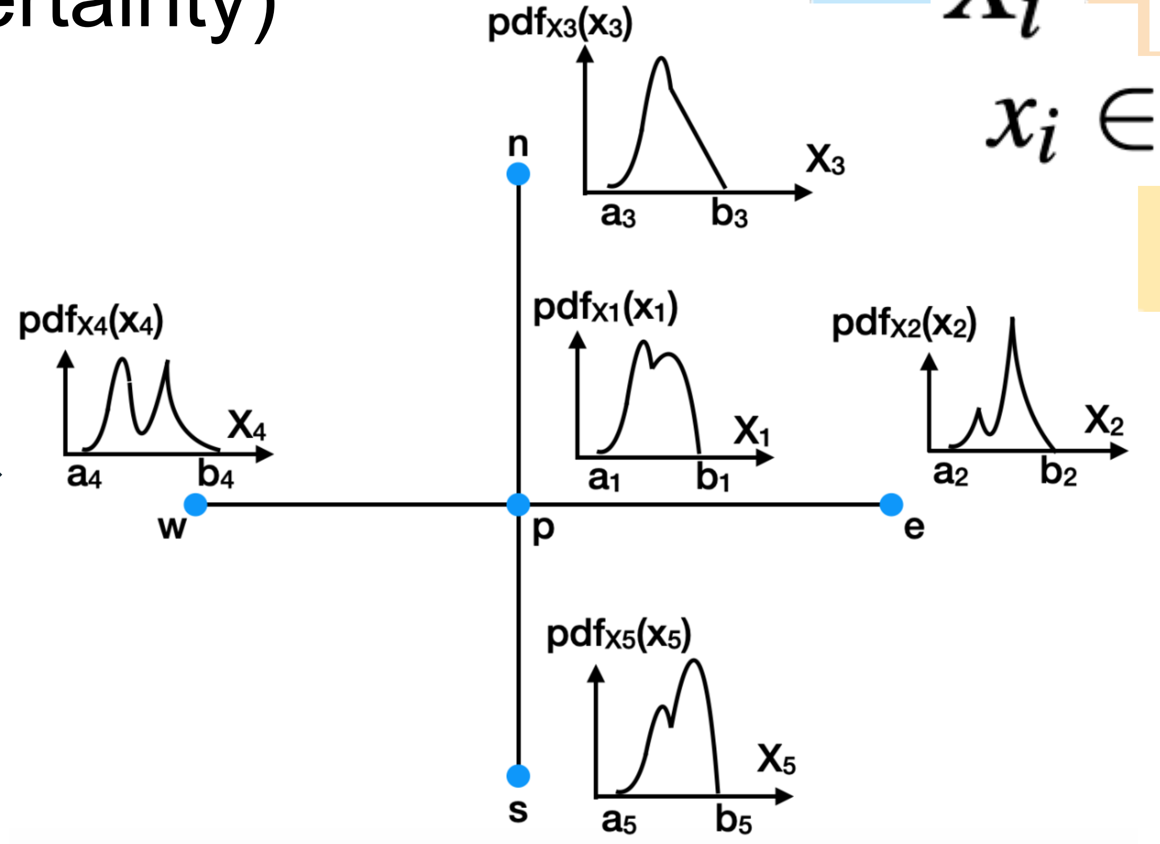
Assumption: uncertainty over finite support

Background and Problem Statement

Critical points in uncertain data?
(Most real data have uncertainty)

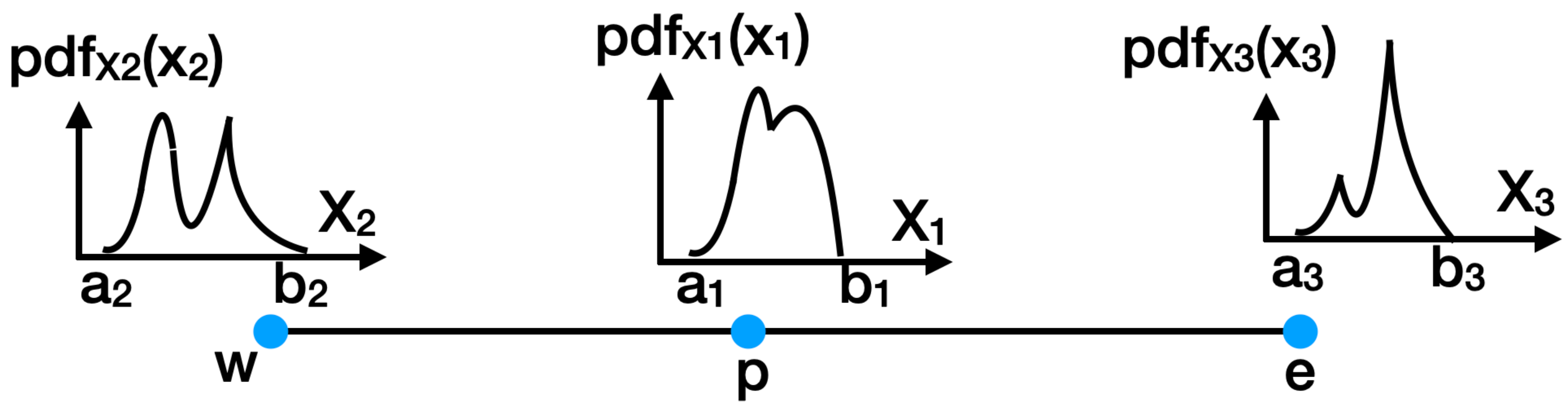
2±0.1	3±0.4	2±0.1	3±0.2	2±0.1	3±0.4	2±0.1
3±0.2	4±0.3	3±0.1	5±0.2	3±0.4	4±0.3	3±0.2
2±0.1	3±0.2	2±0.3	3±0.1	2±0.2	3±0.5	2±0.1
4±0.2	5±0.4	4±0.1	6±0.2	4±0.3	5±0.2	4±0.1
2±0.2	3±0.3	2±0.1	3±0.5	2±0.1	3±0.2	2±0.4
3±0.3	4±0.1	3±0.4	5±0.1	3±0.3	4±0.1	3±0.2
2±0.3	3±0.4	2±0.5	3±0.1	2±0.1	3±0.4	2±0.2

$$X_i \sim \mathrm{Pdf}_{X_i}(x_i)$$
$$x_i \in [a_i, b_i]$$



What is the probability of “point p” to be a local maximum, local minimum, or saddle?

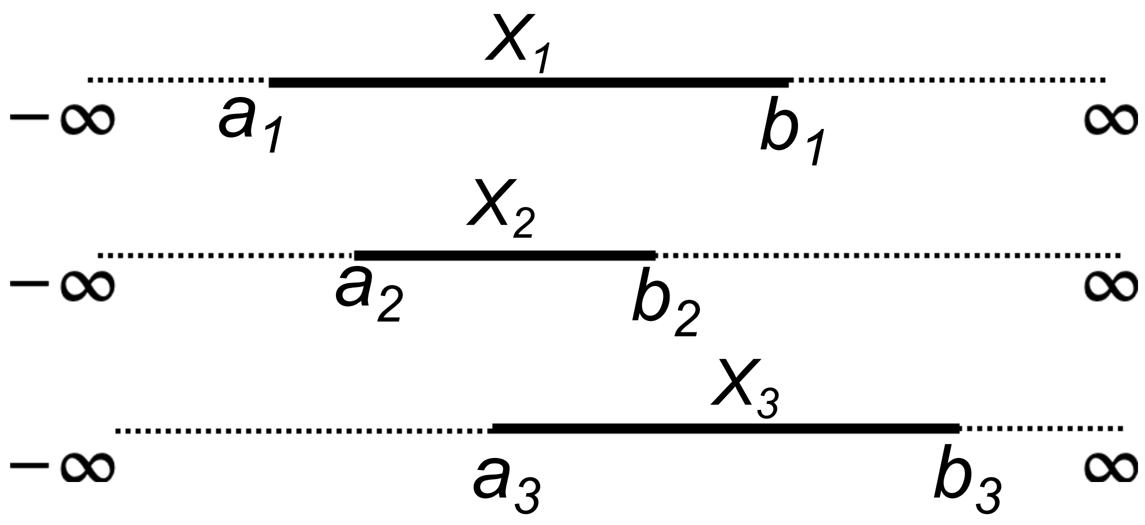
Background and Problem Statement



$$X_i \sim \mathrm{Pdf}_{X_i}(x_i)$$
$$x_i \in [a_i, b_i]$$

What is the probability of "point p" to be a local minimum? (1D case)?

Approach

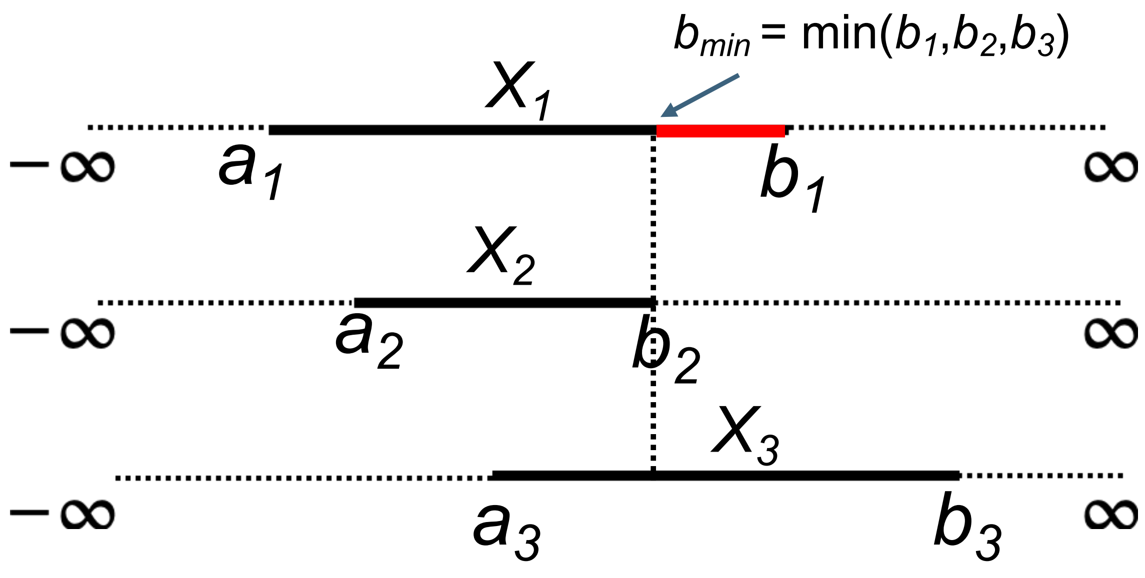


Pr(local minimum)

$$= \int_{X_1} \int_{X_1 < X_2} \int_{X_1 < X_3} (\mathbf{Pdf})_{joint} \mathbf{dx}_1 \mathbf{dx}_2 \mathbf{dx}_3$$

Independence assumption: $\mathbf{Pdf}_{joint} = \mathbf{Pdf}_{X_1}(x_1)\mathbf{Pdf}_{X_2}(x_2)\mathbf{Pdf}_{X_3}(x_3)$

Approach: Piecewise Integral



Pr(local minimum)

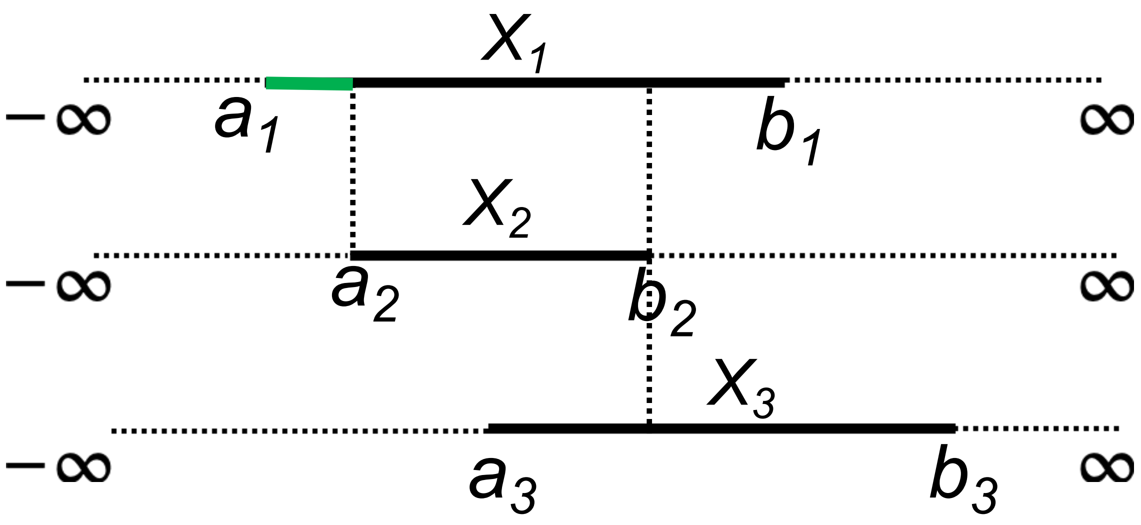
$$= \int_{X_1} \int_{X_1<X_2} \int_{X_1<X_3} (\mathbf{Pdf})_{joint} \mathbf{dx}_1 \mathbf{dx}_2 \mathbf{dx}_3$$

simplifies to

$$= 0$$

The red range is always greater than X_2!

Approach: Piecewise Integral



The green range is always smaller than X_2 and X_3

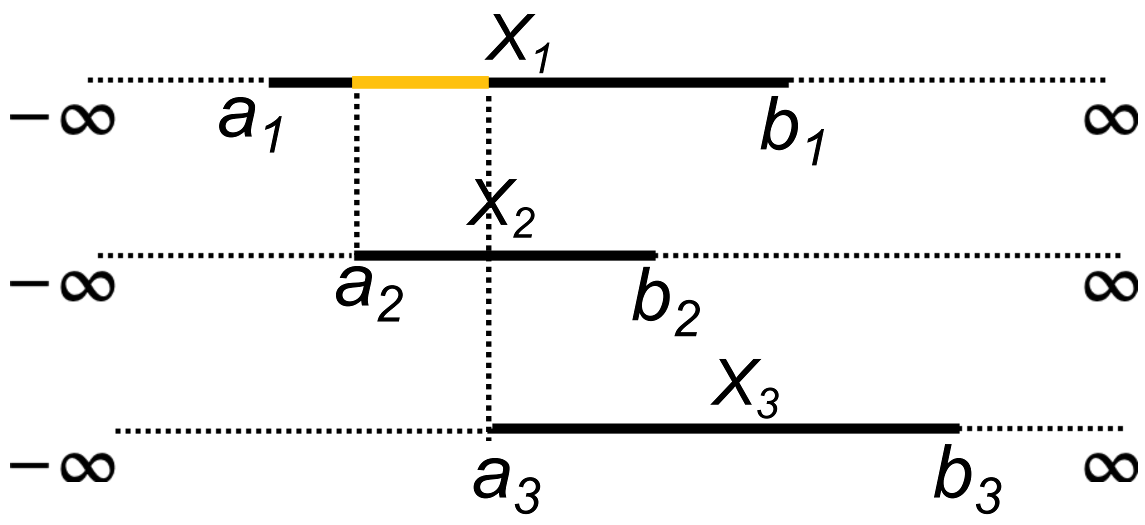
Pr(local minimum)

$$= \int_{X_1} \int_{X_1<X_2} \int_{X_1<X_3} (\mathbf{Pdf})_{joint} \mathbf{dx}_1 \mathbf{dx}_2 \mathbf{dx}_3$$

simplifies to

$$= \int_{X_1} \mathrm{Pdf}_{X1}$$

Approach: Piecewise Integral



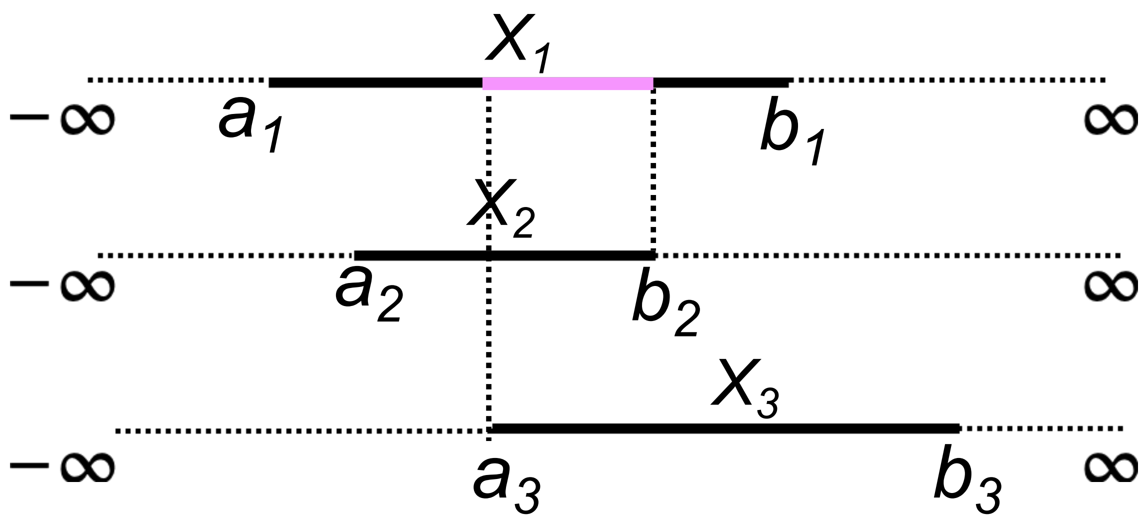
Pr(local minimum)

$$= \int_{X_1} \int_{X_1<X_2} \int_{X_1<X_3} (\mathbf{Pdf})_{joint} \mathbf{dx}_1 \mathbf{dx}_2 \mathbf{dx}_3$$

simplifies to

$$= \int_{X_1} \int_{X_1<X_2} \mathrm{Pdf}_{X1} \ \mathrm{Pdf}_{X2}$$

Approach: Piecewise Integral



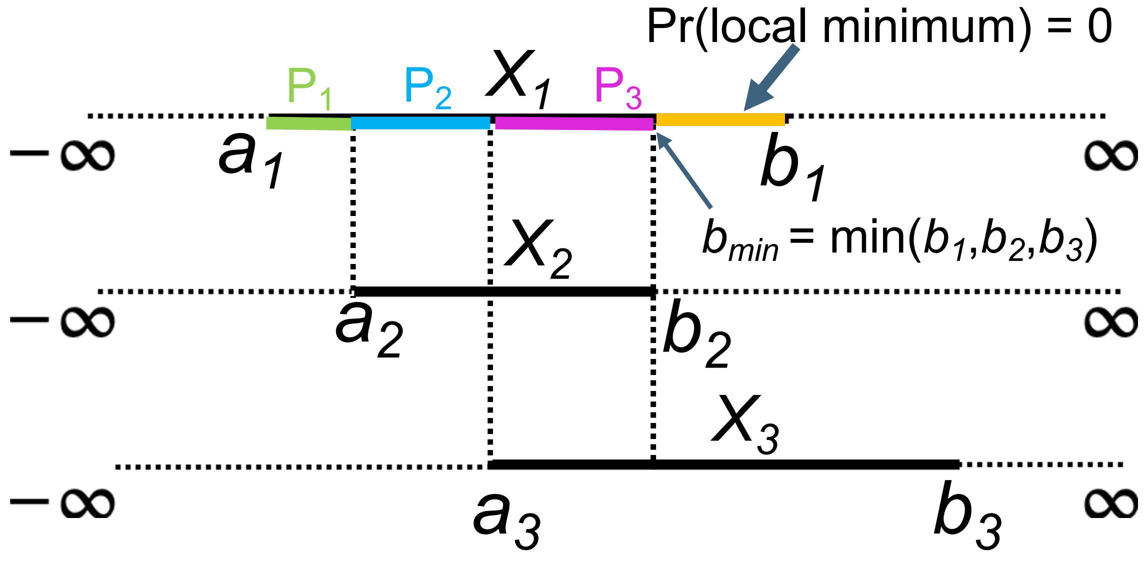
Pr(local minimum)

$$= \int_{X_1} \int_{X_1<X_2} \int_{X_1<X_3} (\mathbf{Pdf})_{joint}\, \mathbf{dx}_1 \mathbf{dx}_2 \mathbf{dx}_3$$

simplifies to

$$= \int_{X_1} \int_{X_1<X_2} \int_{X_1<X_3} \mathrm{Pdf}_{joint}$$

Approach: Observations for the Integral Computation Algorithm



Pr(local minimum) =

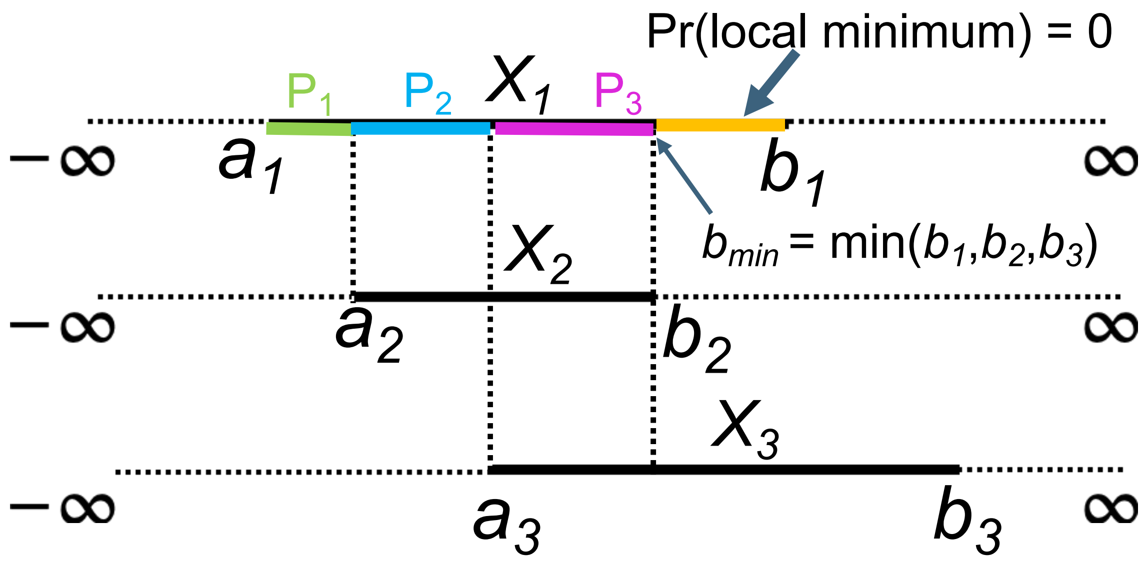
$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} (\mathbf{Pdf})_{joint}\,\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3$$

(1) Pieces depend on the order of start points [a_1, a_2, a_3] and b_{min}

(2) Four types of integration simplifications (integration templates):

$$\int_{X_1} Pdf_{X1}\ , \quad \int_{X_1}\int_{X_1<X_2} Pdf_{X1}\ Pdf_{X2}\ , \quad \int_{X_1}\int_{X_1<X_3} Pdf_{X1}\ Pdf_{X3}\ , \quad \int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} Pdf_{joint}$$

Approach: Observations for the Integral Computation Algorithm



Pr(local minimum) =

$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} (\mathbf{Pdf})_{joint}\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3$$

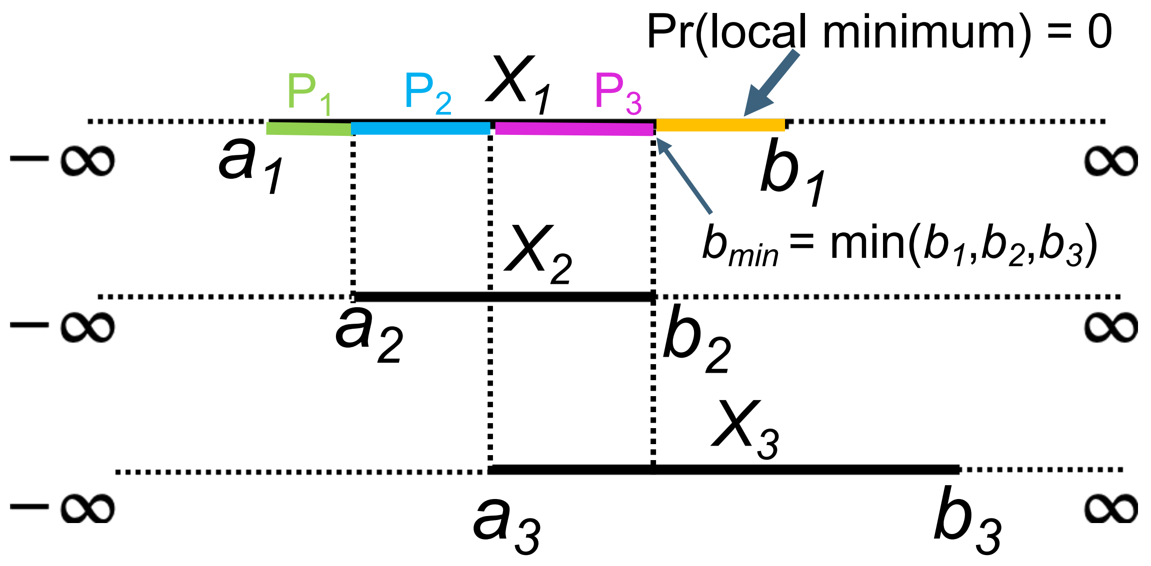
(1) Pieces depend on the order of start points [a_1, a_2, a_3] and b_{min}

(2) Four types of integration simplifications (integration templates):

$$\int_{X_1} Pdf_{X1}, \quad \int_{X_1}\int_{X_1<X_2} Pdf_{X1}\, Pdf_{X2}, \quad \int_{X_1}\int_{X_1<X_3} Pdf_{X1}\, Pdf_{X3}, \quad \int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} Pdf_{joint}$$

Thanks to Wolfram Alpha!!

Approach: Local Minimum Probability Algorithm



Pr(local minimum) =

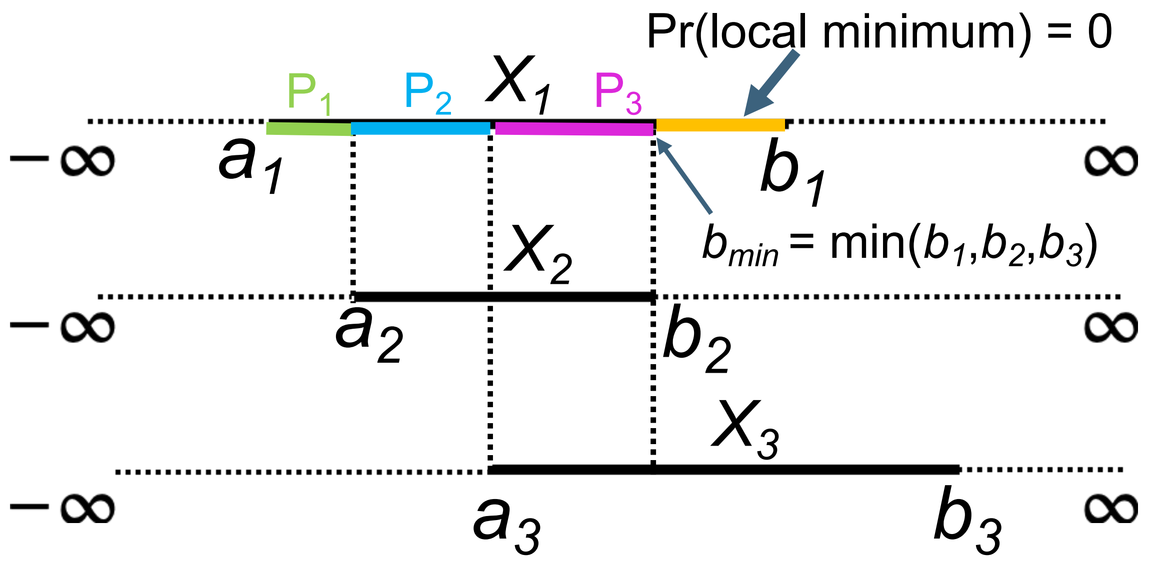
$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} (\mathbf{Pdf})_{joint}\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3$$

<u>Algorithm</u>:

(1) Sort a_1, a_2, $a_{3,}$ and b_{min} to determine pieces P_i

(2) Compute (precomputed) integral templates on the fly and sum them up

Approach: Local Minimum Probability Algorithm



Pr(local minimum) =

$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3} (\mathbf{Pdf})_{joint}\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3$$

Time complexity: O(nlogn), n = # start points a_i **(extremely efficient and accurate than Monte Carlo!)**

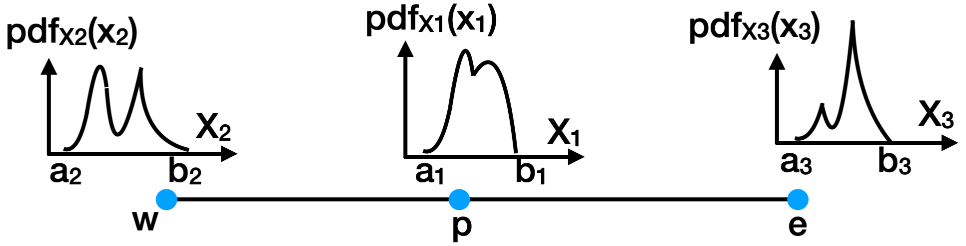
<u>Algorithm</u>:

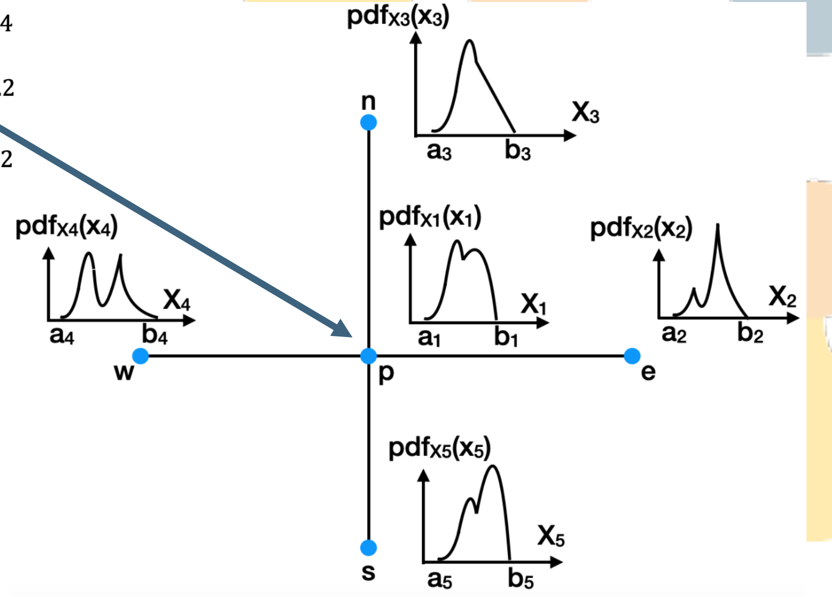
(1) Sort a_1, a_2, a_3, and b_{min} to determine pieces P_i

(2) Compute (precomputed) integral templates on the fly and sum them up

Algorithmic Intricacies

2±0.1	3±0.4	2±0.1	3±0.2	2±0.1	3±0.4	2±0.1
3±0.2	4±0.3	3±0.1	5±0.2	3±0.4	4±0.3	3±0.2
2±0.1	3±0.2	2±0.3	3±0.1	2±0.2	3±0.5	2±0.1
4±0.2	5±0.4	4±0.1	6±0.2	4±0.3	5±0.2	4±0.1
2±0.2	3±0.3	2±0.1	3±0.5	2±0.1	3±0.2	2±0.4
3±0.3	4±0.1	3±0.4	5±0.1	3±0.3	4±0.1	3±0.2
2±0.3	3±0.4	2±0.5	3±0.1	2±0.1	3±0.4	2±0.2





$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3}(\mathbf{Pdf})_{joint}\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3$$

$$\int_{X_1}\int_{X_1<X_2}\int_{X_1<X_3}\int_{X_1<X_4}\int_{X_1<X_5}(\mathbf{Pdf})_{joint}\mathbf{dx}_1\mathbf{dx}_2\mathbf{dx}_3\mathbf{dx}_4\mathbf{dx}_5$$

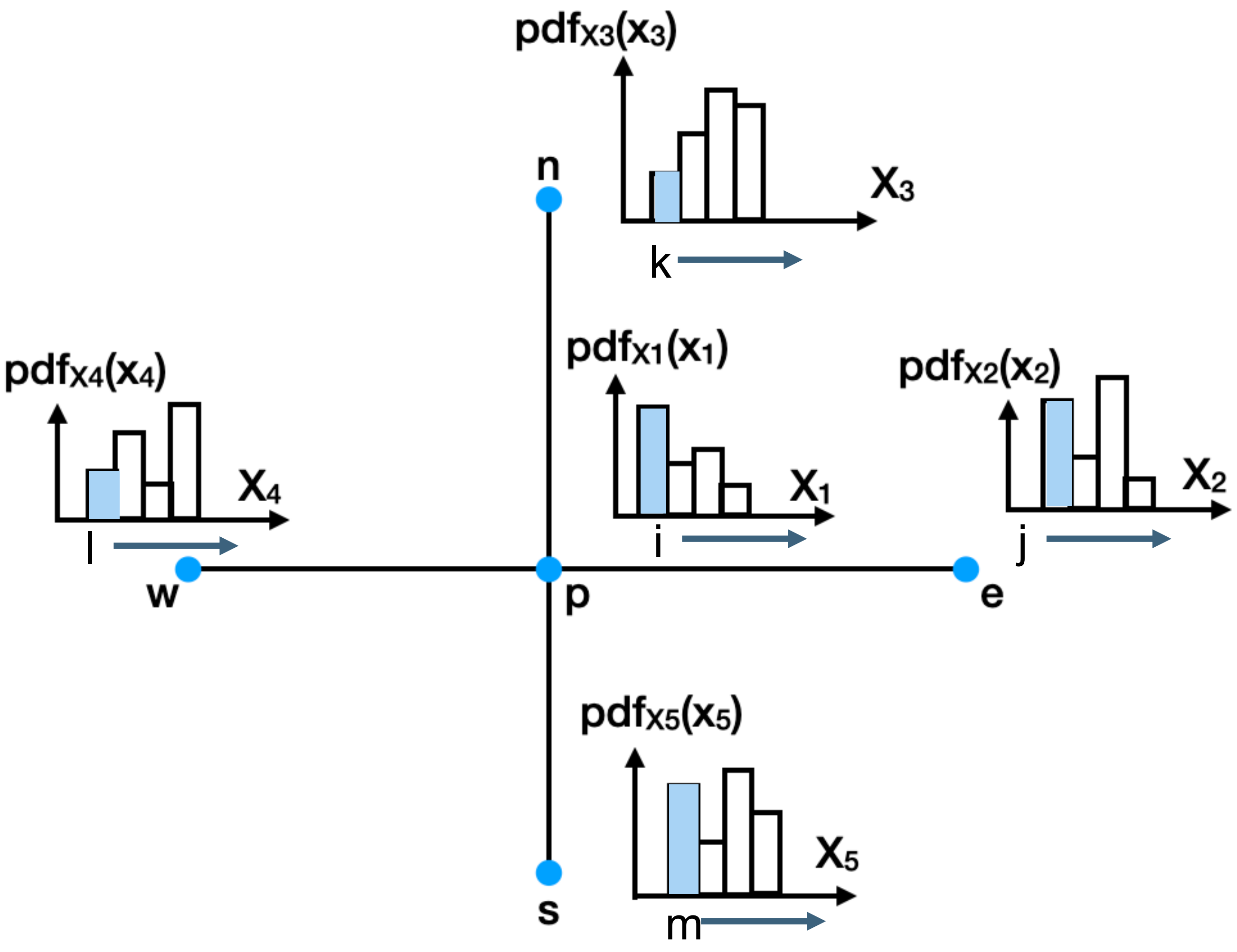
1D Case (2 neighbors)

2D Case (4 neighbors)

More integral templates/simplifications!

Algorithmic Intricacies: Nonparametric Noise Models

Capture more realistic shape of distributions compared to parametric models



Pr(local minimum)

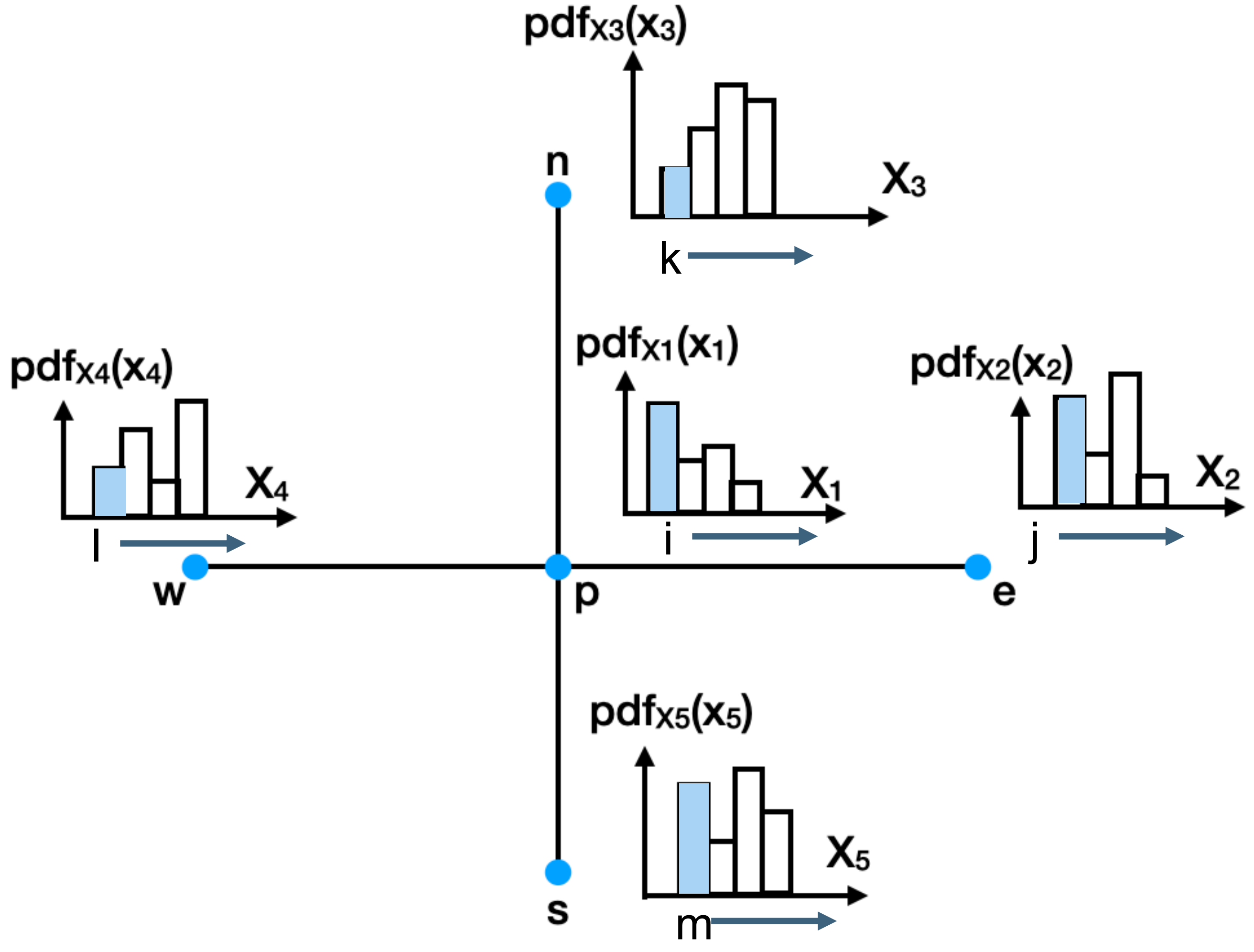
$$= w \sum_{i=1}^{i=h} \cdots \sum_{m=1}^{m=h} \Pr(p = l_{min})_{i,j,k,l,m},$$

$$w = w_i w_j w_k w_l w_m$$

h = # histogram bins

Algorithmic Intricacies: Nonparametric Noise Models

Capture more realistic shape of distributions compared to parametric models



Pr(local minimum)

$$= w \sum_{i=1}^{i=h} \cdots \sum_{m=1}^{m=h} \Pr(p = l_{min})_{i,j,k,l,m},$$

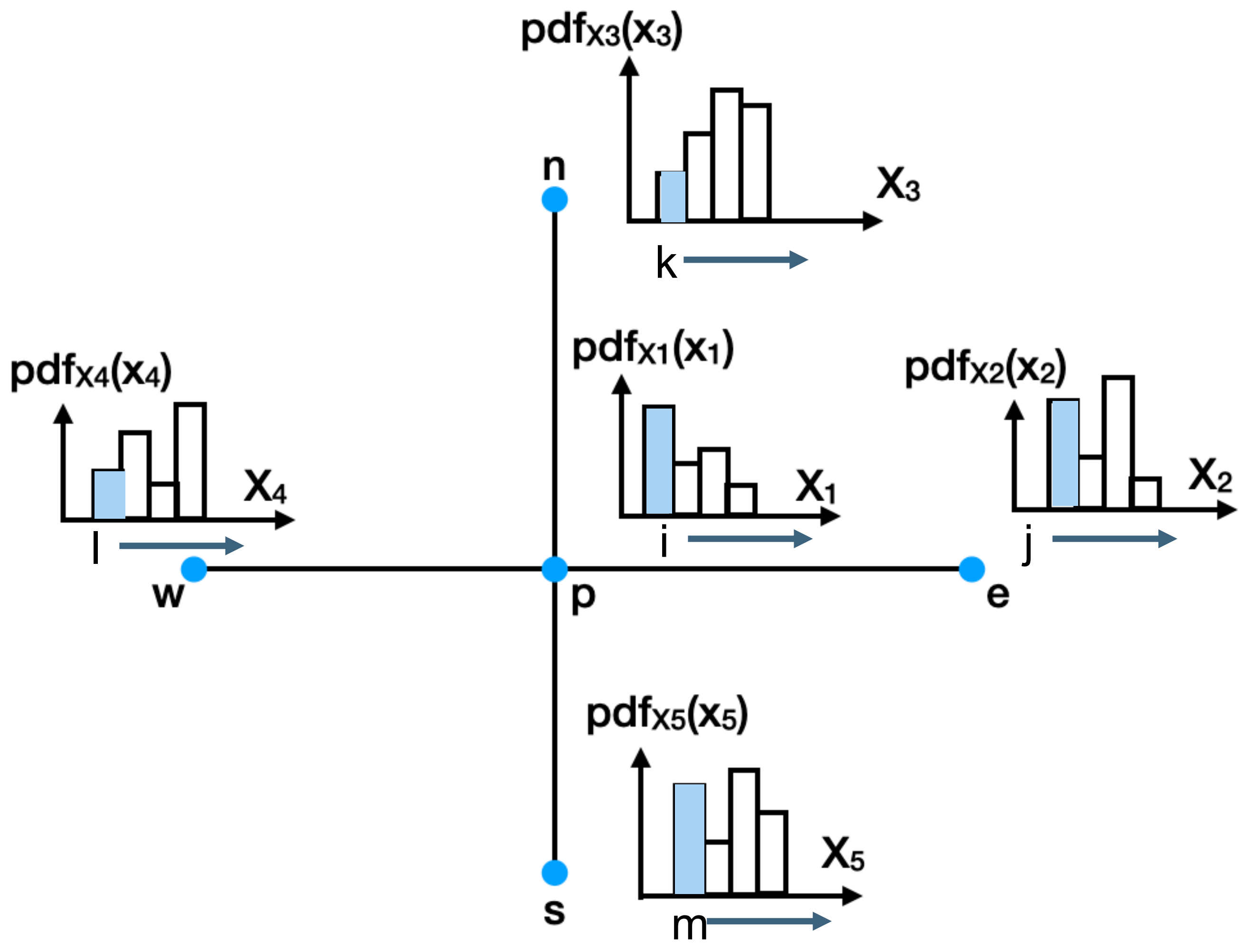
$$w = w_i w_j w_k w_l w_m$$

h = # histogram bins

Time complexity: O(h^5),
(UQ more accurate than parametric, but inefficient!)

Algorithmic Intricacies: Nonparametric Noise Models

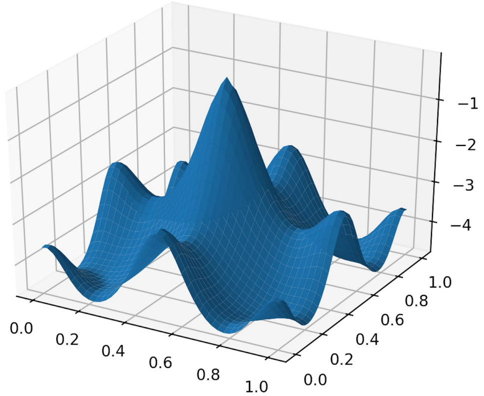
Two approaches to enhance the performance of nonparametric models:



- **Semianalytical solution**
  Time complexity: **O(nh),**
  (n: # Monte Carlo samples,
  h: # histogram bins)

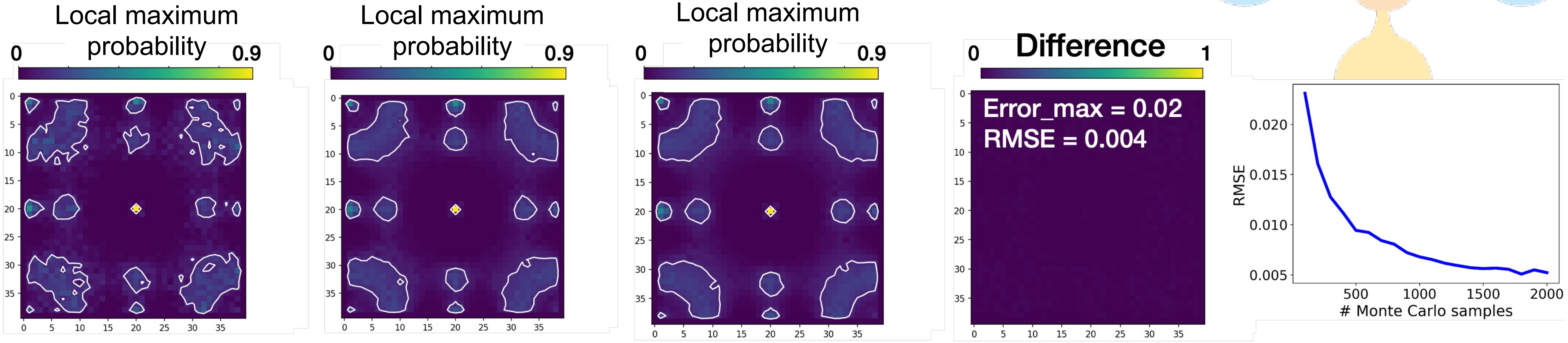
- **VTK-m GPU acceleration**
  Probability computation per point is independent of others

Results: Synthetic Data



Add uniform noise → 50 ensemble members

Ackley function [Ackley, 1987]



0.13 Seconds

(a) MC sampling (100 samples)

1.28 Seconds

(b) MC sampling (2000 samples)

0.02 Seconds

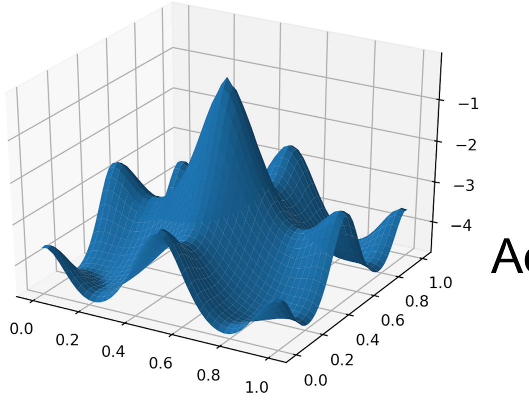
(c) Proposed closed-form solution

(64× speedup with respect to 2000 MC)

(d) Difference between subfigures (b) and (c)

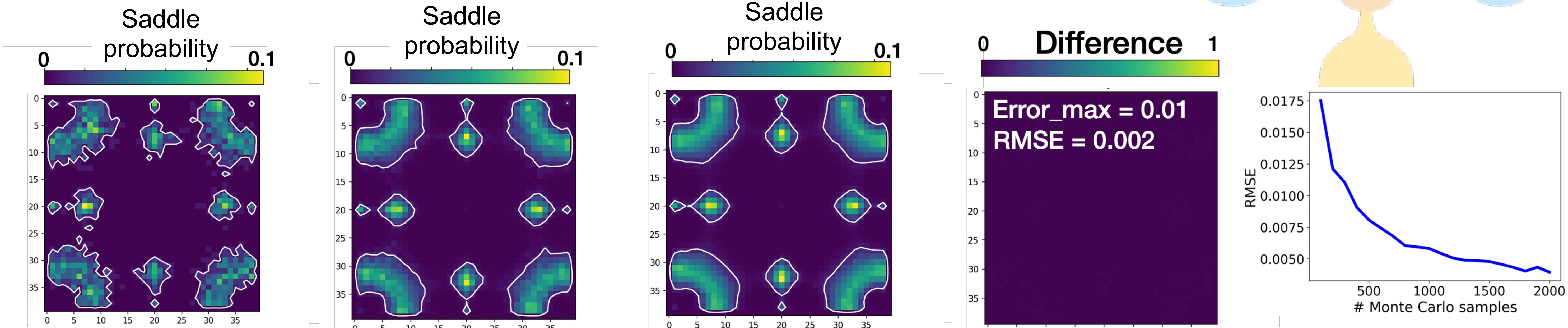
(e) Convergence curve

Results: Synthetic Data



Ackley function [Ackley, 1987]

Add uniform noise → 50 ensemble members



0.13 Seconds

(a) MC sampling (100 samples)

2.38 Seconds

(b) MC sampling (2000 samples)

0.02 Seconds

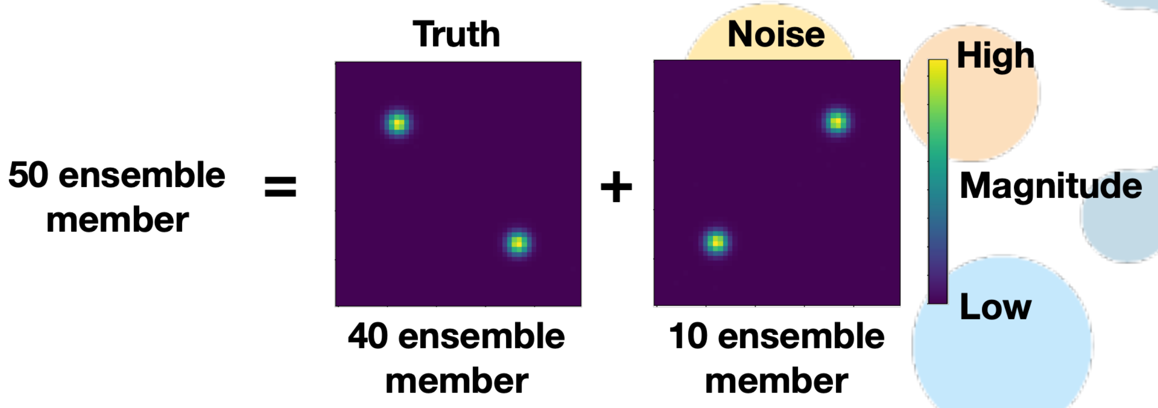
(c) Proposed closed-form solution

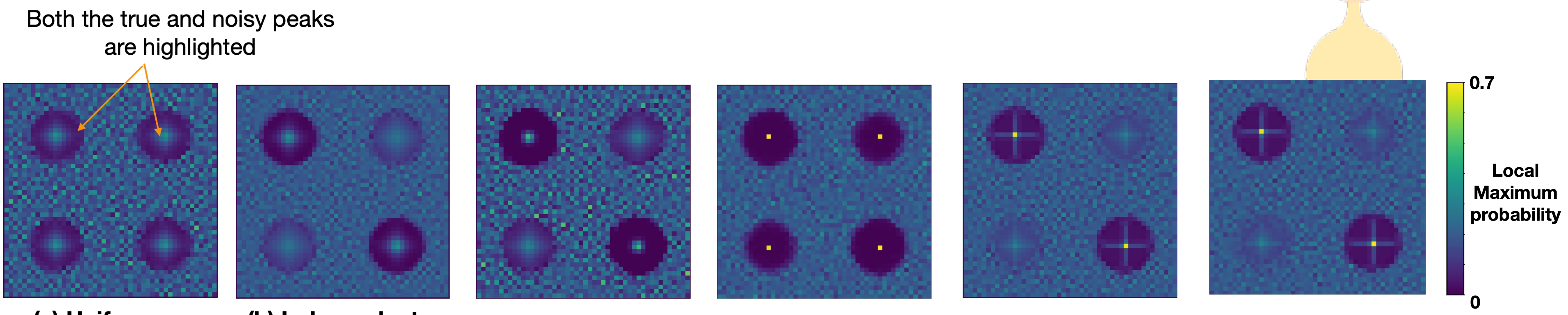
(119× speedup with respect to 2000 MC)

(d) Difference between subfigures (b) and (c)

(e) Convergence curve

Results: Synthetic Data





(a) Uniform (Closed-form)
0.06 s

(b) Independent Gaussian (Monte Carlo)
1.44 s, 1000 samples

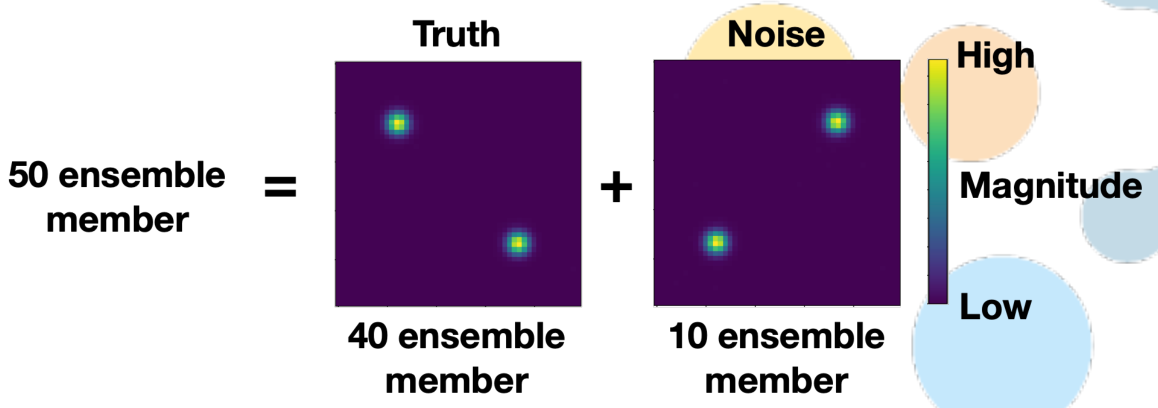
(c) Epanechnikov (Closed-form)
0.58 s

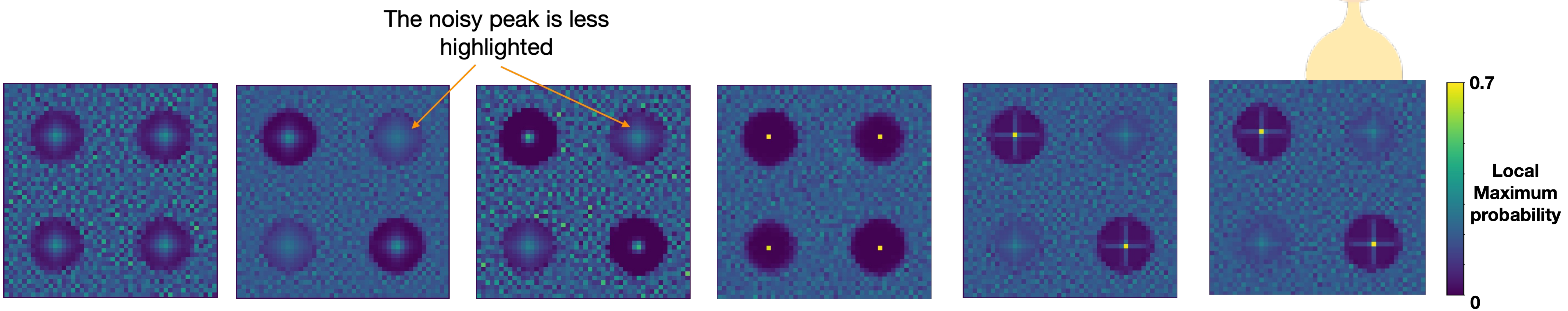
(d) Multivariate Gaussian (Monte Carlo)
1.82 s, 1000 samples

(e) Histogram with 5 bins (Closed-form)
39.17 s

(f) Histogram with 5 bins (Semi-analytical)
16.72 s, 1000 samples

Results: Synthetic Data





(a) Uniform (Closed-form)

0.06 s

(b) Independent Gaussian (Monte Carlo)

1.44 s, 1000 samples

(c) Epanechnikov (Closed-form)

0.58 s

(d) Multivariate Gaussian (Monte Carlo)

1.82 s, 1000 samples

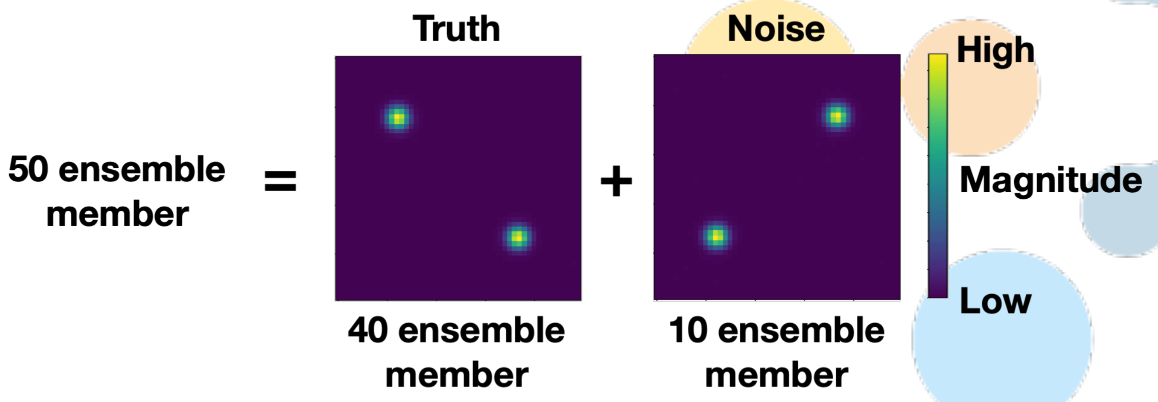
(e) Histogram with 5 bins (Closed-form)

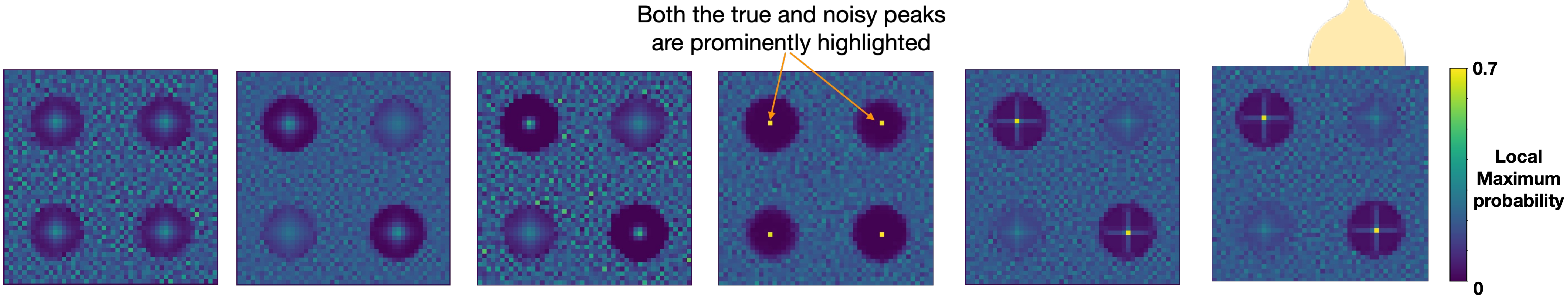
39.17 s

(f) Histogram with 5 bins (Semi-analytical)

16.72 s, 1000 samples

Results: Synthetic Data





(a) Uniform (Closed-form)

0.06 s

(b) Independent Gaussian (Monte Carlo)

1.44 s, 1000 samples

(c) Epanechnikov (Closed-form)

0.58 s

(d) Multivariate Gaussian (Monte Carlo)

1.82 s, 1000 samples

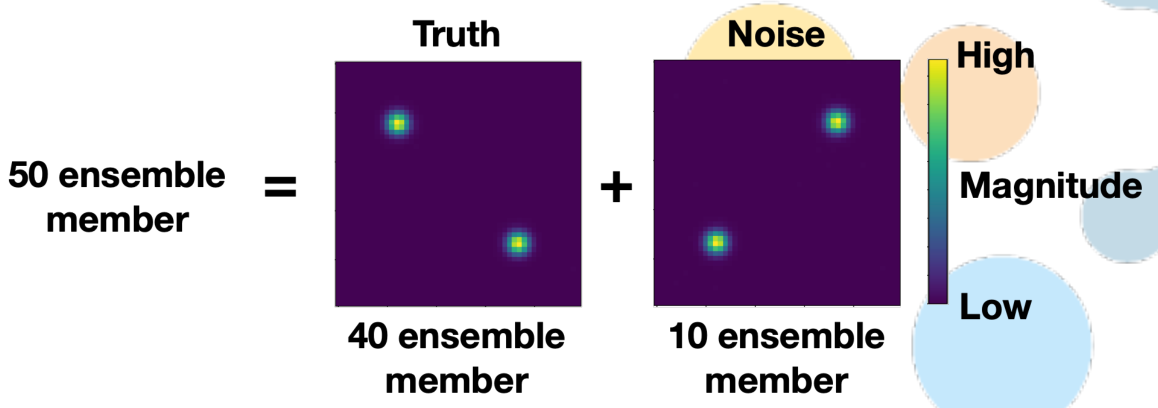
(e) Histogram with 5 bins (Closed-form)

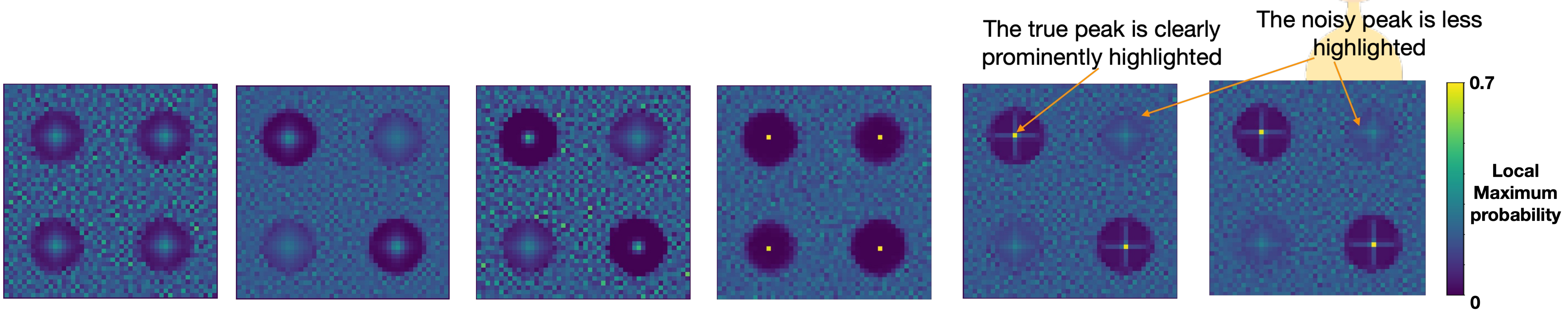
39.17 s

(f) Histogram with 5 bins (Semi-analytical)

16.72 s, 1000 samples

Results: Synthetic Data





(a) Uniform (Closed-form)
0.06 s

(b) Independent Gaussian (Monte Carlo)
1.44 s, 1000 samples

(c) Epanechnikov (Closed-form)
0.58 s

(d) Multivariate Gaussian (Monte Carlo)
1.82 s, 1000 samples

(e) Histogram with 5 bins (Closed-form)
39.17 s

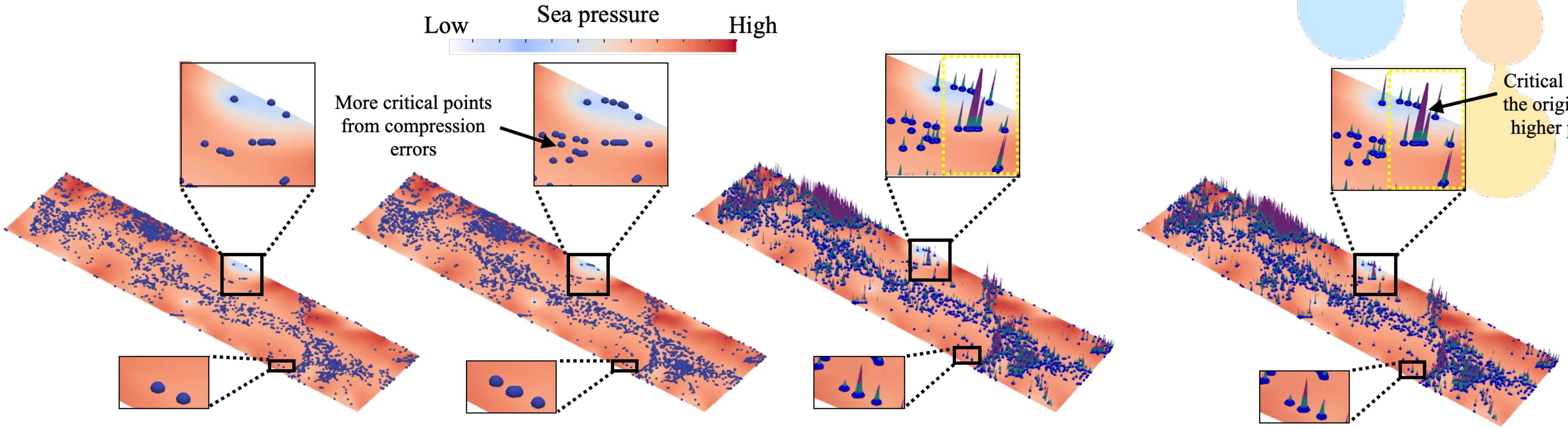
(f) Histogram with 5 bins (Semi-analytical)
16.72 s, 1000 samples

Results: Real Data

Climate Data: Energy Exascale Earth System Model (E3SM)

Data compression use case: compression error bound used to model data uncertainty

Compressor: MGARD [Gong et al., 2023], Compression ratio: 16.68



(a) Original critical points

(b) No uncertainty visualization

(c) Uncertainty visualization using 1000 MC samples **(4.94 seconds)**

(d) Uncertainty visualization using *proposed closed-form* probabilistic algorithms **(0.003 seconds, 1646× faster than MC sampling)**

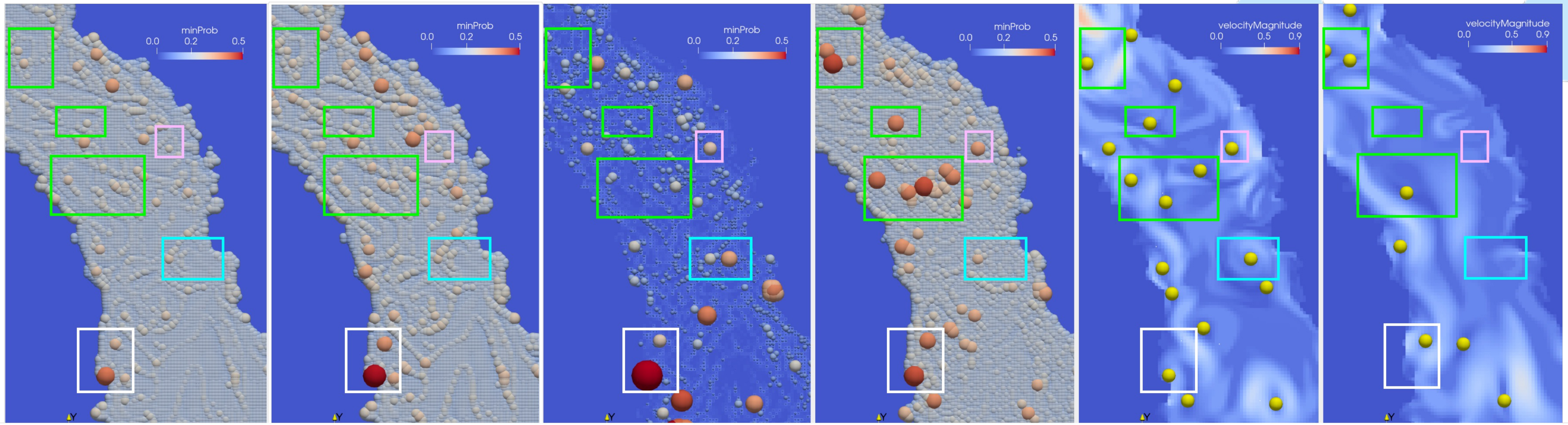
(411×:closed-form solution
4×:VTK-m implementation)

Compute resources: Frontier supercomputer

Results: Real Data

Oceanology: Red Sea simulation data [Sanikommu et al., 2020]

Ensemble data use case: 20 ensemble members of the velocity magnitude dataset each with data resolution 500×500



(a) Independent Uniform (Proposed) 0.094s

(b) Independent Epanechnikov (Proposed) 0.102s

(c) Multivariate Gaussian (Petz et al., 2014 Liebmann et al., 2016) 0.167s

(d) Independent Histogram (Proposed) 0.145s

(e) Ensemble Member 1

(f) Ensemble Member 2

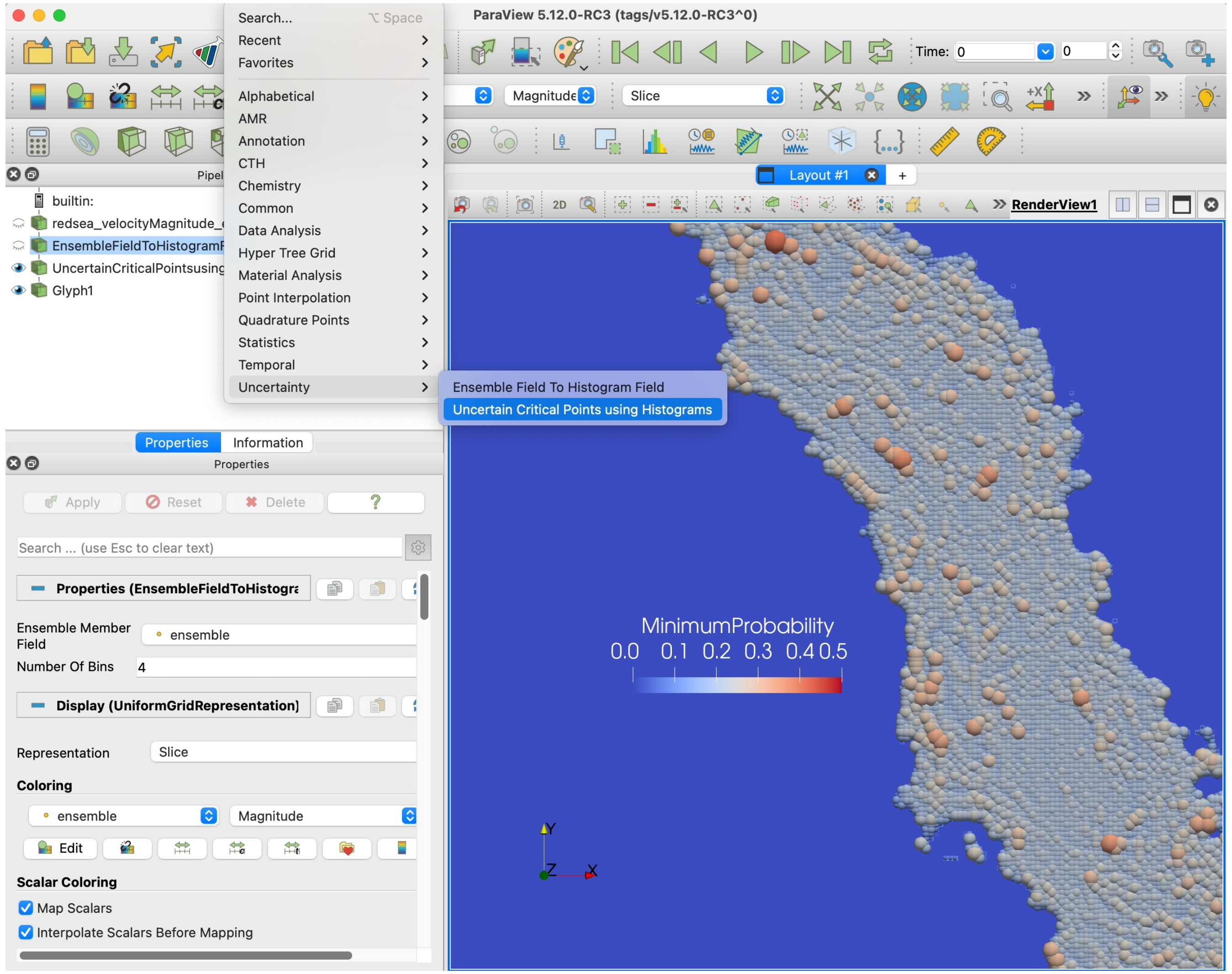
Critical point visualization with the Topology Toolkit (TTK) [Tierny et al., 2018]

Compute resources: Frontier supercomputer

Results: Real Data

Oceanology: Red Sea simulation data [Sanikommu et al., 2020]

VTK-m implementation enables seamless integration of our methods into ParaView for broader accessibility.



ParaView
[Ahrens et al., 2005]

Conclusion and Future Work

- **Closed-form framework** for accurate and efficient critical point probability computation (upto 411× speed-up)
- Integration of closed-form framework with VTK-m library for **near-real-time computation** of critical point uncertainty (upto 1646× speed-up)
- Seamless **integration with ParaView** using VTK-m for broader accessibility
- Future work: Closed-form uncertainty framework for critical points with six/eight neighbors, 3D data, other topological visualizations (e.g., persistence diagrams, contour trees)

Acknowledgements

This work was supported in part by the U.S. Department of Energy (DOE) RAPIDS-2 SciDAC project under contract number DE-AC0500OR22725, NSF III-2316496, the Intel OneAPI CoE, and the DOE Ab-initio Visualization for Innovative Science (AIVIS) grant 2428225. This research used resources of the Oak Ridge Leadership Computing Facility (OLCF), which is a DOE Office of Science User Facility supported under Contract DE-AC05-00OR22725, and National Energy Research Scientific Computing Center (NERSC), which is a DOE National User Facility at the Berkeley Lab.