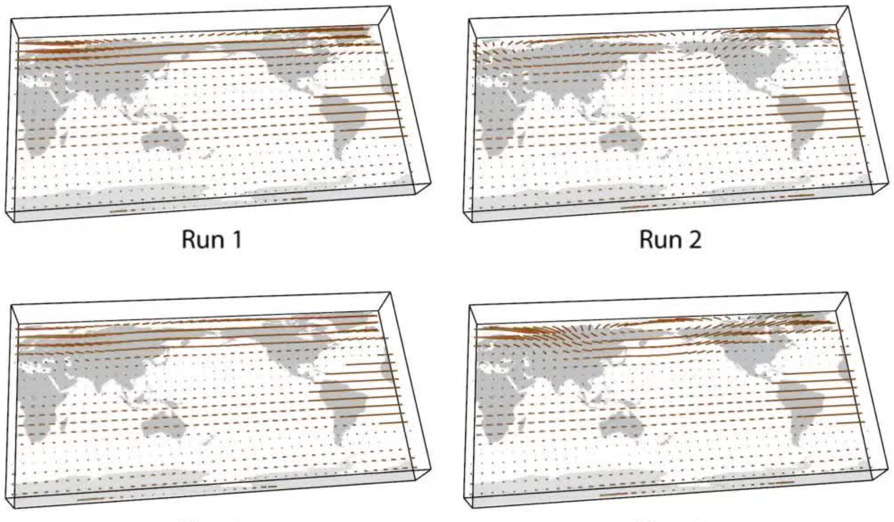
Scalable Ensemble and Uncertain Flow Visualization

Hanqi Guo
Argonne National Laboratory
10/23/2016

VIS2016

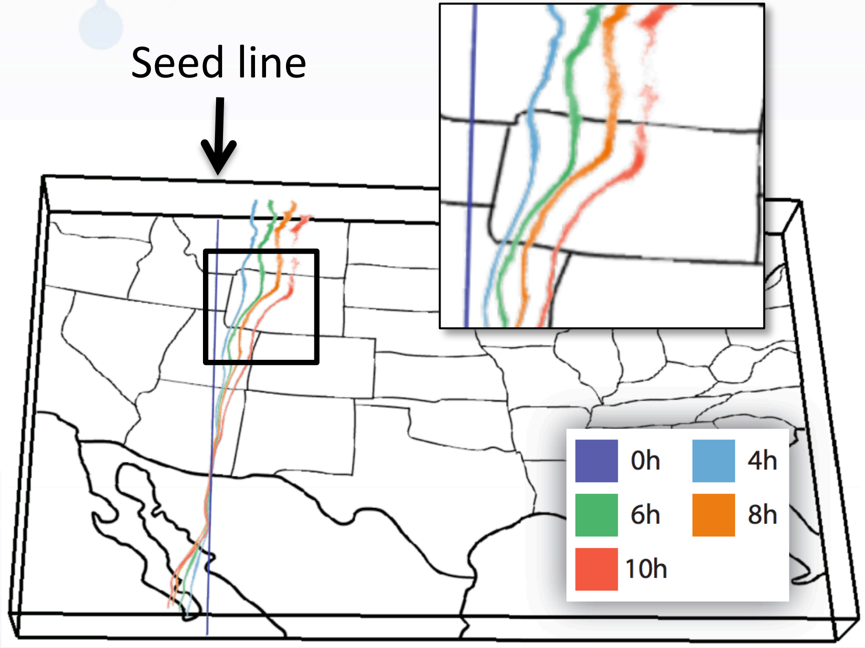
Argonne NATIONAL LABORATORY

Ensemble Flows



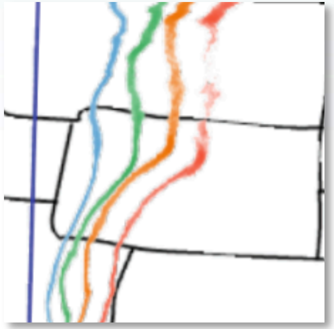
- Ensemble runs are conducted to understand model sensitivities to different parameters and initial conditions
- Each location in a ensemble flow field has a number of different vector values

Uncertain Flows

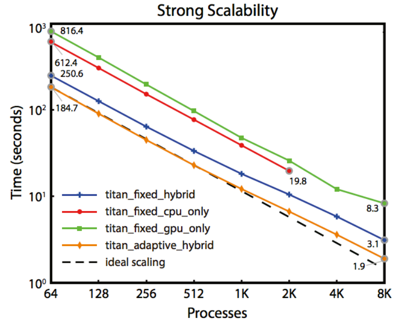


- Uncertain flows can be from experiments, observations, ensemble simulations, and data reduction
- Vector values on each location can be represented as a distribution
- Transport behaviors in uncertain unsteady flows are stochastic

Challenges in Ensemble/Uncertain Flow Visualization



- Theories
  - Visual comparison of ensemble runs
  - Visual encoding of uncertainties
  - Defining features in ensemble/uncertain flows
    - Vortices, topology, finite-time Lyapunov exponents (FTLE), Lagrangian coherent structures (LCS)
  - …
- Scalabilities
  - Handling large ensemble/uncertain data that contains more information than deterministic flows
  - Using supercomputers to visualize and analyze features in ensemble/uncertain flows
  - …



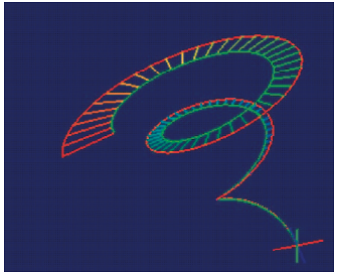
Motivation and Outline

- We must combine the study of theories and scalabilities in visualizing ensemble and uncertain flows
  - Redefine features in ensemble/uncertain flows
  - Scale the ensemble/uncertain flow visualization algorithms for real world problems
- We cover two examples in this tutorial
  - Ensemble flows—scalable flow line analysis
  - Uncertain flows—scalable FTLE/LCS computation

Examples of Ensemble Flow Analysis

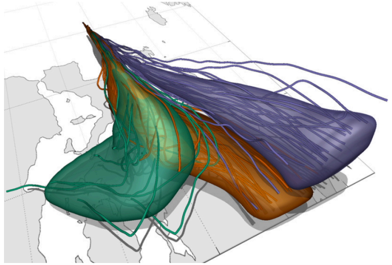
- Direct visualization of ensemble flows

Comparative visualization



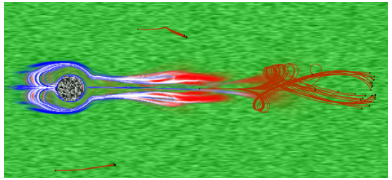
Verma and Pang 2004

Streamline variability plots



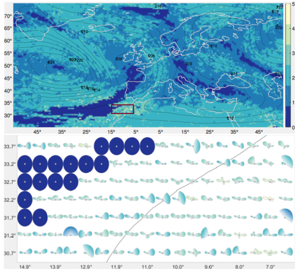
Ferstl et al. 2016

FTLE in ensemble flows

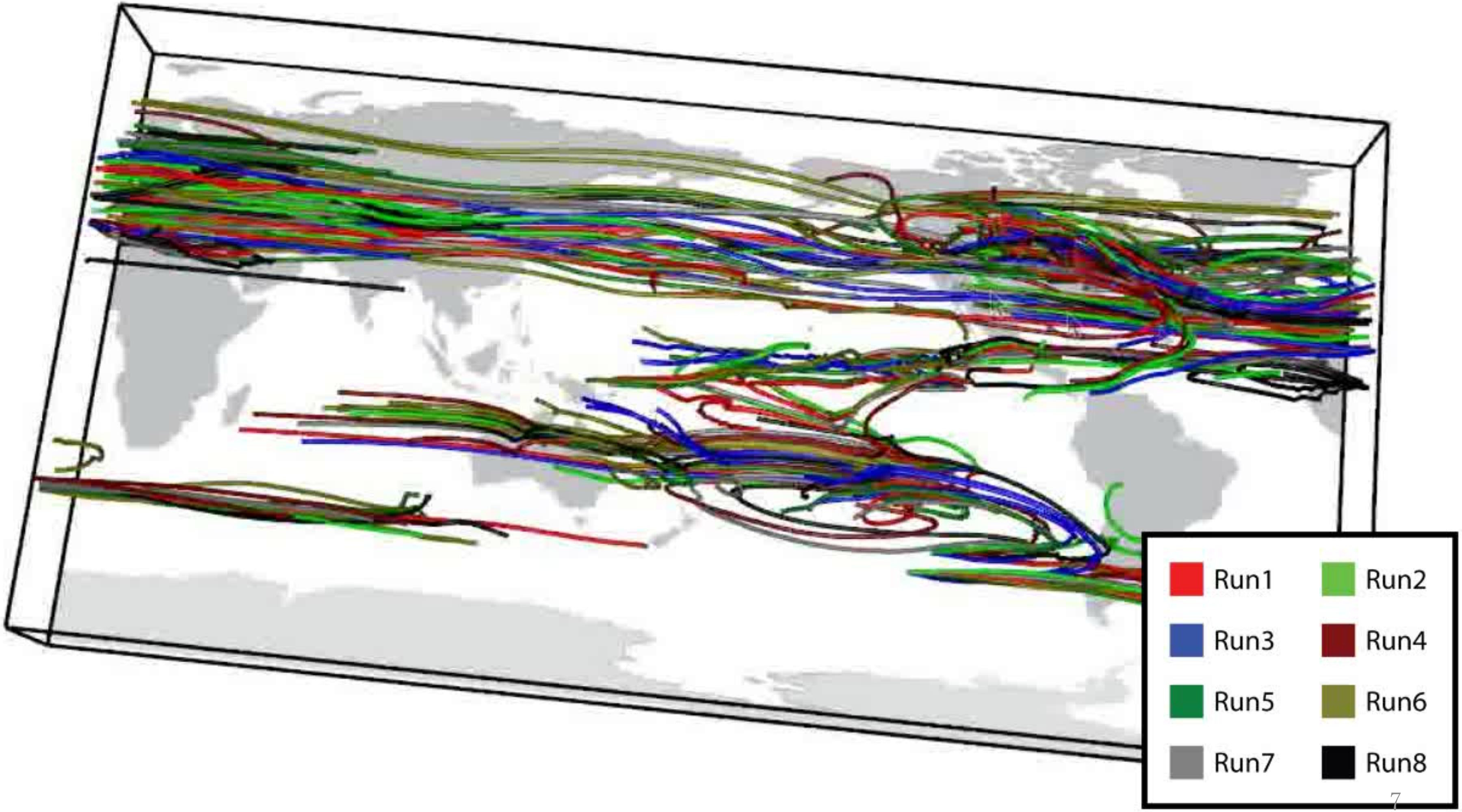


Hummel et al. 2013

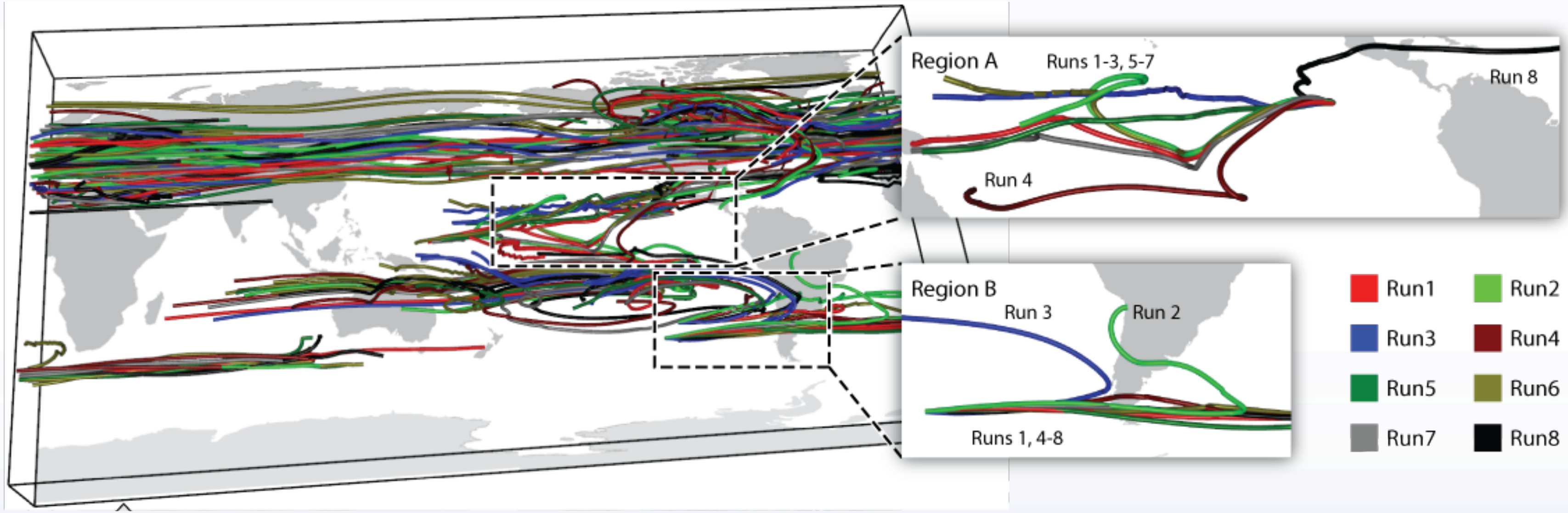
Visual analytics



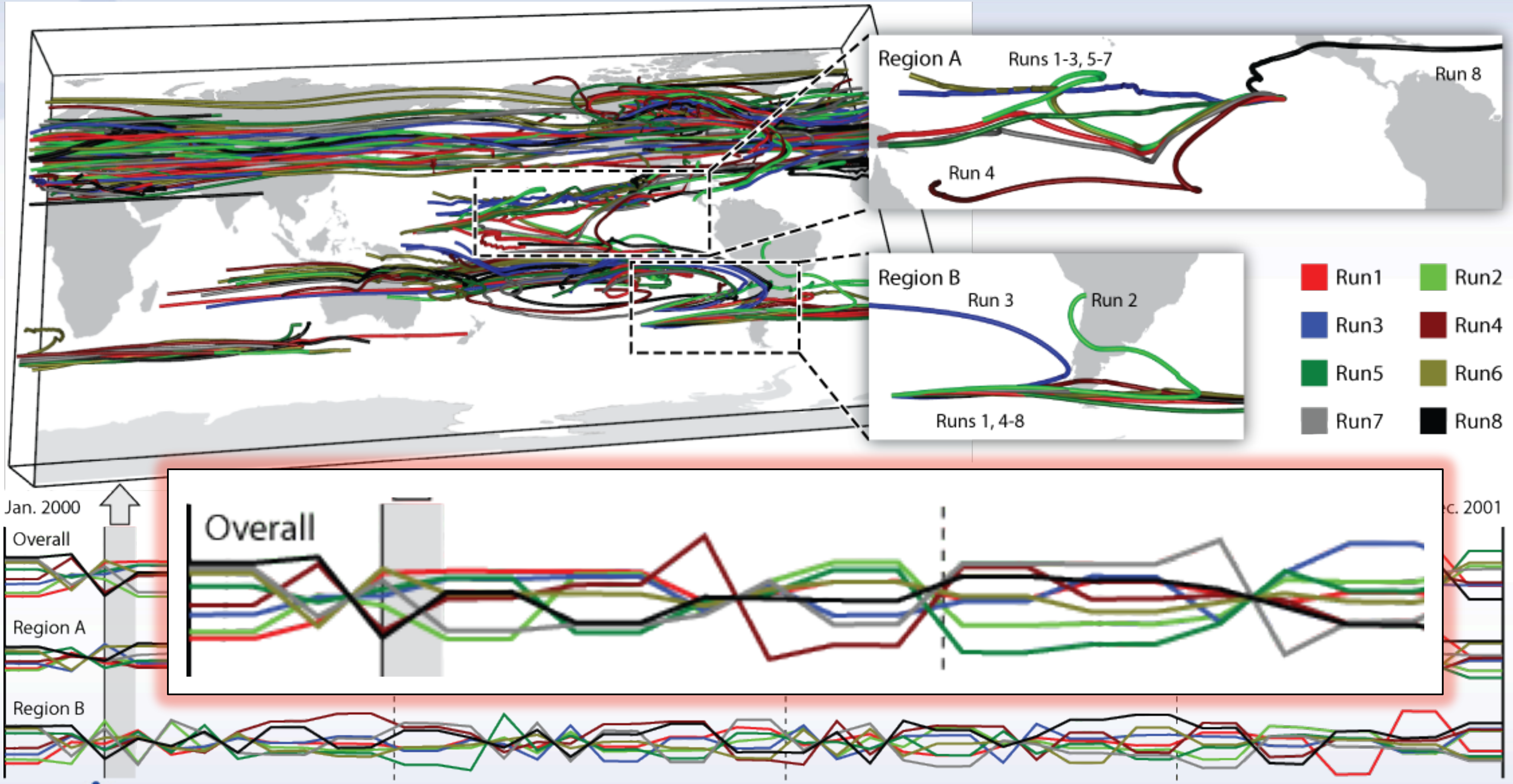
Jarema et al. 2015


Run1
Run2
Run3
Run4
Run5
Run6
Run7
Run8

Ensemble GEOS-5 Simulation



H. Guo, X. Yuan, J. Huang, and X. Zhu, "Coupled ensemble flow line advection and analysis." IEEE TVCG, 19(12):2733—2742, 2013



H. Guo, X. Yuan, J. Huang, and X. Zhu, "Coupled ensemble flow line advection and analysis." IEEE TVCG, 19(12):2733—2742, 2013

Lagrangian-based Distance Metrics for Ensemble Flows

- The flexible distance metric as the distance of flow maps

$$d_{\mathbf{x},t}(\mathbf{U},\mathbf{U}') = \mu(\mathbf{U}(\Phi_t^{t+\tau}(\mathbf{x})),\mathbf{U}'(\Phi_t^{t+\tau}(\mathbf{x}))), \tau \in [t, t+t_0]$$

  – μ can be maximum distance, Hausdorff distance, etc.
  – **U** can be the location, some scalar quantities, etc.

- In our application, we use accumulated difference as μ

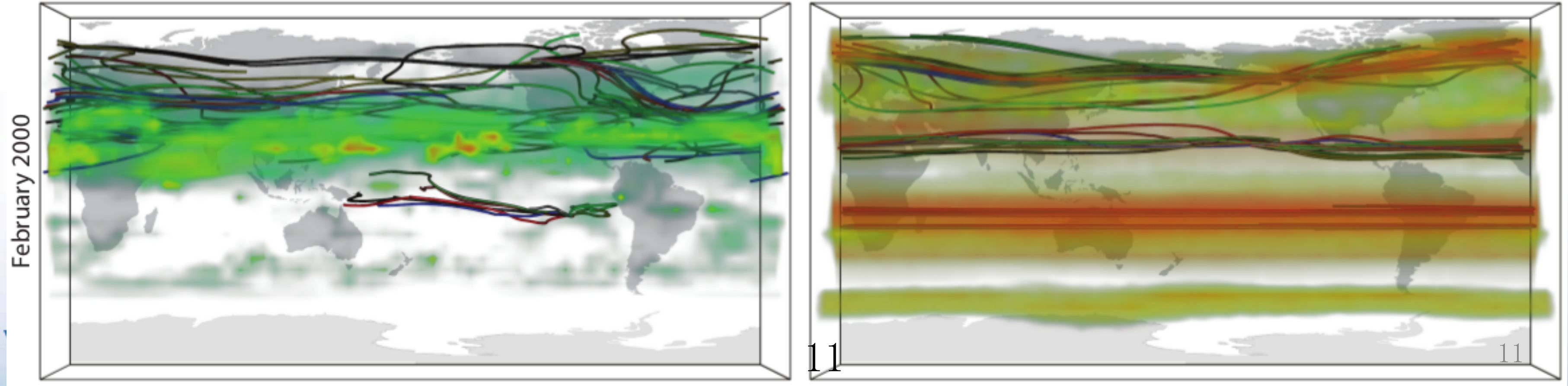
$$\mathscr{V}(\mathbf{x},t) = \frac{1}{N(N-1)} \sum_{i<j} d_{\mathbf{x},t}(\mathbf{U}_i,\mathbf{U}_j)$$

- The variation values of the ensemble run

$$d_{\mathbf{x},t}(U,U') = \int_t^{t+t_0} ||U(\Phi_t^{t+\tau}(\mathbf{x})) - U'(\Phi_t^{t+\tau}(\mathbf{x}))||^2 d\tau$$

CO_2-based Metric in the GEOS-5 Simulations

- The metric: the differences of locations / CO_2 concentration along pathlines
- Findings
  - The variation of the wind field is high in some part the north hemisphere, but CO_2 differences are quite different
  - CO_2 concentration is not sensitive to wind in above regions



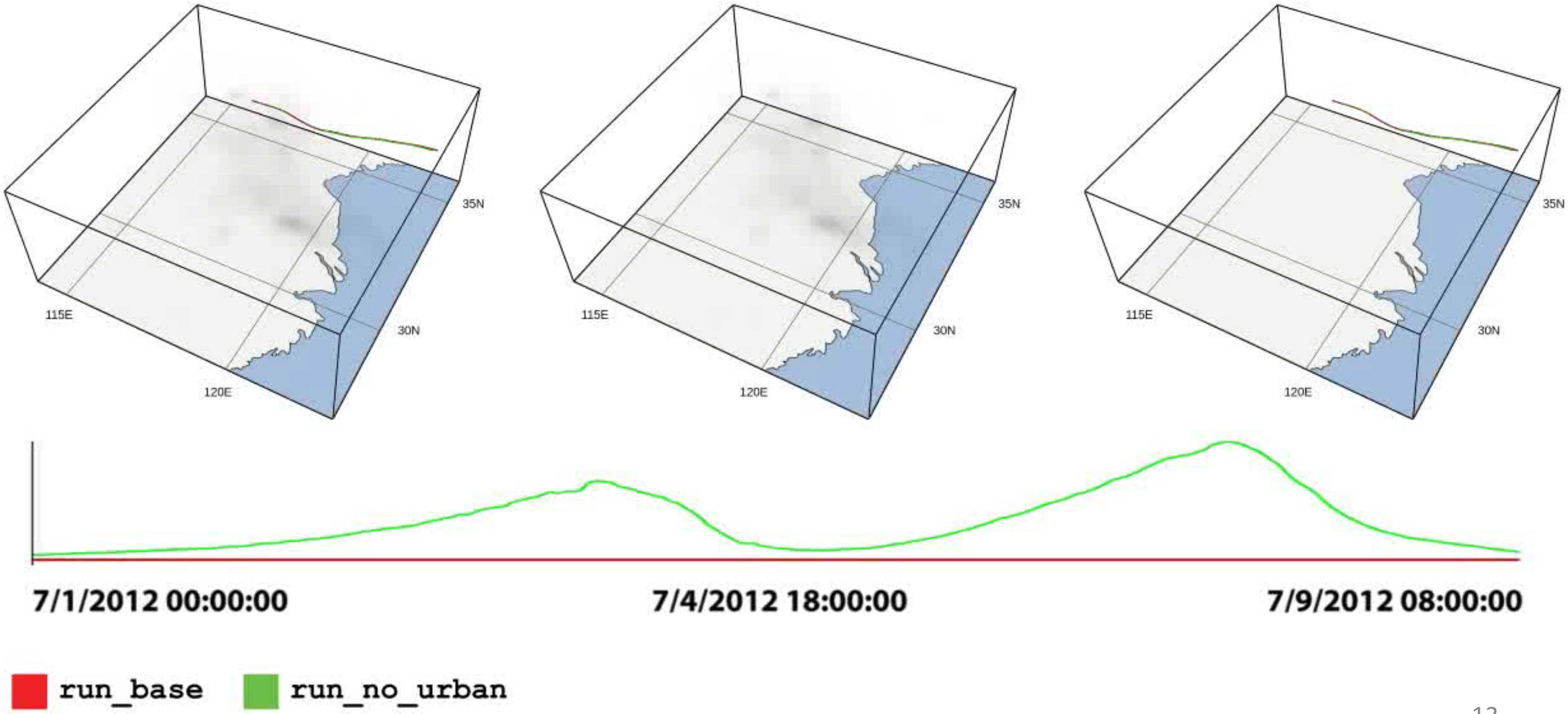
WRF Simulation

- Scientists study how urbanization change the weather with WRF ARW model
  - Ensemble runs: base and no_urban
  - 100x100x27 resolution, East China
  - From 2012-7-1 00:00:00 UTC to 2012-7-10 18:00:00 UTC

Hybrid

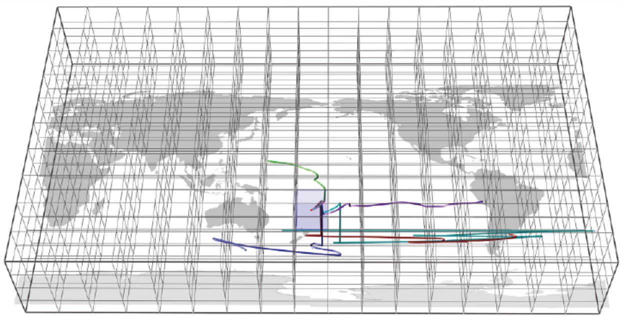
Variation Field

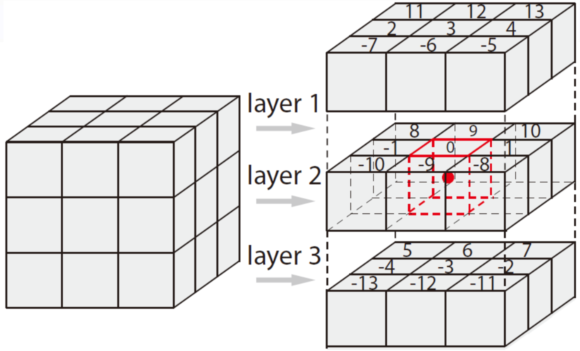
Field Lines

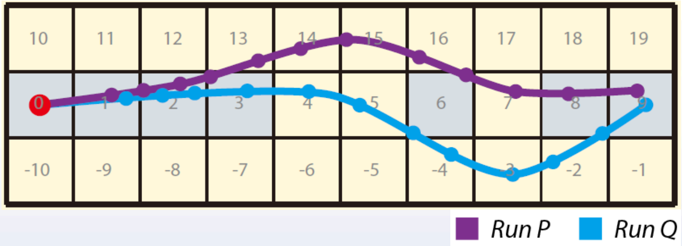


Alternative Distance Metrics for Ensemble Flows

- Distance metric based on subsequent analysis



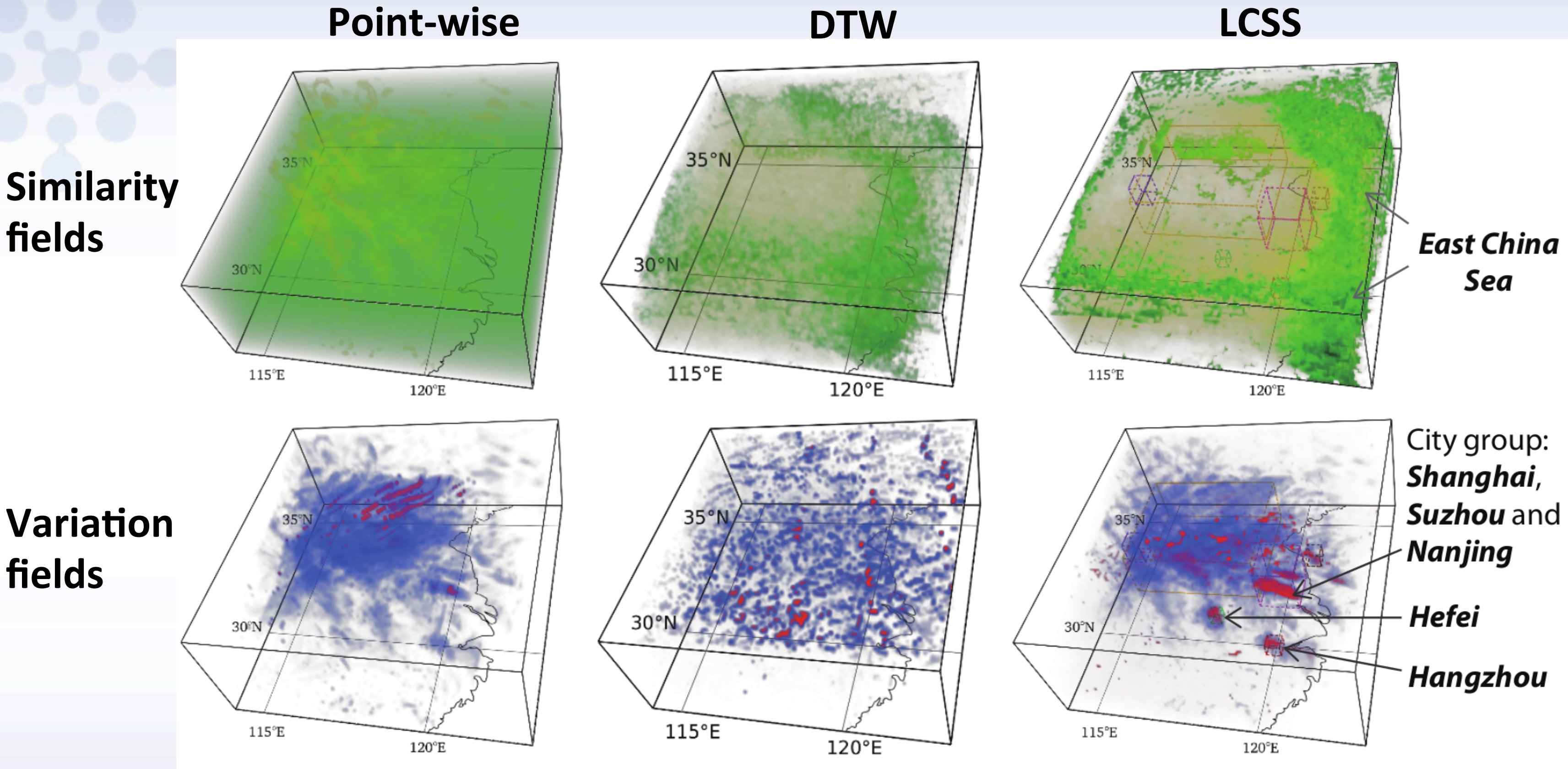




***Run P*: (0, 1, 2, 2, 3, 13, 14, 15, 16, 6, 7, 8, 9);**

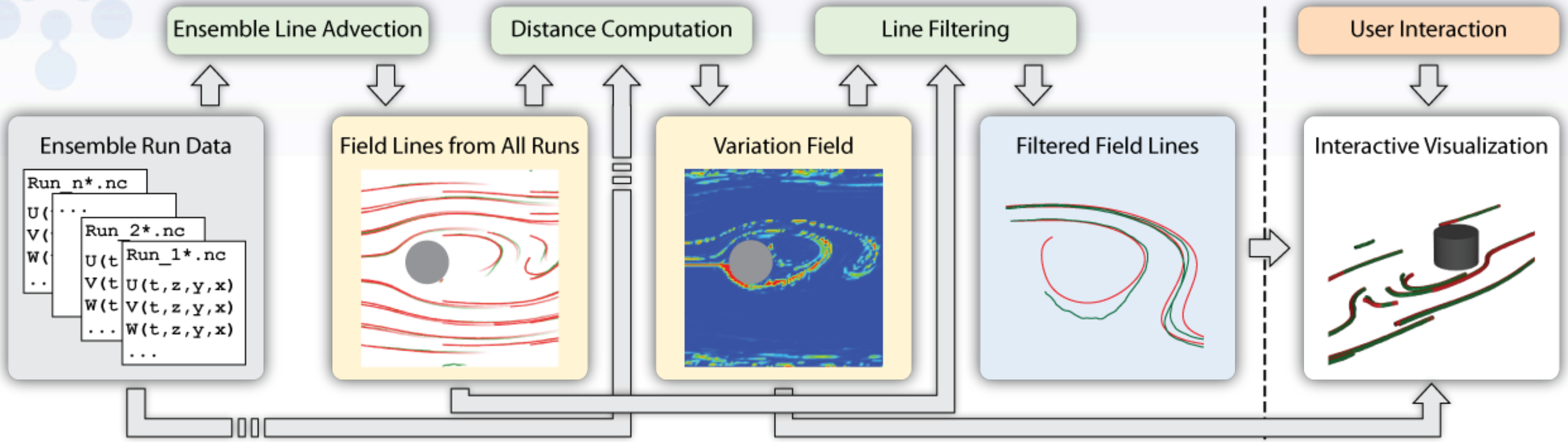
***Run Q*: (0, 1, 2, 2, 3, 4, 5, 6, -4, -3, -2, 8, 9).**

R. Liu, H. Guo, J. Zhang, and X. Yuan, "Comparative visualization of vector field ensembles based on longest common subsequence." In Proc. IEEE PacificVis, pp.96-103, 2016.

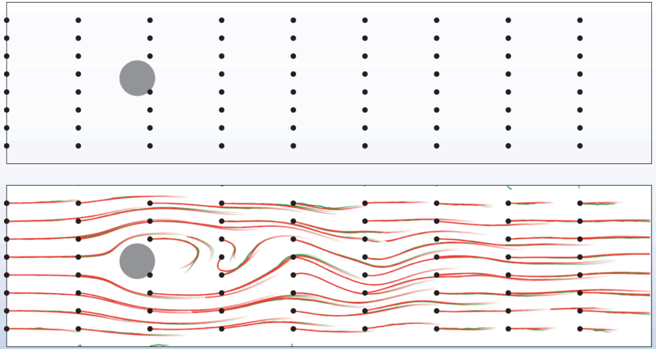


R. Liu, H. Guo, J. Zhang, and X. Yuan, "Comparative visualization of vector field ensembles based on longest common subsequence." In Proc. IEEE PacificVis, pp.96—103, 2016.

Workflow



- Ensemble data (large, 76 GB)
- Field line data (much larger than ensemble data, 5.8 TB)
- Variation field (small, less than 1 GB)
- Filtered lines (even smaller)

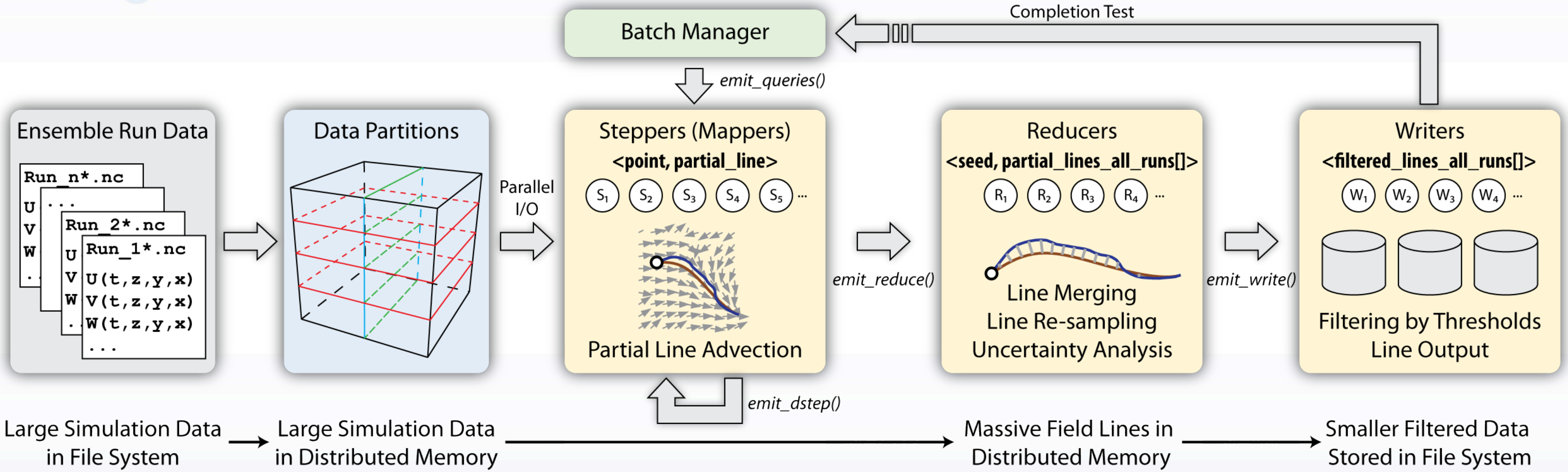


Parallel System Design

- The challenges
  - I/O for huge ensemble run data
  - Massive computation of pathlines
  - Extraordinary memory requirements for intermediate results
- eFLAA (Ensemble Flow Line Advection and Analysis)
  - Coupling flow line advection and analysis for ensemble flow visualization
  - Redesigning DStep (Kendall et al.), a MapReduce-like framework, for parallel ensemble flow analysis

Kendal et al., “Simplified parallel domain traversal.” In Proc. SC11, pp.10:1-10:11, 2011

eFLAA System



Benchmark Platform: NSCCJN

ShenWei-based supercomputer*

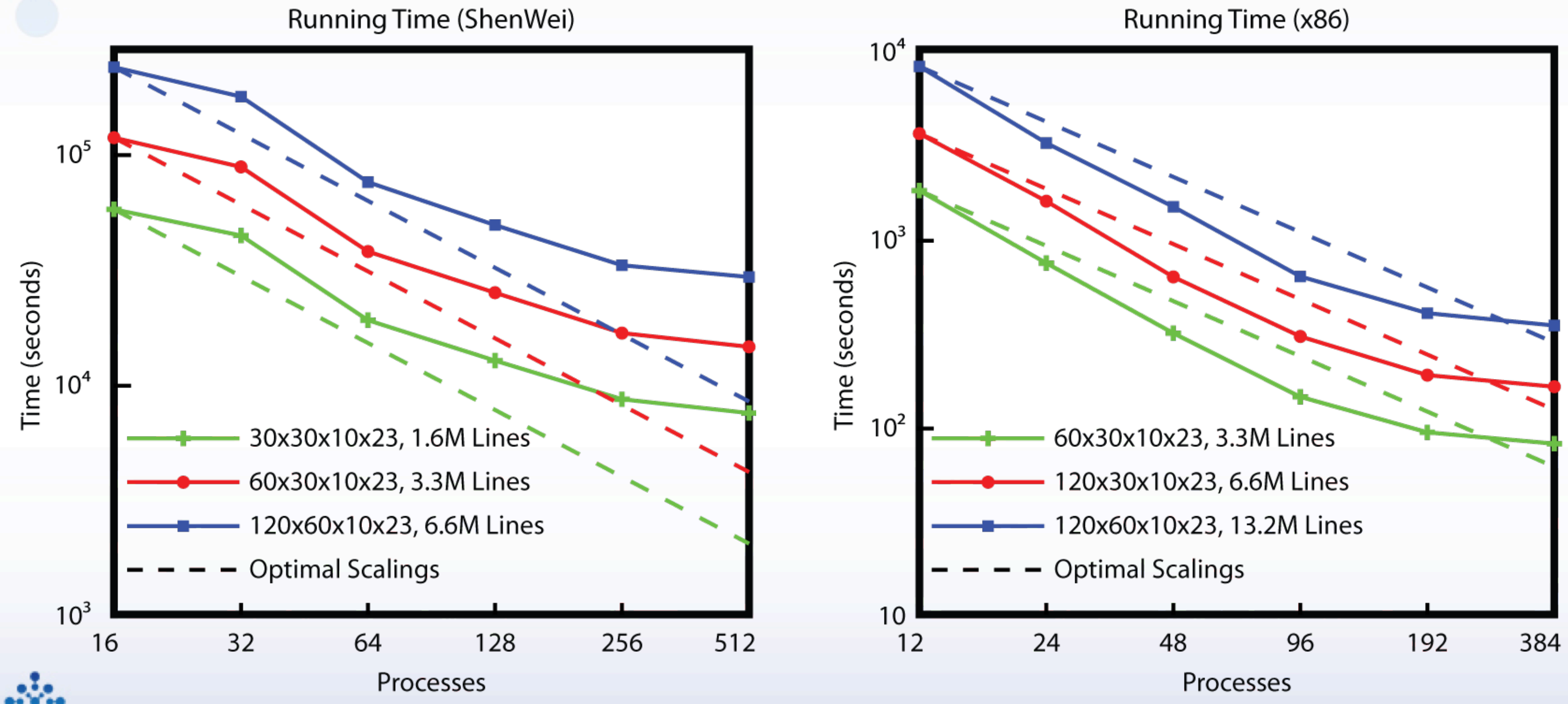
- SW1600 processors, 1.0~1.1GHz
- 1GB memory for each core
- 40Gbps high-speed interconnection

x86-based supercomputer

- Intel Xeon E5675 hexa-core processors, 3.06GHz
- 4GB memory for each core
- QDR 40Gbps Infiniband interconnection

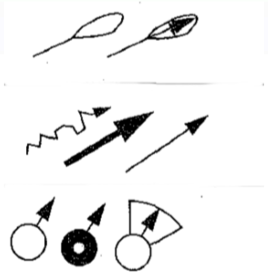
* The same architecture as ShenWei TaihuLight, the new No.1 system in the TOP500 list as of June 2016

Scalability

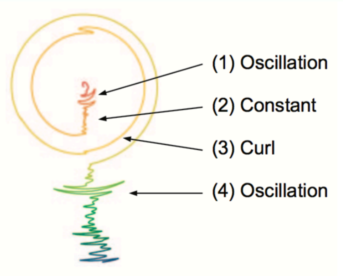


Examples of Uncertain Flow Visualization

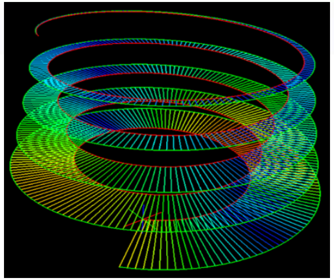
- Uncertain flow visualization based on glyphs and textures



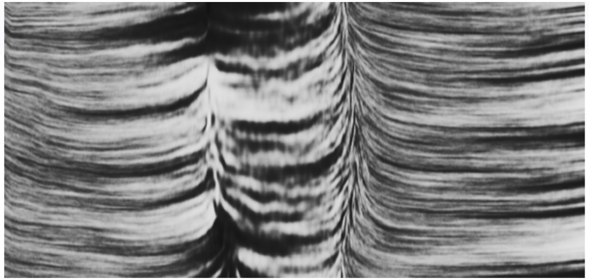
Wittenbrink et al. 1996



Hlawatsch et al. 2011

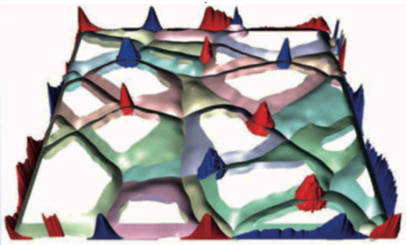


Lodha et al. 1996

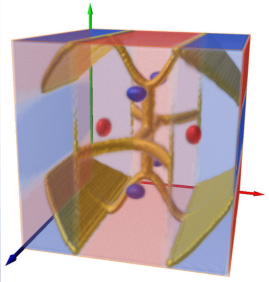


Botchen et al. 2005

- Uncertain flow topologies



Otto et al. 2010



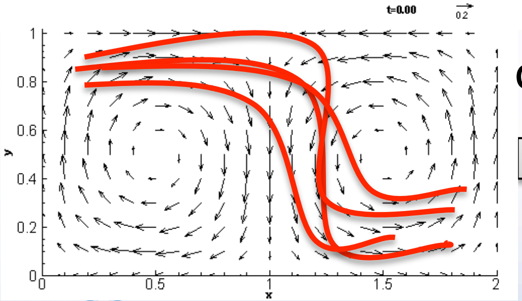
Otto et al. 2011

FTLE and LCS in Deterministic Flows

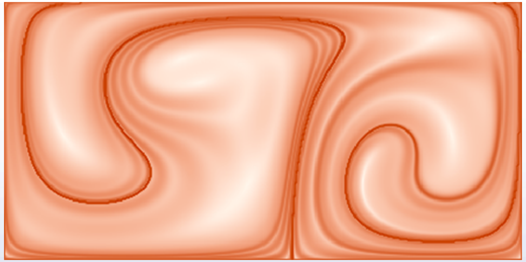
- Finite-Time Lyapunov Exponents (FTLE)

$$\sigma(\mathbf{x}, t, \tau) = \frac{1}{|\tau|} \log \sqrt{\lambda_{\max}((\nabla\phi)^{\intercal}\nabla\phi)}$$

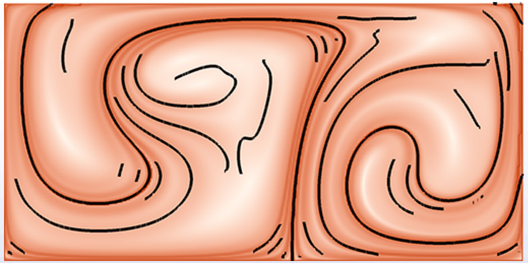
  - The rate of separation of infinitesimally close trajectories over a time interval of interest
- Lagrangian Coherent Structures (LCS)
  - Distinguished boundaries of trajectories over a time interval of interest
  - Usually derived as ridges of FTLE field



σ(•)



R(•)



Unsteady Flow FTLE LCS

FTLE and LCS in Uncertain Unsteady Flows

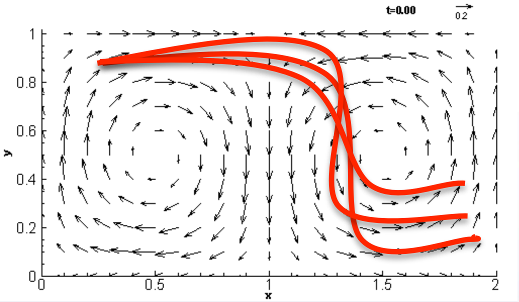
Uncertain Unsteady Flow* "Uncertain" FTLE "Uncertain" LCS

- How to quantify the uncertainty of FTLE and LCS?
  - D-FTLE (distribution of FTLE)
  - U-LCS (uncertain LCS)
- How to measure the separation in uncertain unsteady flows?
  - FTLE-D (FTLE of distributions)

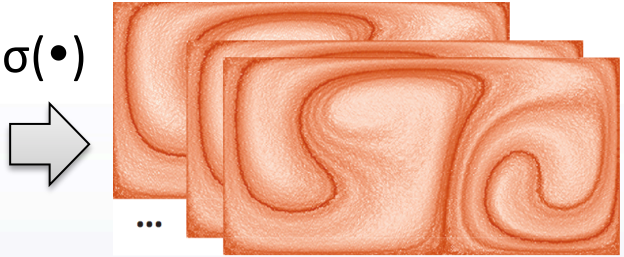
*) We synthesize the "uncertain double-gyre" data by adding Gaussian noises

D-FTLE (Distribution of FTLE)

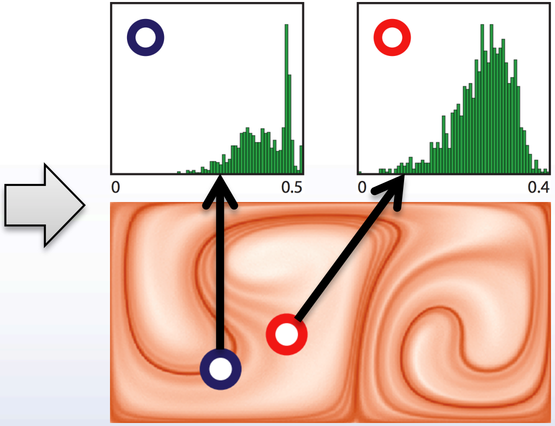
- D-FTLE quantifies the uncertainty of FTLE as distributions of FTLE values
- The PDF of D-FTLE is a (n+3)-dimensional scalar function
- We conduct Monte Carlo simulations to compute D-FTLE
  - Stochastic Runge-Kutta methods
  - Iteration stops if D-FTLE is stable



Stochastic particle tracing
(stochastic flow map)



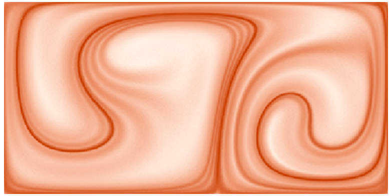
Stochastic FTLE runs



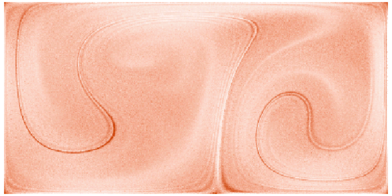
D-FTLE (mean)

Visualization of D-FTLE Field

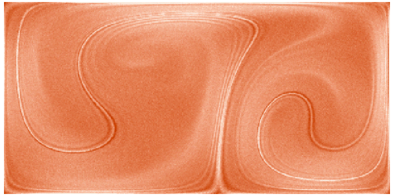
- Interactive queries
- Statistics



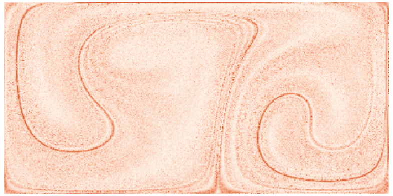
Mean



Standard deviation

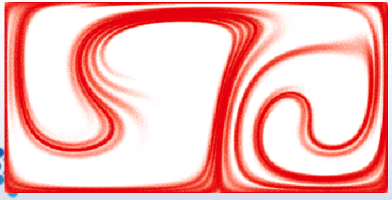


Entropy



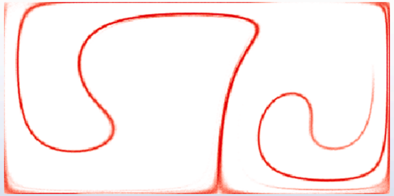
Shapiro-Wilk test

- Statistical threshold $T(\mathbf{x}, t, \tau) = Pr(\Sigma | \Sigma(\mathbf{x}, t, \tau) \geq \gamma) = 1 - \int_{-\infty}^{\gamma} \rho(\mathbf{x}, t, \tau; \sigma) d\sigma$



γ=0.2

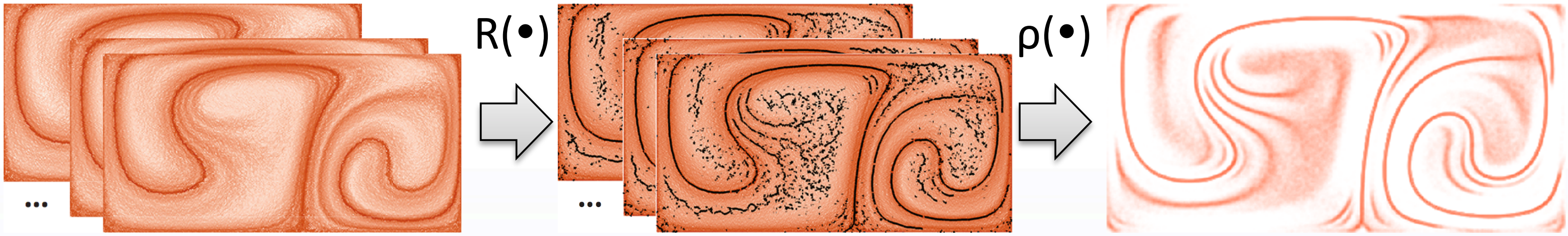
γ=0.25



γ=0.3

U-LCS

- U-LCS quantifies the probability of being LCSs
- U-LCS is a *n*-dimensional scalar field that can be directly visualized
- We propose a line/surface density estimator for U-LCS computing



D-FTLE resampling/
Stochastic FTLE runs

Ridges in stochastic
FTLE runs

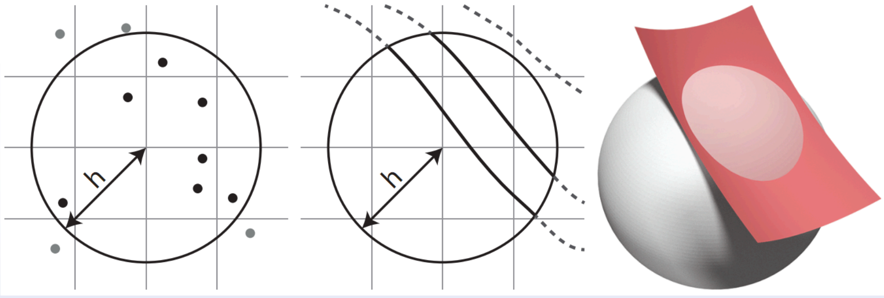
U-LCS

Line/Surface Density Estimation

- We propose a generalized SPH model to estimate the densities of ridge lines/surfaces
- Line densities: the length of lines inside the circle
- Surface densities: the areas of surfaces inside the sphere

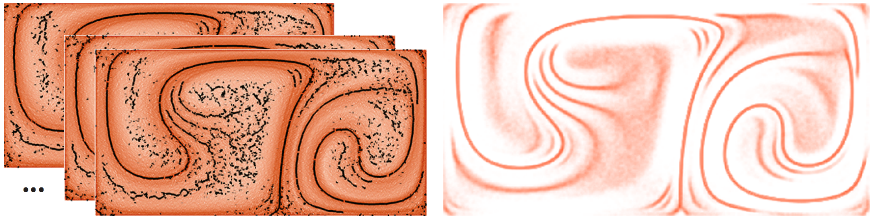
$$\rho(\mathbf{r}) = \sum_{j=0}^{N_b-1} \int_{D_j} \omega(|\mathbf{r} - \mathbf{r}_j|, h) ds$$

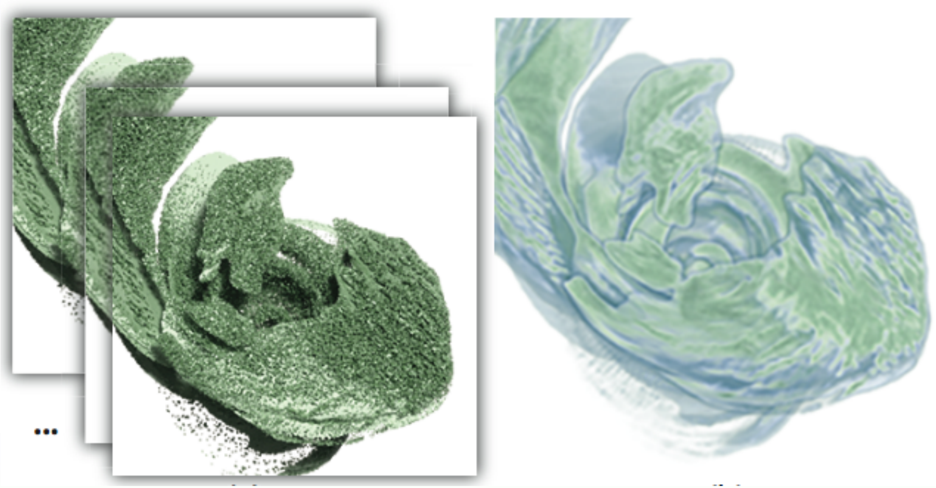
$$\omega(d,h) = \begin{cases} 1/\pi h^2 \text{ or } 3/4\pi h^3, & \text{if } d \leq h \\ 0, & \text{otherwise.} \end{cases}$$



Kernel density Line density Surface density

Line/Surface Density Estimation (cont'd)

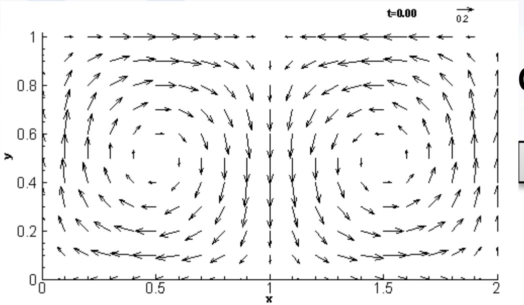




Line density estimation

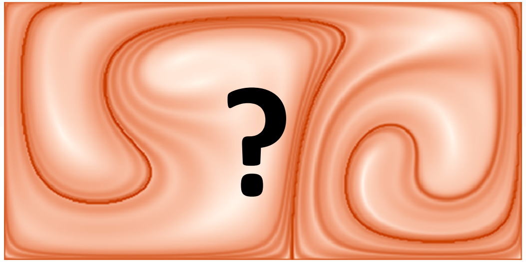
Surface density estimation

FTLE and LCS in Uncertain Unsteady Flows



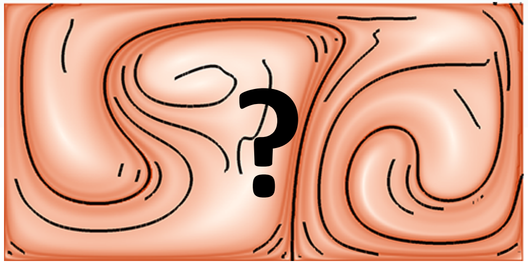
Uncertain Unsteady Flow*

σ(•)



"Uncertain FTLE"

R(•)



"Uncertain LCS"

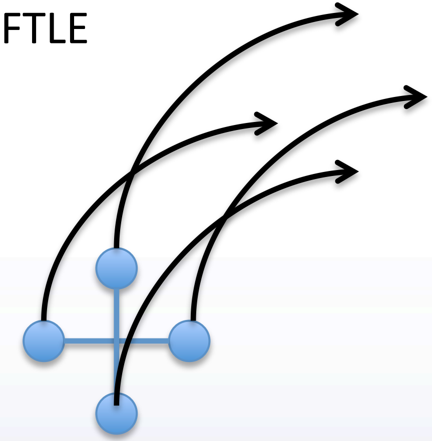
- How to quantify the uncertainty of FTLE and LCS?
  - D-FTLE (distribution of FTLE)
  - U-LCS (uncertain LCS)
- How to measure the separation in uncertain unsteady flows?
  - FTLE-D (FTLE of distributions)

VIS 2016

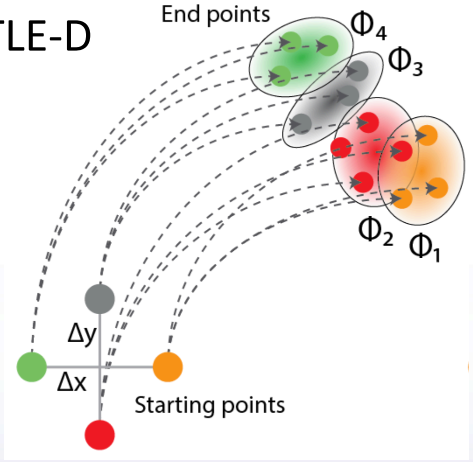
FTLE-D

- FTLE-D generalizes FTLE by measuring the rate of separation in uncertain unsteady flows

FTLE



FTLE-D

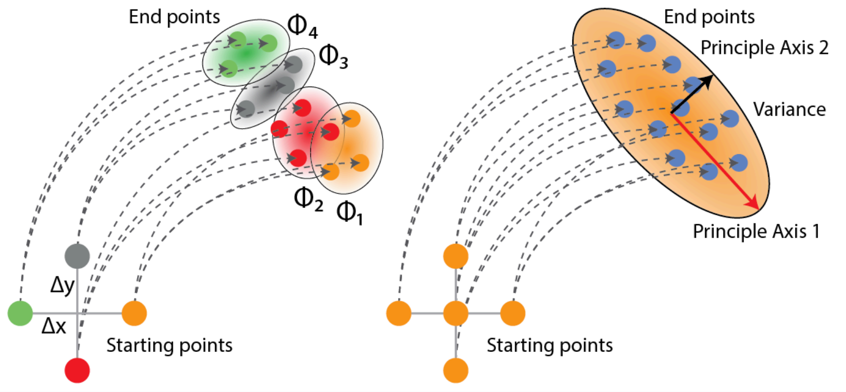


$$\sigma(\mathbf{x}, t, \tau) = \frac{1}{|\tau|} \log \sqrt{\lambda_{\max}((\nabla \phi)^\intercal \nabla \phi)}$$

$$\hat{\sigma}(\mathbf{x}, t, \tau) = \frac{1}{|\tau|} \log \sqrt{\lambda_{\max}(E(\nabla \Phi)^\intercal E(\nabla \Phi))}$$

Comparison between FTLE-D and FTVA

- In FTVA (Finite-Time Variance Analysis), the distribution of end locations are assumed to be Gaussian, because it is based on PCA
- FTLE-D does not make assumptions on the distributions
- FTLE-D values are directly comparable with FTLE and they better preserve features

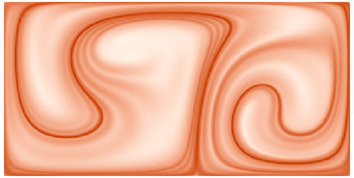


FTLE-D FTVA

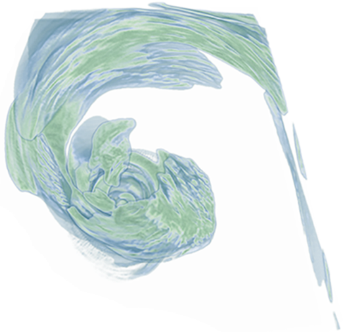
Schneider et al., A variance-based FTLE-like method for unsteady uncertain vector fields. In Proc. TopoInVis 2011, pp.255-268, 2011.

Experiment Results

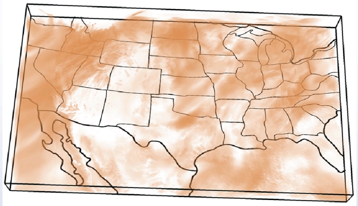
- Synthetic uncertain double-gyre data (2D)
  - The deterministic double-gyre data + Gaussian noise



- Uncertain Hurricane Isabel data (3D)
  - The temporal down-sampling error



- Ensemble WRF simulation data (3D)
  - The average and variances from 3 ensemble members



VIS2016

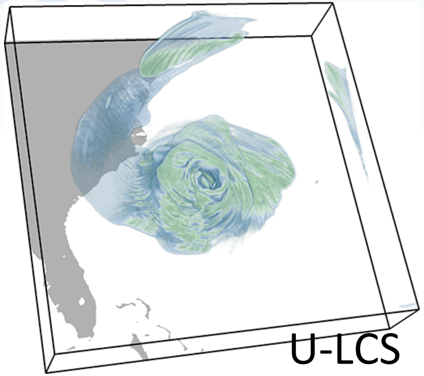
Results: Uncertain Isabel Data

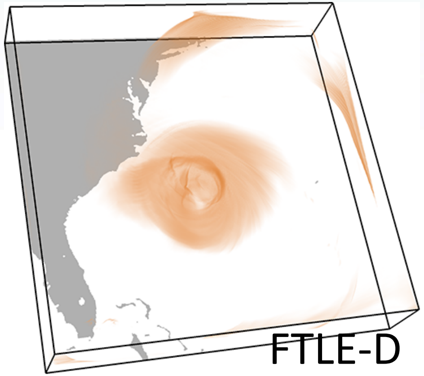
- Temporal down-sampling is a common practice in scientific simulations, but introduces uncertainty
- We follow Chen et al. to down-sample the original Isabel data using quadratic Bezier interpolation; the errors are Gaussian

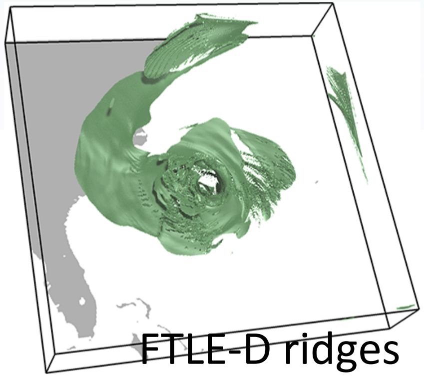
$$V_{\mathbf{x},t} \simeq \mathbf{B}_{\mathbf{x}}(t/n) + \mathrm{N}(0, s_{\mathbf{x}}^2)$$

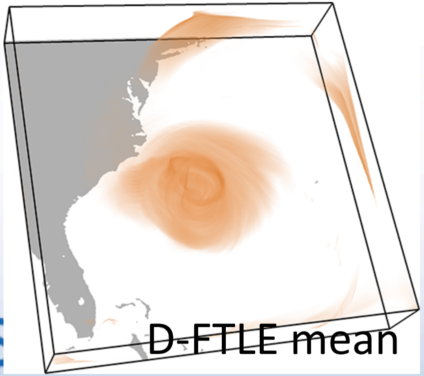
Chen et al., Uncertainty modeling and error reduction for pathline computation in time-varying flow fields. In Proc. IEEE PacificVis 2015, pp.215-222.

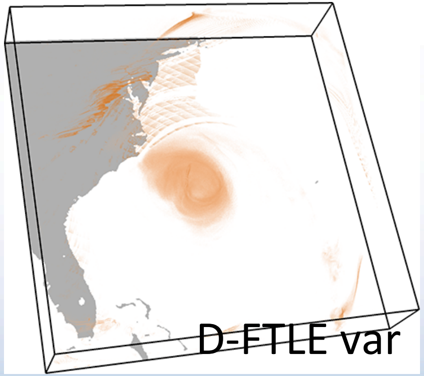
Results: Uncertain Isabel Data (cont'd)

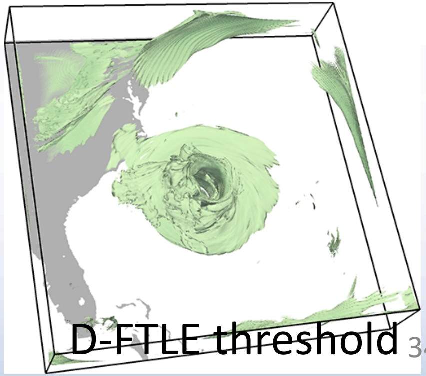






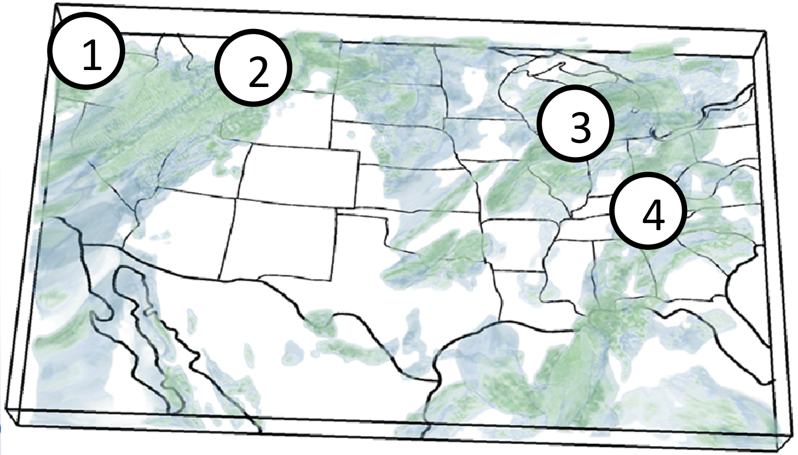




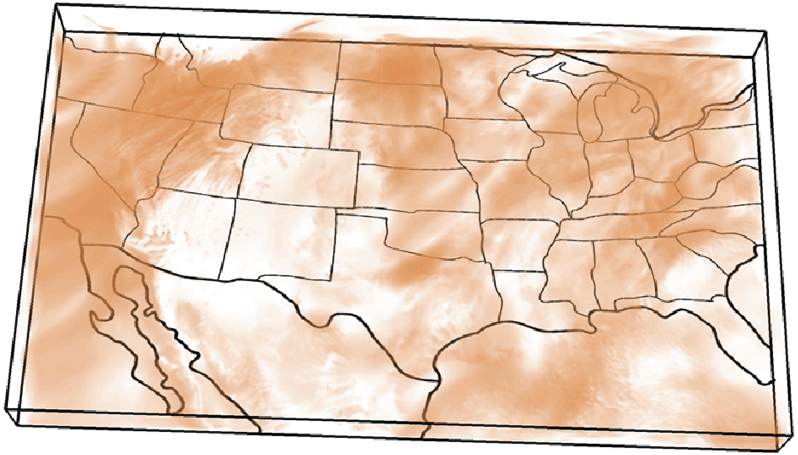


Results: WRF Ensembles

- Four distinct zones in continental U.S.
  - On-shore flow from the Pacific being pushed over Cascade mountains
  - Cold front stretching from Oklahoma up into the Dakotas
  - Two unstable trough regions



U-LCS



FTLE-D

Performance Issues

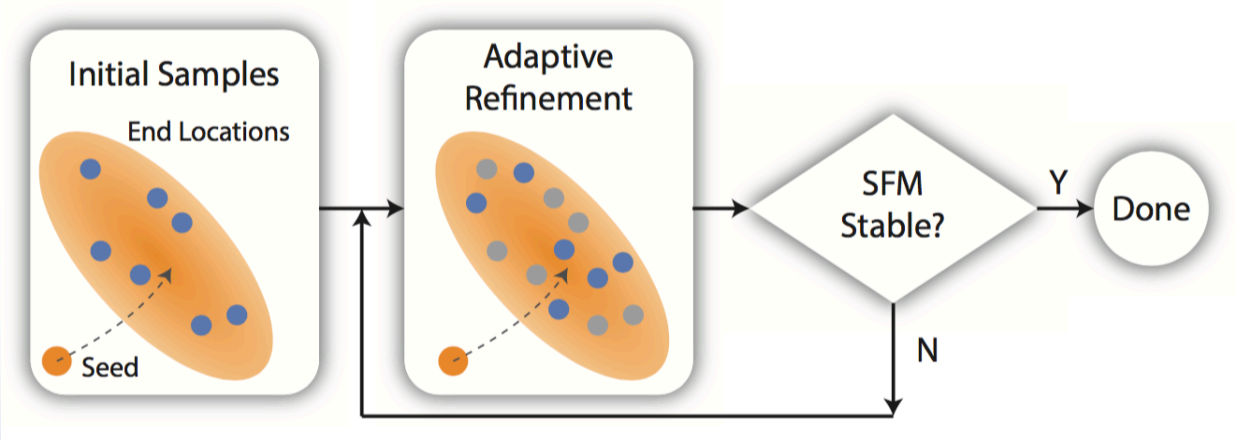
- We have an serial version for a single workstation, but it may take hours to days to run our analysis, even with GPUs
- Approaches to reduce computation time
  - Adaptive refinement
  - Scalable and parallel computation

Dataset	Resolution	Performance (CPU)				Performance (GPU)			
		t_p	t_f	t_l	t_d	t_p	t_f	t_l	t_d
Double gyre	N/A (analytical)	4.60k	18.7	60.0	0.28	10.0	0.05	1.00	0.06
Isabel	$500 \times 500 \times 100 \times 4$	447k	17.2k	6.30k	1.61k	4.76k	25.0	0.56k	25.8
WRF	$1799 \times 1059 \times 40 \times 15$	308k	13.1k	8.36k	1.21k	3.5k	22.0	0.47k	19.7

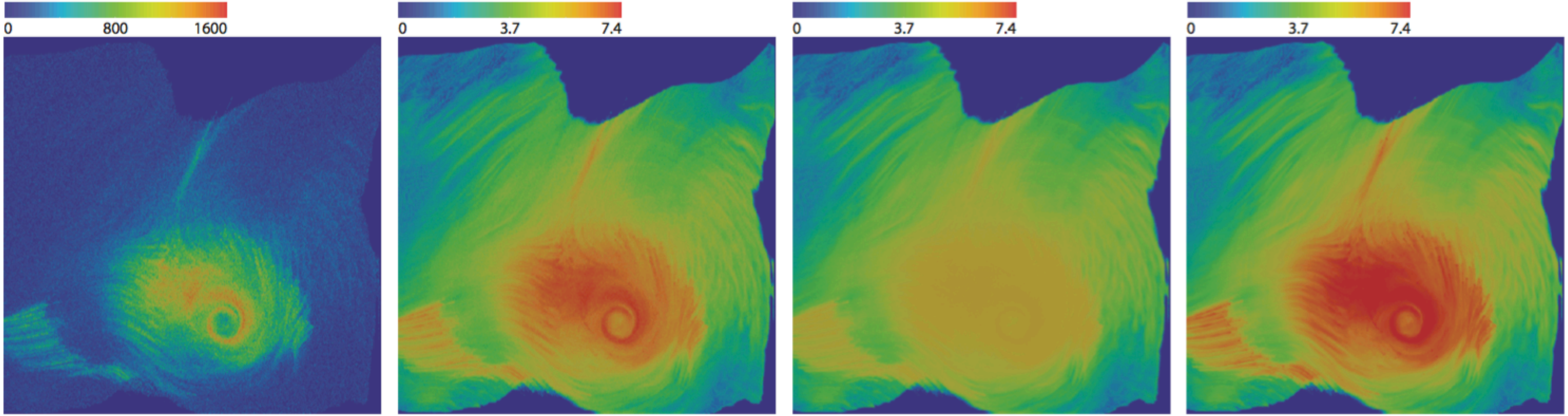
Particle Tracing D-FTLE U-LCS FTLE-D

Adaptive Refinement

- We alleviate the computation cost of SFMs (stochastic flow map) by adaptively determine the number of Monte Carlo runs, instead of fixed numbers of particles

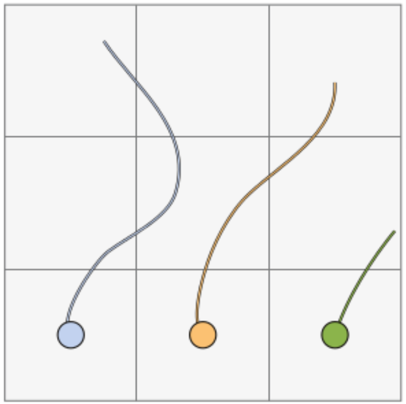


Adaptive Refinements (cont'd)

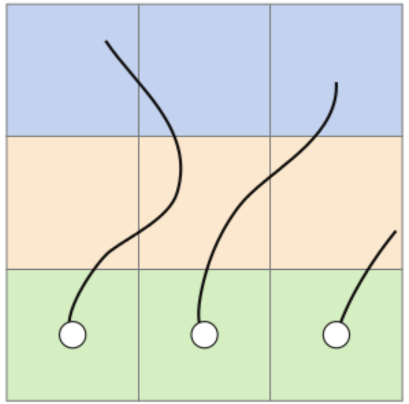


num_particles_adaptive　entropy_adaptive　entropy_fixed_256　entropy_fixed_2048

Parallel Particle Tracing Approaches

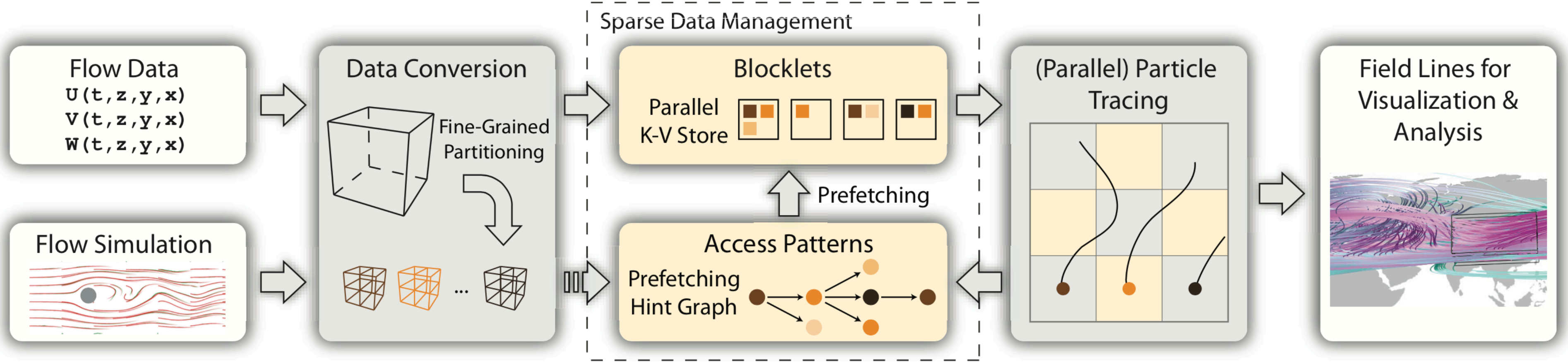


- Task parallelism (parallel over particles, POP)
  - Parallelize-over-seeds
  - High I/O cost
  - Load balancing depends on flow transport



- Data parallelism (parallel over data, POD)
  - Parallelize-over-data
  - Usually requires loading all data
  - Load balancing depends on data partitioning

Sparse Data Management for Flow Field Data



Solution

- Fine-grained data partitioning
- On-demand parallel data access with caching and prefetching
- Parallel key-value store

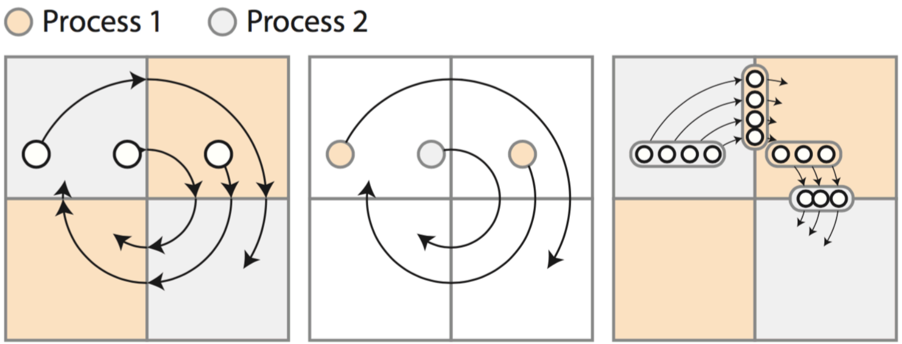
Benefits

- Memory and I/O bandwidth efficient data access with bounded resources
- Improving scalability of task-parallel particle tracing

VIS 2016

Scalable Stochastic Particle Tracing

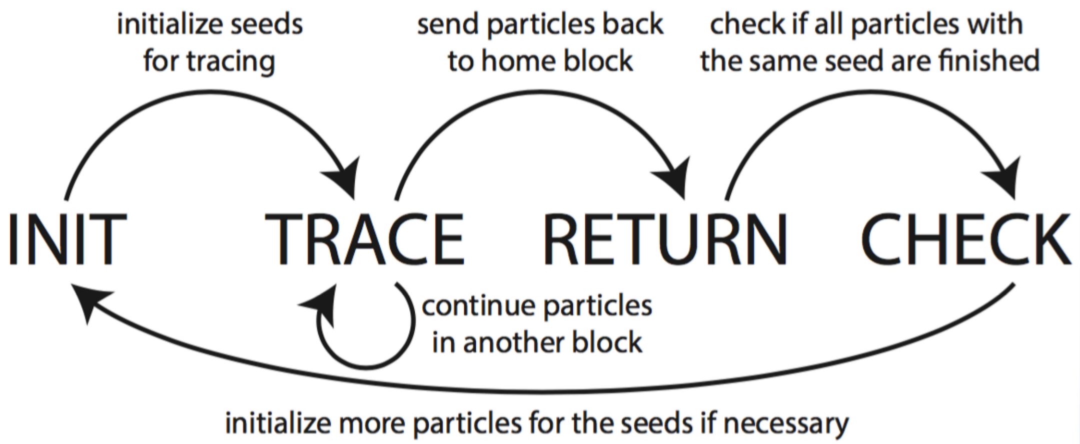
- Our solution: parallel over packets of particles, MPI/thread
- Benefits
  - Improving data locality
  - Reducing context switch
  - Enabling CPU/GPU coprocessing



POD POP Our method

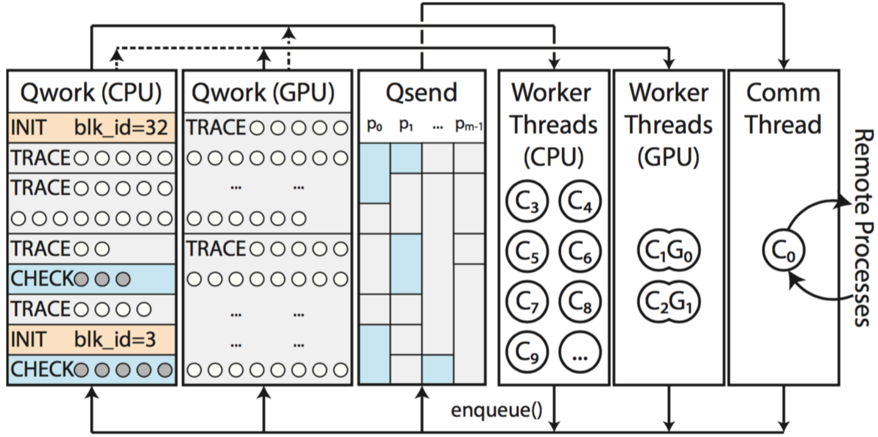
Task Model

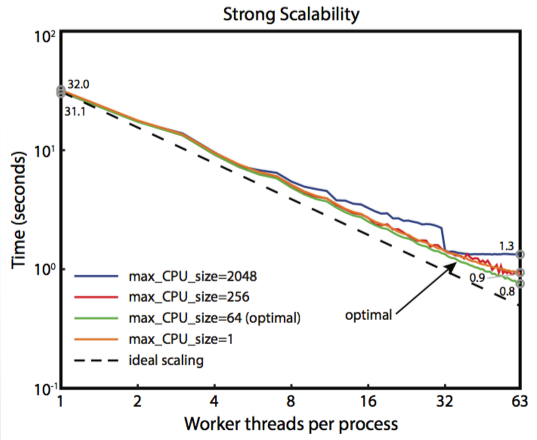
- Each task in our model is associated with a block: (BlkID, type, particles[])
- Four types of tasks: INIT, TRACE, RETURN, CHECK



Thread Model

- Threads: CPU/GPU worker threads, and the communication thread
- Task queues: CPU queues and GPU queues
- The thread model is scalable to number of worker threads
- Important parameter: max_CPU_size



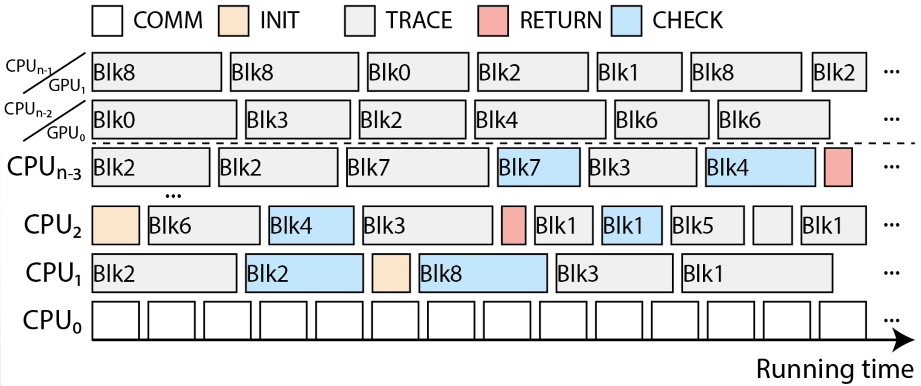


CPU/GPU Coprocessing

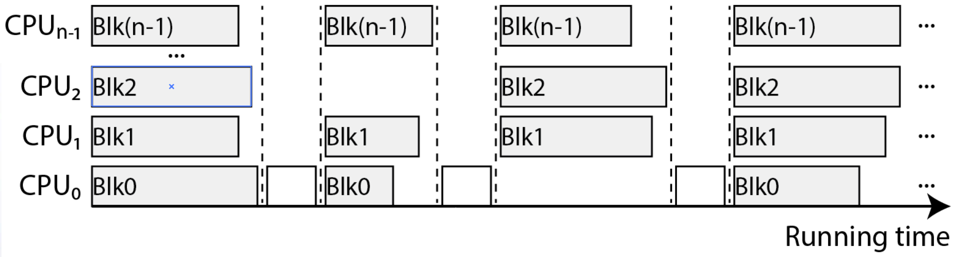
- The task model enables CPU/GPU coprocessing when GPUs are available in compute nodes
- GPU(s) process larger TRACE tasks, while CPUs process all other tasks
- GPU workers dequeues CPU tasks when starving, vice versa
- Important parameters
  - min_GPU_size
  - max_GPU_size

Asynchronous Communication

- The dedicated communication thread fully overlaps the computation and communication
- MPI nonblocking communications are further used to reduce delays



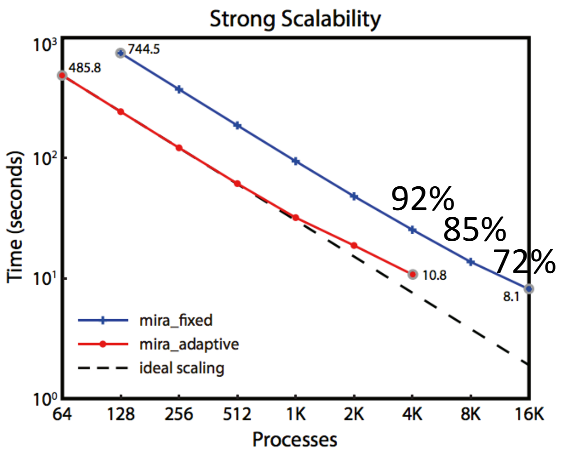
Our model

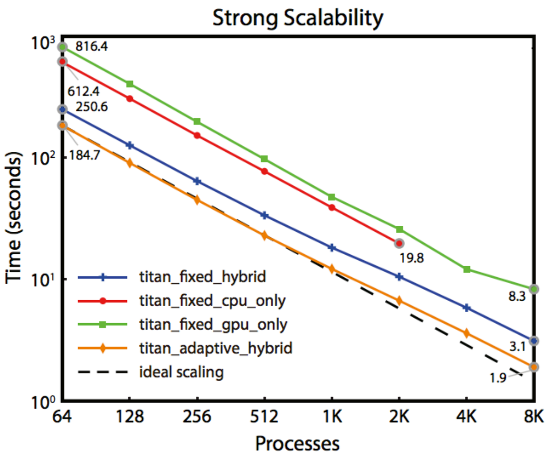


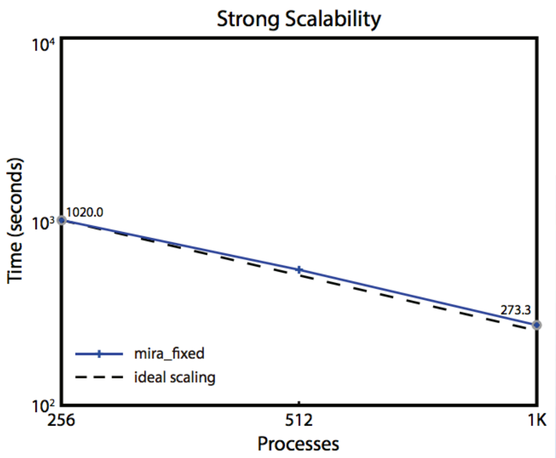
Bulk synchronous processing

Strong Scalability Studies

- Mira, up to 1 million Blue Gene/Q cores
- Titan, up to 128K Opteron cores cooperating with 8K GPUs
  - 2.5x faster, compared with the CPU-only mode







Conclusions

- Combination of theories and scalabilities in visualizing ensemble and uncertain flows
  - Redefining features in ensemble/uncertain flows
  - Scaling the ensemble/uncertain flow visualization algorithms for real world problems
- Two examples
  - Ensemble flows—scalable flow line analysis (Guo et al. 2013, Liu et al. 2016)
  - Uncertain flows—scalable FTLE/LCS computation (Guo et al. 2016, Guo et al. 2013, Guo et al. 2016b)

References

- Guo et al., "Finite-Time Lyapunov Exponents and Lagrangian Coherent Structures in Uncertain Unsteady Flows." IEEE TVCG, 22(6):1672-1682, 2016
- Liu et al., "Comparative Visualization of Vector Field Ensembles Based on Longest Common Subsequence." Proc. IEEE PacificVis, pages 96-103, 2016
- Guo et al., "Advection-based Sparse Data Management for Visualizing Unsteady Flow." IEEE TVCG, 20(12):2555-2564, 2014
- Guo et al., "Coupled Ensemble Flow Line Advection and Analysis." IEEE TVCG, 19(22): 2733-2742, 2013
- Guo et al., "Extreme-scale stochastic particle tracing for uncertain unsteady flows." Technical report, http://www.mcs.anl.gov/publication/extreme-scale-stochastic-particle-tracing-uncertain-unsteady-flow-analysis

Further Readings

- Wittenbrink et al., “Glyphs for visualizing uncertainty in vector fields.” IEEE TVCG, 2(3):266-279, 1996
- Lodha et al., “UFLOW: visualizing uncertainty in fluid flow.” In Proc. IEEE PacificVis, pp.249-254, 1996
- Otto et al., “Uncertain 2D vector field topology.” Comput. Graph. Forum, 29(2):347-356, 2010
- Bothchen and Weiskopf, “Texture-based visualization of uncertainty in flow fields.” In Proc. IEEE Vis, pp.647-654, 2005
- Otto et al., “Uncertain topology of 3D vector field.” In Proc. IEEE PacificVis, pp.67-74, 2011
- Hlawasch et al., “Flow radar glyphs—static visualization of unsteady flow with uncertainty.” IEEE TVCG, 17(12):1949-1958, 2011
- Schneider et al., “A variance based FTLE-like method for unsteady uncertain vector fields.” In Proc. TopoInVis, pp.255-268, 2011
- Petz et al., “Probabilistic local features in uncertain vector fields with spatial correlation.” Comput. Graph. Forum, 31(3): 1045-1054, 2012
- Otto and Theisel, “Vortex analysis in uncertain vector fields.” Comput. Graph. Forum, 31(3):1035-1044, 2012
- Hummel et al., “Comparative visual analysis of Lagrangian transport.” IEEE TVCG, 19(12):2743-2752, 2013
- Chen et al., “Uncertainty modeling and error reduction for pathline computation in time-varying flow fields.” In Proc. IEEE PacificVis, pp.215-222, 2015
- Jarema et al., “Comparative visual analysis of vector field ensembles.” In Proc. IEEE VAST, pp.81-88, 2015
- Ferstl et al., “Streamline variability plots for characterizing the uncertainty in vector field ensembles.” IEEE TVCG, 22(1): 767-776, 2016

Acknowledgments

- U.S. DOE Office of Science, Office of Advanced Scientific Computing
- U.S. DOE OBER ARM Program
- Argonne Leadership Computing Facility
- Oak Ridge Leadership Computing Facility
- National Science Foundation of China, No. 61170204 and 61232012
- Chinese Academy of Science, No. XDA05040205
- National Supercomputing Center in Jinan, China
- National Supercomputing Center in Tianjin, China

Q&A