()

Revisiting the Evaluation of In Situ Lagrangian Analysis

Sudhanshu Sane[†1], Roxana Bujack[2] and Hank Childs[1]

[1]Department of Computer and Information Science, University of Oregon, USA
[2]Los Alamos National Laboratory, USA

Abstract

In situ usage of Lagrangian techniques has proven to be superior with respect to emerging supercomputing trends than the traditional Eulerian approach for scientific flow analysis. However, previous studies have not informed two key points: (1) the accuracy of the post hoc interpolated trajectory as a whole and (2) the spatiotemporal tradeoffs involved when using Lagrangian analysis. With this short paper, we address these points. We first conduct a more comprehensive evaluation via additional accuracy metrics tailored for evaluating Lagrangian trajectories. Second, we provide an understanding of the configurations where the Lagrangian approach works well by studying spatiotemporal tradeoffs. In addition, our study highlights the effects of error propagation and accumulation when performing Lagrangian interpolation for large numbers of steps. We believe our study is significant for better understanding the use of in situ Lagrangian techniques, as well as serving as a baseline for future Lagrangian research.

1. Introduction

In situ processing, through its successful frameworks and usage [FMT*11, WFM11, MOM*11, VHP11, AJO*14, YWG*10, BAA*16], has been demonstrated to be an important approach for large data analysis and visualization on upcoming supercomputers. In this short paper, we discuss the use of in situ Lagrangian techniques for flow field analysis.

The Lagrangian approach consists of two phases. In the first phase, pathlines are extracted in situ. These pathlines are referred to as basis flows. In the second phase, new pathlines can be calculated post hoc by interpolating from the basis flows.

The Lagrangian approach has potential advantages over the traditional (Eulerian) approach. With the Lagrangian approach, the basis flow are calculated in situ, giving it access to all spatiotemporal data. As a result, the basis flows accurately capture an interval in time. This contrasts with the Eulerian method, where vectors are stored and the post hoc analysis requires temporal interpolation and integration between time slices saved to disk. Overall, the Lagrangian approach has the potential to represent more information per byte compared to the Eulerian approach, enabling more accurate analysis for the same storage or enabling the same accuracy with less storage.

The advantages of in situ Lagrangian analysis were established through the study of Agranvosky et al. [ACG*14]. However, their evaluation has two significant issues. First, all of their results were comparative in nature. There was no information provided regarding spatial and temporal tradeoffs. In short, it showed Lagrangian techniques were superior to Eulerian techniques, but did not provide insights into how many basis flows were needed to achieve desired accuracies. Second, their accuracy metric focused on the end location of an interpolated particle trajectory, and did not consider the locations between the seed and the end point. This resulted in a limited overview of the accuracy of particle trajectories as a whole, and in particular for circular flow.

The purpose of this study is to address these two issues with their evaluation. The result both supplements the evaluation of Agranovsky et al. and also provides new understanding of the efficacy of their technique. We believe the evaluations in the current study will be the most useful comparators for future Lagrangian research that endeavors to improve on the work of Agranovsky et al. Specifically, our study focuses on the following points:

- We use an accuracy metric which evaluates the entire particle trajectory.
- Where the previous study considered reduced storage for only the Lagrangian approach, our study considers reduced storage for both approaches.
- We conduct experiments evaluating interpolation steps, which advance a particle forward in time. Specifically, we study the effect of large numbers of interpolation steps, each of which results in further advancement in time. This is important because each interpolation step has an associated error, and so multiple interpolation steps suffer from error propagation and accumulation.

† ssane@cs.uoregon.edu

2. Related Work

Lagrangian analysis has been a prominent technique within the flow visualization community for the past decade. Haller et al. [Hal01, HY00, Hal00] introduced Lagrangian Coherent Structures (LCS), which focuses on the calculation of stable and unstable manifolds to reveal features within a flow field. Additional work has gone on to accelerate the computation and visualization of LCS [GGTH07, GLT*09, SP07, SRP11].

In the remainder of this section, we focus on work related to storing flow field data, pathline interpolation, and error analysis.

Recent work has seen techniques proposed to store Lagrangian representations of flow fields. Bujack et al. [BJ15] suggest using parameter curves such as Bezier curves and cubic Hermite splines instead of polygonal chains. Sauer et al. [SXM16] present a new data representation which combines the Eulerian and Lagrangian reference frames into a joint format. Hlawatsch et al. [HSW11] and Agranovsky et al. [ACG*14] store flow field data directly.

Pathline interpolation can also be done using sparse particles via a method like moving least squares [AGJ11] or barycentric coordinate interpolation [ACG*14]. Chandler et al. [COJ15] demonstrate the use of Smoothed Particle Hydrodynamics (SPH) for pathline calculation. The use of SPH enables considering particles with mass, wherein each particle can have its own smoothing radius.

Research has also been focused on identifying sources of error in advection methods used in the Lagrangian paradigm. Chandler et al. present an error analysis of their interpolation-based pathline tracing system [CBJ16] and find that error roughly correlates with divergence in flow fields. Hummel et al. provide theoretical error estimates, which act as reasonable upper bounds for actual errors and suggest using this information in situ during calculation of flow field data [HBJG16].

3. Theoretical Background

In this section, we will provide a brief recap of the theoretical foundations. We use h_t to denote the resolution in time and h_x for the resolution in space.

Post-hoc advection in the Eulerian setting is typically performed using the fourth order Runge Kutta scheme, which is an iterative numerical integration method that has a total accumulation error of $O(h_t^4)$ [Atk08, Sch02]. Since we use it on top of discrete data interpolated multi-linearly in space, it actually is of the overall order $O(h_t^4 + h_x^2)$.

Previous work [BJ15] showed that the Lagrangian method as described by Agranovsky et al. [ACG*14] is also a numerical integration method with a total accumulation error of $O(h_x^2)$ for each interpolation. Since we approximate the intermediate values between two cycles using linear interpolation in time, it is of overall order $O(h_t^2 + h_x^2)$.

4. Experiment Overview

4.1. Study Configuration

We use the same in situ basis flow extraction and post hoc pathline interpolation technique as Agranovsky et al. [ACG*14]. Like the Agranovsky study, our in situ environment was theoretical, evaluating analytic data sets on the fly or loading simulation results from disk.

4.1.1. Datasets

We used the following data sets for our experiments —

Double Gyre — This data set is a two-dimensional flow field consisting of two counter-rotating gyres with a time dependent perturbation. This data set is simulated for 1024 cycles at a base grid resolution of 512×256. We set the period of the Double Gyre flow to 1000 cycles.

Arnold-Beltrami-Childress (ABC) — This data set is a time-dependent variant of the three-dimensional ABC analytic vector field [BCT01]. This data set is simulated for 400 cycles with a base grid resolution of $128 \times 128 \times 128$. We set the period of the ABC flow to 100 cycles.

Tornado — This data set is from a simulation of the dynamics of an F5 tornado [OWW15]. The base grid resolution is $490 \times 490 \times 280$. A mature tornado vortex exists in the domain during the 512 simulation seconds we considered for our experiments. Our collaborating scientist normally uses a temporal frequency of "every two simulation seconds" for his studies. Thus, we consider 256 time slices, with the time-steps evenly distributed from $t_0 = 8502$s to $t_{256} = 9014$s.

Let N_T denote the total number of time slices or cycles. For the Double Gyre data set, $N_T = 1024$. For the ABC data set, $N_T = 400$. For the Tornado data set, $N_T = 256$.

4.1.2. Storage Budget

We use the term storage budget to denote the allowed amount of data that can be saved to disk for post hoc pathline interpolation. We believe allowing both Lagrangian and Eulerian the same storage budget enables a fair comparison between them. Let N_C denote the number of cycles saved ($N_C \leq N_T$) and let P denote the number of basis flows or vector samples stored per cycle. If B denotes the storage budget for total number of basis flows or vector samples that can be saved, we select combinations of N_C and P such that $N_C \times P = B$. Further, we set the value of B to be equal to the number of points in the base grid resolution of each data set respectively. For the Double Gyre data set, $B = 131,072$ points. For the ABC data set, $B = 2.1$M points. For the Tornado data set, $B = 67.2$M points. For our experiments we consider three storage budgets ($1B, 2B, 4B$) for each data set. For each budget, we select multiple configurations that are combinations of N_C and P.

For example, the Double Gyre $2B$ test used 262,144 points. It further varied N_C with values ranging from 4 to 1024. For $N_C = 4$ we calculate four intervals of basis flows (Lagrangian) or four time slices of vectors (Eulerian), with each set containing $P = 65536$. Similarly, for $N_C = 1024$, there are 1024 sets, with $P = 256$.

4.2. Error Evaluation

For a given seed point, we calculate its corresponding pathline using three different methods.

- **Ground Truth** — We calculate the ground truth trajectories with a fourth-order Runge Kutta scheme [CK90] using the full spatial and temporal resolution available for each data set. The ground truth is considered to be perfectly accurate and have 0% error.
- **Lagrangian** — We calculate basis flows in situ with each selected configuration of N_C and P. We then use the calculated basis flows post hoc to interpolate new Lagrangian trajectories.
- **Eulerian** — We calculate Eulerian trajectories with each selected configuration of N_C and P, for comparison with the Lagrangian approach. Similar to ground truth calculation, a fourth-order Runge Kutta scheme is used to calculate the Eulerian trajectories.

Together, these three sets of trajectories can be used to evaluate and compare the approaches.

For both the ABC and Double Gyre data sets we randomly seed 1000 points over the entire flow field. For the Tornado data set, we place 144 seeds along rakes at locations used by our collaborating scientist to study the phenomena.

In contrast to the error metric used by Agranovsky et al. [ACG*14], we use a standard curve evaluation error metric — the L2-norm. The number of positions of a particle to represent the ground truth is equal to N_T. However, the number of known positions for a Lagrangian trajectory is N_C.

Given a test configuration average L2-norm is calculated as follows —

$$\frac{1}{p}\sum_{i=0}^{p}\frac{1}{n}\sum_{t=0}^{n}||x_{i,t}-g_{i,t}|| \tag{1}$$

where p is the total number of particles, $x_{i,t}$ is the location of a Lagrangian or Eulerian interpolated particle i at time t and $g_{i,t}$ is the location of the ground truth particle i at time t. We use two variants of the L2-norm to calculate the error —

- **Full L2-Norm Metric** When calculating the Full L2-norm, n is equal to N_T (total number of cycles).
- **Select L2-Norm Metric** When calculating the Select L2-norm, n is equal to N_C (number of cycles saved).

Agranovsky et al. used $n = 1$, which is similar in spirit to the Select L2-Norm; we add the Full L2-Norm for our evaluation to capture behavior along the interpolated trajectory at locations between the seed and the end point.

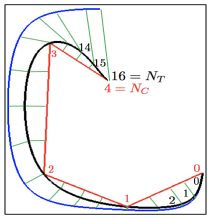


Figure 1: *Notional example of ground truth (black), Lagrangian (red), and Eulerian (blue) trajectories.*

Figure 1 illustrates a notional example of the difference between the accuracy metrics for a simplified sample trajectory. While the Lagrangian trajectory is accurate at its known points, the remainder of the trajectory can significantly deviate from the ground truth because it is linearly interpolated from the known points. We expect the Full L2-norm evaluation to show this error for low N_C configurations. In contrast, the Select L2-norm, which evaluates only at the known points along the trajectory, shows how close a particle is to the ground truth at these locations. Together, these error metrics provide a more comprehensive evaluation and understanding of a Lagrangian trajectory accuracy as a whole.

5. Results

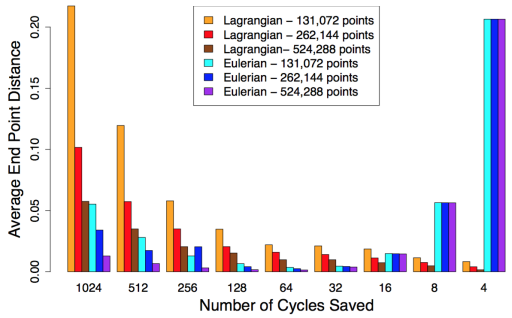


Figure 2: *Double Gyre analysis verifying Agranovsky et al.'s results.*

The first step of our evaluation was to verify the Agranovsky et al. results using their error metric. One of our results, for Double Gyre, is plotted in Figure 2. These results match their findings. Further, the results are similar to the Select L2-norm results in Figure 3b.

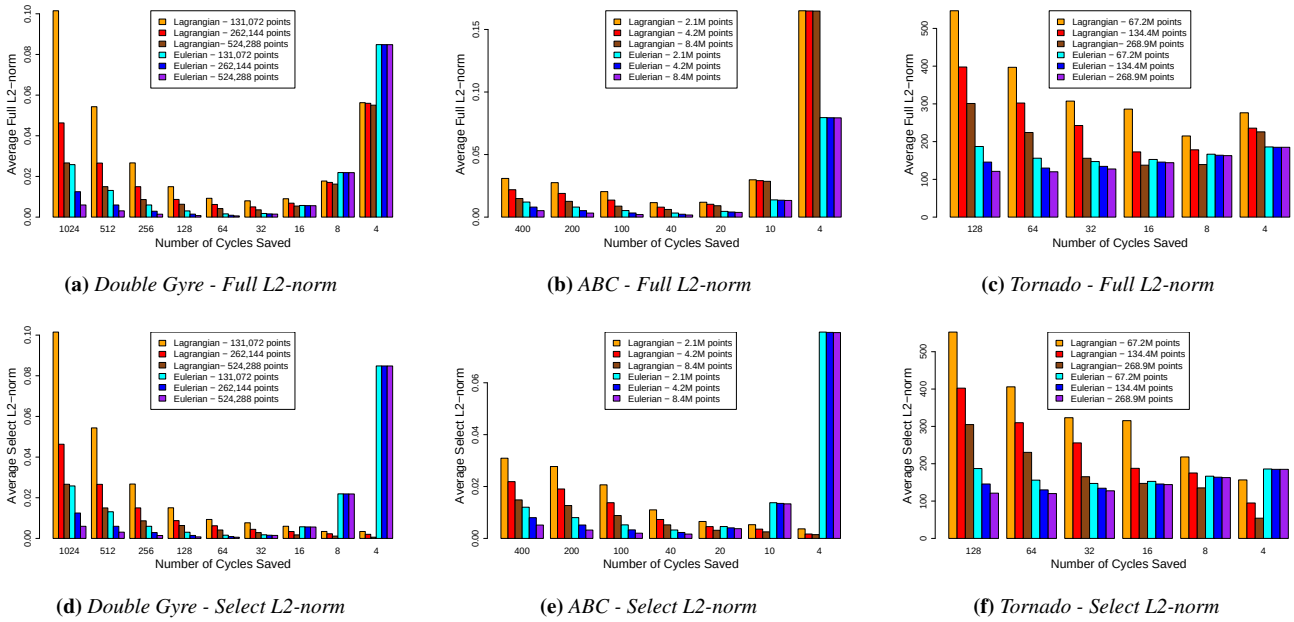
After verifying Agranovsky et al.'s results, we proceeded with our own study. The trends we observe, for both our spatiotemporal and error propagation analyses, are consistent regardless of data set. Figure 3 plots our results.

5.1. Spatiotemporal Tradeoff

For each of the Full L2-norm evaluations, the optimal values fell in between our largest and smallest N_C configurations (i.e., $N_C = 32$ for Double Gyre, 20 for ABC, and 8 for Tornado). This represents configurations using a sufficiently high P, enabling accurate interpolation, and sufficiently high N_C, such that the trajectory is well represented even with linear interpolation being performed between known points.

As N_C gets smaller, the Lagrangian trajectories have poor accuracy as a whole (see Full L2-norm result) even though the interpolated trajectory follows the ground truth closely at known locations (see corresponding Select L2-norm result). For example, in Figures 3a and 3b, for $N_C = 4$ and 8, we observe high error for Full L2-norm but low error for Select L2-norm. As demonstrated by Bujack et al. [BJ15], curve fitting can significantly reduce the Full L2-norm error for Lagrangian trajectories. For Eulerian configurations with a low N_C, the approach suffers from low temporal resolution and this is reflected in the high error for both metrics.

Further, we observe increases in storage benefit the Lagrangian approach more than the Eulerian approach. An increase in the number of basis flows reduces the interpolation error per step. Figure 4



(a) *Double Gyre - Full L2-norm* **(b)** *ABC - Full L2-norm* **(c)** *Tornado - Full L2-norm*

(d) *Double Gyre - Select L2-norm* **(e)** *ABC - Select L2-norm* **(f)** *Tornado - Select L2-norm*

Figure 3: *Evaluation results for Full L2-norm and Select L2-norm. Legends indicate the total data storage budget information.*

shows Double Gyre trajectories for multiple configurations, each using the same total storage.

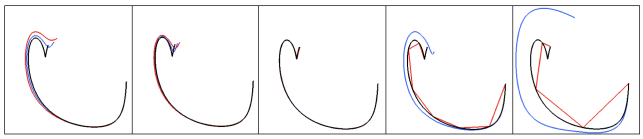


Figure 4: *Series of sample trajectories interpolated in the Double Gyre data set using varying number of cycles saved. Color code: Black - Ground Truth, Red - Lagrangian, Blue - Eulerian. From l-r : 1024, 512, 64, 8 and 4 cycles saved.*

5.2. Error Propagation

High N_C configurations in particular allow us to study the effect of large numbers of interpolations steps, each of which advances a particle forward in time. The performance of the Lagrangian trajectories for high N_C values is poor relative to the Eulerian approach. The first contributing factor is the low value of P (which, for this study, is inversely proportional to N_C to keep total storage constant). We observe large gains in accuracy with an increase in storage budget for these configurations. The second contributing factor is the error propagation which occurs when using the one-step second order integration method [HBJG16] for Lagrangian interpolation. The left-most set of trajectories in Figure 4 show the difference in error accumulation when using the second order integration method for the Lagrangian trajectory and the fourth order integration method used for the Eulerian trajectory.

6. Conclusion

Our study provided information regarding spatial and temporal trade-offs when working with a fixed storage budget. Further, by considering multiple storage budgets, our study informed trade-offs between data reduction and accuracy for the Lagrangian approach. With these results, future researchers can make better informed decisions regarding how many basis flows are needed to achieve reasonably high accuracy. Another takeaway from our study is an increased understanding of best practices for the Lagrangian approach with respect to tradeoffs between number of basis flows and frequency of output. Further, we suggest (and use) two variants of the L2-norm which together provide a more comprehensive evaluation of a Lagrangian trajectory.

Acknowledgment

This research was supported by the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration.

References

[ACG*14] AGRANOVSKY A., CAMP D., GARTH C., BETHEL E. W., JOY K. I., CHILDS H.: Improved post hoc flow analysis via lagrangian representations. In *Large Data Analysis and Visualization (LDAV), 2014 IEEE 4th Symposium on* (2014), IEEE, pp. 67–75. 1, 2, 3

[AGJ11] AGRANOVSKY A., GARTH C., JOY K. I.: Extracting flow structures using sparse particles. In *VMV* (2011), pp. 153–160. 2

[AJO*14] AHRENS J., JOURDAIN S., O'LEARY P., PATCHETT J., ROGERS D. H., PETERSEN M.: An image-based approach to extreme scale in situ visualization and analysis. In *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (2014), IEEE Press, pp. 424–434. 1

[Atk08] ATKINSON K. E.: *An introduction to numerical analysis*. John Wiley & Sons, 2008. 2

[BAA*16] BAUER A. C., ABBASI H., AHRENS J., CHILDS H., GEVECI B., KLASKY S., MORELAND K., O'LEARY P., VISHWANATH V., WHITLOCK B., ET AL.: In situ methods, infrastructures, and applications on high performance computing platforms. In *Computer Graphics Forum* (2016), vol. 35, Wiley Online Library, pp. 577–597. 1

[BCT01] BRUMMELL N., CATTANEO F., TOBIAS S.: Linear and nonlinear dynamo properties of time-dependent abc flows. *Fluid Dynamics Research 28*, 4 (2001), 237–265. 2

[BJ15] BUJACK R., JOY K. I.: Lagrangian representations of flow fields with parameter curves. In *Large Data Analysis and Visualization (LDAV), 2015 IEEE 5th Symposium on* (2015), IEEE, pp. 41–48. 2, 3

[CBJ16] CHANDLER J., BUJACK R., JOY K. I.: Analysis of error in interpolation-based pathline tracing. In *Proceedings of the Eurographics / IEEE VGTC Conference on Visualization: Short Papers* (2016), EuroVis '16, Eurographics Association, pp. 1–5. URL: https://doi.org/10.2312/eurovisshort.20161152. 2

[CK90] CASH J. R., KARP A. H.: A variable order runge-kutta method for initial value problems with rapidly varying right-hand sides. *ACM Transactions on Mathematical Software (TOMS) 16*, 3 (1990), 201–222. 2

[COJ15] CHANDLER J., OBERMAIER H., JOY K. I.: Interpolation-based pathline tracing in particle-based flow visualization. *IEEE transactions on visualization and computer graphics 21*, 1 (2015), 68–80. 2

[FMT*11] FABIAN N., MORELAND K., THOMPSON D., BAUER A. C., MARION P., GEVECIK B., RASQUIN M., JANSEN K. E.: The paraview coprocessing library: A scalable, general purpose in situ visualization library. In *Large Data Analysis and Visualization (LDAV), 2011 IEEE Symposium on* (2011), IEEE, pp. 89–96. 1

[GGTH07] GARTH C., GERHARDT F., TRICOCHE X., HANS H.: Efficient computation and visualization of coherent structures in fluid flow applications. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1464–1471. 2

[GLT*09] GARTH C., LI G.-S., TRICOCHE X., HANSEN C. D., HAGEN H.: Visualization of coherent structures in transient 2d flows. In *Topology-Based Methods in Visualization II*. Springer, 2009, pp. 1–13. 2

[Hal00] HALLER G.: Finding finite-time invariant manifolds in two-dimensional velocity fields. *Chaos: An Interdisciplinary Journal of Nonlinear Science 10*, 1 (2000), 99–108. 2

[Hal01] HALLER G.: Distinguished material surfaces and coherent structures in three-dimensional fluid flows. *Physica D: Nonlinear Phenomena 149*, 4 (2001), 248–277. 2

[HBJG16] HUMMEL M., BUJACK R., JOY K. I., GARTH C.: Error Estimates for Lagrangian Flow Field Representations. In *EuroVis 2016 - Short Papers* (2016), The Eurographics Association. 2, 4

[HSW11] HLAWATSCH M., SADLO F., WEISKOPF D.: Hierarchical line integration. *IEEE transactions on visualization and computer graphics 17*, 8 (2011), 1148–1163. 2

[HY00] HALLER G., YUAN G.: Lagrangian coherent structures and mixing in two-dimensional turbulence. *Physica D: Nonlinear Phenomena 147*, 3 (2000), 352–370. 2

[MOM*11] MORELAND K., OLDFIELD R., MARION P., JOURDAIN S., PODHORSZKI N., VISHWANATH V., FABIAN N., DOCAN C., PARASHAR M., HERELD M., ET AL.: Examples of in transit visualization. In *Proceedings of the 2nd international workshop on Petascal data analytics: challenges and opportunities* (2011), ACM, pp. 1–6. 1

[OWW15] ORF L., WILHELMSON R., WICKER L.: Visualization of a simulated Long-Track EF5 tornado embedded within a supercell thunderstorm. *Parallel Comput. 0*, 0 (2015). in press. 2

[Sch02] SCHATZMANN M.: *Numerical Analysis: A Mathematical Introduction*. Oxford University Press, New York, USA, 2002. 2

[SP07] SADLO F., PEIKERT R.: Efficient visualization of lagrangian coherent structures by filtered amr ridge extraction. *IEEE Transactions on Visualization and Computer Graphics 13*, 6 (2007), 1456–1463. 2

[SRP11] SADLO F., RIGAZZI A., PEIKERT R.: Time-dependent visualization of lagrangian coherent structures by grid advection. In *Topological Methods in Data Analysis and Visualization*. Springer, 2011, pp. 151–165. 2

[SXM16] SAUER F., XIE J., MA K.-L.: A combined eulerian-lagrangian data representation for large-scale applications. *IEEE Transactions on Visualization and Computer Graphics* (2016). 2

[VHP11] VISHWANATH V., HERELD M., PAPKA M. E.: Toward simulation-time data analysis and i/o acceleration on leadership-class systems. In *Large Data Analysis and Visualization (LDAV), 2011 IEEE Symposium on* (2011), IEEE, pp. 9–14. 1

[WFM11] WHITLOCK B., FAVRE J. M., MEREDITH J. S.: Parallel in situ coupling of simulation with a fully featured visualization system. In *Proceedings of the 11th Eurographics Conference on Parallel Graphics and Visualization* (2011), EGPGV '11, Eurographics Association, pp. 101–109. URL: http://dx.doi.org/10.2312/EGPGV/EGPGV11/101-109. 1

[YWG*10] YU H., WANG C., GROUT R. W., CHEN J. H., MA K.-L.: In situ visualization for large-scale combustion simulations. *IEEE computer graphics and applications 30*, 3 (2010), 45–57. 1