SAND2011-4088C

# Conditional Analysis of DNS Combustion Data Using Local and Global Shape Characteristics

**Janine C. Bennett[1], Peer-Timo Bremer[2], Jacqueline H. Chen[1], Ray W. Grout[3], Andrea Gruber[4], Attila Gyulassy[2], Evatt R. Hawkes[5], Hemanth Kolla[1], and Valerio Pascucci[2]**

[1] Sandia National Laboratories, Livermore, CA, USA
[2] University of Utah, Salt Lake City, UT, USA
[3] National Renewable Energy Laboratory, Golden, CO, USA
[4] Sintef, Trondheim, Norway
[5] University of New South Wales, Sydney, Australia

E-mail: jcbenne@sandia.gov, ptbremer@llnl.gov, jhchen@sandia.gov, ray.grout@nrel.gov, andrea.gruber@sintef.no, aggyulassy@ucdavis.edu, evatt.hawkes@unsw.edu.au, hnkolla@sandia.gov, pascucci@acm.org

**Abstract.** Feature-based conditional statistical methods are essential for the analysis of complex, large-scale data. We introduce two shape-based conditional analysis algorithms that can be deployed in complementary settings: local methods are required when the phenomena under study comprises many small intermittent features, while global shape methods are required to study large-scale structures. We present the algorithms in context with their motivating combustion science case studies, but note that the methods are applicable to a broad class of physics-based phenomena.

## 1. Introduction

As high performance computing (HPC) resources continue to improve, scientists are able to study physical phenomena with unprecedented resolution and complexity. For example, direct numerical simulations (DNS) are first principle high-fidelity computational fluid dynamics (CFD) simulations in which Navïer-Stokes equations are numerically solved on a computational mesh in which all of the spatial and temporal scales of turbulence can be resolved. These detailed simulations are performed by the combustion science community to study fundamental turbulence-chemistry interactions in an effort to reliably predict efficiency and pollutant emissions for new engines and fuels. Through computing enabled by SciDAC, DOE INCITE and T2O grants, a library of DNS configurations and parametric studies [HSSC07, CCdS$^+$09, YSC09] have been performed and archived using S3D, a massively parallel simulation code [CCdS$^+$09], developed at the Combustion Research Facility at Sandia National Laboratories. The unique benchmark data generated by S3D is used to glean fundamental insight into combustion processes, and to aid the development and validation of engineering models used in CFD to optimize the design of future combustors.

The data that is output from S3D is extremely large and multi-dimensional and, with increased computational resources, the data will continue to increase in both size and complexity. For example, a recent run that modeled a lifted ethylene jet flame [YSC09] comprised 1.3 billion grid points, 22 chemical species, vector field data, and contained 40 million active particles being tracked within the flow at any time. On the order of 240 TB of raw field data was written to disk along with 50TB of particle data. Feature identification and characterization in such massive data is challenging and is further complicated by the fact that turbulence is a chaotic phenomena characterized by a wide range of scales – 4 decades of spatial scales from microns to centimeters and temporal scales from nanoseconds to milliseconds.

Historically statistical analysis has been used to summarize large data in a succinct manner. However, when the data is very large and/or features exist in a wide range of scales, global aggregation is problematic as features of interest may be averaged towards the mean. In this paper we present two algorithms that were motivated by the need to perform feature-based conditional analysis in two complementary settings. Both algorithms define features according to the shape of level sets structures that have been extracted from one of the scalar fields in the data; however, the algorithm described in Section 2 should be employed when there exist many small features in relation to the size of the domain, while the algorithm presented in Section 3 is applicable when there exist few relatively large features in relation to the size of the domain.

## 2. Conditional Analysis Using Local Shape Characteristics

*Motivation* The scalar dissipation rate, $\chi$, is the rate at which scalar fluctuations decay due to diffusive processes. Thin pancake-like regions of locally high $\chi$ are created by compressive turbulent strains and the thickness of these structures is assumed to be correlated with length scales of turbulence. In an effort to gain insight into the relationship between mechanical strains and chemical processes in turbulent mixing, one interesting characteristic is the relationship between the thickness and the mean temperature within these structures.

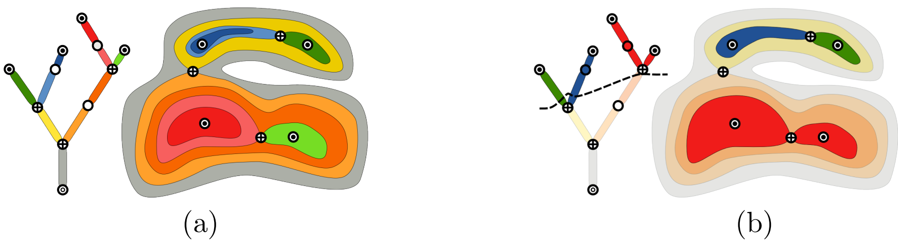


**Figure 1.** (a) The merge tree provides a complete history of how families of level sets interact as a function is swept from global maximum to global minimum. On the right a height function is defined on a 2-dimensional plane with associated merge tree shown on the left, with each arc in the tree corresponding to the region with the same color. (b) A relevance based segmentation for relevance slightly above 0.2 (slightly below 80% of the local maximum per branch).

*Conditional Analysis* There are typically many small regions of high $\chi$, distributed intermittently throughout the simulation domain. The merge tree [CSA03], has a history of being deployed successfully to encode such phenomena, where feature threshold values may vary locally and are not necessarily known *a priori*[MGB$^+$10, BWP$^+$10]. Let $M$ be a simply connected domain with the function $f : M \rightarrow \mathbb{R}$. The level set $L(v)$ of $f$ at isovalue $v$ is defined as the collection of all points on $\mathbb{R}$ with function value equal to $v$ and each connected component of a level set is called a contour. The merge tree of $f$, as shown in Figure 1 (a), represents the merging of contours as the isovalue $v$ is varied from global maximum to global minimum through the range of $f$. Each branch of the tree represents a family of contours that continuously evolve without merging as $v$ is lowered.

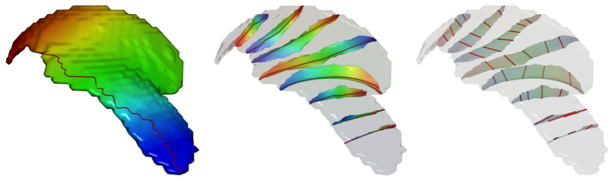


**Figure 2.** The first three length-scales are estimated using a spectral technique in which each shape is parametrized according to its first non-trivial eigenvector to compute its length. The same technique is performed recursively on iso-contours of the first eigenvector to compute the width and thickness.

Regions of high $\chi$ are identified using a a relevance-based [MGB$^+$10] metric in which each node in the tree is scaled by its local maximum – the highest in its corresponding subtree. As shown in Figure 1 (b), features are defined as the subtree above a user-specified relevance value that lies in $[0, 1]$.

For each feature we compute its mean temperature value, temperature variance and first three length scales. Length, width and thickness are estimated using a spectral technique similar to the one introduced by [RWSN09]. We parametrize each shape according to its first non-trivial eigenvector to compute its length, and use the same technique recursively on iso-contours of the first eigenvector to compute the width and thickness, see Figure 2. The plot in Figure 3 shows the distribution of $\chi$-thicknesses, at a relevance of 0.85, computed for structures grouped by the mean temperature in the segment for four bins, each 250 Kelvin wide. To ensure that the results are not influenced by excessive internal temperature variations, only structures with a low temperature variance (below 5% of the maximum variance) are considered.

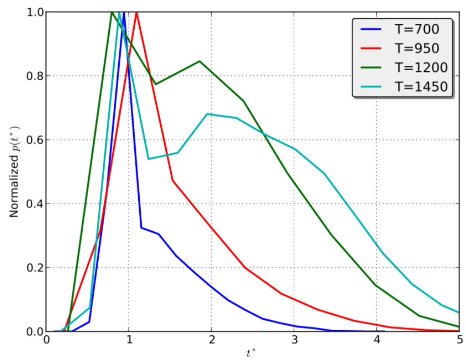



**Figure 3.** The distribution of the $\chi$-thicknesses are computed for segments grouped by the mean temperature in the segment.

## 3. Conditional Analysis using Global Shape Characteristics

*Motivation* Characterizing flame stabilization in transverse jets is complicated by the fact that the flame is affected by large-scale boundary layer structures. When aggregating statistics using traditional methods, the large, slow motions induced by the boundary structures require the simulation to be run for prohibitively many time steps in order to guarantee convergence. We introduce a new method for computing statistics conditional on the bulk flame position, in which convergence of statistics takes place with many fewer simulation timesteps.

*Conditional Analysis* Conceptually we would like to identify slabs normal to the jet centerline and compute conditional statistics within each slab, see Figure 4 (a). However, the centerline is not trivial to define in a rigorous way. Furthermore, the jet is not symmetric around the centerline due to differences in the physics of mixing on the top and bottom of the jet flame. To address these issues, we introduce a parameterization of the jet trajectory and compute statistics based on the resulting coordinate system, see Figure 4 (b). Because the jet isosurface spans most of the simulation domain, we first reduce the resolution of the mesh using QSlim [GH97]. In practice we are able to reduce the size of the mesh many orders of magnitude without degrading

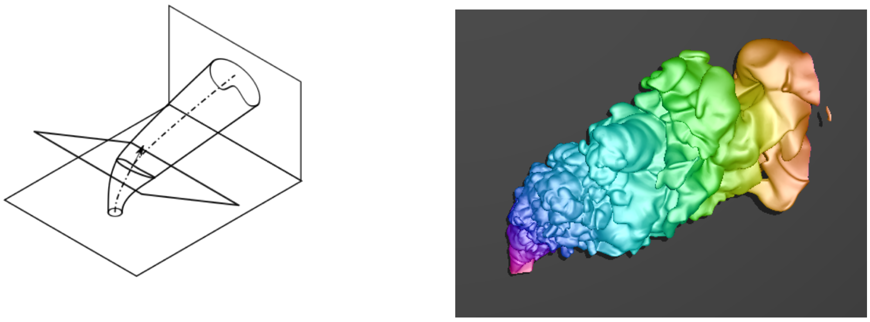


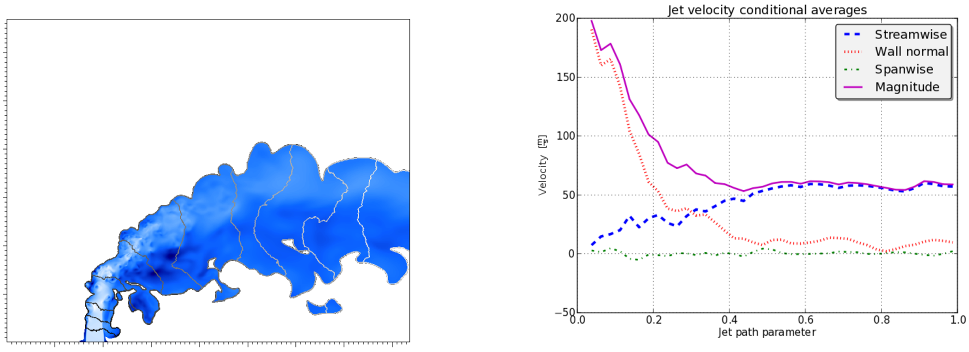

**Figure 4.** (a) This cartoon depicts a the center line of a flame with a slab indicated by the plane. While this is a nice conceptual notion, in practice it is not easy to define rigorously. (b) A jet-based coordinate system allows for the aggregation of averages conditioned on bulk flame position.

the quality of the resulting parameterization. Next, solving Laplace's equation on the simplified jet surface captures the trajectory at boundaries and varies these smoothly from entrance to exit, even for convoluted surfaces. During this step we constrain the jet inlet to zero and constrain the exit boundary to one. We tetrahedralize the interior of the jet using TetGen[Si], extending our surface parameterization to the interior of the jet by solving Laplace's equation again, now using the surface solution as boundary conditions. The tetrahedralization provides a four-dimensional point cloud comprised of position and parameter value. Parameter values are computed for all grid points inside the jet by loading the four-dimensional point cloud into S3D and using Approximate Nearest Neighbors (ANN) code [MA] in-situ to identify the nearest point cloud neighbor for each grid point.

Using this parameterization pipeline, scalar quantities are conditionally averaged on the jet path parameter for the entire jet interior (defined as mixture fraction greater than 0.05). Averages are aggregated across a full flow-through time (0.4ms) for the 202 timesteps that were saved to disk. Figure 5 (a) shows a slice of the velocity magnitude with isolines of the jet path parameter depicted. Figure 5 (b) shows the variation of velocity magnitude and the three components: stream wise (left-to-right w.r.t Figure 4(b)), wall normal (bottom-to-top) and span wise (into plane) velocities, averaged in time and conditioned on the jet path parameter. The average velocity components vary nearly monotonically since the parametrization adapts to the bending and flapping of the jet and provides a natural coordinate system to examine statistical quantities meaningfully. Due to the large scale three dimensional flow motions, statistical quantities computed in a traditional Cartesian coordinate system would not have been as informative.

## 4. Conclusion

Feature-based conditional analysis is increasingly important as the size and complexity of simulations continues to increase. In this paper we introduce two shape-based conditional analysis algorithms that can be deployed in complementary settings. When many small features exist whose threshold definitions are not known *a priori*, a merge tree efficiently encodes features of interest augmented with various statistics (first- through fourth-order moments, minimal and maximal function values) as well as shape metrics. Statistical summaries, including conditional empirical cumulative density functions, histograms, timeseries and parameter studies can be efficiently generating using our data structures. When there are relatively few large features, a parameterization of the large-scale structure(s) is computed and statistics are aggregated




(a) (b)

**Figure 5.** (a) A slice of the velocity magnitude with isolines of the jet path parameter. (b) Various velocity measures plotted versus the jet path parameter.

conditional on a coordinate system derived from the shape of the feature itself. While both algorithms were motivated by needs in combustion science, they can be applied generally to a broad family of physics-based simulation data.


## Acknowledgements

Sandia National Laboratories is a multi-program laboratory managed and operated by Sandia Corporation, a wholly owned subsidiary of Lockheed Martin Corporation, for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-AC04-94AL85000.


## References


[BWP+10] P.-T. Bremer, G. H. Weber, V. Pascucci, M. Day, and J. B. Bell. Analyzing and tracking burning structures in lean premixed hydrogen flames. *Visualization and Computer Graphics, IEEE Transactions on*, 16(2):248–260, March-April 2010.

[CCdS+09] J H Chen, A Choudhary, B de Supinski, M DeVries, E R Hawkes, S Klasky, W K Liao, K L Ma, J Mellor-Crummey, N Podhorszki, R Sankaran, S Shende, and C S Yoo. Terascale direct numerical simulations of turbulent combustion using S3D. *Comput. Science & Discovery*, 2(1):015001, 2009.

[CSA03] H. Carr, J. Snoeyink, and U. Axen. Computing contour trees in all dimensions. *Comput. Geom. Theory Appl.*, 24(3):75–94, 2003.

[GH97] M. Garland and P. Heckbert. Surface simplification using quadric error metrics. *Proceedings of SIGGRAPH'97*, pages 209–215, 1997.

[HSSC07] Evatt R. Hawkes, Ramanan Sankaran, James C. Sutherland, and Jacqueline H. Chen. Scalar mixing in direct numerical simulations of temporally evolving plane jet flames with skeletal co/h2 kinetics. *Proceedings of the Combustion Institute*, 31(1):1633 – 1640, 2007.

[MA] David Mound and Sunil Arya. Ann: A library for approximate nearest neighbor searching. `http://www.cs.umd.edu/~mount/ANN`.

[MGB+10] A. Mascarenhas, R. W. Grout, P.-T. Bremer, E. R. Hawkes, V. Pascucci, and J. H. Chen. *Topological feature extraction for comparison of terascale combustion simulation data*, pages 229–240. Mathematics and Visualization. Springer, 2010. to appear.

[RWSN09] Martin Reuter, Franz-Erich Wolter, Martha Shenton, and Marc Niethammer. Laplace-beltrami eigenvalues and topological features of eigenfunctions for statistical shape analysis. *Computer-Aided Design*, 41(10):739–755, 2009.

[Si] Hang Si. Tetgen. `http://tetgen.berlios.de/`.

[YSC09] C.S. Yoo, R. Sankaran, and J.H. Chen. Three-dimensional direct numerical simulation of a turbulent lifted hydrogen jet flame in heated coflow: Flame stabilization and structure. *J. Fluid Mech.*, 640:453–481, 2009.