Applied Mathematics: Topology for Statistical Modeling of Petascale Data

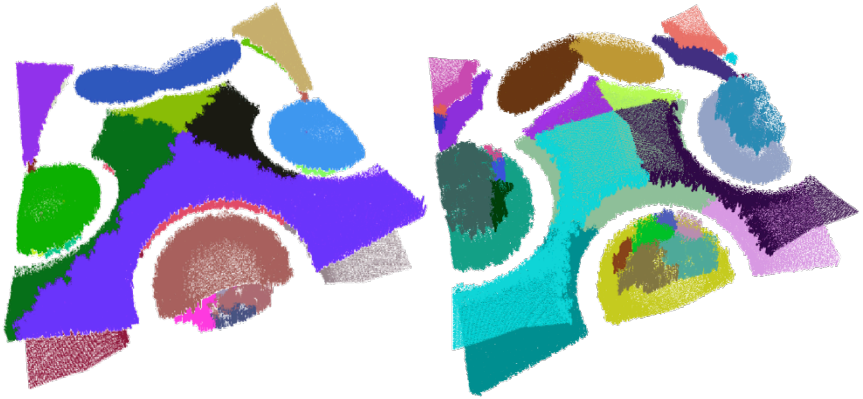
SAND2011-4919P

Objectives

- **Develop new mathematical tools combining statistical and topological methods for the analysis of large-scale, complex data.**
- **Demonstrate that our new techniques preserve topological and statistical features and that the algorithms developed will work at petascale.**

Impact

- **Introduced hixels, a novel, information-rich data representation that trades data size and complexity for scalar-valued uncertainty, to the scientific community.**
- **Developed a suite of tools for the identification of features in hixelated data.**



Our novel hixel technique enables feature identification on extremely large or uncertain data. We statistically associate instances of subsets of the data, and apply topological segmentation methods per subset to recover views of large-scale features that would either be infeasible to compute or ambiguous to infer within the original data.

Accomplishments

- **Developed several hybrid approaches to analyze the results of petascale simulations:**
  - Computing statistics on topologically-defined features, averaging that can obscure insight is avoided.
  - Using statistically accurate subsampling, topological features can be computed at scales otherwise impossible.
  - Examining topologically-defined features of statistical constructs (*e.g.* contingency tables or random variables), information available due to increased notational complexity is communicated.
- **"Analysis of Large-Scale Scalar Data Using Hixels",** by D. Thompson, J. Levine, J. Bennett, P.-T. Bremer, A. Gyulassy, V. Pascucci, & P. Pébay, submitted to IEEE Symp. on Large-Scale Data Analysis and Visualization, 2011.
- **"Design and performance of a scalable, parallel statistics toolkit"**, by P. Pébay, D. Thompson, J. Bennett, & A. Mascarenhas. In Proc. 25th IEEE International Parallel & Distributed Processing Symp., 12th International Workshop on Parallel and Distributed Scientific and Eng. Computing. Anchorage, AK, U.S.A., May 2011.
- **"Optimizing n-variate (n+k)-nomials for small k",** by P. Pébay, M. J. Rojas, & D. Thompson, Theoretical Comp, Science, 412(16):1457--1469, 2011.

P. Pébay, D. Thompson, J. Bennett, A. Mascarenhas, V. Pascucci, A. Gyulassy, J. Levine, J. M. Rojas, K. Rusek

U.S. DEPARTMENT OF ENERGY | Office of Science
Sandia National Laboratories | The University of Utah | SCI Institute | Texas A&M University