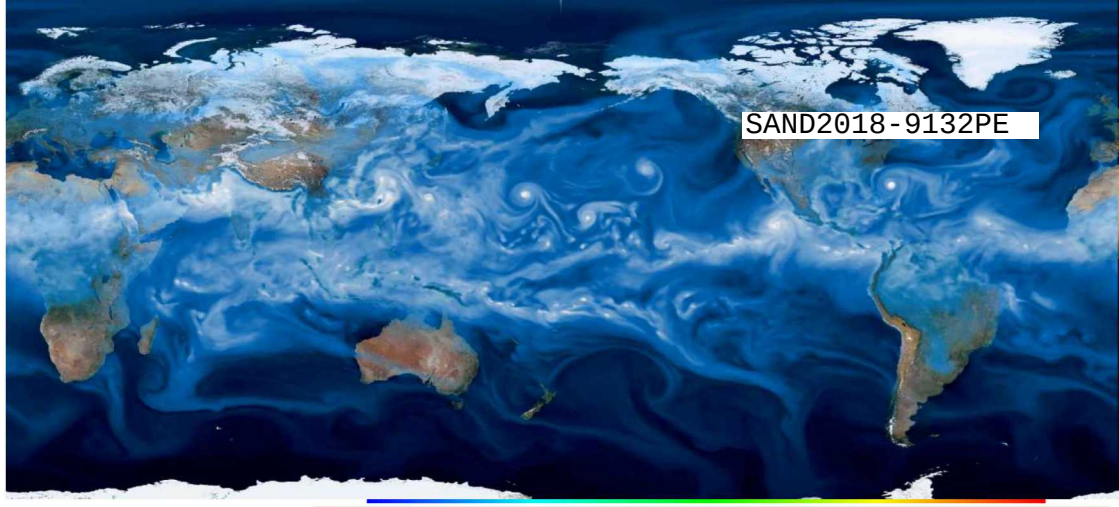
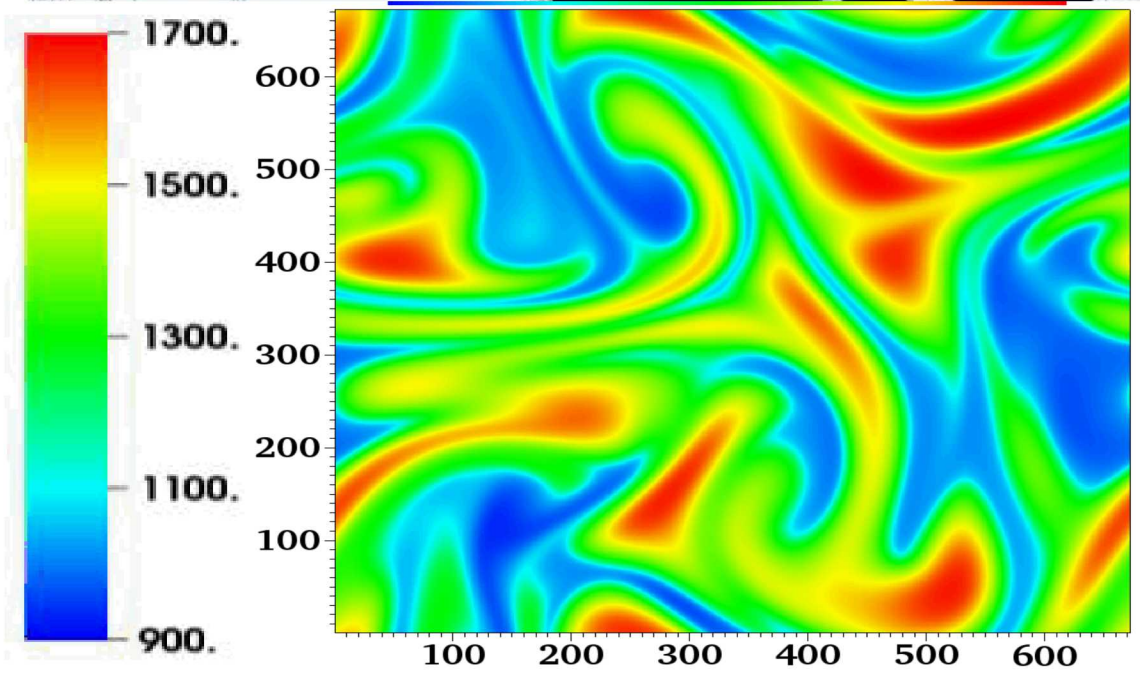
In-Situ Machine Learning for Intelligent Data Capture on Exascale Platforms





Warren L. Davis IV
Kevin A. Reed

August 22, 2018

U.S. DEPARTMENT OF ENERGY | Office of Science

Sandia National Laboratories is a multi-mission 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

Sandia National Laboratories | Stony Brook University

Team

- Warren L. Davis IV (PI-Sandia)
- Kevin Reed (PI-Stony Brook)
- Danny Dunlavy
- Philip Kegelmeyer
- Hemanth Kolla
- Aditya Konduri
- Tim Shead

Problem

- Scientific computation often involves running computationally intense simulations on HPC
- Goal is to find interesting events (e.g., auto-ignition, cyclones)
- Current HPC Simulation strategy for detection of events and anomalies involves saving data to disk at regular intervals.

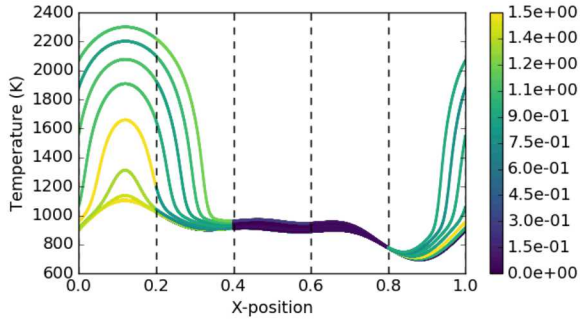
- Overhead for I/O is large
  - Writing everything is too expensive
  - Writing at infrequent intervals may lead to missed events, or loss of critical information
  - Lost information can only be regained by rerunning the simulations and adjusting the save interval.

Research Goals

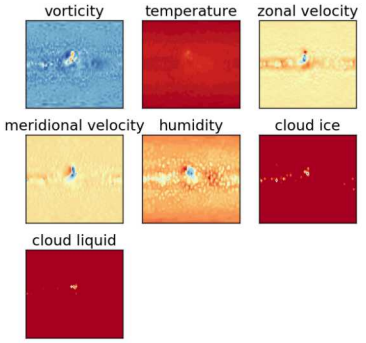
- Develop efficient distributed machine learning and anomaly detection algorithms to enable intelligent data capture.
- These algorithms will be used to determine localized events of interest *in situ*, and the data will be selectively saved at the relevant time steps and spatial locations.
- The machine learning techniques will be implemented and validated on DOE-relevant test cases:
  - Auto-ignition in combustion simulations
  - Extreme weather prediction in climate simulations

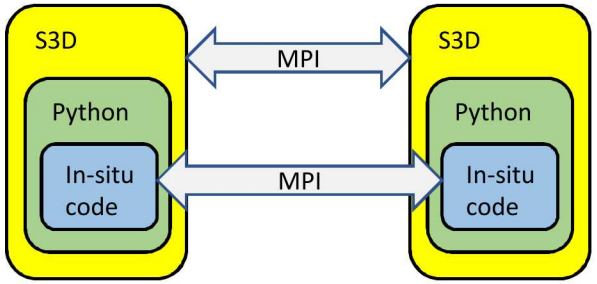
Background (FY17)

- Created a new event detection algorithm, FIEDA, which detects some basic events of interest in combustion and climate.



- Established vehicle for In-Situ machine learning tests on actual scientific simulations on HPC
  - Domain experts/developers actively engaged in making this possible
  - S3D is widely used, increasing the potential applicability of this research

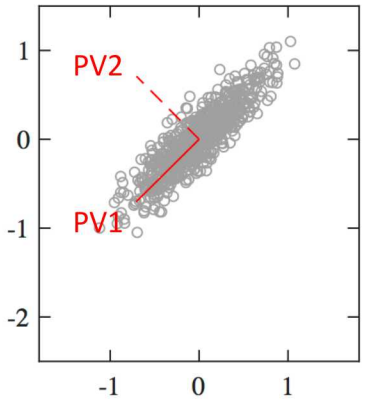




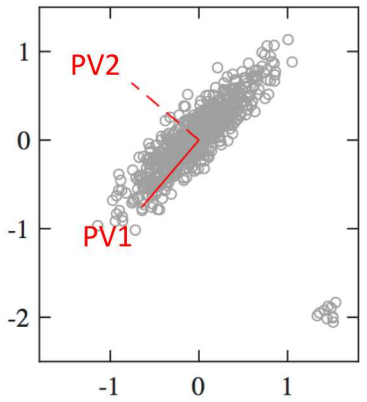
Co-Kurtosis for Anomaly Detection

- Covariance matrix + eigenvalue decomposition

"Normal" Data



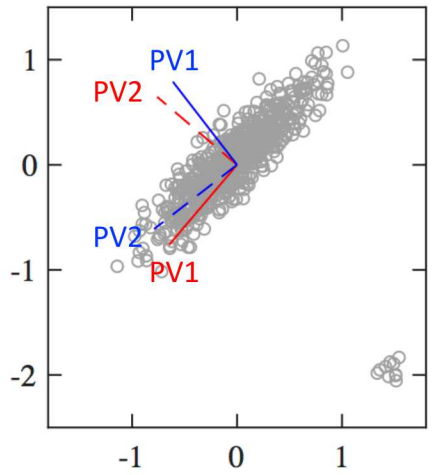
Data with Anomalies



- Principal components (orientation and magnitude) change with changes in distribution
- Vectors do not align with the direction of the anomalies

Co-Kurtosis for Anomaly Detection

- Kurtosis
  - either existing outliers (for the sample kurtosis) or propensity to produce outliers (for the kurtosis of a probability distribution)

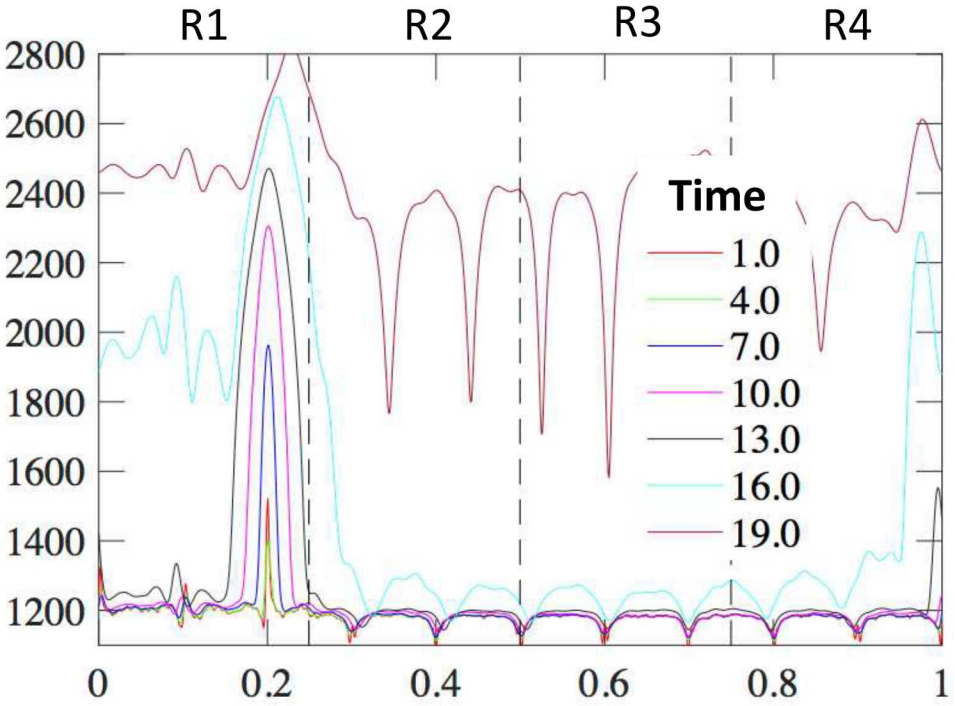


Second Order Moment
Fourth Order Moment

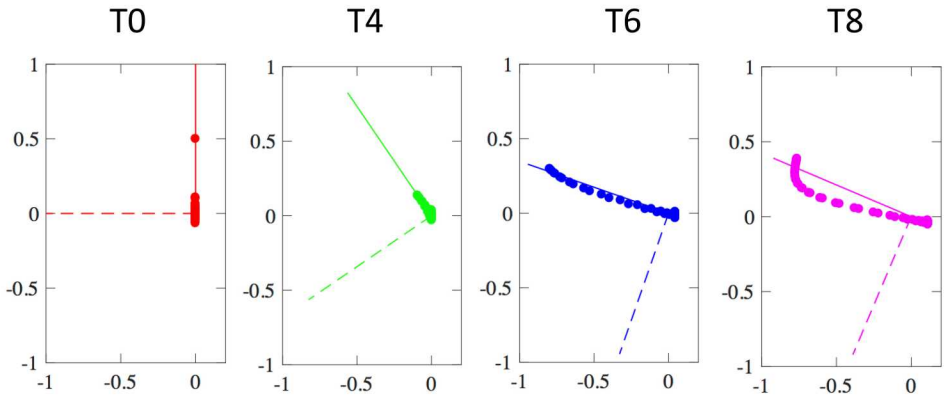
- Principal vectors, weighted by principal values, can be projected onto the features to create a Feature Moment Metric.
- The distance between the FMM distributions can be measured against a threshold to detect anomalous nodes spatially or temporally

Auto-Ignition Test Case

- 1-d Combustion Example
- Early ignition occurs in Region 1
- Spatial anomaly in Region 1
- Eventually temporal anomaly in Regions 2, 3, and 4

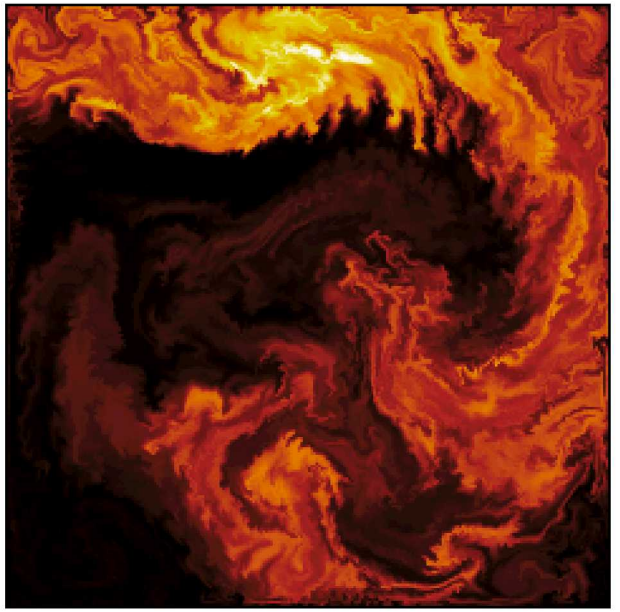


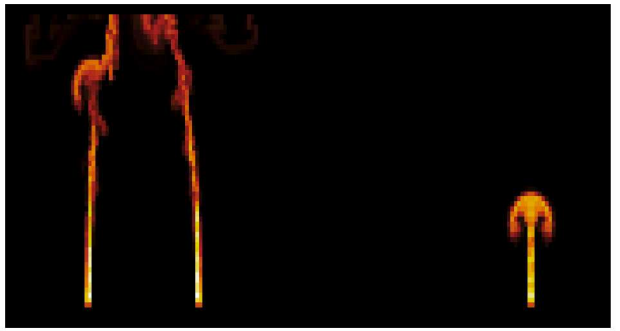
Temporal Evolution of Principal Kurtosis Vectors



Mantaflow

- Additional in-situ platform
  - Simulate HPC environment from analysis algorithm perspective
- Turbulent flows
- Allows Rapid prototyping
  - Developed about 30 new algorithms/approaches
  - Inspired new framework





Signatures, Measures, Decisions

- **Signatures**
  The representation of the simulation data on a node

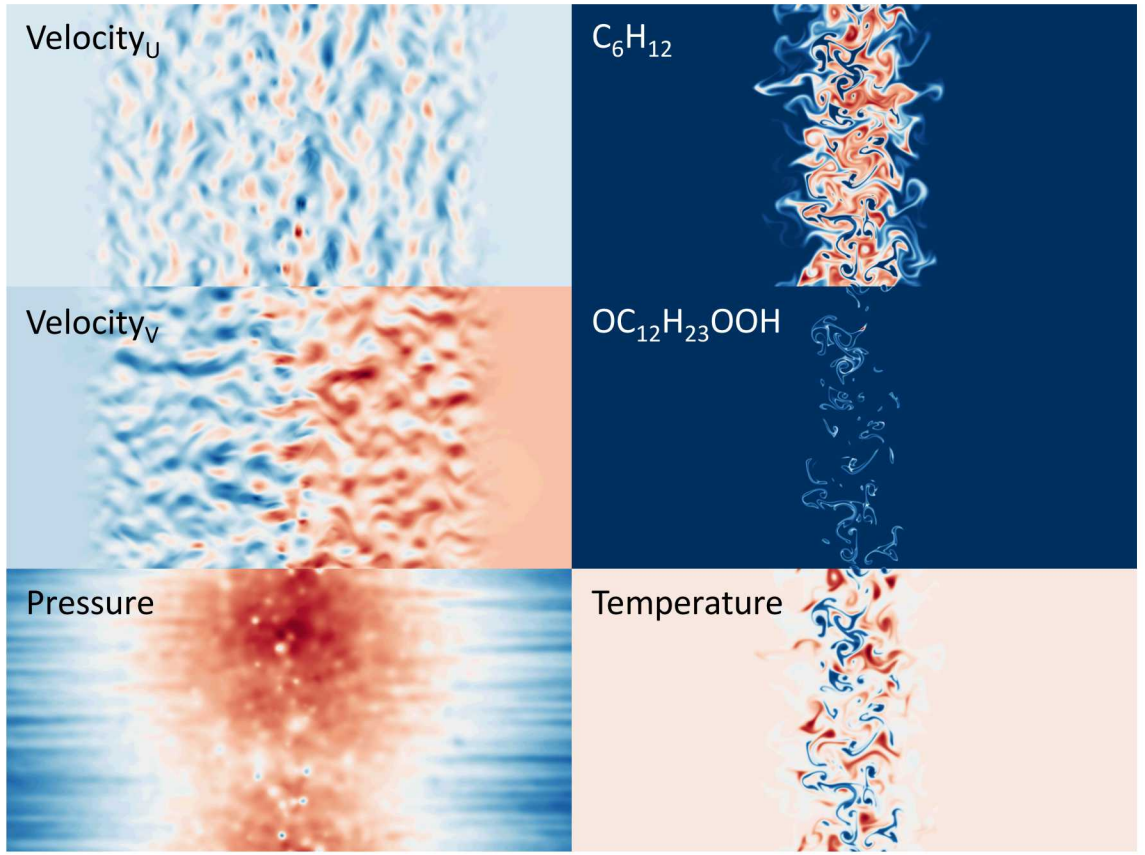
- **Measures**
  A representation of how close a signature is to other signatures

- **Decisions**
  An arbitration of the measures to determine which nodes contain “interesting” data, given the signatures and measures

- This, in combination with Manta Flow, has allowed us to much more rapidly come up with new algorithms (and combinations) to test against various use cases.

Advanced Combustion Simulations

A mixing layer slab of fuel (n-dodecane a.k.a NC12H26) in the middle and air (O2 and N2) on either sides, that are at high temperature and are bound to mix and ignite.



Publications/Presentations

- Ling, Julia, W. Philip Kegelmeyer, Aditya Konduri, Hemanth Kolla, Kevin A. Reed, Timothy M. Shead and Warren L. Davis IV. “Using feature importance metrics to detect events of interest in scientific computing applications.” *2017 IEEE 7th Symposium on Large Data Analysis and Visualization (LDAV)* (2017): 55-63.

- Kolla, Hemanth, Aditya Konduri, Prashant Rai, Tamara G. Kolda, Warren Leon Davis. “Tensor Decomposition to Perform Change of Basis in Multi-Variate HPC Data to Preserve Higher Order Statistical Moments,” *Presentation,* SIAM Parallel Processing 2018, March 2018.

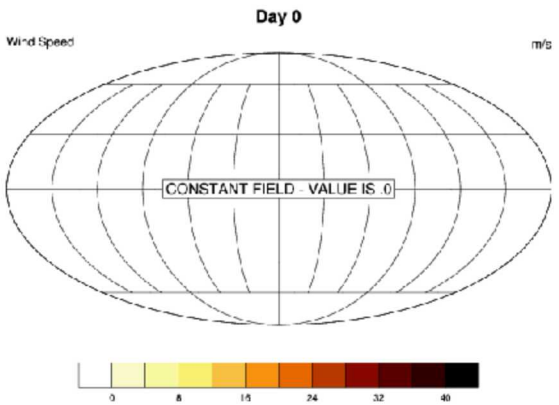
- Konduri, Aditya, Hemanth Kolla, Julia Ling, W. Philip Kegelmeyer, Timothy Shead, Daniel Dunlavy, Warren Leon Davis. Event Detection in Multi-Variate Scientific Simulations Using Feature Anomaly Metrics," *Presentation,* SIAM Parallel Processing 2018, March 2018.

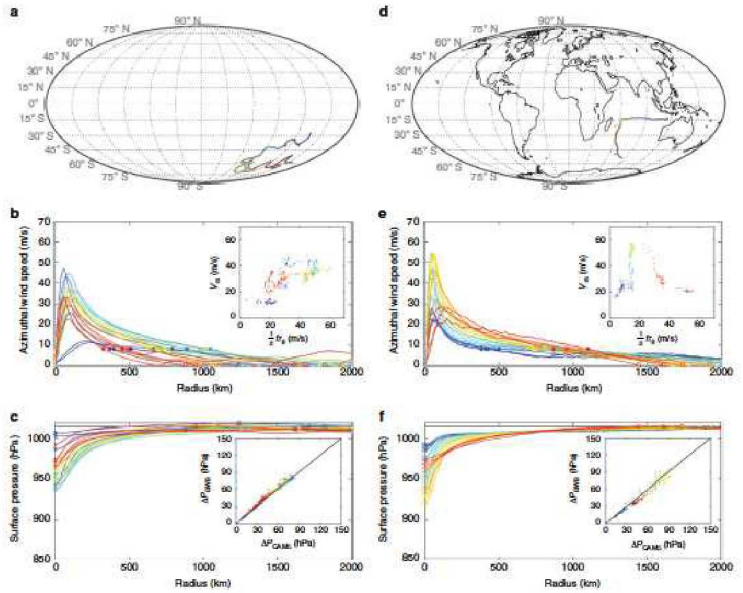
Future directions

- Focus on Temporal Analysis
- Characterization of algorithms/applications/performance
- Explore finite-element simulations
- In-Situ Integration with CESM/E3SM
- Analysis of more advanced combustion, climate, and turbulence simulations

Development and Testing of Controlled Problem

- Does a better climate model setup exist for testing both spatial and temporal anomalies?
- The thought is full rotating radiative convective equilibrium (RCE) where tropical cyclone spin up spontaneously.





Explored initially in Chavas et al. 2017 (Nature Comms), but **science applications** being explored in Chavas and Reed, 2018.