

---

---

# Using Data Mining Techniques in Support of Office of Science Applications

---

---

**Chandrika Kamath**

*Center for Applied Scientific Computing*

**5 October 2005**

UCRL-PRES-216185: This work was performed under the auspices of the U.S.  
Department of Energy by University of California Lawrence Livermore National  
Laboratory under contract No. W-7405-Eng-48.



<http://www.llnl.gov/casc/sapphire/>



---

---

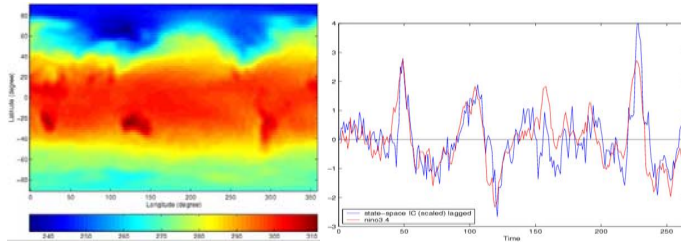
## Part I: The past and the present

---

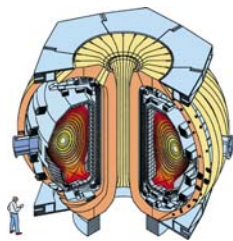
---

Joint work with  
Abraham Bagherjeiran, Erick Cantú-Paz, Imola Fodor,  
Abel Gezahegne, Cyrus Harrison, Siddharth Manay, Nu Ai Tang

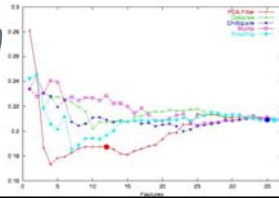
## Our early work was in dimension reduction and feature selection



Separating signals in climate data



Identifying features connected to edge-harmonic oscillations in DIII-D Tokamak



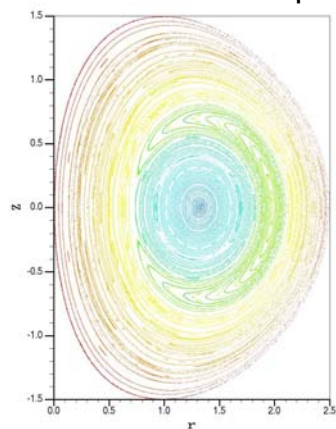
**UPDATE:** Licensed dimension reduction and anomaly detector software to GAT in Aug 2005. Work funded by Scidac and DOE NNSA ASC Program.

CASC

Sapphire/SDM AHM/CK 3

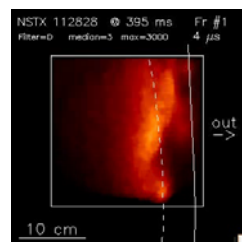
## Our current work focuses on analysis of data from PPPL

Classification of Poincaré plots



Joint work with Neil Pomphrey, Don Monticello, Scott Klasky

Tracking of plasma



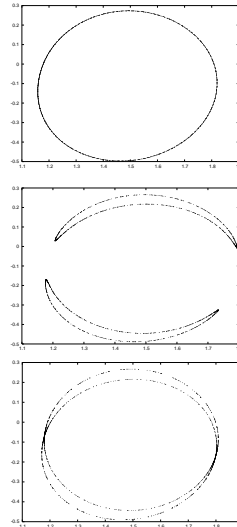
Joint work with Stewart Zweben, Scott Klasky

CASC

Sapphire/SDM AHM/CK 4

## Our progress on analysis on Poincare plots: classification of a single orbit

- Piece-wise polynomial approach
- Graph-based approach
  - the default KAM
  - the customized KAM
  - feature-based classifier
- Current status
  - PPoly and KAM codes ready
  - need to find a way to incorporate the feature-based classifier
  - need more objective comparisons
  - a paper on graph-based approach: submitted

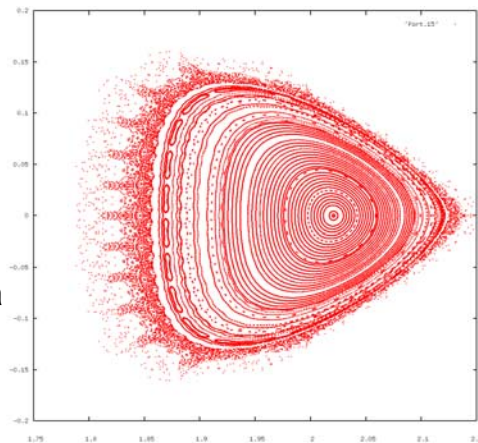


Sapphire/SDM AHMCK 5

CASC

## The real problems in analysis of Poincaré plots and why they are difficult

- Find width of separatrix
- Find width of islands (if no separatrix)
- Find width of ??? (if no islands)
- Handling orbits with few points
- Handling multi-orbit data
- Noisy data



Our next steps

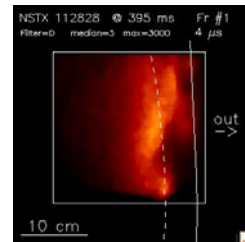
CASC

Sapphire/SDM AHMCK 6

## Tracking plasma in NSTX images: work just started

---

- Problem: track the plasma over time
- IDL code implementing a variant of block matching is too slow
- Implementing other block-matching approaches



National Spherical  
Torus Experiment