Md Mahmudur Rahman
Morgan State University

My expertise in the fields of Computer Vision, Image Processing, Information Retrieval, Machine Learning, and Data Mining and their application to retrieval of Biomedical images from large collections. Prior to joining Morgan last Fall, I significantly contributed to research and development of the image processing and machine learning methods extensively used in the NLM Open-I multimodal biomedical retrieval system, which enables users to search for and retrieve citations that are enriched with relevant images from a collection of 250,000 open access articles and nearly 1 million illustrations from the PubMed Central ® repository. The Open-I system is developed in Java and uses Hadoop to parallelize text processing and image feature extraction.

Md Mahmudur Rahman
Morgan State University

I expect to use SDSC resources both for my teaching and research purposes:

- I will teach a course "Introduction to Data Science" to our UG students from STEM majors in Spring' 2016. I will cover some of the materials, such as Git/GitHub, Predictive Data Analysis, Visualization, HPC etc.
- For my research, I plan to use SPARK MLib for my machine learning (Deep Learning) related tasks and GPU CUDA for some of my image processing related tasks and GitHub to host the project and version controlling.

Nitin Sukhija
Mississippi State University
nitin@hpc.msstate.edu

My Work

Role: Demand Driven

Research: Performance Modeling and Optimization, Resilience and Robustness and many more

Problem Domain: Scientific Applications, Cyber Security and Distributed Analytics

Algorithms: Dynamic Scheduling Algorithms

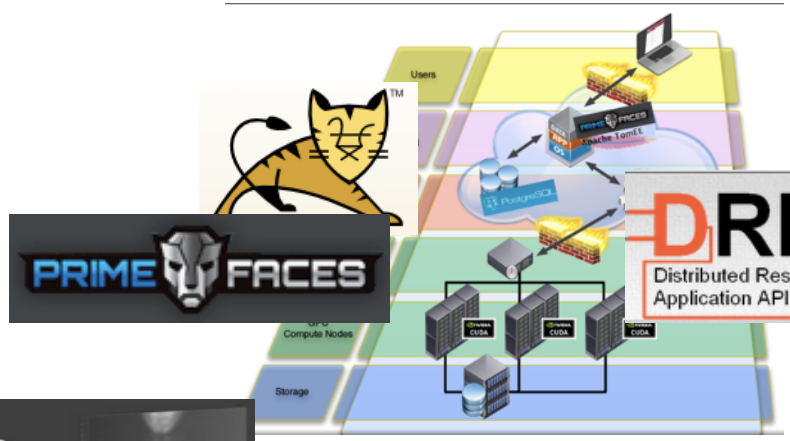
Nitin Sukhija
Mississippi State University
nitin@hpc.msstate.edu

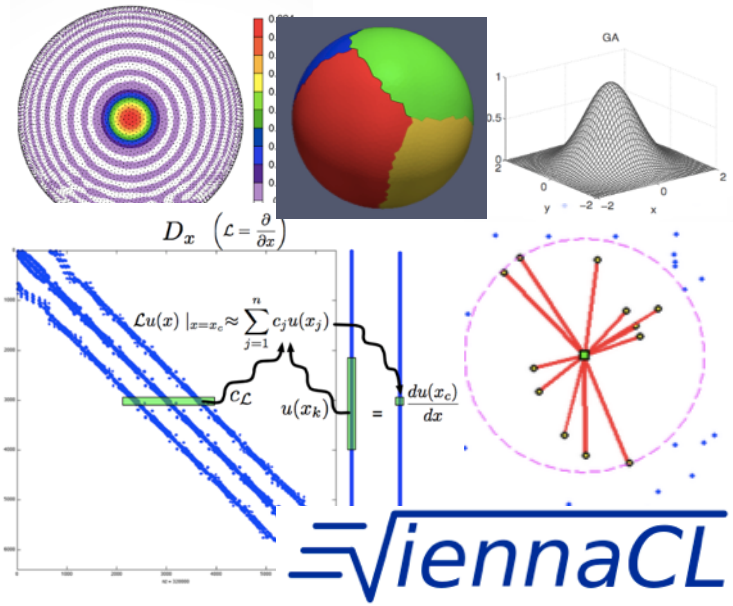
Usage:

- **Intellectual merit**
- **Outreach**
- **Performance optimization is my area**
- **Try to implement some techniques discussed in Graph engines**

Evan Bollig

University of MN (Supercomputing Institute)





Evan Bollig
University of MN (Supercomputing Institute)

- Hadoop/Spark
  - Will target similar user experience @ MSI
- Gordon: SSDs for genomics pipelines
- Weka
- Kepler
- Globus

Johannes Brust

UC Merced

On Mathematical Optimization of Multivariate Functions.

Goal : Development of general, usable methods for the optimization of multivariate functions.

Approach : Develop computational methods from mathematical principles.

Current Research : We developed a Matlab Trust-Region strategy based on a diagonalization of the 2nd derivative matrix. This is to minimize general objective functions. Currently we are testing the method on a standard test set of optimization problems.

We published our method, June' 15.

Johannes Brust
UC Merced

On Mathematical Optimization of Multivariate Functions.

Prospect I : We view the computational experience of the professionals at the SDSC as a vital resource. I spend the summer as a research fellow at the SDSC supervised by a mentor and intent to foster this relation.

Prospect II : I intent to prepare a talk on HPC for a seminar of applied mathematics students at the UCM.

Ronald M. Caplan

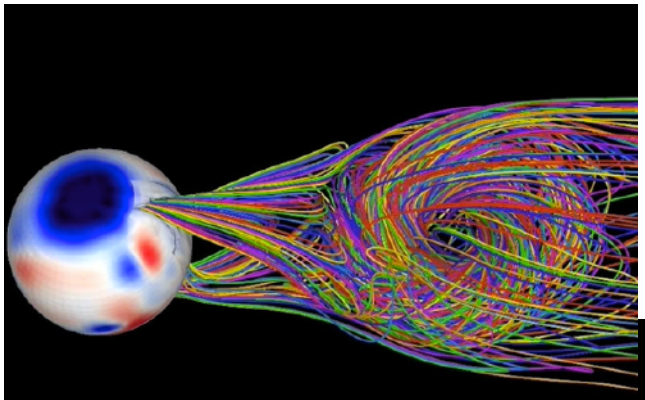
Predictive Science Inc.

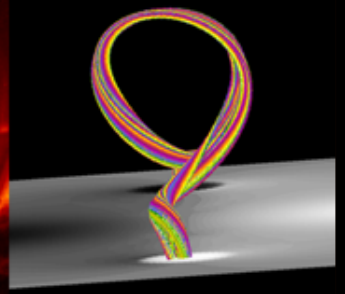
FORTRAN

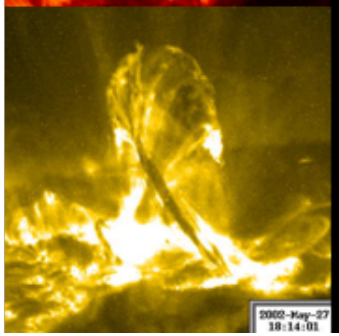
MPI

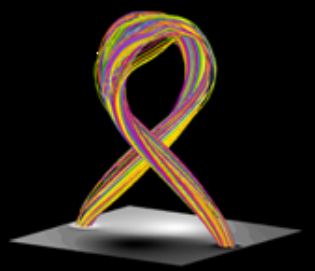
HPC

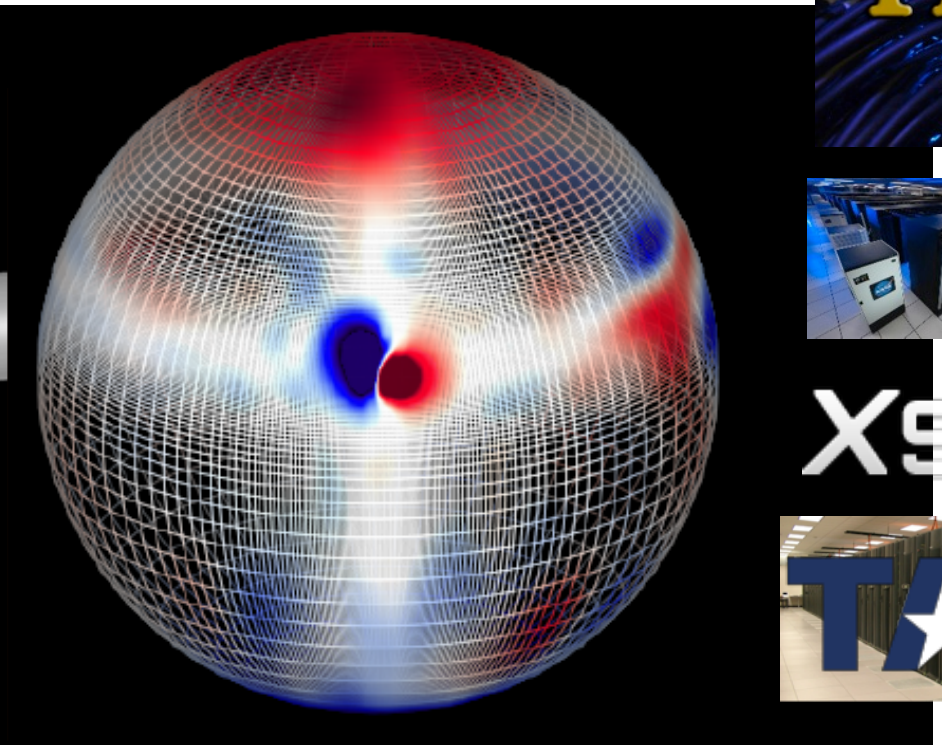
Analyzing structures and dynamics in the solar corona and heliosphere using global magnetohydrodynamic simulations.











XSEDE

Ronald M. Caplan

Predictive Science Inc.

globus

git

Hongmei Chi
Florida A & M University

- Use Weka in CyberSecurity application

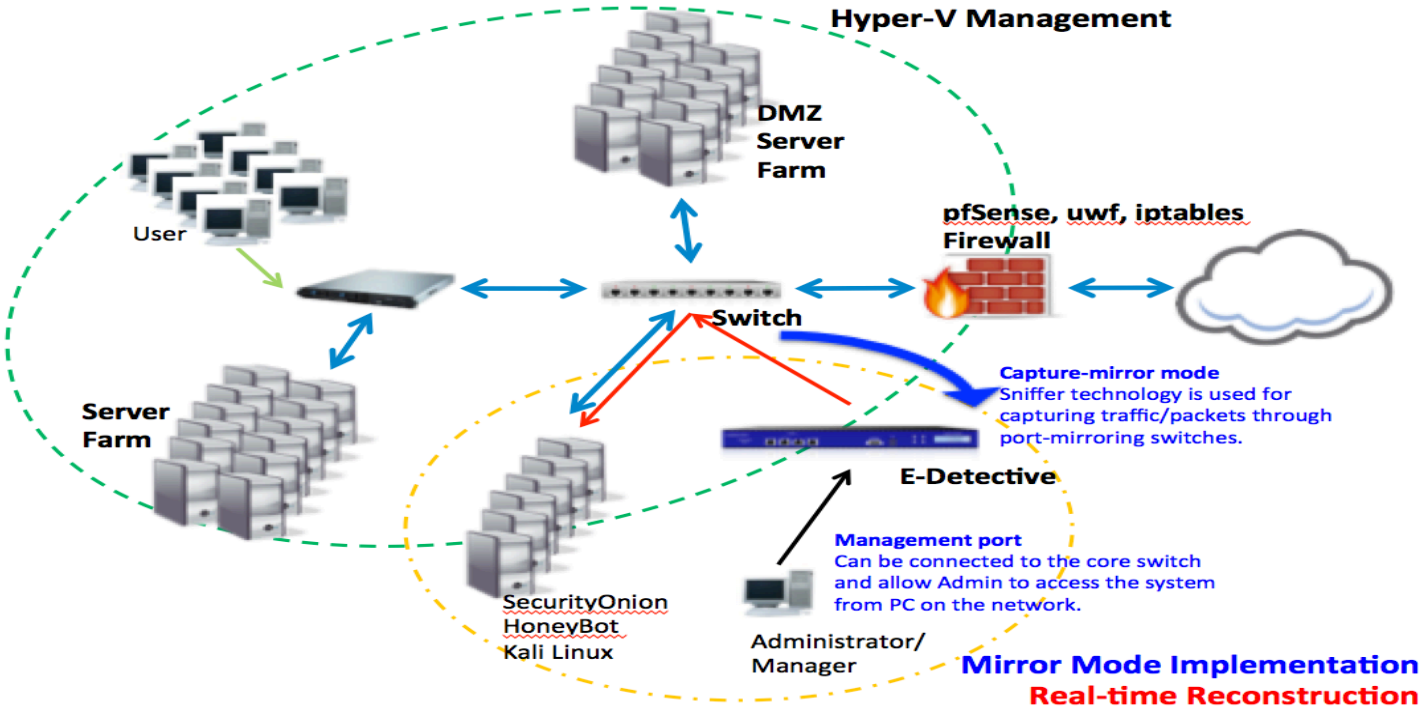
Project title: potential insider threat detection using online game simulation environment

My role: PI

We are planning to do parallel for building and testing our model

Hongmei Chi
Florida A & M University

- Our virtual lab deployment



Hongmei Chi
Florida A & M University

tell us about how you expect to use SDSC resources

- Use Comet to do parallel and hope that my graduate students can use SDSC resources to work on their thesis projects
- Education: I wish that my students have a chance to use Comet for their term project in data mining/ parallel computing courses
- I am planning to write NSF proposal related in data science and I can get support from SDSC

Terry Gaasterland
UCSD – Bioinformatics & Systems Biology
Scripps Institution of Oceanography

We generate & analyze large quantities of genomics data

- **gene expression in stem cells & neural progenitor cells**
  - in time series
  - under different stress conditions
  - signaling pathway response in cell culture
- **human patients with disease (optic nerve degeneration)**
  - DNA variants & tissue specific gene expression
- **unusual organisms' *de novo* genomes & transcriptomes**
  - California squid, medicinal leech, Channel Island fox

Terry Gaasterland
UCSD – Bioinformatics & Systems Biology
Scripps Institution of Oceanography

- Apply predictive analytics to **find association rules over DNA variants and phenotypes** that may **predict onset, progression or severity of disease**

- Invent **new 3D visualizations of gene expression changes** across **time and dose series**

- Invent ways to **visualize gene expression changes resulting from signaling pathway activation** in
  - diseased vs normal cells / tissues
  - stem cells vs derived cells

Garren Gaut

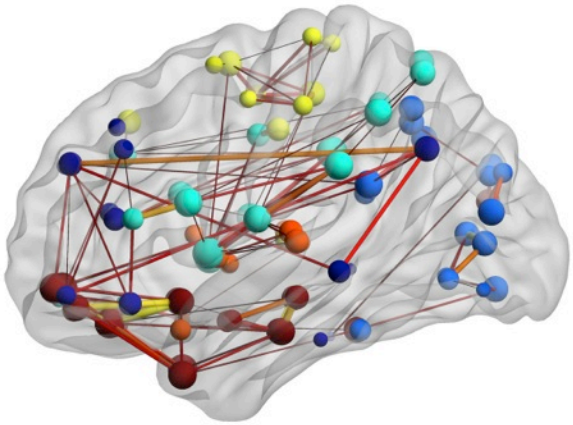
University of California, Irvine

Research Interests: Psychotherapy Process Research, Joint Neural and Behavioral Models, Cognitive Prediction

- ML/Statistics
- Cognitive Models
- NLP

Current: Behavior Classification and Performance Prediction using functional Magnetic Resonance Imaging (fMRI)

- Functional networks
  - Task/Performance-dependent



Garren Gaut

University of California, Irvine

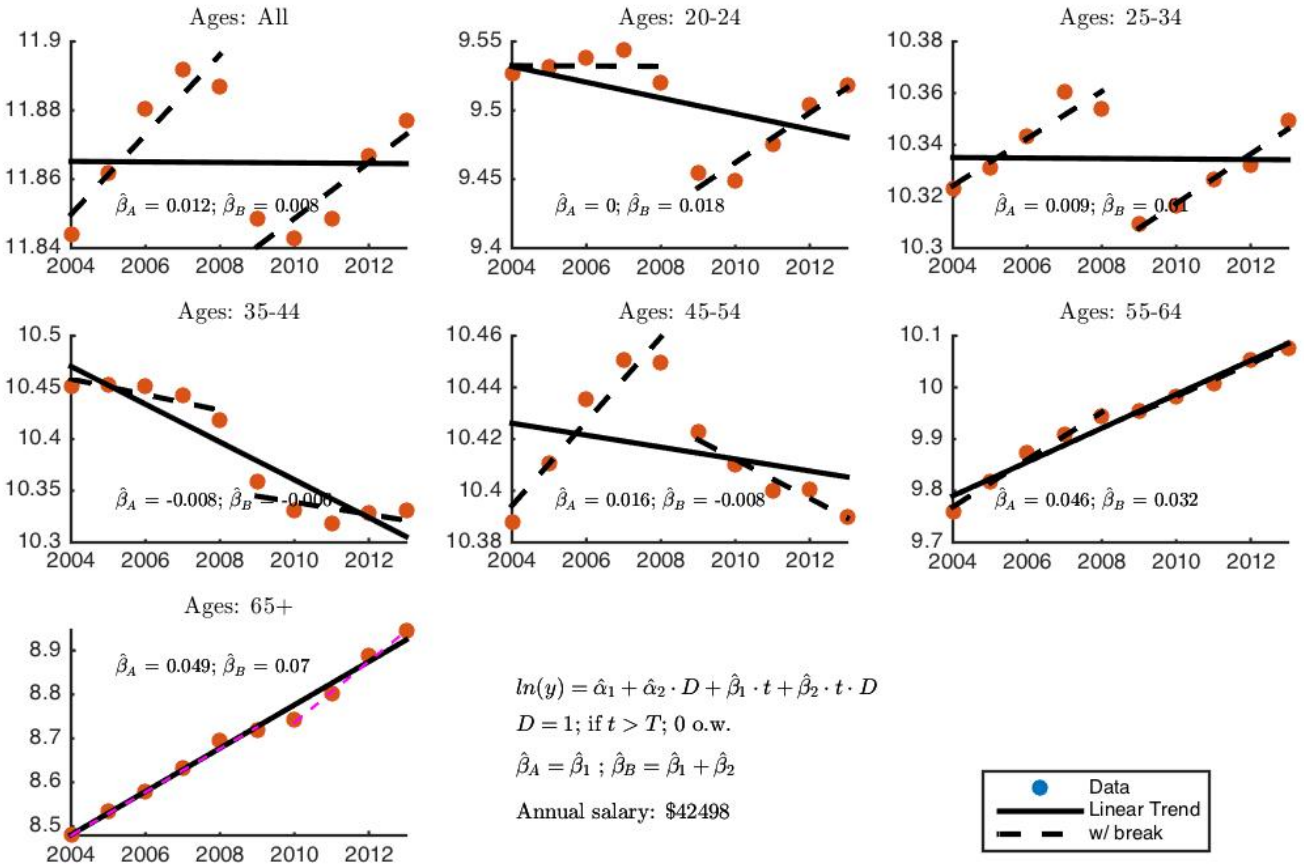
HPC

- Reduced dataset ~1Gb
- Full dataset ~200Gb

Highlights

- Computer architecture
- HPC Basics
- VisIt
- Spark
- Git

Seung Lee
UC Santa Cruz



Seung Lee
UCSC

- I mostly use optimization packages and linear algebra packages.
- I can give each node, a portion of hypercube to search for a solution.
- I can do many different parameterizations.
- When I get bored with this, I will probably move on to something data intensive where I can apply some machine learning algorithms.

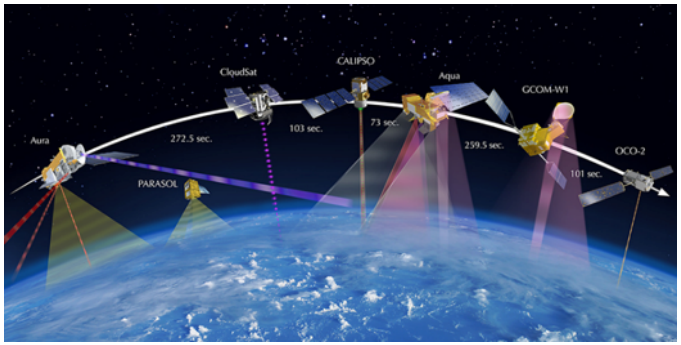
Paul Levine — University of California, Irvine

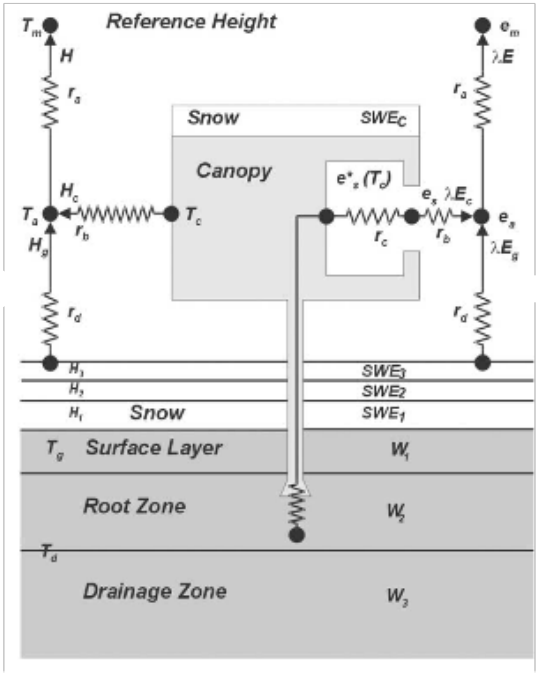
Research interests:

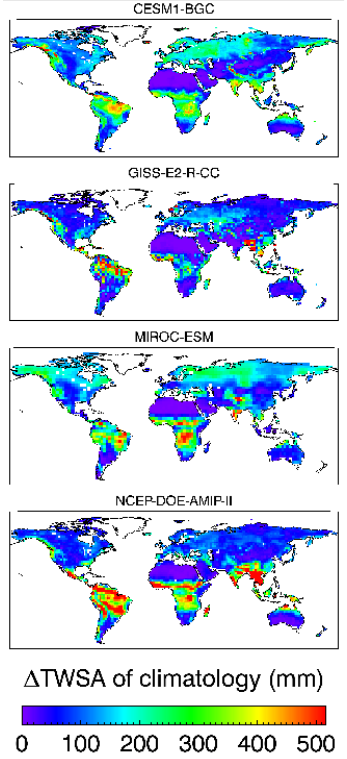
- Modeling of biosphere-atmosphere interactions and the coupled climate system
- Satellite remote sensing for model calibration, validation, and data assimilation

HPC applications:

- Community Earth System Model (CESM) using NCAR supercomputer (Yellowstone)
  - Optimization of terrestrial biogeochemistry models
  - Coupled model evaluation and intercomparison



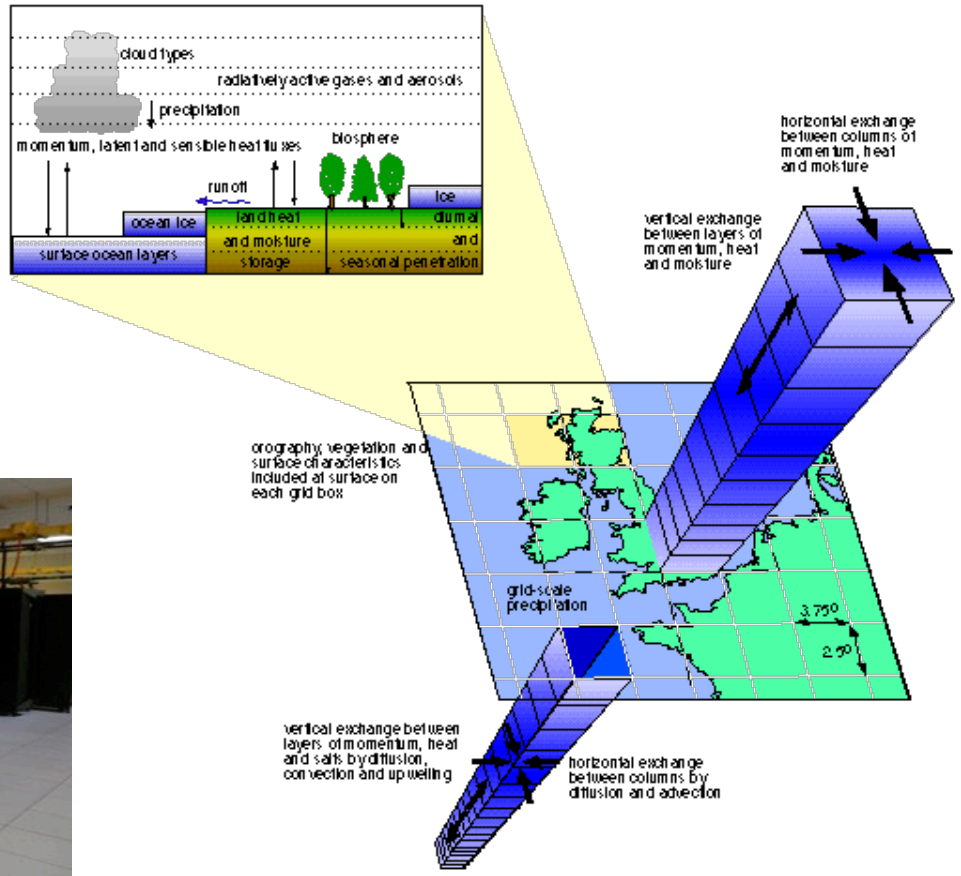




Paul Levine University of California, Irvine

What I learned at SDSC Summer Institute:

- Submitting jobs on HPC systems with scheduler
- Workflow documentation
- Version control
- Code parallelization
- Visualization for data exploration



Dapeng Li
University of Utah

My current research focuses on wildfire evacuation modeling, which involves using fire spread modeling, trigger modeling, and traffic simulation to model the evacuation process as a coupled human-environmental system.

I am working on an open source software system for setting wildfire evacuation triggers. Fire spread modeling is used to calculate the spread rates of the fire in eight directions within a raster cell. And the Dijkstra's algorithm is used to create the evacuation trigger buffer.

Dapeng Li
University of Utah

I am working on wildfire spread modeling, which involves intensive computation over raster datasets. I have used OpenMP to parallelize the code. I plan to use MPI and GPU to implement a better solution so as to perform fire spread modeling and trigger modeling more efficiently.

Jason Marshall - Caltech

- Rigid body particle dynamics (DEM)
- Particle geometry from 3D digital tomography
- Model geometry with level sets
- Interested in thermal physics, mechanical physics, etc
- Currently developing multi-physics hpc code

Video from Reid Kawamoto

Jason Marshall

Caltech

- **SDSC Resources**
  - Hope to get startup allocation on Comet
  - Potentially get ECSS support
  - Perform scaling testing of code on Comet
  - Investigate in-situ visualization on Comet
  - Finish HPC development on Comet
- **What I learned/will apply from SI**
  - Intersection testing is computationally expensive, potentially offload to GPUs (if mappable)
  - Investigate performance and optimization tools
  - Investigate additional libraries (neighbor lists, etc.)

Phu Nguyen

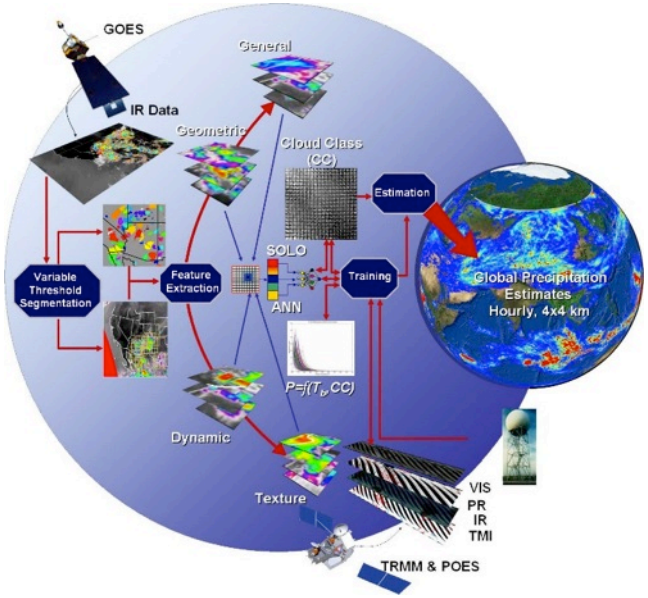
Postdoctoral Scholar, CHRS, UCI

Education:

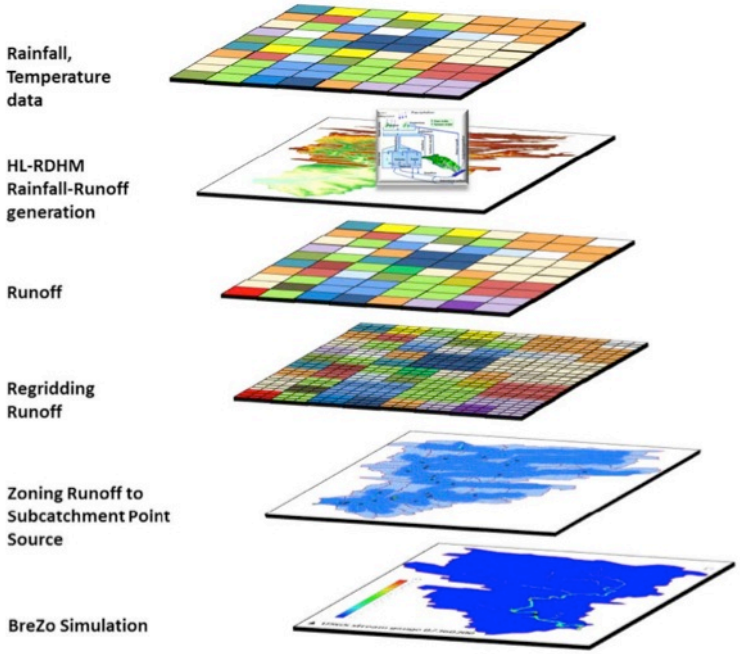
- Water Resources Management, UCI, PhD, 2014
- Water Resources Management, University of Melbourne, MSc., 2008
- Civil Engineering, Bach Khoa University - Hochiminh City, B.S., 2003

Research:

- Flash flood warnings (HiResFlood-UCI)
- Satellite precipitation estimation
- Global extreme precipitation event database (CHRS CONNECT)
- Crowdsourced precipitation observation (iRain)



PERSIANN-CCS



HiResFlood-UCI

Phu Nguyen

Postdoctoral Scholar, CHRS, UCI

Connect.eng.uci.edu

Bipush Osti University of Maine

Current project involves creating a bird flock migration model. The input to the model is hourly weather data like u,v wind, precipitation and pressure and output is position of the bird. Each of the birds in the model has same specific rules regarding when to fly and when to stop depending on the input parameters.

The model assumes that the position of one bird is not dependent on another therefore it is highly parallelizable. The goal is to use CUDA so that each bird can be assigned a thread.

GPU nodes in a supercomputer could be used in the future if the data gets too large.

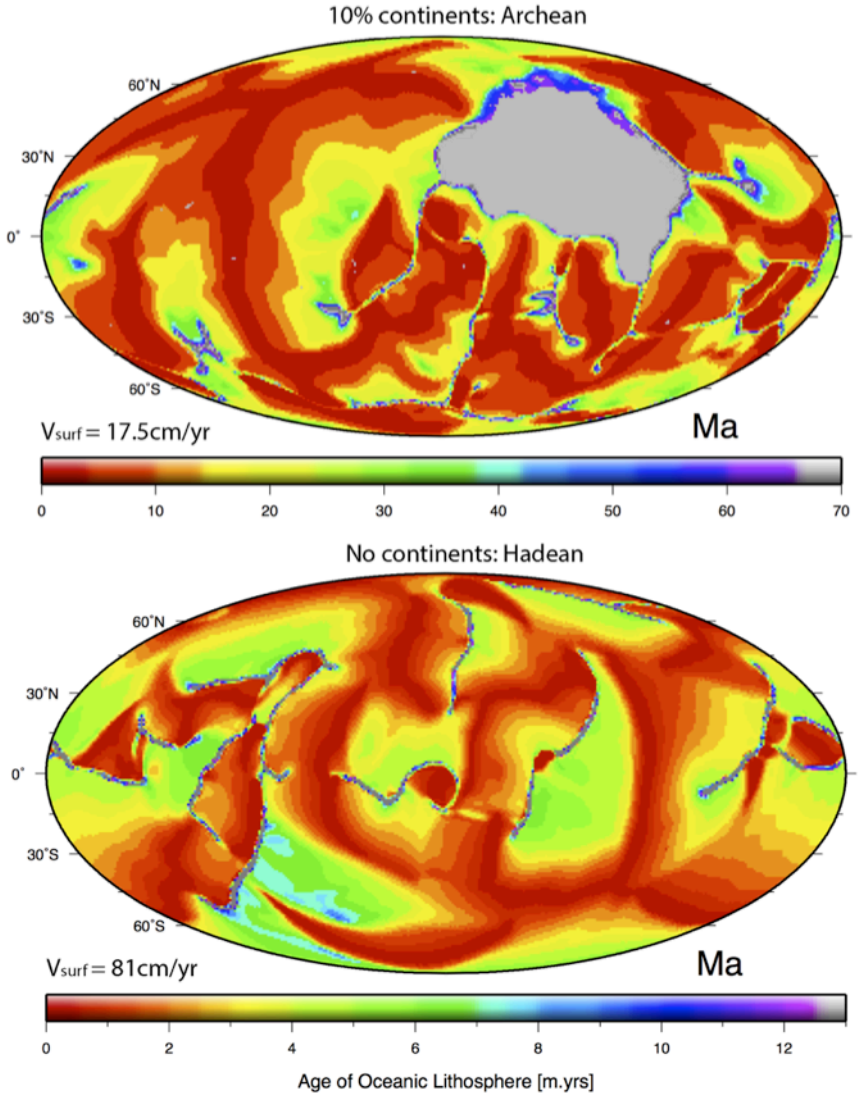
Bipush Osti University of Maine

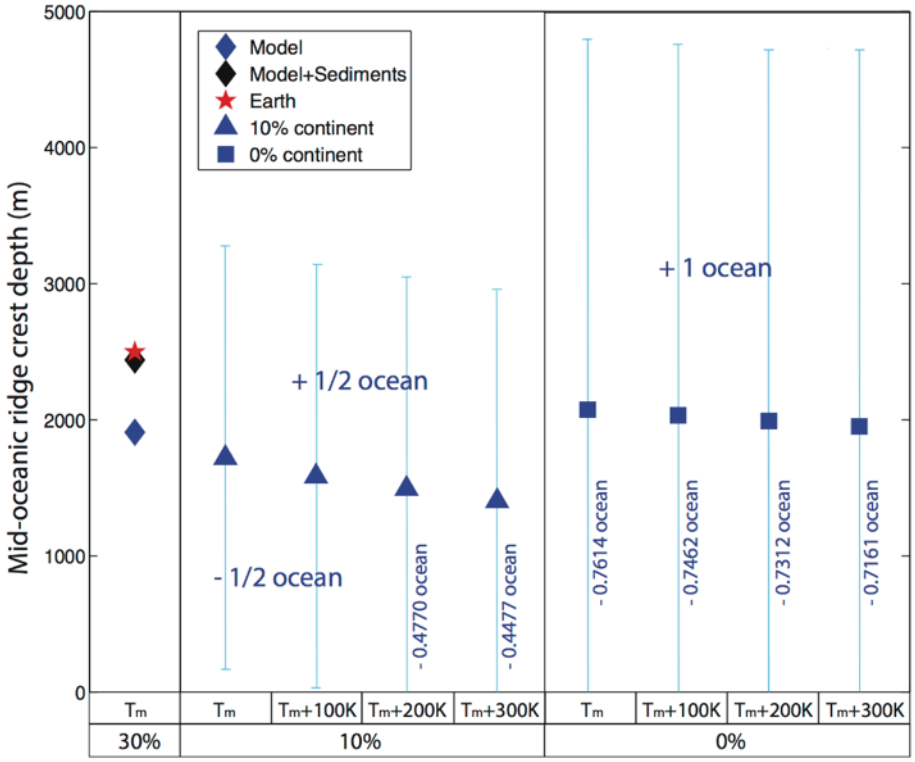
- **I learned about how to access gordon and comet,use workflows,optimize code and also got some insight into the problems I am having in regards to GPU.**

- **I also got the chance to talk to some of the researchers from other universities and from SDSC regarding some aspects of my project and was able to get new ideas to solve them efficiently.**

- **I hope to use the GPU nodes at the Supercomputers at my university or at SDSC once my data gets much larger.**

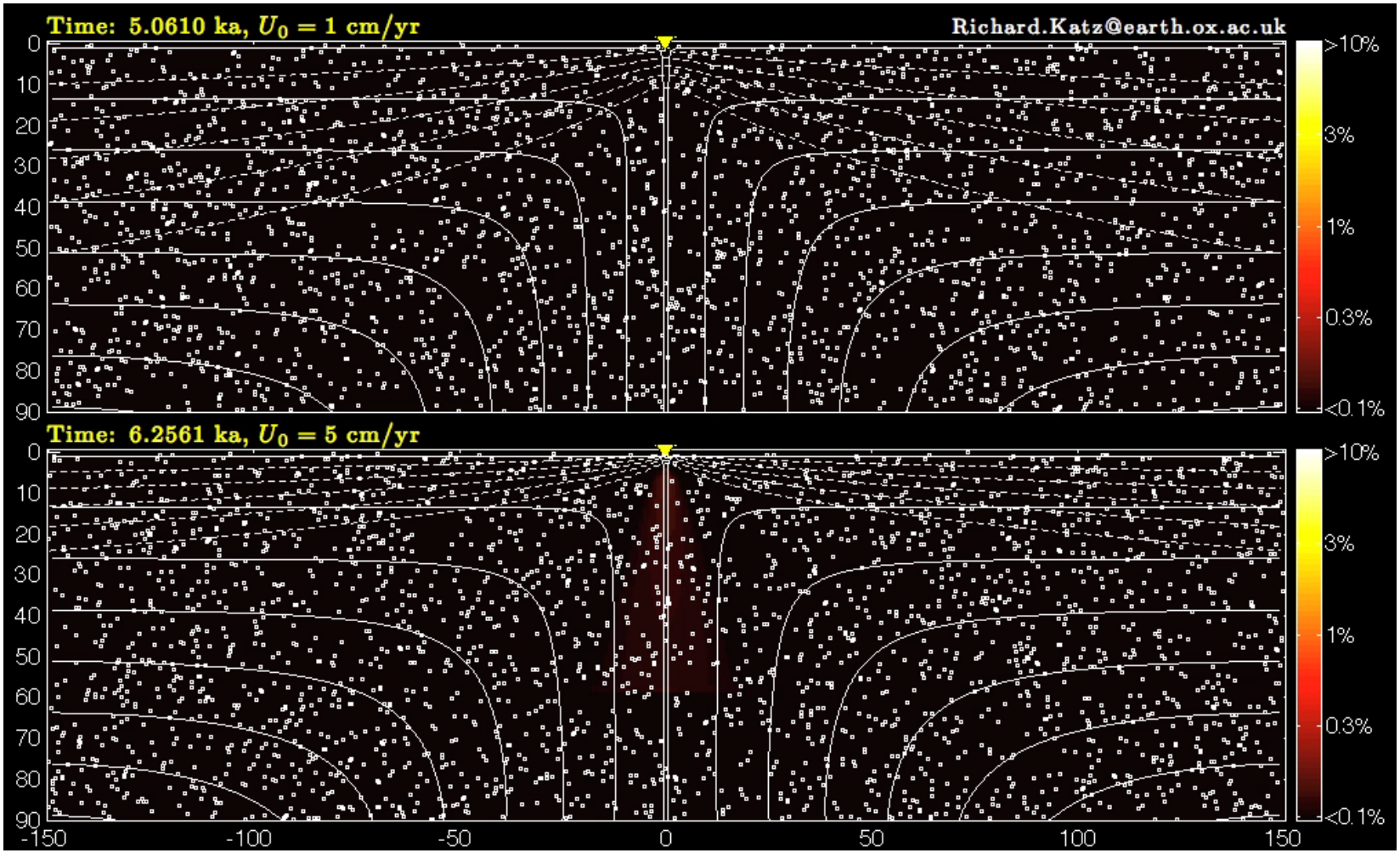
Shi Sim
Scripps Institution of Oceanography
Dave Stegman, James Day





Shi et al, in prep

Shi Sim
Scripps Institution of Oceanography
Dave Stegman, James Day



FEniCS Project: Dorsal: TerraFERMA (SPuD, FEniCS,Gmsh, PETSc)

tdsmith

@biotimylated

Tim D. Smith

UC Irvine

tim@tds.xyz

Grad student, Biomedical Engineering

- "Hey, I bet a computer can do that"
- Aligning, segmenting, measuring fluorescence microscopy images of single cells in microwell devices
  - Tiled high-mag images (~1 GB/set)
  - Low-throughput (< 10 GB/week)
- Trivial processing of time-lapse microscopy image sets
  - ~10 GB/set, 10-20 GB/week
- Inconvenient but not "big" data
- Scipy stack, multiprocessing Pools, GNU parallel

tdsmith

@biotimylated

Tim D. Smith

UC Irvine

tim@tds.xyz

Excited about performant Python, ways to scale laptop computing up towards cluster scale

- Arrayfire: multi-backend GPU operations
- Numba: numpy-aware JITting
- IPython Parallel: scale out across nodes at UCI HPC

Homebrew: brew.sh, https://github.com/Homebrew/homebrew

Use Homebrew on your laptop, Linuxbrew in cluster $HOME

```
brew install open-mpi homebrew/science/trilinos
     homebrew/python/numba homebrew/science/r
```

Send us a pull request!

John Theiring

JPL – Jet Propulsion Laboratory

Group Supervisor Advanced IT Systems

JPL HPC Team

•The JPL Office of the Chief Information Officer provides High-Performance Computing (HPC) resources and services to JPL staff to meet their challenging computing needs.

•The HPC team supports and manages the on premise Message Passing & Embarrassing Parallel Supercomputing clusters

•Common use cases include processing jobs for flight projects, navigation, climate physics, astrophysics and space science, water and carbon cycles, propulsion and fluid flight systems, and oceans and ice.

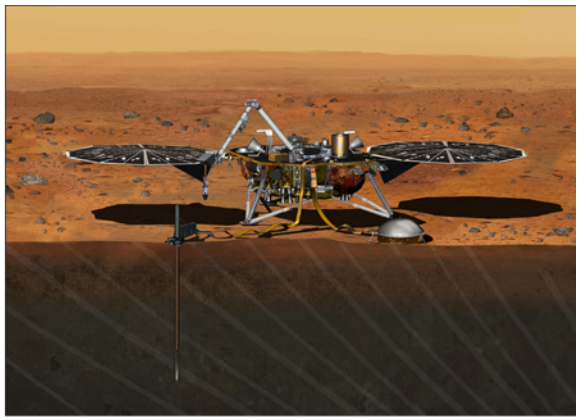
John Theiring

JPL – Jet Propulsion Laboratory

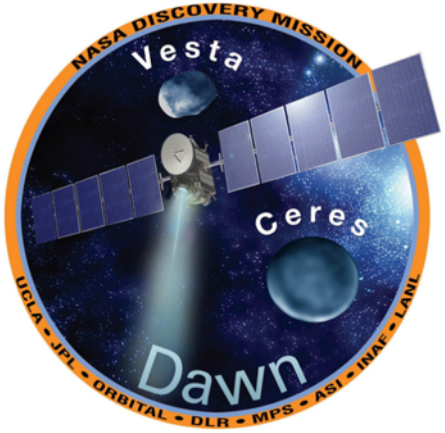
Sample of Projects Supported by JPL HPC Computing

www.jpl.nasa.gov

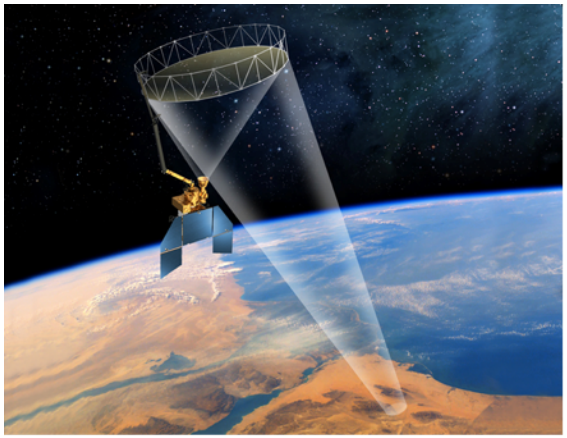
Curiosity (2011)



Insight (2016)

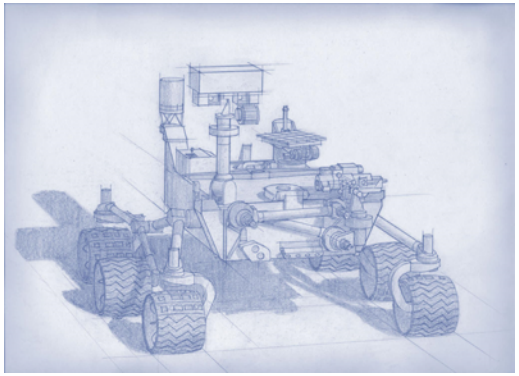


(2007)



SMAP (2015)

(2003)



MARS2020

John Theiring

JPL – Jet Propulsion Laboratory

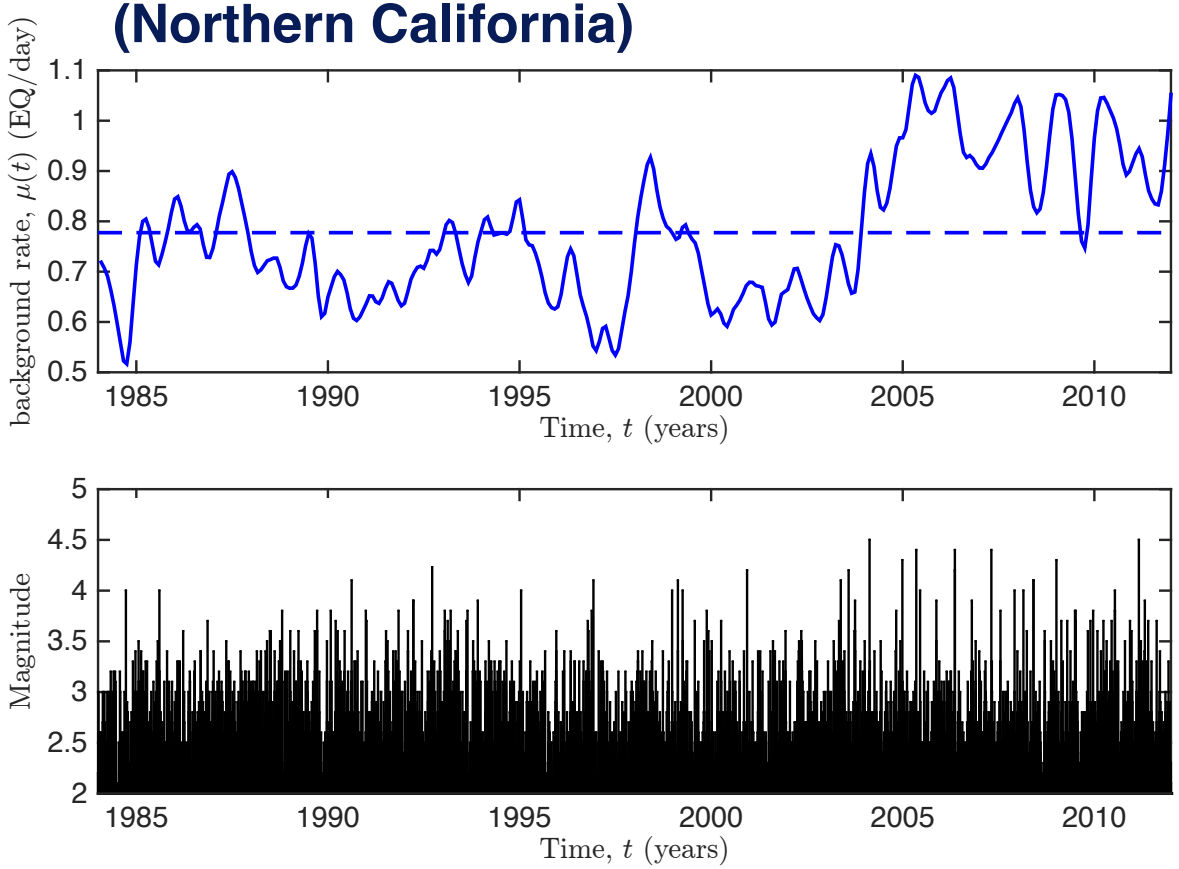
- Thank You SDSC team for this great opportunity to become educated and more aware of SDSC services.
- Really enjoyed meeting the SDSC personnel who provide support for JPL
- Will use this momentum to help the JPL HPC team acquire SDSC accounts and become more familiar with the SDSC environments and services.
- Work with JPL HPC customers to identify Use Cases best fitted for SDSC and assist with the setup and execution of their jobs.

John K. Theiring
Phone: (818) 393-2214
John.K.Theiring@jpl.nasa.gov

Daniel Trugman

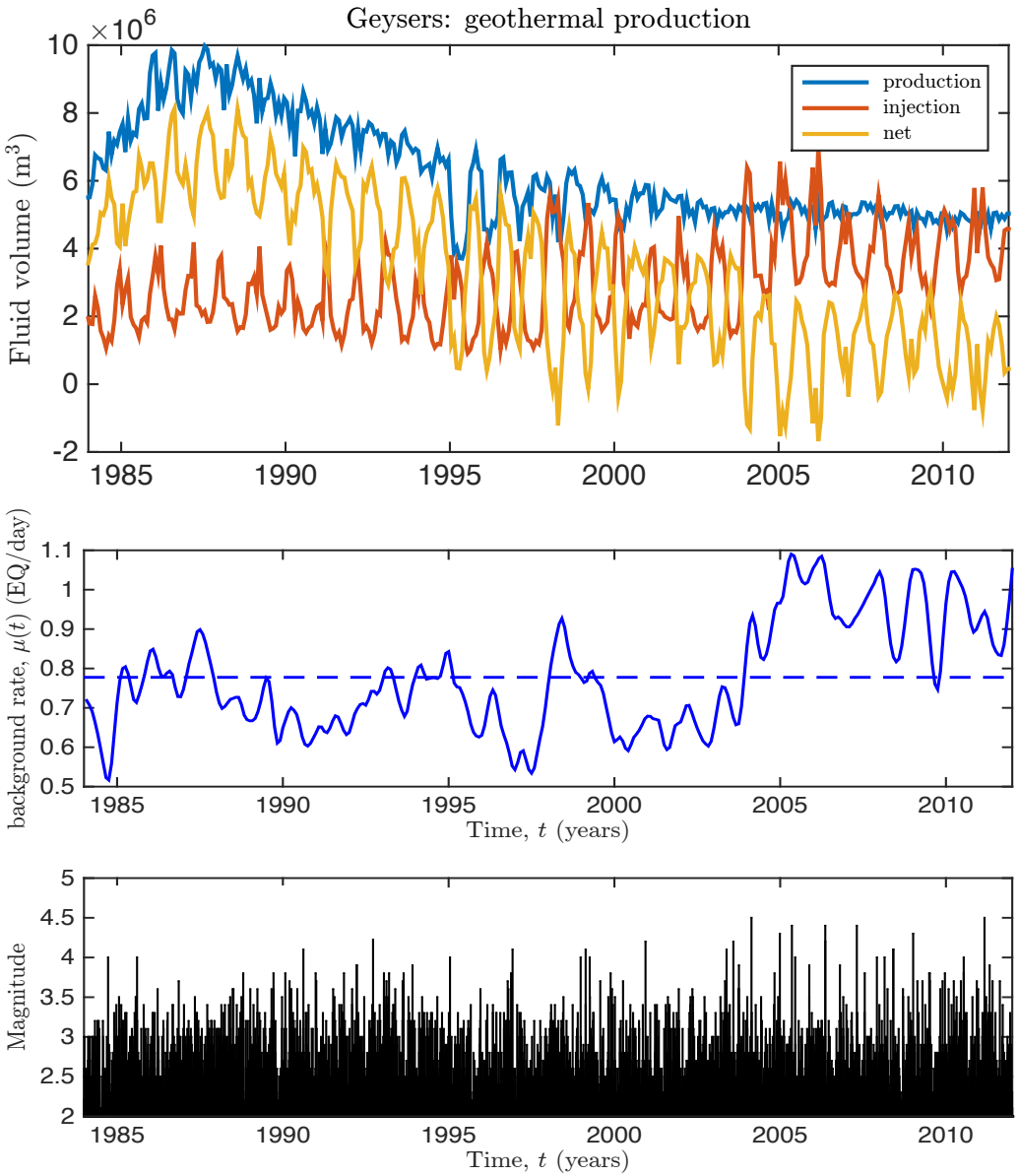
Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography - UCSD

Earthquakes in Geysers Geothermal Field (Northern California)



Research Interests:

- Mechanisms of crustal stress transfer (earthquake seismology, tectonic geodesy)
- Induced seismicity (human-triggered earthquakes)



SDSC Summer Institute: Applications/Takeaways

- Machine learning fundamentals → explore nonlinear relationships in data
- HPC infrastructure → capability to perform realistic and fully dynamic, 4D simulations
- Visualization tools, algorithm optimization and design, IPython, …

Edward Villanueva
Jet Propulsion Laboratory

Software Systems Engineer
JPL High Performance Computing team

- Provide supercomputing resources for JPL projects
- Embarrassingly and massively parallel computing support
- User support
- Open-source and licensed software

Edward Villanueva
Jet Propulsion Laboratory

I learned a great deal this week! (Thanks, SDSC!)

- What SDSC has to offer our users
- Latest HPC technology
- Tools to increase team and user productivity
- Exposure to software – user support
- Best practices – software and storage