MINING WEATHER AND CLIMATE DATA
FOR FUN AND PROFIT

R. Saravanan
Dept. of Atmospheric Sciences
Texas A&M University
sarava@tamu.edu
Weather and climate scientists have been dealing with *big data* long before it was called that!
- Structured data, on a 4-dimensional grid (x, y, z, t)
- A single climate model prediction produces tens of TB of data (100x100km global grid)
- Multiple climate runs from different models/nations are archived using the Earth System Grid Federation: 2 PB of data
- NASA satellite data: over 7 PB, growing by 4 TB daily

Tend to use traditional statistical analysis techniques
- Mean, variance, correlation, ...
- Principal Components/Empirical Orthogonal Functions for variance analysis
- Singular Value Decomposition for covariance analysis
FOR FUN

- *Ongoing project*: Fix errors in climate model simulations of tropical rainfall (with Schumacher, Jun)
- Construct statistical model relating rainfall to atmospheric state variables using satellite measurements
- Important for simulating El Niño and climate prediction

Eye formation during Hurricane Patricia (NASA)
Oct. 2013: Monsanto acquires The Climate Corporation for nearly $1 billion.

Weather Happens
The Climate Corporation monitors the weather locally, using weather stations, radar & satellites.

Get Paid Automatically
So if a covered weather event occurs, The Climate Corporation automatically sends payment to you.

No claims
No adjuster
No waiting for payment

Whether an ideal year or a shortfall year, you meet your profit goal.
FOR PROFIT

Oct. 2015: IBM acquires digital assets of The Weather Company for over $2.5 billion

Plan: Collect weather data from sensors and provide real time information and forecasts using Watson Analytics

2.2 Billion
Weather forecast locations

40 Million
Mobile phones

50,000
Flights per day

7X
The volume of the leading search engine
BIG DATA IN WEATHER AND CLIMATE RESEARCH

- Problems
  - Developing empirical “closures” for weather and climate models
  - Improving short-term weather forecasting at local scales
  - Combining data and models to assess weather and climate risk

- Potential for interdisciplinary research
  - Advanced statistical techniques
  - Machine learning and Artificial Intelligence
  - Parallel I/O for improved throughput in High Performance Computing