15% More Accuracy in Seasonal Hurricane Forecasts through Comparative Climate Networks Analytics

Nagiza Samatova NCSU/ORNL and Fred Semazzi NCSU

Objectives

- Develop predictive forecasting methodology for climate extremes (e.g., hurricanes, droughts, rainfalls)
- Devise scalable algorithms for predictive mining of large-scale climate complex networks
- Provide mechanistic insights about the key factors contributing to extreme events variability
- Demonstrate high predictive skill for North Atlantic seasonal hurricane activity

Impact

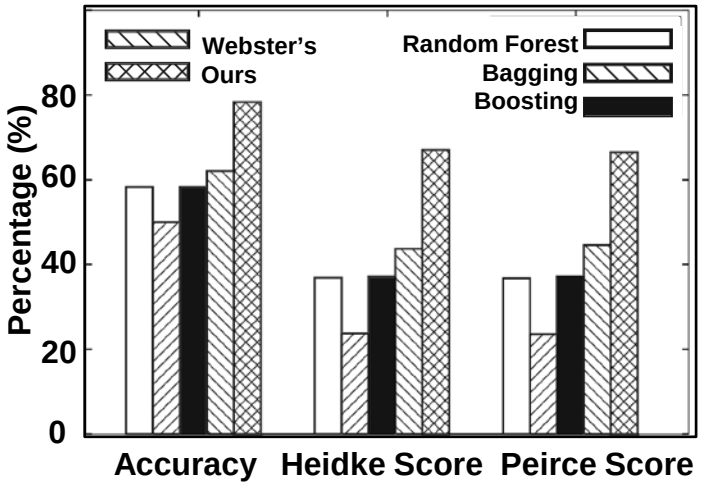
- Provide policy makers more reliable information on seasonal climate extremes
- Scalable large-scale graph mining algorithms of broader applicability (e.g., bioenergy)
- Advance our understanding of the mechanisms that influence hurricane variability and behavior
- International impact managing meningitis epidemic outbreaks driven by climate extremes

2012 Accomplishments

15 percent more accurate forecast of seasonal hurricane activity

Comparative climate networks analytics & machine learning methods

"***Novel data-driven methods promise to excel beyond the traditional methods in climate prediction tools***"
(Fred Semazzi, Nobel Prize co-winner, climate scientist)

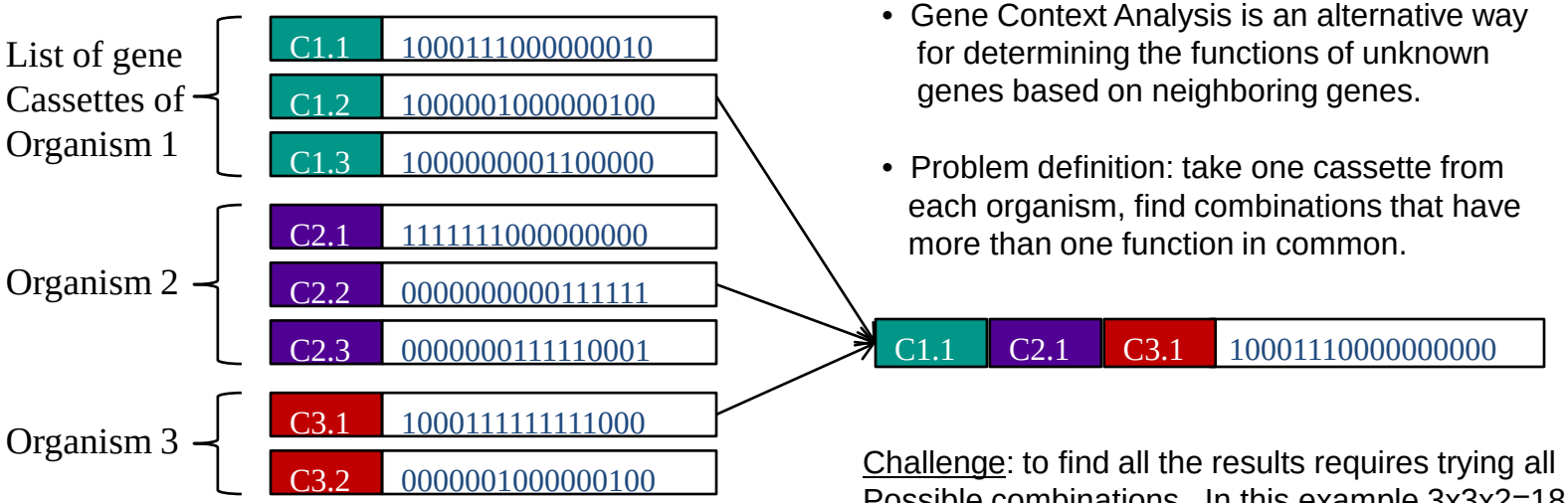


Z. Chen, W. Hendrix, H. Guan, I. Tetteh, A. Choudhary, F. Semazzi, N. Samatova, "Discovery of extreme events-related communities in contrasting groups of physical system networks," Data Mining and Knowledge Discovery, 27(2), p. 225-258, 2012.

U.S. DEPARTMENT OF ENERGY | Office of Science

SDAV
Scalable Data Management, Analysis, and Visualization

Gene Context Analysis Now Performed in Seconds



The figure shows the bitmap representation of cassette properties, and finding matching combinations

- Gene Context Analysis is an alternative way for determining the functions of unknown genes based on neighboring genes.
- Problem definition: take one cassette from each organism, find combinations that have more than one function in common.

Challenge: to find all the results requires trying all Possible combinations. In this example 3x3x2=18. For millions of cassettes this search is exponential.

APPROACH

- Reorganize the list of functions per gene cassette into bitmaps.
- Use FastBit to compress the bitmaps.
- Re-structure the query processing algorithm into bitwise logical operations.
- Remove solution entries that are contained in other entries (maximal solutions only).
- Progressive pruning the possible solutions based using bitmaps as keys for comparisons.

IMPACT

- Providing interactive exploration
  - When more than 5 organisms are involved in a query, the previous system based on a commercial database system takes too long and the GUI times out.
  - Using the new solution, queries involving 600 organisms took **less than 10 seconds.**
- New solution deployed in IMP system at <img.jgi.doe.gov> since May 2013.

A. Romosan, A. Shoshani, K.Wu, V. Markowitz, K. Mavrommatis, Accelerating Gene Context Analysis Using Bitmaps, 25th International Conference on Scientific and Statistical Database Management (SSDBM) 2013

Feature Tracking and Visualization of the Madden-Julian Oscillation in Climate Simulation

Application:

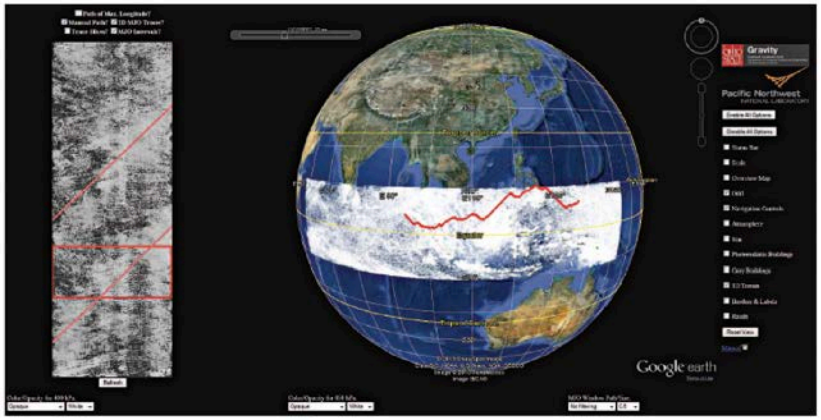
- Madden-Julian Oscillation (MJO) simulations by Samson Hagos and Rubby Leung at Pacific Northwest National Laboratory

Goal:

- Automatic detection and tracking of MJO
- Identify the path and time evolution of MJO phenomena
- Develop an interactive data analysis and browsing tool for the domain scientists

Requirement:

- Automatic Feature Extraction and Tracking
  - Incorporate the domain specific knowledge about the speed and direction of MJO movement
- Visualization and analysis of MJO movement
  - Understand how MJO is related to different physical quantities
- Support fast queries of the simulation data



A web-based interactive MJO data visualization and tracking system

Results

- Software
  - An interactive web-based data visualization and analysis tool (left image)
- Algorithm
  - MJO tracker: A robust MJO detection and tracking method
- Publication
  - Teng-Yok Lee, Xin Tong, Han-Wei Shen, Pack Wong, Samson Hagos, Rubby Leung, Feature Tracking and Visualization of the Madden-Julian Oscillation in Climate Simulation, IEEE CG&A special issue on big data visualization, July/August 2013

SDAV
Scalable Data Management, Analysis, and Visualization