Data Mining in Earth System Sciences

Jitendra Kumar[†], Forrest M. Hoffman[†], Richard T. Mills[†], and William W. Hargrove[‡]

[†]Computer Science and Mathematics Division, Oak Ridge National Laboratory
[‡]Eastern Forest Environmental Threat Assessment Center, USDA Forest Service

jkumar@climatemodeling.org

Tuesday August 30, 2011
CCSI 3x5 seminar

- Earth science data span many orders of magnitude in space and time scales.

- Increasingly large and complex, often representing long time series, making them difficult to analyze, visualize, interpret, and understand.

- Satellite remote sensing data tend to be very large and grow quickly as spatial and temporal resolutions increase.
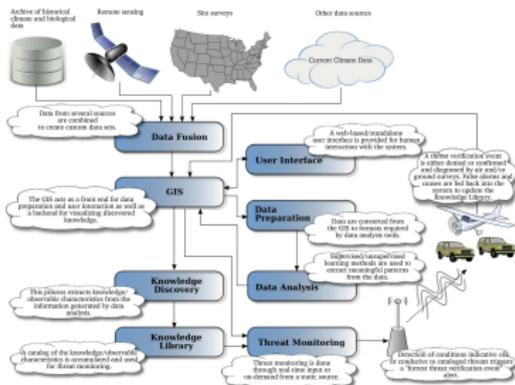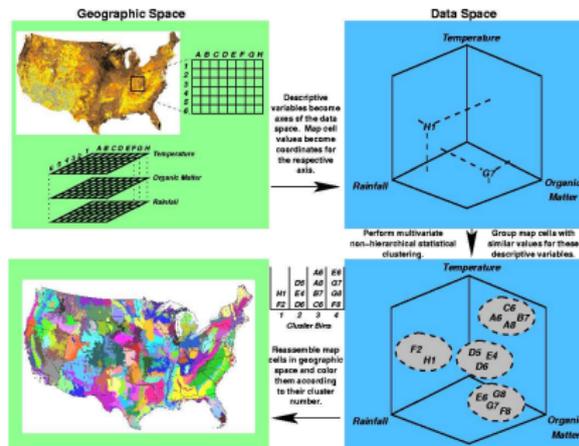

- Forest Incident Response and State Tracking (FIRST) System

- Changes in forest states are captured by the remote sensing (MODIS NDVI).

- Determine the normal seasonal and inter-seasonal variations expected at a geographic location.

- Identify unexpected changes in forest phenology states.

Scalable parallel analysis tool set

- Data mining algorithms/tools, Geospatial analysis tools
- Algorithmic enhancements to k-means clustering
- Scalable, suitable for extremely large data sets
- Data analysis and visualization in same (supercomputing) environment as the models and data
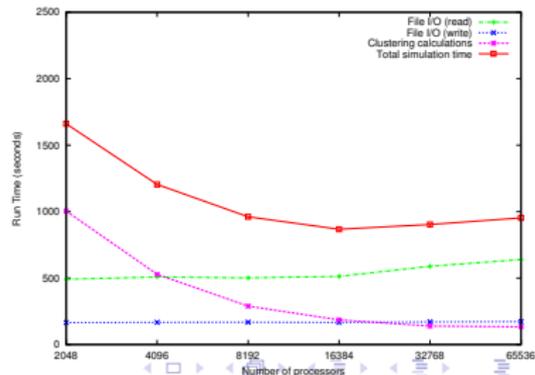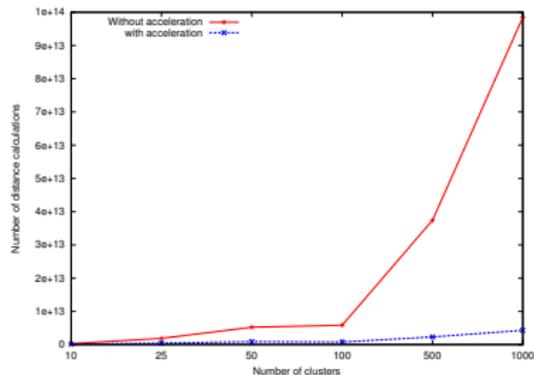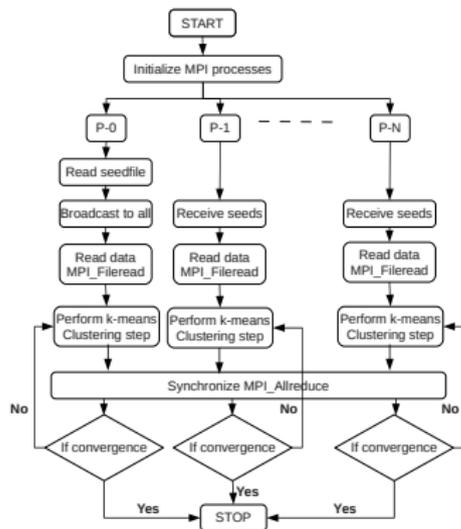
Forest threat detection
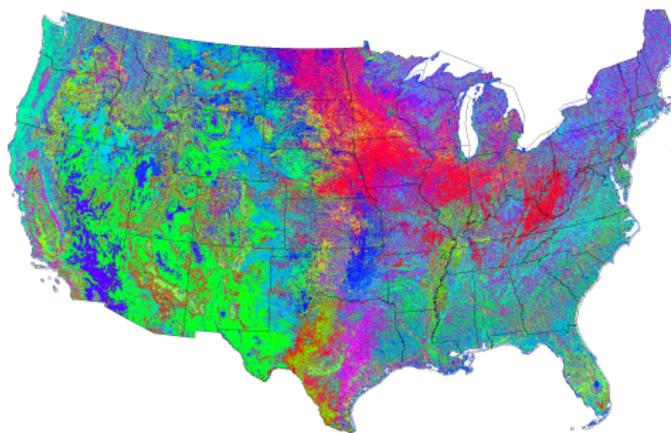
Phenology map (year 2000)
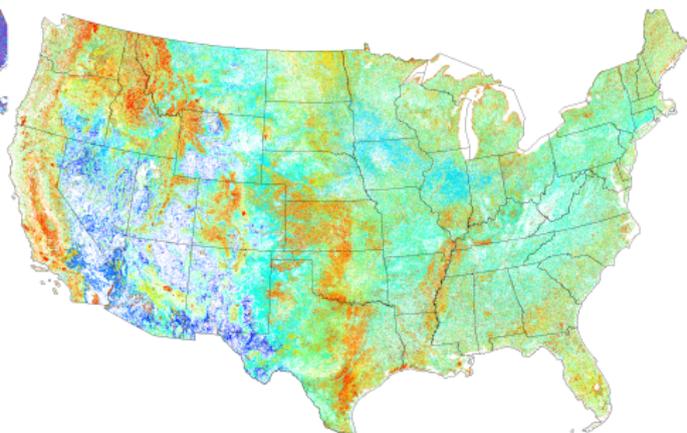
Year 2000-2009 Phenology change map
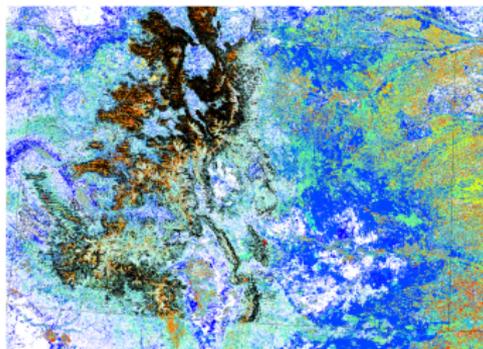
Colorado mountain pine beetle
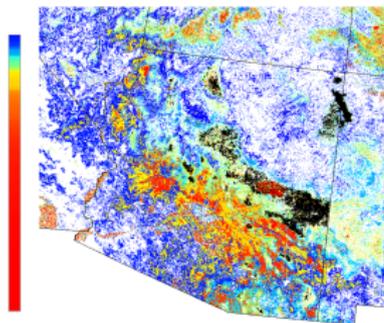
Arizona bark beetle

California wild fires