NSF Expeditions in Computing

Understanding Climate Change: A Data Driven Approach

Vipin Kumar
University of Minnesota
kumar@cs.umn.edu

http://climatechange.cs.umn.edu
Expeditions Team

Vipin Kumar, UM
Auroop Ganguly, NEU
Nagiza Samatova, NCSU
Arindam Banerjee, UM
Fred Semazzi, NCSU
Joe Knight, UM
Shashi Shekhar, UM
Peter Snyder, UM
Jon Foley, UM
Alok Choudhary, NW
Ankit Agrawal, NW
Abdollah Homiafar, NCA&T
Michael Steinbach, UM
Singdhansu Chatterjee, UM
Karsten Steinhaeuser, UM
Stefan Liess, UM
Shyam Boriah, UM
Understanding Climate Change - Motivation

CO2 levels hit new peak at key observatory

- 2100 Higher Emissions Scenario
- 2100 Lower Emissions Scenario

Carbon Dioxide Concentration (ppm)

Year

http://www.ncdc.noaa.gov/indicators/

May 14-16, 2013
Understanding Climate Change – Physics-Based Approach

General Circulation Models: Mathematical models with physical equations based on fluid dynamics

Parameterization and non-linearity of differential equations are sources for uncertainty!

Figure Courtesy: NCAR
Understanding Climate Change - Physics Based Approach

General Circulation Models: Mathematical models with physical equations based on fluid dynamics.

Figure Courtesy: NCAR

Figure Courtesy: ORNL

Ensemble average with observed greenhouse gas concentrations

Ensemble average with pre-industrial greenhouse gas concentrations
Projection of temperature increase under different Special Report on Emissions Scenarios (SRES) by 24 different GCM configurations from 16 research centers used in the Intergovernmental Panel on Climate Change (IPCC) 4th Assessment Report.
Physics based models are essential but insufficient

- Relatively reliable predictions at global scale for ancillary variables such as temperature
- Least reliable predictions for variables that are crucial for impact assessment such as regional precipitation

“The sad truth of climate science is that the most crucial information is the least reliable”
(Nature, 2010)

Disagreement between IPCC models

Regional hydrology exhibits large variations among major IPCC model projections

<table>
<thead>
<tr>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Out of scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Hurricanes</td>
<td>Fires</td>
</tr>
<tr>
<td>Pressure</td>
<td>Extremes</td>
<td>Malaria outbreaks</td>
</tr>
<tr>
<td>Large-scale wind</td>
<td>Precipitation</td>
<td>Landslides</td>
</tr>
</tbody>
</table>
Data-Driven Knowledge Discovery in Climate Science

Transformation from Data-Poor to Data-Rich

- Sensor Observations
- Reanalysis Data
- Model Simulations

A new and transformative data-driven approach that:

- Makes use of wealth of observational and simulation data
- Advances understanding of climate processes
- Informs climate change impacts and adaptation

“Climate change research is now ‘big science,’ comparable in its magnitude, complexity, and societal importance to human genomics and bioinformatics.”

(Nature Climate Change, Oct 2012)
### Need for data driven discovery

<table>
<thead>
<tr>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Out of scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Hurricanes</td>
<td>Fires</td>
</tr>
<tr>
<td>Pressure</td>
<td>Extremes</td>
<td>Malaria outbreaks</td>
</tr>
<tr>
<td>Large-scale wind</td>
<td>Precipitation</td>
<td>Landslides</td>
</tr>
</tbody>
</table>

**Physics based models**

- Global sea surface temperatures
- Atlantic hurricanes
- Global fires
Need for data driven discovery

Physics based models

<table>
<thead>
<tr>
<th>Low uncertainty</th>
<th>High uncertainty</th>
<th>Out of scope</th>
</tr>
</thead>
<tbody>
<tr>
<td>Temperature</td>
<td>Hurricanes</td>
<td>Fires</td>
</tr>
<tr>
<td>Pressure</td>
<td>Extremes</td>
<td>Malaria outbreaks</td>
</tr>
<tr>
<td>Large-scale wind</td>
<td>Precipitation</td>
<td>Landslides</td>
</tr>
</tbody>
</table>

Global sea surface temperatures

Atlantic hurricanes

Global fires

SST Anomaly Time Series in the ENSO region

Correlation with fires in Amazon

Chen et al., Science, 2011

May 14-16, 2013
Challenges in data driven analysis

- Complex dependence
  - Non-IID
  - Spatio-temporal correlation
  - Long memory in time
  - Long range dependence in space
  - Nonlinear relationships

- Data characteristics
  - Heterogeneous, Multivariate
  - Heavy Tailed Distributions
  - Noisy, incl. low frequency variability
  - Paucity of training data

- Complex processes
  - Evolutionary
  - Multi-scale in space and time
  - Non-stationary
# Project vision and scope

**Transformative Computer Science Research**

**Advancing Climate Change Science**

<table>
<thead>
<tr>
<th>Process Understanding</th>
<th>Extreme Events</th>
<th>Change Detection</th>
</tr>
</thead>
<tbody>
<tr>
<td>- Heat Waves</td>
<td>- Abrupt vs. Gradual</td>
<td></td>
</tr>
<tr>
<td>- Rainfall Extremes</td>
<td>- Point vs. Regions/Intervals</td>
<td></td>
</tr>
<tr>
<td>- Droughts</td>
<td>- Change in Extremes</td>
<td></td>
</tr>
<tr>
<td>- Hurricanes</td>
<td>Spatio-Temporal Classification</td>
<td></td>
</tr>
<tr>
<td>Model Evaluation</td>
<td>Sparse/High-Dim. Methods</td>
<td></td>
</tr>
<tr>
<td>Downscaling</td>
<td>Causal Relationships</td>
<td></td>
</tr>
<tr>
<td>- Statistical</td>
<td>Networks/Graphs</td>
<td></td>
</tr>
<tr>
<td>- Dynamical</td>
<td>HPC</td>
<td></td>
</tr>
<tr>
<td>Ocean-Atm.-Land Interactions</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Understanding Climate Change**

**Computational Innovations**
Pattern Mining: Ocean Eddies Monitoring

- Scalable spatio-temporal pattern mining algorithms for noisy and continuous data
- Novel multiple object tracking for uncertain features
- Detect more accurate features and tracks for improved ocean dynamics monitoring
- Open source data base of 20+ years of eddies and eddy tracks available for scientific applications

Faghmous et al. AAAI (2012a)
Faghmous et al. CIDU (2012b) **Best student paper award**
Faghmous et al. AAAI (2013)
NSF Nordic Research Opportunity Grant to conduct research at the Bjerknes Centre for Climate Research in Norway
Network analysis: Climate teleconnections

- Scalable method for discovering anti-correlated graph regions
- Novel dynamic graph clustering for dense directed graphs
- Significance testing for spatio-temporal patterns

- Discovered previously unknown climate teleconnection
- Analyzed climate network properties to better understand global climate dynamics
- Method used to compare climate models

Kawale et al. *SDM*(2011a)
Kawale et al. *CIDU* (2011b) **Best student paper award**
SC’11: Exploration in Science through Computation Award
Grace Hopper ‘12: Best Poster Award (Winner of the ACM Student Research Competition)
Predictive Modeling: Regression, Ensembles, Inference

- Hierarchical sparse regression: rates of convergence with low samples
- Multi-task learning with spatial smoothing
- Primal decomposition based LP solver for max-cut type problems (~10 million+ node graphs)
- Regional land-climate predictions from observations over oceans
- Combining multiple GCM outputs more accurately than state-of-art
- Mega-drought detection, trends over past 100-1000 years

Fu et al. *UAI(2013)*
Subbion et al. *SDM(2013)* Best Application Paper Award
Hsieh et al. *NIPS(2012)*
Wang et al. *ICML(2012)*
Chatterjee et al. *SDM(2012)* Best Student Paper Award
Fu et al. *SDM(2012)*
Relationship mining: Seasonal hurricane activity

- Contrast-based network mining for discriminatory signatures
- Novel dynamic graph clustering for dense directed graphs
- Statistically robust methodology for automatic inference of modulating networks
- Improved forecast skill for seasonal hurricane activity
- Discovered key factors and mechanisms modulating NA hurricane variability
- Discovered novel climate index with much improved correlation with NA hurricane variability: 0.69 vs 0.49

NSF News, DOE Research News, Science360
Sencan et al. IJCAI (2011)
Pendse et al. SIAM SDM (2012)
Chen et al. Data Mining & Knowledge Discovery (2012)
Chen et al. SIAM SDM (2013)
Chen et al. IJCAI (2013)
Semazzi et al. in review at journal (2013)
Extremes and uncertainty: Heat waves, heavy rainfall, ...

- Extreme value theory in space-time and dependence of extremes on covariates
- Mutual information and copula-methods for space-time extremes dependence
- Uncertainty quantification with Bayesian and resampling techniques
- Physics-guided data mining and quantification of uncertainty

- Spatiotemporal trends in heat waves, cold snaps, and heavy rain with climate change
- Climate model evaluation and physics-guided uncertainty quantification
- Covariate-based improvement of extremes projections under climate change
- Translation to adaptation and stakeholder relevant metrics

Press Release 11-266
JOURNAL PIECE REVEALS NEW DATA-DRIVEN METHODS FOR UNDERSTANDING CLIMATE CHANGE

Geographical variability of rainfall extremes in India enhances interpretation of climate change data

Ganguly et al. Climate Extremes & UQ: Book Ch. (2013)
Kodra et al. in revision at journal (2013)
Kumar et al. in review at journal (2013)
High Performance Tools and Methods

• Scalable library and software for
  • data mining / machine learning
  • Input-Output
  • Many algorithms have shown speedups of several orders of magnitude.

• HPC solutions for bootstrapping methods for extreme value prediction and Markov Random Field based abrupt change detection

• Enabled execution of a high-resolution cloud resolving model that is critical to operationalize the next generation of an IPCC GCM
  • Improved I/O throughput using PnetCDF optimizations, massive scalability
  • For 3.5 km grid resolution, grid size is 41.9M cells with 256 vertical layers

Jin et al. EuroMPI (2011)
Patwary et al. SC (2012)
Hentrix et al. HPC (2012)
Kumar et al. IPDPS (2011)
Rangel et al. in review (2013)
Jin et al. in review (2013)
Education/Outreach Activities

• Undergraduate and graduate courses/programs at the intersection of climate and data sciences
• Cross disciplinary training environment
• Extensive research opportunities for students from historically underrepresented groups
• Interdisciplinary workshops and sessions at climate and computer science venues
• Engagement with UNEP (United Nations Environmental Program) and IPCC
Future Directions and Goals

• Climate science problems provide transformative research opportunities
  • Complex dependence and noise structures
  • Nonlinear dynamical spatiotemporal systems
  • Data size from few petabytes to 350 petabytes by 2030
  • Motivates the development of “physics-guided data mining”

• Transformative spatiotemporal methods can generalize to multiple domains
  • Brain science
  • Ecology and biodiversity
  • Social networks
  • Geospatial Intelligence

• Help establish the field of “climate informatics” over the next 5-10 years