UNIVERSITY OF MINNESOTA
Driven to Discover℠

Fifth Workshop on

Understanding Climate Change from Data

The Fifth Annual Meeting of
NSF Expeditions in Computing Award # 1029711

August 4-5, 2015

University of Minnesota
200 Union Street SE
Room 3-180 Keller Hall
Minneapolis, MN 55455

Understanding Climate Change from Data

August 4-5, 2015
3-180 Keller Hall, University of Minnesota

Table of Contents

Agenda: Fifth Workshop on Understanding Climate Change from Data

	Tuesday, August 4, 2015	pg
8:00	Registration opens	
8:30	Vipin Kumar, University of Minnesota *Expeditions in Computing, Understanding Climate Change: A Data-Driven Approach*	7
9:10	Auroop Ganguly, Northeastern University *Climate Stresses on Critical Lifelines and Key Resources*	
9:40	Shafiqul Islam, Tufts University *BIG Data for BIG Floods: Can we BREAK the Predictability Limits?*	
10:10	Break	
	Session 2 Chair: *Varun Chandola*	9
10:40	Padhraic Smyth, University of California Irvine *Graphical Latent Variable Models with Applications to Climate Data*	
11:10	Arindam Banerjee, University of Minnesota *Spatiotemporal predictive modeling for climate science: Are we there yet?*	
11:40	Nagiza Samatova, North Carolina State University *Data-driven discovery of modulatory factors for African rainfall variability*	
12:10	Lunch	
	Session 3 Chair*: Juan Carlos Castilla-Rubio*	10
1:40	Sangram Ganguly, NASA Ames Research Center & BAERI *Climate Resiliency Tools and Workflows Using the NASA Earth Exchange (NEX) and OpenNEX Platforms*	
2:10	Varun Mithal, University of Minnesota *Mapping forest fires from satellite data: A new predictive modeling framework in absence of ground truth labels*	
2:30	Claire Monteleoni, George Washington University *Advances in Climate Informatics: Machine Learning Approaches to Improving the Multi-Model Ensemble and Defining Extreme Events*	
3:00	Break	
	Session 4 Chair: *Kate Brauman*	12
3:30	Soumyadeep Chatterjee, University of Minnesota *Understanding dominant factors for precipitation in great lakes region*	
3:50	Brian Smoliak, The Climate Corportation *Empirical approaches to uncovering teleconnections in global climate data*	

4:20	Stefan Liess, University of Minnesota *Introducing and Finding Tripoles: A Connection Between Central Asia and the Tropical Pacific*	
4:40	Anuj Karpatne, University of Minnesota *Global Monitoring of Inland Surface Water Dynamics using Remote Sensing Data*	
5:30	Poster Session and Dinner	
	Wednesday, August 5, 2015	15
8:00	Registration opens, Breakfast	
	Session 5 Chair: *Kyran Mish*	15
8:30	Raju Vatsavai, North Carolina State University *A View From Space – Remote Sensing Applications in Water, Food, and Energy Security*	
9:00	Shashi Shekhar, University of Minnesota *Spatial Decision Tree: A Novel Approach to Land-Cover Classification*	
9:30	Daniel Jiménez R, International Center for Tropical Agriculture (CIAT) *Big Data for Climate Smart Agriculture - Enhancing Rice Systems for Latin America*	
10:00	Break	
	Session 6 Chair: *Karsten Steinhaeuser*	16
10:30	Fred Semazzi, North Carolina State University *The Role of the Atlantic Ocean in Modulating the Recent Multi-Decadal Drought of East Africa*	
11:00	Tim DelSole, Center for Ocean-Land-Atmosphere Studies, George Mason University *Finding Relations in Climate Data Sets*	
11:30	Abdollah Homaifar- North Carolina Agricultural & Technical University *Multiple change detection in climate time series: a machine learning approach*	
12:00	Lunch	
1:30	Panel Discussion: Data Science and Mechanistic Understanding Moderators: Shashi Shekhar & Arindam Banerjee, University of Minnesota Panelists: Tim Bodin, Cargill Minnesota Tim DelSole, George Mason University Sai Ravela, Massachusetts Institute of Technology John Sharp, University of Texas Brian Smoliak, The Climate Corporation	19

3:30	Closing Reception

Session 1 Presentations – Tuesday, August 4

Vipin Kumar – University of Minnesota
Expeditions in Computing, Understanding Climate Change: A Data Driven Approach

Climate change is the defining environmental challenge facing our planet, yet there is considerable uncertainty regarding the social and environmental impact due to the limited capabilities of existing physics-based models of the Earth system. Consequently, important questions relating to food security, water resources, biodiversity, and other socio-economic issues over relevant temporal and spatial scales remain unresolved. A new and transformative approach is required to understand the potential impact of climate change. Data driven approaches that have been highly successful in other scientific disciplines hold significant potential for application in environmental sciences. This Expeditions project addresses key challenges in the science of climate change by developing methods that take advantage of the wealth of climate and ecosystem data available from satellite and ground-based sensors, the observational record for atmospheric, oceanic, and terrestrial processes, and physics-based climate model simulations. Methodologies developed as part of this project will be used to gain actionable insights and to inform policymakers. This presentation provides an overview of the challenges being addressed in this multi-disciplinary, multi-institutional project and includes highlights of some of the results obtained over the past several years.

Auroop Ganguly – Northeastern University
Climate Stresses on Critical Lifelines and Key Resources

Climate variability and change, as well as associated weather extremes and changes in regional hydro-meteorology, are leading to significant stresses on critical lifeline infrastructure systems and key natural resources. Lifelines ranging from multi-modal transportation, energy and power grids to water and waste water distribution systems, as well as communications networks, are increasingly vulnerable, especially in urban areas and coastlines. Resources such as water, energy, food, ecosystems, as well as the nexus of food-energy-water, are increasingly at risk owing to both gradual change and disruptions. Recent advances in data and network sciences have led to new insights in climate science and hydro-meteorological extremes, in addition to novel methods for statistically downscaling climate information and characterizing uncertainties at stakeholder relevant scales and prediction horizons. Network and data sciences have also been effectively used in recent years to quantify the robustness of infrastructure systems and natural resources as well as to develop strategies for recovery and restoration. Case studies in hazards resilience of transportation infrastructures and water stress on power production are showcased to highlight how climate change acts as a threat multiplier and how data and network science can help guide adaptation decisions and resilience policy.

Shafiqul Islam – Tufts University
BIG Data for BIG Floods: Can we BREAK the Predictability Limits?

Mounting losses from catastrophic floods are driving an intense effort to increase preparedness and improve response to disastrous flood events by providing early warnings. Improvements in the skill of numerical weather prediction (NWP) models have led many recent studies to suggest that flood forecasts might be extended to longer lead-times. However, improvements in the prediction of precipitation have lagged behind the improvements made by operational NWP models in forecasting aspects of the large-scale circulation.

The high-dimensionality and non-linearity in weather, hydrology and hydraulic models produces an "uncertainty cascade" that makes estimation of the forecast uncertainty practically impossible. We suggest a complementary approach that combines dynamical and probabilistic methods within the context of heuristic search aided by machine learning. Our proposed framework synergistically combines "big data" with "big weather" to predict the conditions conducive for "big floods". By combining traditional methods of synoptic typing with advanced machine learning techniques developed to identify patterns from large datasets, we will systematically identify the precursors to major floods. Our strategy is to combine a mechanistic understanding of the precursors to major flooding events with numerical weather forecasts to create probabilistic forecasts of flooding up to two weeks in advance. This collaborative initiative between Tufts University and the University of Massachusetts brings together complementary expertise in hydrology, synoptic meteorology and machine learning to develop algorithms for knowledge discovery and prediction of sequences of extreme precipitation events

Session 2 Presentations – Tuesday, August 4

Padhraic Smyth – University of California, Irvine
Graphical Latent Variable Models with Applications to Climate Data

Graphical models provide a general and systematic framework for describing and working with problems involving large numbers of interacting random variables. In this talk we will begin with a brief review of the basic concepts underlying graphical models, including the specification of joint distributions in the form of sparse graphs, with nodes representing random variables and edges representing dependence relations. The talk will discuss how the structure of the underlying graph can be leveraged for the purposes of efficient computation of unobserved quantities of interest, such as point estimates or distributions of model parameters given observed data, or computation of unobserved latent state variables (examples being smoothing and filtering in hidden Markov and Kalman filter models). Applications of these ideas to a number of different climate-related data sets will be used to illustrate the concepts presented in the talk. The talk will conclude with some brief comments on current and future directions in machine learning of potential relevance to climate modeling, including scalability to large data sets and recent advances in deep learning.

Arindam Banerjee – University of Minnesota
Spatiotemporal predictive modeling for climate science: Are we there yet?

Spatiotemporal predictive modeling has emerged as an active area of research in climate informatics. From the perspective of accuracy as well as interpretability, a key requirement for success in such modeling is the ability to find good predictors for climatological phenomenon of interest. The challenges in finding such good predictors come from the fact that one has to explore a wide range of spatial and temporal features, different temporal lags, spatial teleconnections, etc., using a typically small number of samples of the phenomenon of interest. Further, there could be multiple mechanisms affecting the phenomenon, possibly under different phases or even as a superposition. In this talk, we briefly discuss recent advances in structured predictive modeling in small sample regimes using geometric techniques, along with promising applications for modeling precipitation using feature selection and considering multiple mechanisms.

Nagiza Samatova – North Carolina State University
Data-driven discovery of modulatory factors for African rainfall variability

Discovering the key factors modulating the dynamics and variability of climate phenomena and weather events is an essential task in the climate science domain. In this talk, we will present an overview of two data-driven methodologies for the discovery of climatic modulatory factors and discuss their application to the study of African seasonal rainfall variability. The first methodology discovers climate indices associated with a response variable of interest from multivariate spatiotemporal data by using response-guided community detection. We apply this methodology to the discovery of climate indices associated with seasonal rainfall variability in the Greater Horn of Africa. The second methodology identifies climate indices with potential causal effects on the response variable of interest and constructs new features by capturing the potential causal relationships between these indices. We apply this methodology to the construction of features for seasonal rainfall anomalies forecast in the African Sahel region. The results

obtained suggest that both methodologies are able to capture the underlying patterns known to modulate seasonal rainfall variability in each of these regions and to improve predictability compared to existing methodologies.

Session 3 Presentations – Tuesday, August 4

Sangram Ganguly – Biospheric Sciences, NASA Ames Research Center and Bay Area Environmental Research Institute (BAERI)
Climate Resiliency Tools and Workflows Using the NASA Earth Exchange (NEX) and OpenNEX Platforms

NEX is a NASA collaboration platform for the Earth science community that provides a mechanism for scientific collaboration, knowledge and data sharing. NEX combines state-of-the-art supercomputing, Earth system modeling, NASA remote sensing data feeds, and a scientific social networking platform to deliver a complete work environment in which users can explore and analyze large Earth science data sets, run modeling codes, collaborate on new or existing projects, and quickly share results within and/or among communities. The work environment provides NEX members with community supported modeling, analysis and visualization software in conjunction with datasets that are common to the Earth systems science domain. Providing data, software, and large-scale computing power together in a flexible framework reduces the need for redundantly downloading data, developing data pre-processing software tools, building standard modeling and analysis codes, and expanding local compute infrastructures, thereby accelerating fundamental research, the development of new applications, and reducing project costs. NEX currently supports almost 120 science users and over 30 science teams. In order to better serve Earth science collaborators and increase scientific engagement, NEX has developed a cloud component using Amazon Web Services (AWS) - OpenNEX. OpenNEX currently hosts a large number of satellite data sets (MODIS, Landsat, AVHRR), downscaled climate data sets (NEX-DCP30) and custom virtual machines from which these datasets can be easily accessed and analyzed using different science codes. OpenNEX is now providing virtual workflows hosted in the cloud for rapid analysis of downscaled climate data and satellite data for large-scale climate analytics and insights. Some of the evolving tools include data discovery, rapid access and search, which in turn will increase the efficiency in dealing with processes related to query, subsetting and automatic data feeds to scientific codebases.

Varun Mithal – University of Minnesota
Mapping forest fires from satellite data: A new predictive modeling framework in absence of ground truth labels

This talk will present RAPT, a new predictive modeling framework for identifying rare classes in complete absence of labeled data. The RAPT framework is designed to use imperfectly annotated training data to learn classification models in the absence of expert-annotated training samples. Our results show that, under some reasonable assumptions, the classifiers trained from imperfectly labeled training data using the RAPT approach have performance comparable to the classification models trained using expert-annotated training data. This capability of learning from imperfect supervision is advantageous in a wide range of applications where the target class of interest is relatively rare and obtaining a precise labeling of even a small number of training samples is infeasible.

The talk will present the application of the RAPT framework for creating historical maps of forest fires from satellite data for the tropical forests. This new forest fire product identifies approximately 1 million sq. km. of burned areas in the tropical forests in South America and South-east Asia during years 2001-2014, which is more than double of the total burned area reported by the state-of-art NASA products. We

show validation of these results using burn-scars visible in satellite images, including high resolution Landsat images, to confirm the veracity of the previously unreported forest fires.

Claire Monteleoni – George Washington University
Advances in Climate Informatics: Machine Learning Approaches to Improving the Multi-Model Ensemble and Defining Extreme Events

Despite the scientific consensus on climate change, drastic uncertainties remain. Crucial questions about changes in regional climate and extreme events, such as heat waves, heavy precipitation, and drought, and understanding how climate varied in the distant past, must be addressed. Machine learning can help answer such questions and shed light on climate change. Further, such questions give rise to new challenges for the design of machine learning algorithms. I will survey my research group's progress in this emerging field of *Climate Informatics*. In particular, I will discuss our work on improving the predictions of the IPCC multi-model ensemble, and using unsupervised learning to detect climate phenomena, as a technique to define and detect extreme events.

Session 4 Presentations – Tuesday, August 4

Soumyadeep Chatterjee – University of Minnesota

Understanding dominant factors for precipitation in great lakes region

Statistical modeling of local precipitation involves understanding local, regional and global factors that carry information about precipitation variability in a region. Modern machine learning methods for feature selection can potentially be explored for identifying statistically significant features from a pool of potential predictors of precipitation. In this work, we consider sparse regression methods, which simultaneously perform feature selection and regression. We experiment on seasonal precipitation over Great Lakes Region in order to identify the dominant factors for each season. Significant features, identified using randomized permutation tests, offer hypotheses over possible mechanisms for seasonal regional precipitation, which can be further analyzed by domain scientists for physical validity. Such feature selection methods can thus be useful for hypothesis generation from pools of possible predictors, which can be invesrigated for further validation and possible novel insights.

Brian Smoliak – The Climate Corporation

Empirical approaches to uncovering teleconnections in global climate data

Teleconnections represent relationships between distant climate anomalies. Spatial teleconnection patterns play an important role in organizing atmospheric variability on month-to-month time scales and several recent studies have suggested that particular teleconnections become increasingly dominant at longer and longer time scales. The study of teleconnections has progressed from origins in heuristic analyses of station data beginning in the late 19th century to contemporary objective analyses of globally gridded data, the latter yielding a multiplicity of patterns. Such a diversity of patterns calls their robustness into question, in terms of reproducibility, stationarity, and governing dynamics. This presentation addresses these questions using the monthly-mean Northern Hemisphere sea-level pressure field to demonstrate methods for objectively identifying teleconnection patterns. Approaches to extending these analyses from univariate to multivariate geospatial fields are discussed, as are applications in climate diagnostics and prediction.

Stefan Liess – University of Minnesota

Introducing and Finding Tripoles: A Connection Between Central Asia and the Tropical Pacific

We introduce a novel long distant climate science teleconnection pattern called tripole. A tripole involves three regions 1,2, and 3, such that the anomaly time series at region 3 is more strongly correlated with either addition or subtraction of anomaly time series observed at region 1 and region 2, as compared to that with any of the anomaly time series at region 1 or 2 alone. A shared nearest neighbor (SNN) graph-based approach was used to find tripoles on DJF-monthly SLP data for 1979-2011 (2.5 x 2.5 degree resolution). The details of the algorithm are described in an accompanying poster. In this talk, we present one of the tripoles that we found across northwestern Russia and the two ends of ENSO. The tripole indicates a much stronger negative correlation between the anomaly time series of northwestern Russia and the sum of the anomaly time series of the two ends of ENSO, as compared to the pairwise correlations with either of the ENSO ends. ENSO is forced by an anomalous SST pattern between the west Pacific warm pool and the eastern Pacific Ocean. We take monthly time series of mean sea level pressure for the two

ends of ENSO (around Darwin, Australia and Tahiti) and add the z-scored area mean values of both regions to describe the background state of ENSO. We find that the background state of ENSO is linked to a pressure pattern over northwestern Russia. We identify a wave train that connects northwestern Russia via central Asia and eastern China to the ENSO region around Darwin, Australia. The link to the ENSO region can be detected in the geopotential height fields around 500 hPa.

Anuj Karpatne – University of Minnesota

Global Monitoring of Inland Surface Water Dynamics using Remote Sensing Data

Freshwater, which is only available in inland water bodies such as lakes, reservoirs, and rivers, is increasingly becoming scarce across the world and this scarcity is posing a global threat not only to human sustainability but also to the Earth's ecosystem. As a result, managing inland water has become one of the major 21st century challenges for the U.S. and the world. To address this challenge, we present a global monitoring system that provides timely and accurate information about the available surface water stocks and their dynamics across the world at regular intervals of time using remote sensing data. This system makes use of novel machine learning algorithms for handling a variety of issues in using remote sensing data, e.g. presence of heterogeneity in the data across space and time, and high degree of noise and missing values due to cloud and aerosol obstructions. In this talk, we will demonstrate some of the capabilities of our water monitoring system that is able to capture a variety of surface water dynamics, e.g. construction of new dams and reservoirs across the world, on-going droughts in Brazil and California, and changes in river morphology such as river meandering and delta erosion. For more information, please visit: http://z.umn.edu/watermrv.

Session 5 Presentations – Wednesday, August 5

Raju Vatsavai– North Carolina State University
A View From Space – Remote Sensing Applications in Water, Food, and Energy Security

Recent advances in remote sensing sensors and the increased number of satellites that are orbiting around the earth allowing us to map, monitor and characterize both natural and man-made resources at global scales. This ability to continuously monitor the Earth from space can help us better understand the linkages between the water, food, and energy. In the water sector, remote sending data has been widely exploited for rainfall/drought monitoring to daily evapotranspiration mapping. High spatial and spectral resolution imagery is the key resource behind national land cover mapping programs like NLCD and CDL, and has been commercially exploited by private sector in precision agriculture. With the ability to acquire daily remote sensing imagery covering the entire globe (e.g., MODIS), we are in a position to monitor biomass changes in near real-time. In addition to exploring these global scale applications, this presentation alludes to the new spatiotemporal data mining challenges in generating information products from the multi-sensor, multi-resolution, and multi-temporal remote sensing imagery.

Shashi Shekhar – University of Minnesota
Spatial Decision Tree: A Novel Approach to Land-Cover Classification

Accurate classification of land-cover (e.g., wetland) from overhead imagery is critical for understanding climate change. However, current machine-learning methods (e.g., decision trees, random forest) suffer from salt-and pepper noise and classification errors requiring manual substantial post-processing. We propose a novel spatial decision tree (and forest) learning approach based on the core ideas of focal-features, spatial objective function, and space partitioning to address challenges of spatial autocorrelation, heterogeneity and anisotropy. Experiments and a case study of wetland mapping from aerial imagery show that the proposed approach outperforms traditional decision trees. More details are available in a poster and a June 2015 publication (Ref. 1) in the IEEE Transactions on Knowledge and Data Eng.

[1] Z. Jiang, S. Shekhar, X. Zhou, J. Knight, J. Corcoran, Focal-Test-Based Spatial Decision Tree Learning, IEEE Transactions on Knowledge and Data Eng., 27(6), June 2015. (A summary in Proc. IEEE Intl. Conf. on Data Mining, 2013.)

Daniel Jiménez R – International Center for Tropical Agriculture (CIAT)
Big Data for Climate Smart Agriculture - Enhancing Rice Systems for Latin America

Climate change is mostly characterized by increasing probabilities of extreme weather patterns, such as temperature or precipitation reaching extremely high value. These time series data usually exhibit a heavy-tailed distribution rather than a Gaussian distribution. This poses great challenges to existing approaches due to the significantly different assumptions on the data distributions and the lack of sufficient past data on extreme events. In this talk, we present the Sparse-GEV model, a latent state model based on the theory of extreme value modeling to automatically learn sparse temporal dependence and make predictions. Our model is theoretically significant because it is among the first models to learn sparse temporal dependencies among multivariate extreme value time series. We demonstrate the

superior performance of our algorithm to the state-of-art methods, including Granger causality, copula approach, and transfer entropy, on both synthetic data and climate data.

Session 6 Presentations – Wednesday, August 5

Fred Semazzi – North Carolina State University
The Role of the Atlantic Ocean in Modulating the Recent Multi-Decadal Drought of East Africa

East Africa has been experiencing persistent decline of the March-April-May Long Rains for multiple decades. Although the connection between the decline and the Indo-Pacific Ocean has received much attention the role of the Atlantic Ocean has not been recognized. This study was motivated by the recent NSF funded collaborative Expedition research of the computer and climate scientists which has identified strong connection between the variability of the East African climate and the origin of Atlantic hurricanes. Here we show the previously unrecognized role of stationary atmospheric wave forms which link the northern Atlantic Ocean basin source region and the East African Long Rains. The Atlantic Multi-decadal Oscillation (AMO) variability completely dominates the variability during the cessation of the Long Rains (CLR) in May. The negative phase of the AMO is associated with enhanced rainfall during the cessation. In contrast, reduced rainfall occurs during the positive phase of AMO and it has contributed to the ongoing multi-decadal decline. The projected continuation of the positive phase of AMO for several more decades by recent studies imply the likelihood of the Atlantic Ocean's potential contribution to prolong the ongoing drought conditions over East Africa.

Tim DelSole – Center for Ocean-Land-Atmosphere Studies, George Mason University
Finding Relations in Climate Data Sets

In this talk, I first illustrate some common pitfalls in determining the relation between the seasonally averaged weather over a geographic region and associated sea surface temperatures. This example illustrates a few reasons why climate scientists might be skeptical of results from "big data" and what data scientists need to do to derive more convincing climate relations. Next, I frame data science as a problem of determining parameters in a "sparse" framework. Therefore, the major problem in combining data science and climate science is defining the correct sparse framework. In seasonal-to-interannual variability, I suggest that one part of this framework is the principle that small-scale structure is less trustworthy than large-scale structure. This principal can be expressed as saying that if variability is represented by a basis set ordered by a measure of length scale, then most of the amplitudes of the basis vectors are zero. This principle is tested in a practical example and compared with the standard methodology used in climate science, based on empirical orthogonal functions (EOFs). We find that the EOF method performs as well as the sparse methods in a cross validated sense. However, when the methods are performed on model simulations, and the resulting relations then used to predict observations, the EOFs are less successful, whereas the sparse methods produce skillful predictions from all dynamical models. It is suggested that the reason for this discrepancy is that apparent skill of the EOF method is inflated because the EOFs are sample-dependent and do not generalize well to independent samples.

Abdollah Homaifar – North Carolina Agricultural & Technical University
Multiple change detection in climate time series: a machine learning approach

Climate time series are generally non-stationary; as such, their statistical properties change with time. Any analysis of non-stationary time series requires detecting the structural change points between a set of clusters, where the time series model in each cluster has different statistical parameters. The change detection problem can be written as a non-convex constrained optimization that is generally ill-conditioned. Common change detection methods have limitations in the presence of autocorrelation, need restrictive statistical assumptions and may become trapped in local solutions. In this work, a new change detection technique based on the genetic algorithm (GA) is developed. The GA is a method for solving complex optimization problems based on the process of natural selection. The proposed method can find the change points in a climate time series with various statistical and regression models. The Bayesian information criterion has been used to find the optimal number of change points. The method has been tested in trend detection of average temperatures in North America as well as modeling the annual number of hurricanes in the North Atlantic.

Panel Discussion – Wednesday, August 5

Title: *Understanding and narrowing gaps between Data Science (e.g., correlations, frequent patterns) and Mechanistic Understanding (e.g., underlying processes driving patterns, extrapolating beyond observed conditions)*

Moderators: **Arindam Banerjee & Shashi Shekhar** - University of Minnesota

Questions for Panelists:
1. What are strengths and weaknesses of data science in context of climate, water and other physical sciences?
2. What are best practices for moving from statistical patterns (e.g., correlations) to mechanistic understanding (e.g., underlying processes driving patterns)?
3. How may data science methods model very rare but high-impact phenomena in face of data paucity and non-stationarity?
4. How may data science help decision making despite challenge of uncertainty and coupling (e.g., nexus of food, energy and water security)?

Background: Although data science methods have been applied quite extensively to analyze some large and complicated systems, such as social networks, data science efforts in complex natural systems (e.g., climate, water) have been fewer. Recent articles in Science [2], Nature [3], and PLOS [4] noted major failures of Google flu trend. Consequently, a 2014 New York Times article [5] said, "no scientist thinks you can solve this problem by crunching data alone, no matter how powerful the statistical analysis; you will always need to start with an analysis that relies on an understanding of physics and biochemistry". A 2014 Geo-Physical Letters paper [1] added: "failure to account for dependence between [Physical] models, variables, locations and seasons yield misleading results".

[1] Statistical significance of climate sensitivity predictors obtained by data mining, P. M. Caldwell et al., Geophys. Res. Lett., 41:1803-1808, 2014.
[2] The Parable of Google Flu: Traps in Big Data Analysis, David Lazer et al., AAAS Science, 343, March 2014.
[3] When Google got flu wrong, D. Butler, Nature, 494(7436), 155-6, 2013 (doi:10.1038/4941551a).
[4] Reassessing Google Flu Trends Data for Detection of Seasonal and Pandemic Influenza: A Comparative Epidemiological Study at Three Geographic Scales, D . Olson et al., PLOS Comp. Biology, 9, Oct. 17th, 2013.
[5] Eight (No, Nine!) Problems With Big Data, G. Marcus et al., New York Times, April 6th, 2014.
www.nytimes.com/2014/04/07/opinion/eight-no-nine-problems-with-big-data.html

Panelists:
Tim Bodin, Cargill, Inc. Minnesota
Tim DelSole, George Mason University
Sai Ravela, Massachusetts Institute of Technology
John Sharp, University of Texas
Brian Smoliak, The Climate Corporation

Participant Bios (alphabetical order):

Robert Andrade – University of Minnesota
andra065@umn.edu

Robert Andrade's primary research interest is impact assessment of technological change in agriculture. Robert is also interested in evaluating changes in rural household livelihood strategies, wellbeing, and agricultural production due to climate change effects. The overarching goal of his research is to design better policy and coping strategies for rural households in less-developed regions of the world. Robert is currently working toward his Ph.D. at the University of Minnesota where he is a graduate research assistant, jointly for the International Center for Tropical Agriculture (CIAT) and InSTePP. One target outcome of his time at InSTePP is to inject HarvestChoice methods and data into research strategy analyses and deliberations in Latin America and the Caribbean (LAC). Robert is also investigating the impacts of the uptake of improved rice varieties and updating and reassessing the evidence on the rates of return to agricultural research in LAC. Prior to his arrival at the University of Minnesota, he obtained his master's degree in agricultural and applied economics at Virginia Tech and his graduate degree in economics at the Pontificia Universidad Catolica del Ecuador in Quito. For the past decade he worked closely with national and international agricultural research institutes through his former position as an impact assessment officer at CIAT.

Gowtham Atluri – University of Minnesota
gowtham@cs.umn.edu

Gowtham Atluri is a postdoctoral associate in the Department of Computer Science at the University of Minnesota. His research focus is on mining large and complex data sets from neuroimaging and biomedical domains. He received his Ph.D. degree in Computer Science from the University of Minnesota and his M.S. degree in Computer Science from IIT Roorkee.

Tim Bodin – Cargill, Inc. Minnesota

Tim Bodin is an economist at Cargill, Inc. and a Board Member of the Minnesota Council of Economic Education. He graduated with his BS degree from the University of Minnesota. Tim is a member of the Sustainable Ag Intensification Team through SNAP, a joint initiative of The Nature Conservancy, NCEAS, and the Wildlife Conservancy. This team's focus and concerns lie in the expansion of agriculture into wild lands, and how that expansion poses an enormous risk to conservation efforts. An alternative may be to intensify agriculture in specific places, growing more food on less land and sparing natural areas. Is such intensification sustainable, and what will it look like?

Kate A. Brauman – University of Minnesota Institute on the Environment
kbrauman@umn.edu Web: *z.umn.edu/brauman*

Kate Brauman is the Lead Scientist for the Global Water Initiative at the University of Minnesota's Institute on the Environment. Kate's research integrates hydrology and plant-water relationships with economics and policy to better understand how water use by people affects the environment and our ability to live well in it. Through projects as diverse as payments for watershed services, global variation in "crop per drop", and worldwide trends in water consumption and availability, Kate works with the Global Water Initiative to find sustainable solutions to pressing water issues. Kate received her doctorate from Stanford University and her undergraduate degree from Columbia University.

Juan Carlos Castilla-Rubio – Planetary Skin Institute
jc@planetaryskin.org

Founder & CEO of non-profit research & development corporation Planetary Skin Institute (PSI) co-founded with NASA tasked to incubate advanced risk management and resource productivity decision-support platforms that will enable communities, businesses and governments to drastically improve the productivity of their resource bases and to proactively manage the exposures and vulnerabilities of their assets, supply chains and related infrastructure to accelerating climate change (www.planetaryskin.org). PSI was awarded one of TIME Magazine's Top Innovation Awards for 2009 and has been profiled in The Economist, the New York Times, Fast Company, amongst others.

Prior to his PSI role, he led the Global Resource/ Risk Management Innovations Group for Cisco Inc with global responsibility for programs incubating Big Data & Analytics, Internet of Things and Cloud-based risk management/ resource productivity platforms across multiple industries and sectors globally. Prior to that, Juan Carlos led Cisco's Emerging Markets Innovation Group focused on designing and operationalizing large scale/ complex technology-enabled innovation programs for governments and corporations emerging markets wide, with a focus in Brazil, Chile, India, South Africa, Mexico, South Africa, Gulf States, Turkey and China.

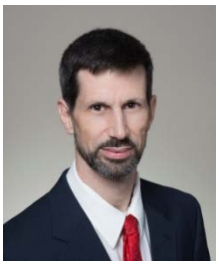
Prior to joining Cisco, he was a Partner for Oliver Wyman where he led the financial services and telecom practices in Latin America. Prior to Oliver Wyman, he was CEO for an innovative biotechnology start-up focused on next generation bio-fuel technologies and pharma research in the Amazon (Bioingenieria Aplicada S.A.), an EVP Operations for a brewery group in Latin America (now SABMiller), a Managing Director for Corporate Strategic Development of a diversified financial services group in Latin America (Intercorp Group) and a senior consultant for McKinsey & Company in Brazil.

He is currently member of the World Economic Forum GAC on Forests, was the co-chair of the World Economic Forum's Global Advisory Council on Measuring Sustainability and was a member of the World Economic Forum's Global Advisory Council on Climate Change; member of Duke Energy's International Strategy Board, a member of the World Economic Forum's Steering Boards for the Water Security,

Resource Scarcity, New Energy Architecture and Big Data initiatives. He led the World Economic Forum's groundbreaking digital inclusion initiative IT Access for Everyone (ITAFE) in partnership with leading high tech players and governments around the world.

Varun Chandola– State University of New York (SUNY) at Buffalo
chandola@buffalo.edu

Varun Chandola is an Assistant Professor at University at Buffalo (SUNY) in the Computer Science Department and in the center for Computational and Data-Enabled Science and Engineering (CDSE) since 2013. Previously, Chandola worked as a scientist at Oak Ridge National Laboratory in the computational sciences and engineering division for 4 years from 2009-2013. His research covers the application of data mining and machine learning to problems involving big and complex data, focusing on anomaly detection – finding surprising patterns, connections, associations, and trends in data. He has led or co-led projects funded by NSF, DOE, and ORNL. He is a recipient of a significant achievement award at ORNL for his work in the area of settlement mapping from remote sensing data. He is also the lead developer of iGlobe, which is an award winning system for analysis and visualization of geospatial data. He has a PhD from University of Minnesota from their Computer Science Department.



Tim DelSole –Center for Ocean-Land-Atmosphere Studies, George Mason University
delsole@cola.iges.org

Timothy DelSole conducts research on climate variations and prediction. Tim works at the intersection of climate science and multivariate statistics to address questions in climate change and prediction. He currently serves as editor for Journal of Climate and was a contributing author to the fifth assessment report of the Intergovernmental Panel on Climate Change. Dr. DelSole received a doctorate in applied physics from Harvard University in 1993 and worked as a postdoc at the NASA Goddard Space Flight Center. Currently, he is a full professor in the Department of Atmospheric, Oceanic, and Earth Sciences at George Mason University, and a senior research scientist at the Center for Ocean-Land-Atmosphere Studies.

Peder Engstrom – University of Minnesota Institute on the Environment
engs0074@umn.edu

Peder Engstrom is a scientist with the Global Landscapes Initiative. Peder has a background in geographic information science (GIS) and a passion for agriculture and food systems. He focuses on stakeholder engagement, data visualization, and map and data production. Peder also maintains the group's geo-data sharing platform, EarthStat.org. Research interests include crop species distribution patterns, better understanding gaps in agricultural data, smallholder agriculture, and equitable farm systems.

Sangram Ganguly – Biospheric Sciences, NASA Ames Research Center and Bay Area Environmental Research Institute (BAERI)

Sangram Ganguly received his PhD in 2008, and has since then worked at NASA Ames Research Center (CA, USA) and Bay Area Environmental Research Institute as a senior research scientist. Dr. Ganguly has made significant contributions in advanced remote sensing techniques for carbon modeling and climate dynamics, and in the development of high performance computing resources in Earth sciences. Dr. Ganguly's activities at the NASA's Advanced Supercomputing Division and Earth Science Division are focused on high end computing technologies for big data computation, physical algorithms in remote sensing, and scalable solutions for Earth science and climate research. Apart from doing research, Dr. Ganguly is involved in web application development and is, for example, a co-founder of the NASA Earth Exchange (NEX) platform (https://nex.nasa.gov/nex/) and the OpenNEX platform in collaboration with Amazon Web Services (AWS). Dr. Ganguly is one of the 'young talents' in NASA: he has won the NASA achievement award several times and has been involved as principal Investigator and co-investigator in several NASA funded projects. He is an active panel member for both National Science Foundation (NSF) and NASA Geoscience/Earth Science Programs.

James Gerber– University of Minnesota Institute on the Environment
jsgerber@umn.edu

James Gerber is the co-director and lead scientist of IonE's Global Landscapes Initiative (GLI) a program at the Institute on the Environment, University of Minnesota which develops solutions for meeting current and future global food needs while sustaining our planet. His research develops solutions for improving global food security while minimizing agriculture's impact on earth's ecosystems. A particular research focus is the interrelation of climate variability, crop yields, and systemic trends in food security. Prior to joining IonE in 2009, he was a lead scientist at Ocean Power Technologies, developing devices to convert ocean wave energy to electricity. Jamie has a doctorate in physics from the University of California, Santa Cruz.

Shafiqul Islam – Tufts University

Shafiqul ("Shafik") Islam is Professor of Civil and Environmental Engineering and Professor of Water Diplomacy at the Fletcher School of Law and Diplomacy at Tufts. He was the first Bernard M. Gordon Senior Faculty Fellow in Engineering at Tufts University. Professor Islam's teaching and research interests are to understand characterize, measure, and model water issues ranging from climate to cholera to water diplomacy with a focus on scale issues and remote sensing. His research group WE REASoN integrates "theory and practice" and "think and do" to create actionable water knowledge.

He maintains a diverse network of national and international partnerships including MIT, Columbia University, Purdue University, Penn State University, Princeton, BUET in Bangladesh, University of Tokyo, ETH in Switzerland, ICDDRB in Bangladesh, IIT in India, and SaciWATERs to conduct multi-year and multi-million dollar interdisciplinary collaborative research for a wide range of problems focusing on water, health, and

climate. His major research sponsors include the National Science Foundation, National Institute of Health, and the National Aeronautics and Space Administration.

Dr. Islam maintains an active national and international consulting and training practice including: flood forecasting in India; national water planning in Bangladesh; water policy planning for ExxonMobil; and advising South Asian Consortium of Interdisciplinary Water initiatives. He acted as consultant to the World Bank, United States Geological Survey, Proctor and Gamble, and several other governmental and non-governmental organizations. He has published more than 100 refereed journal and other publications. His research findings have been featured in numerous media outlets including the BBC World Service, Voice of America, Boston Globe, Huffington Post, Nature, and Yale E360. To learn more about Dr. Islam's academic interests and research, please visit WE REASoN and Water Diplomacy.

Daniel Jiménez R – International Center for Tropical Agriculture (CIAT)
d.jimenez@cgiar.org

In many parts of the world, farmers have been using traditional calendar landmarks to make climate-related decisions such as what, where and when to grow. Unfortunately, climate is not stable or highly predictable, and farmers are faced with a reality that the next cropping season is more likely to be different than the past one. Over the last years, national average rice yields have dropped in Colombia (from 6t/ha in irrigated rice before 2009 to 5t/ha today) and rice growers have not managed to recover since then. New approaches are required to provide growers with updated and relevant information that can support the decision making process and make them more resilient to climate variability. Novel use of ICTs and the possibility nowadays of capturing, analyzing and sharing large amounts of information in agriculture offer an alternative to approaches based on small scale field-based studies. We used empirical modelling techniques mostly based on machine learning, to analyze commercial harvest monitoring data from the Colombian National Rice Growers Association (FEDEARROZ), combined with climate information at daily resolution. We identified main climatic limiting factors in several rice producing areas and linked our results to seasonal forecasts to generate recommendations on best variety to grow.

Kyran D. Mish – Sandia National Laboratories
kdmish@sandia.gov

Kyran D. Mish is a principal member of the technical staff at Sandia National Laboratories in Albuquerque, New Mexico. At Sandia, Dr. Mish serves as the technical liaison between the Department of Defense computational analyst community and the Sandia engineering code groups funded under the NNSA's Advanced Simulation and Computing (ASC) initiative. Dr. Mish has four decades of experience in computational engineering in national laboratory, private engineering practice, and academic venues. Dr. Mish's professional experience includes his current work at Sandia, a senior management tenure at Lawrence Livermore National Laboratory as the founding director of the Center for Computational Engineering, and service on the engineering faculty of the University of California, Davis and the University of Oklahoma. Dr. Mish's research interests lie at the interface of critical infrastructure and information technology, and his body of research work includes interests in subsurface mechanics, structural engineering, fluid-structure coupling, soil-structure interaction, scalable computing, and scientific visualization.

Claire Monteleoni – George Washington University

Claire Monteleoni is an assistant professor of Computer Science at George Washington University. Previously, she was research faculty at the Center for Computational Learning Systems, at Columbia University. She did a postdoc in Computer Science and Engineering at the University of California, San Diego, and completed her PhD and Masters in Computer Science, at MIT. She holds a Bachelors in Earth and Planetary Sciences from Harvard. Her research focuses on machine learning algorithms and theory for problems including learning from data streams, learning from raw (unlabeled) data, learning from private data, and climate informatics: accelerating discovery in climate science with machine learning. Her work on climate informatics received the Best Application Paper Award at NASA CIDU 2010. In 2011, she co-founded the International Workshop on Climate Informatics, which is now in its fifth year, attracting climate scientists and data scientists from over 16 countries and 28 states. She presented an invited tutorial on climate informatics at NIPS 2014. She currently serves as Area Chair for NIPS 2015 and ICML 2015, and on the Senior PC of UAI 2015.

Sai Ravela – Massachusetts Institute of Technology
ravela@mit.edu Web: *essg.mit.edu*

I studied Computer Vision and Robotics at the University of Massachusetts at Amherst. As a graduate student, I became interested in Earth's sustainability and therefore joined the Earth, Atmospheric and Planetary Sciences department as a post-doc studying Geophysical Fluids. I continue as a Principal Research Scientist.

My enduring research interest is to develop succinct representations of our Earth's Stochastic Signals and Systems. The projects I work on emphasize dynamic coupling between phenomenology and physics. My current results suggest new algorithms to overcome the curses of nonlinearity, dimensionality and uncertainty in inference problems characteristic of the Earth system. In particular, physically-based Bayesian learning improves Hurricane Risk estimation, Ensemble Learning treats Model Error better, non-parametric Information theoretic learning deals with non-Gaussianity tractably, manifold methods improve uncertainty quantification, deformable models enable efficient representation of coherent fluids, Lightfields recover atmospheric turbulence, and reinforcement learning enables mitigation planning under climate change.

Some of my research has made it out of academia, specifically to a company (WindRiskTech) I co-founded in 2005 with Kerry Emanuel, and E5 Aerospace, which I co-founded in 2013. I have also served as a board member for Sustainable-step New England (2000-2002) and remain deeply committed to sustainability science and engineering.

Deepak Ray – University of Minnesota Institute on the Environment
dray@umn.edu

Deepak Ray is a Senior Scientist with the Institute on the Environment's Global Landscapes Initiative. He conducts research on how global crop production is changing over time. He answers questions such as where the green revolution has stopped and where crop productivity gains continue and why. To answer these broad questions he builds a crop statistics database and then analyzes the "Big Data." He is a climate scientist by training and is now focusing on problems connected with climate and agriculture.

John (Jack) Sharp – The University of Texas
jmsharp@jsg.utexas.edu

Jack is the Carlton Professor of Geology in Department of Geological Sciences at The University of Texas (UT) and was recently on a one-year leave of absence at the National Science Foundation as Program Director in the Hydrologic Sciences. Jack has a Bachelor of Geological Engineering from the University of Minnesota and a MS and a PhD in geology from the University of Illinois. He is a Fellow of the Geological Society of America and the Alexander von Humboldt Stiftung. His research covers flow and transport in fractured and carbonate rocks, thermohaline free convection, sedimentary basin hydrogeology, subsidence and coastal land loss, groundwater management, and the effects of urbanization on groundwater systems. He has been President of the Geological Society of America (GSA) and the Austin Geological Society and Chairman of the US Chapter of International Association of Hydrogeologists (IAH). He has received a number of honors including the O.E. Meinzer and Hydrogeology Division Distinguished Service Awards (GSA), the C.V. Theis and Founders' Awards (American Institute of Hydrology), the Presidents' Award (IAH), the Publication Award (Association of Engineering Geologists), the Farvolden Lecturer (U. Waterloo), Phi Kappa Phi, and Tau Beta Pi. Hobbies include gardening, genealogy, fishing, duck hunting, Australia, opera, UT football, and (before bad knees) handball.

Brian Smoliak – The Climate Corporation
bsmoliak@climate.com

Dr. Smoliak is an atmospheric scientist at The Climate Corporation in Seattle, WA. Previously he was a postdoctoral research associate in the Department of Soil, Water, and Climate at the University of Minnesota, where he was involved in the field and analysis portions of a project investigating the Twin Cities urban heat island led by Professors Peter Snyder and Tracy Twine. Brian completed his PhD in Atmospheric Sciences under the supervision of Prof. John "Mike" Wallace at the University of Washington; his thesis work focused on detection and attribution of surface air temperature change in the instrumental record. His scientific interests span spatial scales from local to global, and from basic research into the fundamental behavior of the climate system to applied research tied directly to particular human and environmental outcomes in areas such as human health, agriculture, and natural ecosystems.

Name: **Padhraic Smyth** – University of California, Irvine
smyth@ics.uci.edu

Padhraic Smyth is a Professor in the Department of Computer Science with a joint appointment in Statistics, and Director of the UCI Data Science Initiative, all at the University of California, Irvine. His research interests include machine learning, data mining, pattern recognition, and applied statistics and he has published over 150 papers on these topics, with best paper awards at the 2002 and 1997 ACM SIGKDD conferences. He is an ACM Fellow (2013), a AAAI Fellow (2010), and a recipient of the ACM SIGKDD Innovation Award (2009). He is co-author of the text Modeling the Internet and the Web: Probabilistic Methods and Algorithms (with Pierre Baldi and Paolo Frasconi in 2003) and Principles of Data Mining, MIT Press (with David Hand and Heikki Mannila in 2001), and he served as program chair of the ACM SIGKDD 2011 conference and the UAI 2013 conference.

Padhraic has served in editorial and advisory positions for journals such as the Journal of Machine Learning Research, the Journal of the American Statistical Association, and the IEEE Transactions on Knowledge and Data Engineering. While at UC Irvine he has received research funding from agencies such as NSF, NIH, IARPA, NIST, NASA, and DOE, and from companies such as Google, IBM, Yahoo!, Experian, and Microsoft. In addition to his academic research he is also active in industry consulting, working with companies such as eBay, Samsung, Yahoo!, Microsoft, Oracle, Nokia, and AT&T, as well as serving as scientific advisor to local startups in Orange County. He also served as an academic advisor to Netflix for the Netflix prize competition from 2006 to 2009.

Padhraic received a first class honors degree in Electronic Engineering from National University of Ireland (Galway) in 1984, and the MSEE and PhD degrees (in 1985 and 1988 respectively) in Electrical Engineering from the California Institute of Technology. From 1988 to 1996 he was a Technical Group Leader at the Jet Propulsion Laboratory, Pasadena, and has been on the faculty at UC Irvine since 1996.

Karsten Steinhaeuser – Progeny Systems Corporation and University of Minnesota
ksteinha@umn.edu

Karsten Steinhaeuser is a Data Scientist with Progeny Systems Corp. and a Research Associate in the Department of Computer Science & Engineering at the University of Minnesota. His research interests are centered around data mining and machine learning, in particular the construction and analysis of complex networks, with applications in diverse domains including climate, ecology, and social networks. He has made significant contributions to shaping the emerging research area of climate informatics, which lines at the intersection of multiple disciplines including computer science and climate sciences, and his interests are more generally in interdisciplinary research and scientific problems relating to climate change and sustainability. He is a founding organizer of the IEEE ICDM Workshop on KDD from Climate Data, the International Workshop on Climate Informatics, and numerous AGU and EGU sessions on related topics. Dr. Steinhaeuser earned his PhD in Computer Science & Engineering from the University of Notre Dame in 2011, where he was a member of the Interdisciplinary Center for Network Science and Applications (iCeNSA). He previously received an MS in Computer Science & Engineering and a BS, Summa Cum Laude in Computer Science, both from the University of Notre Dame.

Raju Vatsavai– North Carolina State University
rrvatsav@nscu.edu

Raju Vatsavai joined the Department of Computer Science at the North Carolina State University in August 2014 as a Chancellor's Faculty Excellence Program Cluster Associate Professor in Geospatial Analytics. Raju is an interdisciplinary scientist known for innovative contributions to large scale spatial and spatiotemporal data management and spatial data mining. Prior to joining the NC State, Raju was the Lead Data Scientist for the Computational Sciences and Engineering Division of the Oak Ridge National Laboratory (ORNL). He has more than 20 years of research and development experience in large-scale data management and knowledge discovery by working at the University of Minnesota, IBM-Research, AT&T Labs, and the Center for Development of Advanced Computing (C-DAC, India). He has authored or co-authored over 75 peer-reviewed publications, edited two books, served on several NSF and DOE panels, developed and co-organized several workshops with leading international conferences including ACM SIGKDD, ACM SIGSPATIAL, IEEE ICDM, and ACM/IEEE Supercomputing. He holds MS and PhD degrees in computer science from the University of Minnesota.

Expeditions in Computing: Understanding Climate Change, A Data Driven Approach

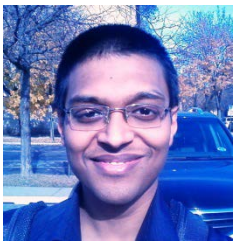
NSF Awards: 1029711, 1029166, 1029731, 1028746
Project Homepage: *climatechange.cs.umn.edu*

Expeditions in Computing Team:

The project team, led by the University of Minnesota, includes faculty and researchers from Minnesota's College of Science and Engineering, College of Food, Agricultural and Natural Resource Sciences, College of Liberal Arts, and the Institute on the Environment, as well as researchers from North Carolina A & T State University, North Carolina State University, Northwestern University, and Northeastern University.

Vipin Kumar – University of Minnesota
PI of Expeditions in Computing Project
kumar@cs.umn.edu, www.cs.umn.edu/~kumar

Vipin Kumar is currently Regents Professor and William Norris Endowed Chair of the Computer Science and Engineering Department at the University of Minnesota. Kumar received the B.E. degree in Electronics & Communication Engineering from Indian Institute of Technology Roorkee, India, in 1977, the M.E. degree in Electronics Engineering from Philips International Institute, Eindhoven, Netherlands, in 1979, and the Ph.D. degree in Computer Science from University of Maryland, College Park, in 1982. Kumar's current research interests include data mining, high-performance computing, and their applications in Climate/Ecosystems and Biomedical domains. Kumar is the Lead PI of a 5-year, $10 Million project, "Understanding Climate Change - A Data Driven Approach", funded by the NSF's Expeditions in Computing program that is aimed at pushing the boundaries of computer science research. He served as the Head of the Computer Science and Engineering Department from 2005 to 2015 and the Director of Army High Performance Computing Research Center (AHPCRC) from 1998 to 2005. He has authored over 250 research articles, and has coedited or coauthored 11 books including widely used text books "Introduction to Parallel Computing" and "Introduction to Data Mining", both published by Addison Wesley. Kumar has served as chair/co-chair for many international conferences and workshops in the area of data mining and parallel computing, including IEEE International Conference on Data Mining (2002) and International Parallel and Distributed Processing Symposium (2001). Kumar co-founded SIAM International Conference on Data Mining and served as a founding co-editor-in-chief of Journal of Statistical Analysis and Data Mining (an official journal of the American Statistical Association). Currently, Kumar serves on the steering committees of the SIAM International Conference on Data Mining and the IEEE International Conference on Data Mining, and is series editor for the Data Mining and Knowledge Discovery Book Series published by CRC Press/Chapman Hall. Kumar is a Fellow of the ACM, IEEE and AAAS. Kumar received the 2009 Distinguished Alumnus Award from the Computer Science Department, University of Maryland College Park, 2005 IEEE Computer Society's Technical Achievement Award, and ACM SIGKDD 2012 Innovation Award for his foundational research in data mining as well as its applications to mining scientific data.



Ankit Agrawal– Northwestern University
ankitag@eecs.northwestern.edu

Ankit Agrawal is a Research Associate Professor in the Dept. of Electrical Engineering and Computer Science at Northwestern University. He received his PhD in Computer Science from Iowa State University, USA in 2009, and was awarded the Research Excellence Award and Peer Research Award for outstanding research accomplishments. He received B.Tech in Computer Science and Engineering from the Indian Institute of Technology, Roorkee, India in 2006, where he was the graduating topper of his batch and was awarded Institute Silver medals for obtaining the highest GPA and the best B.Tech project. His research interests include high performance data mining and its applications in bioinformatics, climate science, social media, materials science, etc., and has published more than 80 papers in various peer-reviewed journals and conferences. His research is supported by National Science Foundation, Department of Energy, Air Force Office of Sponsored Research, and National Institute of Standards and Technology.

Arindam Banerjee – University of Minnesota
banerjee@cs.umn.edu

Arindam Banerjee is an Associate Professor at the Department of Computer and Engineering and a Resident Fellow at the Institute on the Environment at the University of Minnesota, Twin Cities. His research interests are in machine learning, data mining, convex analysis and optimization, and their applications in complex real-world problems, including climate science, ecology, aviation, and the web. He has won several awards, including the IBM Faculty Award (2013), the Yahoo Faculty Research Engagement Program Award (2013), the NSF CAREER award (2010), the McKnight Land-Grant Professorship at the University of Minnesota, Twin Cities (2009 &2011), the IBM PhD fellowship (2003-05), and six Best Paper awards in top-tier conferences.

Gonzalo Bello – North Carolina State University
gabellol@ncsu.edu

Gonzalo Bello is a PhD student in the Department of Computer Science of North Carolina State University working with Dr. Nagiza F. Samatova. His main research interests are in the areas of data mining and machine learning. His current work focuses on the design and analysis of graph mining algorithms and their application to real-world problems.

Snigdhansu Chatterjee – University of Minnesota
chatterjee@stat.umn.edu

Ansu Chatterjee is Associate Professor in the School of Statistics, University of Minnesota. After graduating from the Indian Statistical Institute, he worked at the University of Manchester in England and at University of Nebraska-Lincoln before joining University of Minnesota, where he is currently tenured. He has published in the Annals of Statistics, Annals of Applied Statistics, Annals of the Institute of Statistical Mathematics, Bioinformatics, and other journals. His research interests include climate statistics, small area statistics, Bayesian statistics, change detection methods, resampling techniques and other nonparametric methodology.

Soumyadeep Chatterjee – University of Minnesota
chat0129@umn.edu

Soumyadeep Chatterjee is a Ph.D. candidate at the University of Minnesota, Twin Cities, where he is advised by Dr. Arindam Banerjee. His primary interests are in the fields of Machine Learning and High Dimensional Statistics and their applications to climate data analysis. He is a part of the NSF funded project "Expeditions in Computing: Understanding Climate Change: A Data Driven Approach", where he is using structured regression methods for understanding statistical relationships among climate variables.

Name: **Mandar Chaudhary** – North Carolina State University
mschaudh@ncsu.edu

Mandar Chaudhary is a PhD student working with Dr. Samatova at North Carolina State University (NCSU). He joined the PhD program in 2014 after receiving his M.S. in Computer Science at NCSU. His research focuses on developing methods to discover features from causal-driven methods to construct a new feature space. This work has an application towards improving the forecasting performance of the African Sahel seasonal rainfall anomalies.

Alok Choudhary – Northwestern University
choudhar@eecs.northwestern.edu

Alok Choudhary is a John G. Searle Professor of Electrical Engineering and Computer Science at Northwestern University. He is the founding director of the Center for Ultra-scale Computing and Information Security (CUCIS). Prof. Choudhary was a co-founder and VP of Technology of Accelchip Inc. in 2000. Accelchip, Inc., was eventually acquired by Xilinx. He received the National Science Foundation's Young Investigator Award in 1993. He also received an IEEE Engineering Foundation award, an IBM Faculty Development award and an Intel Research Council award. He is a fellow of IEEE, ACM and AAAS. His research interests are in high-performance computing, data intensive computing, scalable data mining, computer architecture, high-performance I/O systems and software and their applications in many domains including information processing (e.g., data mining, CRM, BI) and scientific computing (e.g., scientific discoveries). Alok Choudhary has published more than 350 papers in various journals and

conferences and has graduated 27 PhD students. Techniques developed by his group can be found on every modern processor and scalable software developed by his group can be found on most supercomputers. Alok received his Ph.D. degree in Electrical and Computer Engineering from the University of Illinois, Urbana-Champaign, in 1989.

James Faghmous – University of Minnesota
jfagh@cs.umn.edu

James H. Faghmous obtained his Ph.D. in computer science from the University of Minnesota -Twin Cities. As part of the Expeditions team, James developed theory-guided data mining algorithms to analyze large climate datasets with applications to tropical cyclone forecasting and mesoscale ocean eddy monitoring. James' Ph.D. thesis was awarded the 2014 Best Dissertation Award in Physical Sciences and Engineering at the University of Minnesota. James was amongst the earliest data scientists to publish on "theory-guided data science" which earned him Best Student Paper Award at the 2011 NASA Conference on Intelligent Data Understanding.

James' research has been generously funded by an NIH Neuro-Physical-Computational Graduate Fellowship, an NSF Graduate Research Fellowship, an NSF Nordic Research Opportunity Fellowship, and a University of Minnesota Doctoral Dissertation Fellowship. James graduated in 2006 with a B.Sc. in computer science from the City of College of New York where he was a Rhodes and a Gates Scholar nominee.

Jonathan Foley – University of Minnesota
jfoley@umn.edu

Jonathan Foley is the director of the Institute on the Environment (IonE) at the University of the Minnesota, where he is a professor and McKnight Presidential Chair in the Department of Ecology, Evolution and Behavior. He also leads the IonE's Global Landscapes Initiative. Foley's work focuses on the sustainability of our civilization and the global environment. He and his students have contributed to our understanding of global food security, global patterns of land use, the behavior of the planet's climate, ecosystems and water cycle, and the sustainability of the biosphere. This work has led him to be a regular advisor to large corporations, NGOs and governments around the world.

Poulomi Ganguli – Northeastern University
p.ganguli@neu.edu

Poulomi Ganguli is a post-doctoral research fellow in the Civil and Environmental Engineering department at Northeastern University. She earned PhD in Civil Engineering from Indian Institute of Technology Bombay in 2013. Her research interests include hydrological extremes, hydro-climatology and assessment of climate change and variability in surface and sub-surface hydrology.

Auroop Ganguly – Northeastern University
a.ganguly@neu.edu

Auroop Ganguly directs the Sustainability and Data Sciences Laboratory (SDS Lab) at Northeastern University in Boston, MA, where he joined in Fall 2011, as an associate professor of civil and environmental engineering. He was at the Oak Ridge National Laboratory in their computational sciences and engineering division for exactly 7 years from 2004-2011, most recently as a senior scientist. He led or co-led projects funded by DARPA, DOD, DHS, DOE, ONR, and ORNL, and received three significant event awards and two outstanding mentor awards at ORNL. He received an outstanding faculty award from the University of Tennessee in Knoxville where he held a joint appointment with ORNL. His research on climate extremes and uncertainty has been published in Proceedings of the National Academy of Sciences, Nature Climate Change, Geophysical Research Letters, Journal of Geophysical Research, and on hydrology in Water Resources Research, Advances in Water Resources, Journal of Hydrometeorology, and Nonlinear Processes in Geophysics. His work on computational methods and complex systems has been published in Physical Review E, IEEE Transactions on Intelligent Transportation Systems, and Intelligent Data Analysis, as well as in peer-reviewed conferences in computer science such as SIAM Data Mining, besides ACM KDD and IEEE ICDM workshops. He was the primary founding organizer of a 2006-2012 workshop series on sensor data mining at the ACM KDD, the first of which resulted in an edited book by CRC Press called "Knowledge Discovery from Sensor Data". He holds associate editor positions in the journal Water Resources Research published by the American Geophysical Union and Journal of Computing in Civil Engineering published by the American Society of Civil Engineers. In addition, he is an elected member of the Artificial Intelligence Committee of the American Meteorological Society. He was on a visiting faculty position at the University of South Florida in Tampa, FL, for about 10 months, and currently holds a visiting faculty position at the Indian Institute of Technology Bombay in Mumbai, India, within their climate change interdisciplinary program. He was employed at Oracle Corporation for about 5 years, first as a time series software developer for a year in their core database kernel, and then as the product manager of their demand forecasting and planning product, which he managed from inception to market acceptance. For a year, he was the product manager for analytics and strategy at a best-of-breed semi-startup company on demand-driven supply chain, the company later got acquired by Oracle Corporation. Ganguly has a PhD from MIT in hydrology from their civil and environmental engineering department, and research experience at their Sloan school of management in supply chain, information sciences, and data mining.

William Hendrix – Northwestern University
whendrix@northwestern.edu

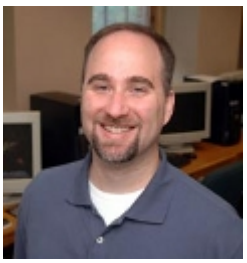
William Hendrix is a Postdoctoral Research Fellow in the Electrical Engineering and Computer Science department at Northwestern University. He earned his PhD in Computer Science from North Carolina State University in 2010, and his research interests include graph algorithms, high performance computing, and data mining.

Abdollah Homaifar – North Carolina Agricultural & Technical University
homaifar@ncat.edu

Abdollah Homaifar received his B.S. and M.S. degrees from the State University of New York at Stony Brook in 1979 and 1980, respectively, and his Ph.D. degree from the University of Alabama in 1987, all in electrical engineering. He is currently the Duke Energy Eminent professor in the Department of Electrical and Computer Engineering at North Carolina A&T State University (NCA&TSU). He is also the director of the Autonomous Control and Information Technology center at NCA&TSU, and Thrust Area Leader for Data Fusion, data mining and Distributed Architecture, NOAA ISET Center, at NCA&TSU. His research interests include machine learning, climate data processing, optimization, optimal control, flexible robotics, signal processing, soft computing and modeling. He is the author and co-author of over 200 articles in journals and conference proceedings, one book, and three chapters of books. He has participated in six short courses, serves as an associate editor of the Journal of Intelligent Automation and Soft Computing, and is a reviewer for IEEE Transactions on Fuzzy Systems, Man Machines & Cybernetics, and Neural Networks. He is a member of the IEEE Control Society, Sigma Xi, Tau Beta Pi, and Eta Kapa Nu.

Anuj Karpatne – University of Minnesota
anuj@cs.umn.edu

Anuj Karpatne is a PhD candidate in the Department of Computer Science and Engineering at the University of Minnesota, Twin Cities. As part of the Expeditions team, Anuj has been working with his advisor, Prof. Vipin Kumar, on analyzing remote sensing data for Earth science applications such as monitoring water dynamics, mapping forest cover, and detecting forest fires. Anuj's research has focused on developing data-driven approaches that can handle the presence of heterogeneity in the data, which is commonly experienced in a number of Earth system monitoring applications. Anuj's thesis has been generously funded by a University of Minnesota Doctoral Dissertation Fellowship and a University of Minnesota Informatics Institute Fellowship. Before joining the PhD program at the University of Minnesota, Anuj completed his undergraduate studies at the Indian Institute of Technology Delhi, with an Integrated B.Tech-M.Tech degree in Mathematics and Computing.



Joseph F. Knight – University of Minnesota
jknight@umn.edu

Joseph Knight is an Assistant Professor of Remote Sensing in the Department of Forest Resources at the University of Minnesota, Twin Cities. Dr. Knight studies how changing land use affects both natural resources and humans. He uses geospatial science methods such as remote sensing, image processing, and geographic information systems (GIS) in applications such as: identifying and characterizing natural and anthropogenic landscape change to assess impacts on natural resources, wetlands mapping and characterization, describing landscape-human interactions that lead to exposure to infectious diseases, and thematic accuracy assessment methods development.

Dr. Knight teaches three courses at the University of Minnesota: Remote Sensing of Natural Resources and Environment, Field Remote Sensing and Resource Survey, and Issues in the Environment. He holds a Ph.D. from North Carolina State University and previously worked as a Biologist with the United States Environmental Protection Agency. He is an author of numerous publications, including peer-reviewed

journal articles, book chapters, and technical reports. Dr. Knight is a recipient of the 2007 U.S. Environmental Protection Agency Science and Technology Achievement Award.

Wei-keng Liao– Northwestern University
wkliao@ece.northwestern.edu

Wei-keng Liao is a Research Professor in the Electrical Engineering and Computer Science Department. He received a Ph.D. in computer and information science from Syracuse University in 1999. Prof. Liao's research interests are in the area of high-performance computing, parallel I/O, parallel file systems, data mining, and data management for large-scale scientific applications.

Stefan Liess – University of Minnesota
liess@umn.edu

Stefan Liess is a Research Associate in Atmospheric Sciences at the University of Minnesota. He analyzes climate dynamics and climate change with observations, general circulation models, and high-resolution regional models. His specific research interests are intraseasonal variability and predictability on the order of a few weeks to multiple months, interactions of climate and vegetation including regional projections of future climate and vegetation pattern, and teleconnections in global and tropical climate. Previously, he worked as the responsible scientist and administrator for the SPARC (Stratospheric Processes and Their Role in Climate) Data Center and studied the impacts of tropopause characteristics on tropical convection. Stefan's research has been funded by the National Science Foundation, the National Aeronautics and Space Administration, the National Oceanic and Atmospheric Administration, and the Max-Planck Institute for Meteorology in Germany. He received his PhD in Atmospheric Sciences from Max-Planck Institute for Meteorology in 2002. In 1997 he earned a MS in Meteorology from the University of Hamburg in Germany.

Rachindra Mawalagedara – Northeastern University

Rachindra Mawalagedara is a Postdoctoral Research Associate in the Sustainability and Data Sciences Lab at Northeastern University. She earned her Ph.D. in Earth and Atmospheric Sciences from University of Nebraska-Lincoln in 2013. Her work is primarily focused on regional climate modeling and climate change with particular emphasis on studying the climatic impacts due to changes in land-use.

Varun Mithal – University of Minnesota
mitha011@umn.edu

Varun Mithal is a PhD candidate in Department of Computer Science at University of Minnesota- Twin Cities. As part of the NSF Expeditions team, he has worked on computational techniques for mining rare phenomena in spatio-temporal data with applications to monitoring land surface. This research has led to development of global-scale land surface change products from satellite data without manual supervision. Varuns' research has also been funded by a University of Minnesota Doctoral Dissertation Fellowship and a NASA grant. Before joining University of Minnesota, Varun completed his B.S. in Computer Science from Indian Institute of Technology, Kanpur in 2009.

Nagiza Samatova – North Carolina State University
samatova@csc.ncsu.edu

Dr. Nagiza F. Samatova is a Professor in Computer Science Department of North Carolina State University and a Senior Research Scientist in Computer Science and Mathematics Division of Oak Ridge National Laboratory. She received the B.S. degree in applied mathematics from Tashkent State University, Uzbekistan, in 1991 and her Ph.D. degree in mathematics from the Computing Center of Russian Academy of Sciences (CCAS), Moscow, in 1993. She also obtained an M.S. degree in Computer Science in 1998 from the University of Tennessee, Knoxville, USA. Dr. Samatova specializes in Graph Theory and Algorithms, High Performance Data Analytics, Climate Informatics, Bioinformatics, Systems Biology, Data Management, Scientific and High Performance Computing, and Machine Learning. She is the author of over 200 publications in peer-reviewed journals and conference proceedings. She co-edited the book, “Practical Graph Mining with R.” She is the recipient of various awards and honors, including the R&D 100 Award for ADIOS, R&D 100 Award for VIPAR, IEEE Distinguished Contribution to Public Service, Euro-Par Distinguished Paper Award, HPDC Best Paper Awards, NASA CIDU Best Paper Award. See her Dossier at http://www.csc.ncsu.edu/directories/faculty_info.php?id=2362.

Fred Semazzi – North Carolina State University
fred_semazzi@ncsu.edu

Dr. Semazzi has served in several senior positions of scientific organizations in the US, Europe, and Africa. He was a lecturer in the department of meteorology at the University of Nairobi, Kenya; Research Associate Scientist at NASA Goddard Space Flight Center, Greenbelt Maryland; US National Science Foundation (NSF) Climate Dynamics Program Associate Program Director, Washington DC; Founding Director of Climate Information & Prediction Services Program of the World Meteorological Organization at the United Nations, Geneva Switzerland; Senior Scientist, World Climate Research Program (WCRP) CLIVAR International Project Office, Southampton, England; Director of the Climate-PSM degree program & Professor at the North Carolina State University, Raleigh NC, USA. Dr. Semazzi has served in capacities of review editor & author for the Intergovernmental Panel on Climate Change (IPCC) climate change assessment. In March 2009 he received a certificate of special recognition from Dr. Rajendra K. Pachauri, Chairman of IPCC, for distinguished contribution resulting in the award of the Nobel Peace Prize for 2007 to the IPCC. This recognition was extended, ‘only to those who have contributed substantially to the work of the IPCC over the years since the inception of the organization’. Dr. Semazzi is a member of the Joint

Scientific Committee (JSC) for the World Climate Research Programme (WCRP; 2009-present). WCRP is sponsored by the World Meteorological Organization, the International Council for Science (ICSU) and the Intergovernmental Oceanographic Commission (IOC) of UNESCO. Dr. Semazzi has directed over 20 Masters and PhD degree theses. He has authored over sixty scientific publications & served as principal investigator and co-investigator on many grants, with total funding of more than $25 million.

Shashi Shekhar – University of Minnesota
shekhar@cs.umn.edu http://www.cs.umn.edu/~shekhar

Shashi Shekhar is a Mcknight Distinguished University Professor at the University of Minnesota (Computer Science faculty). For contributions to geographic information systems (GIS), spatial databases, and spatial data mining, he received the IEEE-CS Technical Achievement Award and was elected an IEEE Fellow as well as an AAAS Fellow. He was also named a key difference-maker for the field of GIS by the most popular GIS textbook. He has a distinguished academic record that includes 280+ refereed papers, a popular textbook on Spatial Databases (Prentice Hall, 2003) and an authoritative Encyclopedia of GIS (Springer, 2008). Shashi is serving as a member of the Computing Community Consortium Council (2012-15), a co-Editor-in-Chief of Geo-Informatica : An International Journal on Advances in Computer Sciences for GIS (Springer), a series editor for the Springer-Briefs on GIS, and as a member of the National Research Council (NRC) committee on Geo-targeted Disaster Alerts and Warning (2013). Earlier, he served on multiple NRC committees including Future Workforce for Geospatial Intelligence (2011), Mapping Sciences (2004-2009) and Priorities for GEOINT Research (2004-2005). He also served as a general or program co-chair for the Intl. Conference on Geographic Information Science (2012), the Intl. Symposium on Spatial and Temporal Databases (2011) and ACM Intl. Conf. on Geographic Information Systems (1996). He also served on the Board of Directors of University Consortium on GIS (2003-4), as well as the editorial boards of IEEE Transactions on Knowledge and Data Eng. and IEEE-CS Computer Sc. & Eng. Practice Board. In early 1990s, Shashi's research developed core technologies behind in-vehicle navigation devices as well as web-based routing services, which revolutionized outdoor navigation in urban environment in the last decade. His recent research results played a critical role in evacuation route planning for homeland security and received multiple recognitions including the CTS Partnership Award for significant impact on transportation. He pioneered the research area of spatial data mining via pattern families (e.g. collocation, mixed-drove co-occurrence, cascade), keynote speeches, survey papers and workshop organization. Shashi received a Ph.D. degree in Computer Science from the University of California (Berkeley, CA).

Peter Snyder – University of Minnesota
pksnyder@umn.edu

Peter Snyder is an atmospheric scientist studying an array of research problems related to atmospheric physics, land-atmosphere interactions, hydrometeorology, climate change, and the biosphere. His research areas span the Arctic, the tropics, and North America. Particular research problems include the role of the Great Plains Low Level Jet on moisture transport and precipitation events in the upper Midwest, the role of climate change on the frequency of extreme events, the influence of Arctic warming on the boreal forest and feedback mechanisms, monitoring and mitigation of urban heat islands around the world, and the climate response to boreal afforestation for carbon sequestration. He uses observations as well as global climate models, regional weather models, and land surface models to investigate these problems.

Michael Steinbach – University of Minnesota
steinbac@cs.umn.edu

Michael Steinbach earned his B.S. degree in Mathematics, a M.S. degree in Statistics, and M.S. and Ph.D. degrees in Computer Science from the University of Minnesota. He is currently a research associate in the Department of Computer Science and Engineering at the University of Minnesota, Twin Cities. Previously, he held a variety of software engineering, analysis, and design positions in industry at Silicon Biology, Racotek, and NCR. His research interests are in the area of data mining, bioinformatics, and statistics. He has authored over 30 research articles, and is a co-author of the data mining textbook, Introduction to Data Mining, published by Addison-Wesley. He is a member of the IEEE Computer Society and the ACM.

Poster Session

Tuesday, August 4, 2015
Listed in Alphabetical order of Last Name of Presenter

Title: *Generalized Additive Modeling of Nonstationary Extreme Events with Application to Precipitation Extremes in Southeastern United States*

Contributors: Agana, Gorji-Sefidmazgi, Homaifar (NCAT)

The variability and seasonal fluctuation in frequency and magnitude of extreme events strongly influence the social and natural environments throughout the world with consequent impacts on natural resources and the economy. As a result, there have been continuous efforts from both engineers and scientists to understand this variation in order to develop accurate models that are capable of explaining the variations. The extreme value theory (EVT) has provided the basis for the use of statistical models to study extreme events. Usually, most of these models assume stationarity of the observed process but this assumption is not usually fulfilled in many real life applications. In this research, we model extreme events in a nonstationary framework using a flexible data-driven approach called the Generalized Extreme Value Vector Generalized Additive Model (GEV-VGAM). The proposed GEV-VGAM model is developed to study the GEV model with covariates where the dependence structure is represented by smooth functions. The method is applied to 100-year monthly precipitation records in the Southeastern US. We characterize the nonstationarities in the precipitation events and the related climate variables by expressing the distribution parameters of the GEV distribution as smooth functions of explanatory variables such as time, the El Nino Southern Oscillation (ENSO) and the North Atlantic Oscillation Index (NAO). Results obtained using the GEV-VGAM model show an improvement over previous work where linear trend in covariates were used.

Presenter: **Saurabh Agrawal, University of Minnesota**
Title: *Introducing and Finding Tripoles: A New Climate Teleconnection Pattern*

Contributors: Agrawal, Liess, Kumar (UMN)

Climate teleconnections are the relationships between long distant regions. Dipoles are one such category of climate teleconnections that have been extensively studied in previous works. They are characterized by a strong negative correlation between the anomaly time series of a climate variable observed at the two distant regions, e.g. NAO, El-Nino oscillation. Here, we introduce a novel climate teleconnection pattern called tripole. A tripole involves three regions A, B, and C, such that the anomaly time series at region C is more strongly correlated with either addition or subtraction of anomaly time series observed at region A and region B, as compared to that with any of the anomaly time series at region A or B alone. We further categorize tripoles based on the correlations between each pair of regions. An interesting category of tripoles are negative tripoles in which the anomaly time series at all the three regions A, B, and C are negatively correlated with each other, and A+B has a stronger negative correlation with C compared to either of the pairs, A and C or B and C. We propose a shared nearest neighbor (SNN) graph-based approach to find such tripoles. We also discuss our findings on the monthly sea-level pressure dataset covering 1979-2011 for winter season (DJF), which include a prominent negative tripole that was found across the ends of SOI and northwestern Russia. The tripole indicates a much stronger negative correlation between the anomaly time series of northwestern Russia and the sum of the anomaly time series of the two ends of ENSO, as compared to the pairwise correlations with either of the ENSO ends.

Presenter: **Gonzalo Bello, North Carolina State University**
Title: *Response-Guided Community Detection: Application to Climate Index Discovery*

Contributors: Bello, Angus, Pedemane, Harlalka, Semazzi (NCSU), Kumar (UMN), Samatova (NCSU)

Discovering climate indices—time series that summarize spatiotemporal climate patterns—is a key task in the climate science domain. In this work, we approach this task as a problem of response-guided community detection; that is, identifying communities in a graph associated with a response variable of interest. To this end, we propose a general strategy for response-guided community detection that explicitly incorporates information of the response variable during the community detection process, and introduce a graph representation of spatiotemporal data that leverages information from multiple variables.

We apply our proposed methodology to the discovery of climate indices associated with seasonal rainfall variability. Our results suggest that our methodology is able to capture the underlying patterns known to be associated with the response variable of interest and to improve its predictability compared to existing methodologies for data-driven climate index discovery and official forecasts.

Presenter: **Soumyadeep Chatterjee, University of Minnesota**
Title: *Understanding dominant factors for precipitation in great lakes region*

Contributors: Chatterjee, Liess, Banerjee (UMN)

Statistical modeling of local precipitation involves understanding local, regional and global factors that carry information about precipitation variability in a region. Modern machine learning methods for feature selection can potentially be explored for identifying statistically significant features from a pool of potential predictors of precipitation. In this work, we consider structured regression methods with sparse and group-sparse regularizers, which simultaneously perform feature selection and regression. In particular, we consider group structures that are learnt from analyzing correlations among predictors, and encode such structures in the regularizer function. We experiment on seasonal precipitation over Great Lakes Region, which is further separated into three sub-regions, in order to identify the dominant factors for each season. Our results show that structured regression offers comparable accuracy to ordinary least squares, while controlling model complexity with parsimony of selected features. Moreover, significant features, identified using randomization tests, offer a hypothesis over possible mechanisms for seasonal regional precipitation, which can be further analyzed by domain scientists for physical validity. Such feature selection methods can thus be useful for hypothesis generation from pools of possible predictors, which can be lent for further validation and possible novel insights.

Presenter: **Mandar Chaudhary, North Carolina State University**

Title: *Toward Discovery of Key Factors Causally Affecting Climate Extremes: Application to African Sahel Rainfall Anomaly Forecasts*

Contributors: Chaudhary, Gonzalez, Bello, Angus, Muralidharan, Hsu, Semazzi (NCSU), Kumar (UMN), Samatova (NCSU)

Understanding core mechanisms causally affecting climate extremes is the Holy Grail of Climate Informatics. The challenge is a combinatorial explosion of the space of putative climate factors that may drive the dynamics and variability of climate extremes. Traditional methods such as those based on EOF analysis, while effective in identifying primary modes as linear combinations of such factors, often fail in discovery of key factors that are causally associated with the climate system's response of interest such as high and low rainfall anomalies (i.e., floods and droughts).

Starting with a set of 35 NOAA climate indices defined over the 6-month period preceding the Sahel July-August-September rainfall season, the proposed method probabilistically explores the combinatorial space defined over these 210 (i.e., 35x6) climate factors to construct a causal network that likely drives the variability of Sahel rainfall extremes. Using subnetworks of this causal network as predictors allows for improving seasonal forecast of rainfall anomalies in the African Sahel region up to 23% compared to state-of-the-art feature selection methods, 19% compared to feature extraction method, and by 29% compared to 10-year climatology forecasts. Specifically it allows for one-month lead time forecasts of seasonal floods with 63% precision and 73% recall and of seasonal droughts with 65% precision and 68% recall upon leave-one-out cross-validation. Furthermore, the derived subnetworks are shown to be not only statistically robust or stable but also physically relevant climatic factors.

Presenter: **Xi C. Chen, University of Minnesota**
Title: *A spatial-temporal clustering algorithm for global in-land surface water monitoring*

Contributors: Chen, Yao, Shi, Khandelwal, Kumar, Faghmous (UMN)

Lack of a global system that can monitor and record all changes of water bodies limits our understanding of the hydrologic cycle, hinders water resource management and compounds water related risks. Remote sensing provides an opportunity to monitor changes of the water surface area at an affordable cost, but the unique properties of remote sensing data, such as noise, the large number of missing values, data heterogeneity and lack of training labels, make this task challenging. In this work, we propose a spatial-temporal clustering algorithm to extract water bodies using remote sensing data. This approach overcomes the above challenges by fully utilizing both spatial autocorrelation and the temporal information. Specifically, we use the proposed method for monitoring the surface area of lakes by clustering data (of each time step) into two groups: water cluster and land cluster. Using a combination of independent validation data and physics-guided labeling, we compare the proposed method with two cluster algorithms (i.e., K-MEANS and EM) and one image segmentation algorithm (i.e., normal-cut). We show that the proposed method is better than other methods in tracking the changes of water/land clusters.

Presenter: **Lindsey Dietz, University of Minnesota**
Title: *Spatial-Temporal Hypothesis Testing in Model Residuals*

Contributors: Dietz, Chatterjee (UMN)

Spatiotemporal correlation in climate-related data must be prioritized in modeling efforts. However, before undertaking complicated spatiotemporal modeling, it is advantageous to confirm the benefits of the effort. One possibly useful method, the Space-Time Index (STI), was introduced in Griffith [*Dynamic Spatial Models*, 258-287 (1981)] to detect the presence of spatiotemporal correlation for a space-time process in vector autoregressive (VAR) form.

To assess the usefulness of the test, we simulated space-time series with various spatial and temporal correlations. Simulations indicated the power of the test was low, especially as the number of spatial neighbors increased. Next, we applied STI to logit-normal model residuals produced in our previous work. The original model assessed the probability of exceeding certain daily precipitation thresholds in summer monsoon seasons from 1973-2013. Results indicated most rainfall thresholds did not show evidence of spatiotemporal correlation in the model residuals, yet were influenced by the choices of spatial neighbors. However, higher rainfall levels failed to detect spatiotemporal correlation in all neighborhoods. Based on this test, a more complicated spatiotemporal model structure appears unnecessary in this application.

Although STI provides a useful first effort in identifying spatiotemporal correlation in residuals, the testing is currently subject to assumptions of normality, specification of spatial neighborhoods, and low power. Future work will add robustness to the test making it a more useful tool for detection of spatiotemporal correlation.

Presenter: **Mohammad Gorji Sefidmazgi, North Carolina A&T State University**
Title: *A genetic algorithm based approach for analyzing abrupt climate change*

Contributors: Homaifar (NCAT), Liess (UMN)

The climate has changed gradually in response to both natural and human-induced processes. However, it is known that climate may have abrupt change, i.e. a large shift may happen in climate that persists for years or decades. Many studies have analyzed climate time series in the stationary framework, which is not valid considering various internal dynamics and externally forcings. Thus, statistical techniques based on stationary assumptions should be modified to reveal the characteristics of the abrupt climate change. The modeling of non-stationary time series can be represented as a non-convex constrained optimization. This optimization should be solved to find the change points between some clusters, such that the model of each cluster is stationary. Several approaches exist for modeling of non-stationary climate time series. However, they suffer from restrictive statistical assumptions, inefficiency in longer time series representation, and limited applicability in correlated time series. We have developed a new method for analyzing time series in non-stationary framework based on the Genetic algorithm (GA). The GA is an optimization method that resembles the natural selection for solving complex optimization problems. The Bayesian Information Criterion is used to find optimum number of change points between the clusters. The proposed modeling approach can be combined with maximum likelihood and least-mean-square techniques. Thus, it is possible to model the non-stationary time series with various statistical models such as regression, generalized linear model and statistical distributions. This technique is utilized for linear trend analysis, and building predictive modeling for climate variables. The simulation results show the accuracy of the method in carious hydro-climatological applications.

Presenter: **Dianwei Han, Northwestern University**
Title: *Estimating Uncertainty in Precipitation Extremes*

Contributors: Hendrix (NWU), Kumar, Das (NEU), Daga, Liu, Han (NWU), Ganguly (NEU), Choudhary (NWU)

Precipitation extremes, such as hurricanes, floods, and droughts, cost millions of dollars and kill thousands annually. Consequently, predicting and understanding the behavior of these extreme phenomena are of acute interest to the scientific community, but global circulation models (GCMs), which are some of our basic tools in understanding and predicting future climate, exhibit key uncertainties in predicting these extreme events. In this ongoing work, we attempt to analyze and quantify some of the sources of uncertainty inherent in predicting precipitation extremes.

In this work, we estimate the 100-year return level; i.e., the intensity of a storm that occurs once in 100 years. In particular, we focus on estimating the uncertainty due to statistical parameter estimation, which is a natural consequence of estimating the extreme behavior of a distribution. By combining statistically-rigorous Extreme Value Theory with bootstrapping, we can produce a set of estimates for the 100-year precipitation return level and quantify the uncertainty in our prediction by computing the 95% confidence interval of this distribution.

One of the main challenges of this approach is its computational cost: this analysis would take an estimated 30 years to run on a single processor using the original MATLAB-based code. In order to run this analysis, we've implemented our solution in MPI C++ so that we can scale our problem to hundreds or thousands of processes. We hope that this work can be extended to measure uncertainty due to model parameters or multi-model ensembles.

Presenter: **Megan Heyman, University of Minnesota**
Title: *The WiSE Bootstrap for Climate Model Evaluation*

Contributors: Chatterjee (UMN), Braverman, Gunson (NASA-JPL), Cressie (Univ. of Wollongong)

At a given grid cell location, climate models, such as the MIROC5 or IPSL, produce time series which claim to reflect the actual average climate conditions within that location. These output series may be considered as a single observation for phenomena like precipitation, temperature, or specific humidity at a given time. Although we only have a single observation for a specified location, time, and phenomena, there is error associated with this observation. Understanding this error allows for model-to-model and model-to-observed climate comparisons, but estimating the error structure is difficult. A novel re-sampling technique called the wild scale-enhanced (WiSE) bootstrap is proposed to aid in error estimation and signal-to-signal comparison. This method is flexible, producing consistent estimators of model parameters and variance components.

The partial linear model formulation for climate time series and WiSE bootstrap methodology is discussed. For a single grid cell over Brazil, the atmospheric infrared sounder (AIRS) climate observations are compared to a single run of each of the MIROC5 and IPSL climate models. Results of a formal hypothesis test for signal equality are presented. This methodology may be extended to any collection of grid cells and number of climate model runs.

Presenter: **Zhe Jiang, University of Minnesota**
Title: *Spatial Classification Techniques for Land Cover Mapping*

Contributors: Jiang, Zhou, Shekhar, Knight, Corcoran

Land cover mapping is a societally important task with many applications such as understanding climate change, natural resource management and disaster management. The rich remote sensing imagery collected from satellite and aerial planes brings a unique opportunity for mapping land cover at a large scale. However, classifying remote sensing images into land cover types is a challenging task due to the existence of spatial autocorrelation and spatial heterogeneity. Many existing classification techniques assume that learning samples follow an identical and independent distribution, and thus produce land cover maps with salt-and-pepper noise and class confusion. To address these limitations, we develop a novel spatial classification technique called spatial decision trees, which incorporate spatial autocorrelation information in decision trees so that a sample's tree traversal direction is based on both local and focal (neighborhood) information. We also investigate a spatial-slicing-and-ensemble approach to address the challenges of spatial heterogeneity. Evaluation on real world wet land mapping datasets show that proposed spatial decision trees outperform traditional decision trees in both classification accuracy and salt-and-pepper noise level, and proposed spatial ensemble approach reduces class confusion when the same spectral features correspond to distinct classes in different regions. To improve the scalability of proposed algorithms, we conduct computational refinement. Theoretical analysis and experimental evaluation show that the refined algorithm is correct and significantly reduces computational cost.

Presenter: **Anuj Karpatne, University of Minnesota**
Title: *Global Monitoring of Inland Water Surface Area Dynamics using Remote Sensing Data*

Contributors: Khandelwal, Ding, Leuenberger, McCaleb, Cai, Chen, Mithal, Faghmous, Kumar (UMN)

Freshwater, which is only available in inland water bodies such as lakes, reservoirs, and rivers, is increasingly becoming scarce across the world and this scarcity is posing a global threat not only to human sustainability but also to the Earth's ecosystem. As a result, managing inland water has become one of the major 21st century challenges for the U.S. and the world. To address this challenge, we present a global monitoring system that provides timely and accurate information about the available surface water stocks and their dynamics across the world at regular intervals of time using remote sensing data. This system makes use of novel machine learning algorithms for handling a variety of issues in using remote sensing data, e.g. presence of heterogeneity in the data across space and time, and high degree of noise and missing values due to cloud and aerosol obstructions. In this poster, we will demonstrate some of the preliminary capabilities of our water monitoring system that is able to capture a variety of surface water dynamics, e.g. construction of new dams and reservoirs across the world, on-going droughts in Brazil and California, and changes in river morphology such as river meandering and delta erosion. For more information, please visit: http://z.umn.edu/watermrv.

Presenter: **Phu Nguyen, Center for Hydrometeorology and Remote Sensing, University of California, Irvine**
Title: *CHRS Connect - a global extreme precipitation event database using Object-ORIENTED approach*

Contributors: Thorstensen, Liu, Sellars, Ashouri, Braithwaite, Hsu, Gao, Sorooshian (CHRS, UC Irvine)

Precipitation is a key variable in hydrological processes and varies within time and space. Extreme precipitation events may cause natural disasters and these events vary in many different regions of the world. This poster presents the recently developed CHRS CONNECT (Center for Hydrometeorology & Remote Sensing CONNected precipitation objECT) system – a global extreme precipitation event database derived from CHRS's satellite precipitation data products using object-oriented algorithm. The datasets include the PERSIANN (Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks) and PERSIANN-CDR (Climate Data Record). The extreme precipitation events with their attributes are stored in a searchable database. The user-friendly interface (connect.eng.uci.edu) allows the user to query the events of interest defined by certain criteria such as spatiotemporal domain, maximum intensity, minimum duration and climatology index. The system also provides various functionalities for visualization including total precipitation, event tracking, and animation of event evolution. The system is designed to be a useful tool for climatology research and water resources management, especially for extreme precipitation events caused by atmospheric rivers and monsoon systems.

Presenter: **Vidyashankar Sivakumar, University of Minnesota**
Title: *Predicting Indian Summer Monsoon Rainfall (ISMR) using a mixture of sparse regression models*

Contributors: Sivakumar (UMN), Saha, Mitra (IIT Kgp), Banerjee (UMN)

We consider the problem of predicting total Indian summer monsoon rainfall (ISMR). A popular approach in prior literature (Rajeevan et al. 2006; DelSole and Shukla 2002) has been to fit a regression model with the precipitation as predictand and various climatological indices and parameters as predictors. The predictor climatological indices and parameters are detected through an analysis of their linear correlations with the Indian monsoon precipitation. Due to limited success of such prior work based on a fixed regression model, in this work we investigate ISMR prediction based on the hypothesis that Indian monsoon operates in a few different regimes, where different predictors become relevant and influential. We model such a multi-regime setting as a finite mixture of regression (FMR) model (McLachlan and Peel 2000; Stadler et al. 2010), with a sparse regression model for each regime of operation. The parameters of the model are determined using the Expectation Maximization (EM) algorithm. The prediction procedure consists of identifying the regime of operation and then applying the corresponding regression model. The FMR model seems to improve overall prediction accuracy compared to a single fixed regression model. We also investigate the relationship between the different regimes of operation and physical climate events like drought vs flood and El Nino vs La Nina years.

Presenter: **Yumeng Tao, University of California, Irvine**
Title: *Precipitation Estimation from Remotely Sensed Data Using Deep Neural Networks*

Contributors: Gao, Hsu, Sorooshian, Ihler (UC Irvine)

This research develops a precipitation estimation system from remote sensed data using state-of-the-art machine learning algorithms. Compared to ground-based precipitation measurements, satellite-based precipitation estimation products have advantages of temporal resolution and spatial coverage. Also, the massive amount of satellite data contains various measures related to precipitation formation and development. On the other hand, deep learning algorithms were newly developed in the area of machine learning, which was a breakthrough to deal with large and complex dataset, especially to image data.

Here, we attempts to engage deep learning techniques to provide hourly precipitation estimation from satellite information, such as long wave infrared data. The brightness temperature data from infrared data is considered to contain cloud information. Radar stage IV dataset is used as ground measurement for parameter calibration. Denoising stacked auto-encoders (DSAE) is applied here to build a 4-layer neural network with 1000 hidden nodes for each hidden layer. DSAE involves two major steps: (1) greedily pre-training each layer as an auto-encoder using the outputs of previous trained hidden layer output starting from visible layer to initialize parameters; (2) fine-tuning the whole deep neural network with supervised criteria.

The results are compared with satellite precipitation product PERSIANN-CCS (Precipitation Estimation from Remotely Sensed Imagery using an Artificial Neural Network Cloud Classification System). Significant improvements are achieved in both rain/no-rain (R/NR) detection and precipitation rate quantification: the results make 33% and 43% corrections on false-alarm pixels and 98% and 78% bias reductions in precipitation rates over the validation summer and winter seasons, respectively.

Attendee Contact Information

Listed in Alphabetical order by Last Name

Norbert Agana
North Carolina Agricultural & Technical State University
naagana@aggies.ncat.edu

Ankit Agrawal
Northwestern University
ankitag@eecs.northwestern.edu

Saurabh Agrawal
University of Minnesota
agraw066@umn.edu

Reid Anderson
University of Minnesota
and02911@umn.edu

Robert Andrade
University of Minnesota
andra065@umn.edu

Michael Angus
North Carolina State University
mpangus@ncsu.edu

Gowtham Atluri
University of Minnesota
gowtham@cs.umn.edu

Arindam Banerjee
University of Minnesota
banerjee@cs.umn.edu

Gonzalo Bello
North Carolina State University
gabellol@ncsu.edu

Tim Bodin
Cargill
tim_bodin@cargill.com

Kate Brauman
University of Minnesota Institute on the Environment
kbrauman@umn.edu
z.umn.edu/brauman

Kenson Cai
University of Minnesota
caix32@umn.edu

Juan Carlos Castilla-Rubio
Planetary Skin Institute
www.planetaryskin.org

Varun Chandola
University at Buffalo
www.cse.buffalo.edu/~chandola

Soumyadeep Chatterjee
University of Minnesota
soumyachat@gmail.com

Ansu Chatterjee
University of Minnesota
chatterjee@stat.umn.edu

Mandar Chaudhary
North Carolina State University
mschaudh@ncsu.edu

Xi Chen
University of Minnesota
chenx645@umn.edu

Alok Choudhary
Northwestern University
choudhar@eecs.northwestern.edu

Conner Cowling
University of Minnesota
cowl0058@umn.edu

Greyson Dehn
University of Minnesota
dehnx058@umn.edu

Timothy DelSole
George Mason University
tdelsole@gmu.edu

Lindsey Dietz
University of Minnesota
diet0146@umn.edu

Yizheng Ding
University of Minnesota
dingx292@umn.edu

Peder Engstrom
University of Minnesota Institute on the Environment
engs0074@umn.edu

Auroop Ganguly
Northeastern University
a.ganguly@neu.edu

Sangram Ganguly
NASA Ames & Bay Area Environmental Institute
sangram.ganguly@nasa.gov

James (Jamie) Gerber
University of Minnesota Institute on the Environment
jsgerber@umn.edu

Mohammad Gorji-Sefidmazgi
North Carolina Agricultural & Technical State University
mgorjise@aggies.ncat.edu
acitcenter.ncat.edu/gorji.html

Dianwei Han
Northwestern University
dianweih@eecs.northwestern.edu

Megan Heyman
University of Minnesota
heyma029@umn.edu

Abdollah Homaifar
North Carolina Agricultural & Technical State University
homaifar@ncat.edu

Shafiqul Islam
Tufts University
shafiqul.islam@tufts.edu
waterdiplomacy.org

Zhe Jiang
University of Minnesota
zhe@cs.umn.edu
www-users.cs.umn.edu/~zhe/

Daniel Jiménez R.
International Center for Tropical Agriculture
d.jimenez@cgiar.org

Anuj Karpatne
University of Minnesota
karpa009@umn.edu

Woodrow Keifenheim
University of Minnesota
Kefi0006@umn.edu

Ankush Khandelwal
University of Minnesota
khand035@umn.edu

Joseph F. Knight
University of Minnesota
jknight@umn.edu

Christopher Koshiol
University of Minnesota
koshi012@umn.edu

Vipin Kumar
University of Minnesota
kumar@cs.umn.edu
www.cs.umn.edu/~kumar

Robert Leuenberger
University of Minnesota
leuen002@umn.edu

Stefan Liess
University of Minnesota
liess@umn.edu

Mengqian Lu
Columbia University
ml3074@columbia.edu

Subhabrata Majumdar
University of Minnesota
majum010@umn.edu

Eric McCaleb
University of Minnesota
mcca0782@umn.edu

Kyran D. Mish
Sandia National Laboratories
kdmish@sandia.gov

Varun Mithal
University of Minnesota
mitha011@umn.edu

Claire Monteleoni
George Washington University

Guruprasad Nayak
University of Minnesota
nayak013@umn.edu

Dr. Phu Nguyen
University of California, Irvine
ndphu@uci.edu

Sai Ravela
Massachusetts Institute of Technology
ravela@mit.edu
essg.mit.edu

Deepak Ray
University of Minnesota Institute on the Environment
dray@umn.edu

Nagiza Samatova
North Carolina State University
samatova@csc.ncsu.edu

Fred Semazzi
North Carolina State University
semazzi@ncsu.edu

John (Jack) Sharp
The University of Texas
jmsharp@jsg.utexas.edu

Shashi Shekhar
University of Minnesota
shekhar@cs.umn.edu

Vidyashankar Sivakumar
University of Minnesota
sivak017@umn.edu

Brian Smoliak
The Climate Corporation
brian.smoliak@climate.com
Padhraic Smyth
University of California, Irvine
smyth@ics.uci.edu

Peter Snyder
University of Minnesota
pksnyder@umn.edu

Michael Steinbach
University of Minnesota
steinbac@cs.umn.edu

Karsten Steinhaeuser
Progeny Systems Corp.
ksteinha@umn.edu

Xun Tang
University of Minnesota
tangx456@umn.edu

Yumeng Tao
University of California, Irvine
yumengt@uci.edu

Raju Vatsavai
North Carolina State University
rvatsavai@ncsu.edu
http://people.engr.ncsu.edu/rrvatsav/

Robert Warmka
University of Minnesota
warm0086@umn.edu

Brian Zhang
University of Minnesota
zhan4136@umn.edu

Wireless access during the workshop:

There are more than 3,700 wireless access points deployed on the system at the University of Minnesota, and room 3-180 Keller Hall has a strong working wireless signal. The network runs on 802.11n technology and improves rogue Access Point (AP) and intrusion detection, central management operations, guest access services and more.

Guests to campus may use the "UofM Guest" network for free. You can use any existing e-mail address to log in to this network. The U of M Guest Wireless Login Page should display immediately when you open your browser on your laptop or other device.