UNIVERSITY OF MINNESOTA
Driven to Discover℠

Third Workshop on

Understanding Climate Change from Data

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

August 15-16, 2013

Workshop Venue:
Northwestern University
2145 Sheridan Road
Room L361 Technological Institute
Evanston, IL 60208

www.climatechange.cs.umn.edu

(This page intentionally left blank)

Understanding Climate Change from Data

Northwestern University – Evanston, IL
Room L361 Technological Institute

August 15-16, 2013

Table of Contents

(This page left intentionally blank)

Agenda: Third Workshop on Understanding Climate Change from Data

Northwestern University, Evanston, IL — Room L361 Technological Institute

	Thursday, August 15, 2013	pg
8:00	Registration & Breakfast	
8:30	Vipin Kumar, University of Minnesota *Introduction to the NSF Expeditions in Computing on Understanding Climate Change: A Data Driven Approach*	9
	Session 1 *Chair: Auroop Ganguly*	11
9:15	Kevin Trenberth, National Center for Atmospheric Research *Challenges of a sustained climate observing system*	
9:40	Ben Santer, Lawrence Livermore National Laboratory *The Search for a Human Fingerprint in the Changing Thermal Structure of the Atmosphere*	
10:05	Stefan Liess, University of Minnesota *Detecting and Describing Teleconnections*	
10:25	Break	
	Session 2 *Chair: Shashi Shekhar*	12
10:50	Amy Braverman, Jet Propulsion Laboratory, California Institute of Technology *Likelihood-based Climate Model Evaluation*	
11:15	Snigdhansu Chatterjee, University of Minnesota *Quantiles and data-depth: the next generation*	
11:35	Richard L. Smith, University of North Carolina and SAMSI *Influence of Climate Change on Extreme Weather Events*	
12:00	Lunch Break	
	Session 3 *Chair: Arindam Banerjee*	13
13:30	Barbara Ransom, National Science Foundation *EarthCube*	
13:55	Ian Foster, Argonne National Laboratory *FACE-IT: A Framework for the Management and Analysis of Environmental Data*	
14:20	Alok Choudhary, Northwestern University *High Performance Computing and Data Analytics*	

14:40	Shashi Shekhar, University of Minnesota *Discovering Persistent Change Windows in Spatiotemporal Datasets*	
15:00	Break	
	Session 4 *Chair: Abdollah Homaifar*	15
15:30	Doug Nychka, National Center for Atmospheric Research *Regional climate informatics: A statistical perspective*	
15:55	Arindam Banerjee, University of Minnesota *Probabilistic Graphical Models for Climate Data Analysis*	
16:15	Fred Semazzi, North Carolina State University *Outreach Opportunities for Expedition's Data Driven Discoveries*	
17:00	Discussions	
17:30	Poster Session & Dinner, Cohen Commons	37

Friday, August 16, 2013

8:30	Breakfast	
	Session 5 *Chair: Stefan Liess*	16
9:00	Greg Holland, University Corporation for Atmospheric Research *Assessing uncertainty in regional scale climate variability and change*	
9:25	James Faghmous, University of Minnesota *Sea surface temperature (SST) spatial patterns and their impact on Atlantic tropical cyclone activity*	
9:45	Abdollah Homaifar, North Carolina Agricultural & Technical University *Tropical Cyclone Tracking and Intensity Estimation, Image Search and Climate Trend Quantification*	
10:05	Auroop Ganguly, Northeastern University *Physics-guided statistical approach to uncertainty quantification from climate model ensembles: Expedition funded doctoral research at the SDS Lab in Northeastern University*	

10:25	Coffee Break	
	Session 6 *Chair: Fred Semazzi*	18
10:50	Soroosh Sorooshian, University of California- Irvine *Understanding Climate Change From Data - Perspectives from Hydroclimate Modeling and Data Assimilation*	
11:15	Jagdish Shukla, George Mason University *Fidelity and Predictability of Models for Weather and Climate Prediction*	
11:40	Stefan Liess, University of Minnesota *Statistical Analysis for Prediction of Monsoon Variability*	
12:00	Lunch Break	
13:30	Panel Discussion: Big Data for Climate Science Leonard P. Hirsch, Smithsonian Institution Praveen Kumar, University of Illinois at Urbana-Champaign Doug Nychka, National Center for Atmospheric Research Nikunj C. Oza , NASA Ames Research Center Ben Santer, Lawrence Livermore National Laboratory Jagdish Shukla, George Mason University Kevin Trenberth, National Center for Atmospheric Research Moderator: Auroop Ganguly, Northeastern University	19
16:00	Closing Remarks	
17:30	Dinner	

(This page intentionally left blank)

Vipin Kumar – University of Minnesota

Title: *Introduction to the NSF Expeditions in Computing on 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. These innovative approaches help provide an improved understanding of the complex nature of the Earth system and the mechanisms contributing to the adverse consequences of climate change, such as increased frequency and intensity of hurricanes, precipitation regime shifts, and the propensity for extreme weather events that result in environmental disasters. 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 year.

(This page intentionally left blank)

Presentations, Session 1, August 15

Kevin E. Trenberth – National Center for Atmospheric Research
Title: *Challenges of a sustained climate observing system*

Observations of planet Earth and especially all climate system components and forcings are increasingly needed for planning and informed decision making related to climate services in the broadest sense. Although significant progress has been made, much more remains to be done before a fully functional and dependable climate observing system exists. Observations are needed on spatial scales from local to global, and all time scales, especially to understand and document changes in extreme events. Climate change caused by human activities adds a new dimension and a vital imperative: to acquire climate observations of sufficient quality and coverage, and analyze them into products for multiple purposes to inform decisions for mitigation, adaptation, assessing vulnerability and impacts, possible geo-engineering, and predicting climate variability and change and their consequences. A major challenge is to adequately deal with the continually changing observing system, especially from satellites and other remote sensing platforms such as in the ocean, in order to provide a continuous climate record. GPS-Radio Occultation measurements provide a benchmark that help enormously with these other changes and for anchoring analyses of variability and change. Coordinated international actions are essential to provide for information-based actions and decisions related to climate variability and change. A Climate Data Guide http://climatedataguide.ucar.edu) developed at NCAR provides a community-authored guide to the world of climate data sets. The Web site provides concise and reliable information on data strengths and limitations, enabling users to make more informed selections of data sets for their own research and to better understand the data and analysis methods that are encountered in numerous contexts.

Ben Santer – Lawrence Livermore National Laboratory
Title: *The Search for a Human Fingerprint in the Changing Thermal Structure of the Atmosphere*

Satellite temperature measurements reveal multi-decadal tropospheric warming and stratospheric cooling, punctuated by short-term volcanic signals of reverse sign. Similar long- and short-term temperature signals occur in model simulations driven by human-caused changes in atmospheric composition and natural variations in volcanic aerosols. Previous research attempted to discriminate a human-caused latitude/altitude pattern of atmospheric temperature change ("fingerprint") from the background noise of internal variability. We present the first evidence that a human fingerprint can also be identified relative to the larger "total" noise arising from internal variability, solar irradiance changes, and volcanic forcing.

Stefan Liess – University of Minnesota
Title: *Detecting and Describing Teleconnections*

A graph-based approach to detect atmospheric teleconnections can identify multiple atmospheric teleconnections simultaneously without being bound by orthogonality (such as EOF analysis). We employ a Shared Reciprocal Nearest Neighbor (SRNN) method that not only finds common teleconnections such as ENSO, NAO, PNA, WP, and SAM, but also reveals smaller, more regional pattern such as a teleconnection between the Tasman Sea and the Southern Ocean. A detailed analysis of this teleconnection describes its relationship to ENSO and SAM during all four seasons, as well as the impact on regional climate over Australia and New Zealand.

Presentations, Session 2, August 15

**Amy Braverman – ** Jet Propulsion Laboratory, California Institute of Technology
(with Noel Cressie, University of Wollongong)
Title: *Likelihood-based Climate Model Evaluation*

Climate models can be evaluated by comparing their output to observations. Remote sensing data provide new possibilities for such comparisons because they are spatially and temporally dense enough to go beyond simple moments and estimate distributions. We evaluate climate model fidelity to observations by the likelihood that a summary statistic computed from an observational time series arises from a sampling distribution of that same statistic calculated from a given climate model's time series. We demonstrate using models from the CMIP5 archive and observations from NASA's Atmospheric Infrared Sounder mission.

**Snigdhansu Chatterjee – ** University of Minnesota
Title: *Quantiles and data-depth: the next generation*

Quantiles can be useful for studying the properties of a random variate in the interior of its support, or near the boundary. Also, they may be used to handle robustness and sensitivity concerns when the assumptions underlying a statistical model may not hold, or when there are possible outliers and influential observations in the data. A challenge for current statistical research is how to extend the notion of quantiles to multi-dimensional and functional data, in a way such that (i) the useful properties of quantiles are retained irrespective of the dimension of the random variate under study, (ii) they are computationally feasible, (iii) statistical inference relating to population characteristics based on sample quantiles is viable. In this talk we propose a multivariate quantile function that has several properties that make it useful both computationally and for statistical inferential purposes. Multivariate quantiles also inform about the shape of the data-cloud. We discuss how data-depth functions, derived from multivariate quantiles, may be useful for studying features of high dimensional data. We present several applications of the proposed multivariate quantiles and data-depth functions in the context of climate data analysis.

**Richard L. Smith – ** University of North Carolina and SAMSI
Title: *Influence of Climate Change on Extreme Weather Events.*

The increasing frequency of extreme weather events raises the question of to what extent such events can be attributed to human causes. Within the climate literature, an approach has been developed based on a quantity known as the fraction of attributable risk, or FAR. The essence of this approach is to estimate the probability of the extreme event of interest from parallel runs of climate models under either anthropogenic or natural conditions; the two probabilities are then combined to produce the FAR. However, a number of existing approaches either make questionable assumptions about estimating extreme event probabilites (e.g. inappropriate assumption of the normal distribution) or ignore the differences between climate models and observational data. Here, we propose an approach based on extreme value theory, incorporated into a hierarchical model to account for differences among climate models. A related technique, based on the same modeling approach, leads to quantitative estimates of how the probability of an extreme event will change under future projected climate change. We illustrate the method with examples related to the European heatwave of 2003, the Russian heatwave of 2010, and the Texas/Oklahoma heatwave and drought of 2011. This is joint work with Michael Wehner (Lawrence Berkeley Lab).

Presentations, Session 3, August 15

Barbara Ransom – National Science Foundation
Title: *EarthCube*

EarthCube is a bold new NSF activity to create a data and geoscience knowledge management system for the 21st Century. This is a joint effort of the Directorate of Geosciences (GEO) and the Division of Advanced Cyberinfrastructure (ACI) in the Computer Information Science and Engineering Directorate at the National Science Foundation. Its goals are to develop a framework over the next decade to, among other things, understand and predict the Earth system from the sun to the center of the Earth by teaming geoscientists with computer/cyber scientists to accelerating the development of the tools, technologies, and theories needed to dramatically accelerate our ability to link disparate data types and structures; develop new community-driven, yet state-of-the-art semantic and automated meta-data generating approaches; create new 3D and/or 4D visualization approaches and methodologies; and devise ways to bring heterogeneous, distributed, dark data to light. In this regard, NSF is facilitating a community dialog and providing funding opportunities to define the research and educational framework for this new effort and with the expectation that it will transform the conduct of research and education for the geo- and environmental sciences. These ambitious goals required close interaction between the end-users and those who will create the innovations, theories, and software to make sure the resulting codes, approaches, and cyberinfrastructure are responsive to changing user needs in an ever changing technological landscape. The motivation for EarthCube is to create a framework that enables us to understand and predict the Earth System and increase the productive and capability of researchers and educators to undertake the science required to tackle the pressing problems facing Earth and society.

Ian Foster – Argonne National Laboratory
Title: *FACE-IT: A Framework for the Management and Analysis of Environmental Data*

Issues relating to climate change and food security require an understanding of the interaction between the natural world and human society over long time scales. Understanding climate change, its impacts on the natural world and society, and the tradeoffs inherent in societal responses demands an unprecedented degree of cooperation across academic fields. New data sources on expected future climate, soil characteristics, economic activity, historical weather, population, and land cover, provide a potential basis for this cooperation. New methods are needed for sharing within and across communities not only data but the software used to generate, synthesize, and analyze it. Progress on these research challenges is hindered by the extreme difficulties that researchers, collaborators, and the community experiences when they collaborate around data. Multiplicity of data formats; inadequate computational tools; difficulty in sharing data and programs, lack of incentives for pro-social behavior and large data volumes are among the technology barriers. The FACE-IT project at the University of Chicago, NASA, and University of Florida employs an integrated approach to cyberinfrastructure to advance the characterization of vulnerabilities, impacts, mitigation, and adaptation to climate change in human and environmental systems. Leveraging existing research cyberinfrastructure, the project is creating a full-featured FACE-IT Platform prototype with new capabilities for ingesting, organizing, managing, analyzing and using large quantities of diverse data. The project team collaborates with two distinct interdisciplinary communities to create community specific FACE-IT Instances to both advance their research and enable at-scale evaluation of the utility of the FACE-IT approach. In this talk I will introduce the FACE-IT system and discuss early applications.

Alok Choudhary – Northwestern University
Title: *High Performance Computing and Data Analytics*

The scale of scientific computing is exploding – the amount and complexity of data produced by scientific applications has grown by orders of magnitude – and this trend is expected to continue. In particular, climate simulations may generate tens to hundreds of terabytes of data, and climate data have spatio-temporal, high-dimensional, multivariate, noisy, and dynamic characteristics. Scalable techniques and software need to be developed in order to enable a data-driven knowledge discovery paradigm for large-scale analysis of massive datasets. Advances in high-performance computing (HPC) related to data and hardware issues are causing an increasing gap between the opportunities created by the growing capabilities of HPC systems and the ability of current data analytics approaches to harness these technologies. In order to help meet the future computational needs of analyzing large and complex climate data, we present our recent progress in the following areas:

- Parallel algorithms for climate applications using hundreds to thousands of processes,
- Scalable MPI/OpenMP implementations of data clustering algorithms using thousands of processes,
- Development of a collection of data mining kernels for acceleration on hybrid GPU and multicore architectures,
- Development of the Parallel NetCDF library, which achieved close to the peak I/O bandwidth on a state-of-the-art high performance computer.

Shashi Shekhar – University of Minnesota
Title: *Discovering Persistent Change Windows in Spatiotemporal Datasets*

Given a region S comprised of locations that each have a time series of length |T|, the Persistent Change Windows (PCW) discovery problem aims to find all spatial window and temporal interval pairs <Si , Ti> that exhibit persistent change of attribute values over time. PCW discovery is important for critical societal applications such as detecting desertification, deforestation, and monitoring urban sprawl. The PCW discovery problem is challenging due to the large number of candidate patterns, the lack of monotonicity where sub-regions of a PCW may not show persistent change, the lack of predefined window sizes for the ST windows, and large datasets of detailed resolution and high volume. Previous approaches in ST change footprint discovery have focused on local spatial footprints for persistent change discovery and may not guarantee completeness. In contrast, we propose a space-time window enumeration and pruning (SWEP) approach that considers zonal spatial footprints when finding persistent change patterns. We provide theoretical analysis of SWEP's correctness, completeness, and space-time complexity. We also present a case study on vegetation data that demonstrates the usefulness of the proposed approach. Experimental evaluation on synthetic data show that the SWEP approach is orders of magnitude faster than the naive approach.

Presentations, Session 4, August 15

Douglas Nychka – National Center for Atmospheric Research
(with Stephan Sain and Tamara Greasby)
Title: *Regional climate informatics: A statistical perspective*

As attention shifts from broad global summaries of climate change to more specific regional impacts there is a need for data sciences to quantify the uncertainty in regional predictions. This talk will provide an overview on regional climate experiments with an emphasis on the data science problems for interpreting these large and complex simulations. A regional climate model (RCM) is a computer code based on physics that simulates the detailed flow of the atmosphere in a particular region from the large scale information of a global climate model. One intent is to compare simulations under current climate to future scenarios to infer the nature of climate change expected at a location. The output of these experiments include surface variables such as temperature and rainfall that can be compared to observations but also simulate more complex variables such as soil moisture or wind patterns that are informative for underlying ecological or geophysical processes. Finally, it is important to attach measures of uncertainty to the model predictions in order to make these geophysical results relevant for forming local policy and making economic decisions. This talk will use the recent North American Regional Climate Change and Assessment Program to illustrate some analyses of a multi-factor design where the response is one or more spatial fields of geophysical variables. Here a flexible spatial model based on fixed rank Kriging is implemented to handle a large number of spatial locations (LatticeKrig) and also include nonstationary spatial dependence. We estimate the change in the seasonal cycle of temperature over the US from climate simulations from the North American Regional Climate Change and Assessment Program (NARCCAP). Part of this analysis is to take into account topography and other covariates and to determine the effect of individual regional models on the synthesized results.

Arindam Banerjee – University of Minnesota
Title: *Probabilistic Graphical Models for Climate Data Analysis*

Probabilistic graphical models provide a powerful formalism for modeling high-dimensional discrete/continuous distributions involving sparsely dependent variables, i.e., where each variable only depends on a few others. In this talk, we discuss two applications of graphical models: first, combining climate model outputs using spatially smooth multivariate regression, and second, meteorological drought detection using Markov random fields. Further, we briefly discuss dependency learning for high-dimensional multivariate Gaussian graphical models and Gaussian copulas along with potential applications.

Fred Semazzi – North Carolina State University
Title: *Outreach Opportunities for Expedition's Data Driven Discoveries*

Expedition project's novel data-driven approaches already promise to excel in performance beyond the traditional methods in climate prediction, climate change projections and the search for fundamental understanding of the climate system. The Expedition project is creating bridges and special partnerships with research and operational climate communities to maximize its broader impacts. The presentation will review opportunities and examples for exploiting Expedition's earth breaking data-driven discoveries.

Presentations, Session 5, August 16

Greg Holland – National Center for Atmospheric Research
(with James Done, Cindy Bruyere and Mari Jones)
Title: *Assessing uncertainty in regional scale climate variability and change*

Moving to the explicit prediction of regional climate over decadal time scales presents formidable computing and scientific challenges. The requirement for high resolution is well known, but of equal and perhaps greater importance is the requirement for adequate assimilation of the initial climate state, assessment of prediction uncertainty, downscaling beyond the model resolution to extract the high-impact weather information, and adequately incorporating upscaling from mesoscale organization. In this presentation I will describe an approach to assessment of uncertainty and weather extremes. In this approach ensemble model, simulations are designed to provide a data set suited to further statistical applications rather than purely end-point predictions. I also will comment on the limiting predictability of regional climate at decadal time scales. Use will be made of recent simulations with the next-generation NCAR Model for Prediction Across Scales.

James Faghmous – University of Minnesota
Title: *Sea surface temperature (SST) spatial patterns and their impact on Atlantic tropical cyclone activity*

Seasonal tropical cyclone (TC) forecasting has become an active field of research with significant scientific and societal impacts. Since the earliest empirical studies, sea surface temperatures (SST) have been intimately linked to Atlantic tropical cyclone (TC) activity. As a result, several SST-based indices have been developed by averaging the SST of fixed regions, with the goal of predicting various phenomena, including Atlantic TCs. We investigate the impact of global SST spatial patterns on Atlantic TCs and show that monitoring both the intensity and spatial distribution of global SSTs provides better predictive insights than SST warming anomalies alone. We demonstrate our results by predicting Atlantic TC activity more precisely several months in advance compared to traditional indices. Given the impact SST warming anomalies have on the globe, such an approach may be useful to researchers forecasting other phenomena associated with SST anomalies.

Abodallah Homaifar – North Carolina Agricultural & Technical University
Title: *Tropical Cyclone Tracking and Intensity Estimation,* Image *Search and Climate Trend Quantification*

In this study, we are interested in determining whether cloud clusters (CCs) will develop into a tropical cyclone (TC), which continues to be an unresolved problem. Forecasters need new techniques to determine whether a TC will develop from a loosely organized cluster of clouds. This research investigates whether a CC will develop into a TC without using numerical weather prediction models but rather using global gridded satellite data, which is readily available. The proposed identification and tracking technique provides satisfactory results when compared to actual tracks that are usually completed by experts. In addition, we have developed an objective technique known as **F**eature **A**nalogs in **S**atellite **I**magery (FASI) to estimate TC intensity using historical hurricane satellite data. FASI requires a TC's center location to extract azimuthal brightness temperature (BT) profiles from current imagery as well as BT profiles from imagery 6, 12 and 24 hours prior as features. When tested and validated with intensity data from aircraft reconnaissance in the North Atlantic Ocean, the accuracy of the FASI technique is on par with other objective techniques. The simplicity, objectivity and consistency of the FASI technique are promising and could provide a valuable tool for estimating the intensity of TCs in an operational environment.

In other efforts, we have investigated image search in large dimensions and quantification of climate trend. Locality Sensitive Hashing (LSH) is an approximate nearest neighbor search technique, which is capable of finding similar records in high dimensional large dataset. We have significantly improved the LSH, and it is twice as fast as the original LSH; reduces the memory space by 50% and it is highly scalable. We have also developed non-stationary time series clustering to identify persistent regimes and also construct a dynamical model for each regime. A finite elements method is used to find regimes in time series that contain information on temperature from different stations in North Carolina for 60 years and its relation to well-known indices.

Auroop Ganguly – Northeastern University

Title: *Physics-guided statistical approach to uncertainty quantification from climate model ensembles: Expedition funded doctoral research at the SDS Lab in Northeastern University*

The Sustainability and Data Sciences Laboratory (SDS Lab) at Northeastern University has been exploring new insights in climate and hydrology, with particular emphases on extremes, uncertainty, and engineering adaptation. We attempt to explore the connection from climate models to water resources and natural hazards, with physics-guided data science methods. We briefly discuss our general research theme, and then showcase the dissertation research of Expedition student Evan Kodra. A statistician by training, Kodra has been working with climate scientists, hydrologists, statisticians, and computer scientists in the SDS Lab and within the Expedition. He is developing a physics-driven approach to uncertainty quantification in the context of precipitation extremes from multi-model ensembles in climate. His prior work with the group has been highlighted in Nature and Nature Climate Change, and published or presented in climate, hydrology, statistics, and computer science venues.

Presentations, Session 6, August 16

Soroosh Sorooshian – University of California, Irvine
(with Jialun Li, Wei Chu, Kuolin Hsu, Xiaogang Gao)
Title: *Understanding Climate Change From Data - Perspectives from Hydroclimate Modeling and Data Assimilation*

Much effort has been devoted to the use of both remotely-sensed and in-situ observations to model hydroclimate processes at a wide range of spatial and temporal scales. The purposes have also ranged from studying water balance over a region to prediction of future hydrologic variables such as precipitation. A number of data sets have been produced using data assimilation techniques and notable among them are LDAS and GLDAS.

In this presentation a number of recent modeling studies conducted by our research group and studying the water balance and regional hydroclimate over California's agricultural region will be presented. The study will highlight the differences between models and data assimilated products and emphasizes the need for care in applying such information in modeling application examining future regional hydroclimate conditions.

Jagadish Shukla – George Mason University
Title: *Fidelity and Predictability of Models for Weather and Climate Prediction*

Does the ability of a model to simulate the observed climate (mean, variances, co-variances, etc) affect that model's ability to make accurate predictions? By comparing the fidelity and skill of seasonal prediction models, and IPCC models, we conclude that the answer to this question should be yes.

We find that models with higher fidelity in simulating the present climate produce higher values of global warming due to increased greenhouse gases. We also find that the forecast skill of dynamical seasonal prediction by coupled ocean-atmosphere models is related to their ability to simulate observed climate. The models with higher fidelity have higher skill. We show that after 50 years of climate modeling, the current climate models can now produce skillful prediction of summer monsoon rainfall. Therefore, an important objective of climate research should be to improve the fidelity of climate models.

Stefan Liess – University of Minnesota
Title: *Statistical Analysis for Prediction of Monsoon Variability*

Interannual variability of monsoon systems is related to multiple factors in the general circulation. A statistical approach based on multiple linear regression is presented that connects relationships of the climate system to variability of the Indian summer monsoon circulation. We predict strength of the monsoon over 30 individual subdivisions in India based on remote influences related to long-term teleconnections such as ENSO, IOD, NAO, QBO, and SAM. In a second approach we use the above approach for statistical downscaling of results from dynamical models. Our method uses k-fold cross validation to determine the spatio-temporal state of the atmosphere before and during the monsoon season, and to group subdivisions based on influencing factors.

Panel Discussion: Big Data for Climate Science

Moderator: Auroop Ganguly, Northeastern University

Panel topic: Enabler of next-generation science or a passing fad: Is the "Big Data" revolution relevant for climate science?

The role of statistics in climate change and impacts research is well acknowledged. However, as climate data from models and sensors grows from a few terbaytes over the last several years to hundreds of petabytes in the coming decades, a crucial question emerges. Does data science in climate need to scale, and what is the urgency? What are the major challenges in the interpretation of the highly complex relationships revealed by data mining to improve understanding of the climate system? What other major challenges in the physical science basis, or in translation of the science to impacts, may benefit from data mining? Conversely, where should computer and data-driven scientists, working in climate, focus their efforts to be able to add value in the next 5, 10, and 20 years? Is the most immediate requirement that of scaling existing methods in statistics to large distributed and parallel computing systems, or is it to develop new methods in data sciences? How should historically disparate data science disciplines, such as information theory, nonlinear dynamics, statistics, machine learning, artificial intelligence, signal processing, and econometrics come together in the context of large climate data? Are climate scientists better off examining well-defined hypothesis with selected data, and are adaptation decisions best informed through focused research and case studies? Does Big Data point to a need for semi-automated hypothesis generation in the physical science basis of climate, or decision support systems for translational climate science?

Panelists:
Leonard P. Hirsch – Smithsonian Institution
Praveen Kumar – University of Illinois at Urbana-Champaign
Doug Nychka – National Center for Atmospheric Research
Nikunj C. Oza – NASA Ames Research Center
Ben Santer – Lawrence Livermore National Laboratory
Jagdish Shukla – George Mason University
Kevin Trenberth – National Center for Atmospheric Research

(This page left intentionally blank)

Invited Panelists and Speaker Bios (alphabetical order):

Amy Braverman – Jet Propulsion Laboratory, California Institute of Technology
Amy.Braverman@jpl.nasa.gov

Amy Braverman is Principal Statistician at NASA's Jet Propulsion Laboratory, California Institute of Technology, in Pasadena California. She received her Ph.D. in Statistics from UCLA in 1999, and was a Caltech Post-doctoral Scholar at JPL from 1999-2001. She joined JPL full-time in 2001, first in the Science Division, and then in the Instruments and Science Data Systems Division. Her research interests revolve around the analysis of massive, remote sensing data sets for climate and Earth science. These include data reduction techniques, spatio-temporal statistical methods for data fusion, statistical methods for evaluating climate models, and the analysis of distributed data sources. Amy is a Fellow of the American Statistical Association, and a founding member of its Advisory Committee on Climate Change Policy.

Ian Foster – Argonne National Laboratory
foster@anl.gov

Dr. Ian Foster is Director of the Computation Institute, a joint institute of the University of Chicago and Argonne National Laboratory. He is also an Argonne Senior Scientist and Distinguished Fellow, and the Arthur Holly Compton Distinguished Service Professor of Computer Science at U. Chicago. His research deals with distributed, parallel, and data-intensive computing technologies, and innovative applications of those technologies to scientific problems. He has led the development of methods and software that underpin many large national and international cyberinfrastructures. Dr. Foster is a fellow of the American Association for the Advancement of Science, the Association for Computing Machinery, and the British Computer Society. His awards include the British Computer Society's Lovelace Medal, honorary doctorates from the University of Canterbury, New Zealand, and CINVESTAV, Mexico, and the IEEE Tsutomu Kanai award.

Leonard P. Hirsch – Smithsonian Institution
LPH@si.edu

Dr. Leonard P. Hirsch (BA 1976 Pomona College, International Relations; MA 1978, PhD 1980, Northwestern University Political Science) is Senior Policy Advisor at the Smithsonian Institution where he supports the scientific programs of the Smithsonian, with emphasis on global environmental concerns. Dr. Hirsch is a co-chair of the International Working Group of the US Group on Earth Observations and the Smithsonian principal to the US Global Change Research Program. He participates in the White House Committee on Environment and Natural Resources of the National Science and Technology Council. He is engaged on issues of biological diversity informatics and facilitating the coordination of government research on land and seascape change. He is involved in the Global Biodiversity Information Facility.

Greg Holland – National Center for Atmospheric Research

Greg Holland is Leader of the Regional Climate Section in MMM. He has had a long career in meteorology including forecasting, teaching, research and community service. He has a MS and PhD from Colorado State University and received their Distinguished Alumnus Award in 2007. He has been working on weather extremes throughout his career. His most recent scientific activities have focused on the impact of climate variability and change on high-impact weather and the resulting societal implications. His publications have included major contributions to six textbooks and forecast manuals, together with over 120 research papers in atmospheric sciences and UAVs.

Praveen Kumar – University of Illinois at Urbana-Champaign
kumar1@illinois.edu, http://www.hydrocomplexity.net

Praveen Kumar is the Lovell Professor of Civil and Environmental Engineering. Dr. Kumar has served as the Editor-in-Chief for Water Resources Research, the major scientific journal in the field, published by American Geophysical Union (AGU). Prior to this he was the Editor for Geophysical Research Letters (2007-2009) also published by AGU. He holds a B.Tech. (Indian Institute of Technology, Bombay, India 1987), M.S. (Iowa State University 1989), and Ph.D. (University of Minnesota 1993), all in civil engineering. He has been on the faculty of the department of Civil and Environmental Engineering at the University of Illinois since 1995. Prior to joining University of Illinois, he was a research scientist (January 1993 to July 1995) at the Universities Space Research Association (USRA) and Hydrologic Sciences Branch, NASA- Goddard Space Flight Center, Greenbelt, Maryland, USA. Dr. Kumar's research includes study of complexity in hydrologic processes including hydroclimatology, ecohydrology, geomorphology, and hydroinformatics. The overall goal of Dr. Kumar's research is to develop quantitative understanding of emergent behaviors that arise from complex non-linear multi-scale interactions between water, climate and vegetation systems.

Douglas Nychka – National Center for Atmospheric Research

Douglas Nychka is a statistical scientist with an interest in the problems posed by geophysical data sets. His Ph. D. (1983) is from the University of Wisconsin and he subsequently spent 14 years as a faculty member at North Carolina State University. His research background in fitting curves and surfaces lead to an interest in the analysis of spatial and environmental data. Pursuing this area of application, he assumed leadership of the Geophysical Statistics Project at the National Center for Atmospheric Research (NCAR) in 1997, an NSF program to build collaborative research and training between statistics and the geosciences. In 2004, he became Director of the Institute of Mathematics Applied to Geosciences, an interdisciplinary component at NCAR with a focus on transferring innovative mathematical models and tools to the geosciences. His current interests are in quantifying the uncertainty of numerical experiments that simulate the Earth's present and possible future climate and spatial statistics applied to large data sets. He has received the Jerry Sacks Award for Multidisciplinary Research (2004), the Achievement Award given by the Committee for the 12th International Meetings on Statistical Climatology (2013) and is a Fellow of the American Statistical Association.

Nikunj C. Oza – NASA Ames Research Center

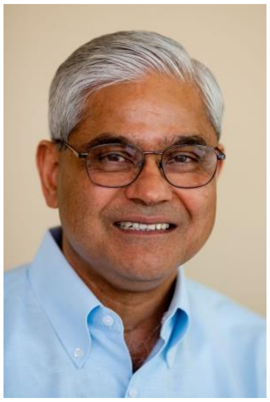
Nikunj Oza is the leader of the Data Sciences Group at NASA Ames Research Center. He additionally leads the Discovery of Precursors to Safety Incidents (DPSI) team which applies data mining to aviation safety. He is an Ames Center Program Manager for the Advanced Information Systems Technology (AIST) program, which funds research and development efforts in Information Technology with application to Earth science. Dr. Oza's 40+ research papers represent his research interests which include data mining, machine learning, fault detection, and their applications to Aeronautics and Earth Science. He received the Arch T. Colwell Award for co-authoring one of the five most innovative technical papers selected from 3300+ SAE technical papers in 2005, and the 2010 NASA Aeronautics Research Mission Directorate Associate Administrator team award for Technology and Innovation, . Dr. Oza is on the editorial board of the peer-reviewed journal Information Fusion (Elsevier). He received his B.S. in Mathematics with Computer Science from MIT in 1994, and M.S. (in 1998) and Ph.D. (in 2001) in Computer Science from the University of California at Berkeley.

Barbara Ransom – National Science Foundation

Barbara Ransom is a Program Director at the National Science Foundation and part of the EarthCube implementation team. She has served at NSF for over 9 years and is familiar with many disciplinary and cross-cutting programs, having run competitions for programs and solicitations as diverse as Cyberenabled Discovery and Innovation (CDI), Software Institutes (SI2), Science and Technology Centers (STC), National Nanotechnology Infrastructure Network (NNIN), IGERT, Emerging Topics in Biogeochemical Cycles, and a host of programs within the earth and ocean sciences. Her scientific interests and expertise range broadly across the geosciences and have expanded to include software and data intensive computing as well as data discoverability, interoperability, and model development and integration. She is particularly interested in building close working partnerships between geoscience and the computer science/cyberinfrastructure communities.

Ben Santer – Lawrence Livermore National Laboratory

Ben Santer is an atmospheric scientist at Lawrence Livermore National Laboratory (LLNL). His research focuses on such topics as climate model evaluation, the use of statistical methods in climate science, and identification of natural and anthropogenic "fingerprints" in observed climate records. Santer's early research on the climatic effects of combined changes in greenhouse gases and sulfate aerosols contributed to the historic "discernible human influence" conclusion of the 1995 Report by the Intergovernmental Panel on Climate Change (IPCC). His recent work has attempted to identify anthropogenic fingerprints in a number of different climate variables, such as tropopause height, atmospheric water vapor, the temperature of the stratosphere and troposphere, ocean heat content, and ocean surface temperatures in hurricane formation regions. Santer is a MacArthur Fellow and a member of the U.S. National Academy of Sciences.

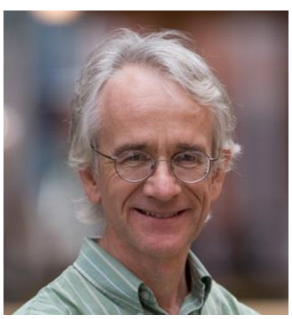


Jagadish Shukla – George Mason University

J. Shukla is a University Professor and founding Chairman of the Climate Dynamics Ph.D. program and the Department of Atmospheric, Oceanic and Earth Sciences at George Mason University (GMU). He is President of the Institute of Global Environment and Society (IGES). He has made contributions to the understanding of the predictability of weather and climate. His research has established that there is predictability in the midst of chaos, and that there is a scientific basis for short-term climate prediction.

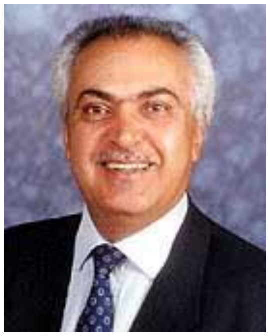
He received Padma Shri from the President of India in April 2012. In 2007, he received the International Meteorological Organization (IMO) Prize. He was a lead author of IPCC-2007 report. In 2009 he was elected to be a Fellow of the American Geophysical Union. In 2005, he received the Rossby Medal; in 2001, he received the Walker Gold Medal; in 1982 he received the Exceptional Scientific Achievement Medal of NASA.

He is a Fellow of the American Geophysical Union, the American Meteorology Society, India Meteorology Society and an Associate Fellow of TWAS (the academy of sciences for the developing world). He is the author/co-author of 250 scientific papers, and the editor/contributor of four books. He has been Ph. D. thesis adviser for about 20 students at M.I.T., Univ. of Maryland, and GMU. He has helped in the establishment of research centers in India and Italy. He founded Gandhi College for education of rural women in the village of his birth in India.



Richard L. Smith – University of North Carolina and SAMSI
http://www.unc.edu/~rls/

Richard L. Smith is Mark L. Reed III Distinguished Professor of Statistics and Professor of Biostatistics in the University of North Carolina, Chapel Hill. He is also Director of the Statistical and Applied Mathematical Sciences Institute, a Mathematical Sciences Institute supported by the National Science Foundation. He obtained his PhD from Cornell University and previously held academic positions at Imperial College (London), the University of Surrey (Guildford, England) and Cambridge University. His main research interest is environmental statistics and associated areas of methodological research such as spatial statistics, time series analysis and extreme value theory. He is particularly interested in statistical aspects of climate change research, and in air pollution including its health effects. He is a Fellow of the American Statistical Association and the Institute of Mathematical Statistics, an Elected Member of the International Statistical Institute, and has won the Guy Medal in Silver of the Royal Statistical Society, and the Distinguished Achievement Medal of the Section on Statistics and the Environment, American Statistical Association. In 2004 he was the J. Stuart Hunter Lecturer of The International Environmetrics Society (TIES). He is also a Chartered Statistician of the Royal Statistical Society.

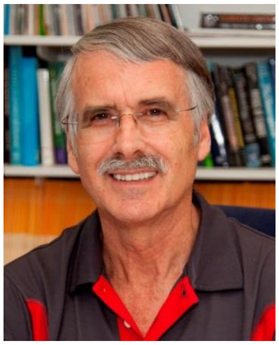


Soroosh Sorooshian – University of California – Irvine
soroosh@uci.edu

Soroosh Sorooshian is the Director of the Center for Hydrometeorology & Remote Sensing (CHRS) and Distinguished Professor of Civil & Environmental Engineering and Earth System Science Departments at UC Irvine. Prior to 2003 he was a faculty at the University of Arizona for 20 years. His area of expertise is Hydrometeorology, water resources systems, climate studies and application of remote sensing to water resources and hydrologic issues.

He is a member of the US National Academy of Engineering (NAE); Member of the International Academy of Astronautics (IAA); and Fellow of AAAS, AGU, AMS, IWRA; Member, Joint Scientific Committee (JSC) of the World Climate Research Programme (WCRP); Member, Independent Expert Committee (IEC) in the field of Engineering, Third World Academy of Sciences(TWAS); Member, Expert Review Panel Water and Green Growth (WGG) joint research Project by the Ministry of Land, Infrastructure and Transport of the Republic of Korea (MOLIT), the Korea Water Resources Corporation (K-water), and the World Water Council (WWC); Past Chair, Science Steering Group (SSG) of Global Energy and Water Cycle Experiment (GEWEX) of the World Climate Research Programme (WCRP); U.S. Member of the Hydrology Commission for WMO; Emeritus member of UCAR Board of Trustees and NOAA Science Advisory Board; Past President of AGU's Hydrology Section; member of five editorial boards and former editor of AGU's Water Resources Research. He has served on numerous advisory committees, including those of NASA, NOAA, DOE, USDA, NSF, EPA, and UNESCO. Current member of the National Research Council's (NRC) Space Study Board (SSB) and past member of the Water Science and Technology Board (WSTB). He has testified to both Senate and House subcommittees on earth observations from space and water resources issues.

Among his most recent honors are: Recipient of the 4th Prince Sultan Bin Abdulaziz International Prize for Water Resources Management & Protection 2010; elected Associate Fellow of TWAS, the Academy of Sciences for developing countries; named an honorary Professor at Beijing Normal University, China (2010); named the Walter Orr Roberts Lecturer, American Meteorological Society (AMS), 2009; recipient of AMS Robert E. Horton Memorial Lectureship, 2006; recipient of the NASA Distinguished Public Service Medal in 2005 and the William Nordberg Memorial Lecture at the NASA Goddard Space Flight Center in 2004.



Kevin E. Trenberth – National Center for Atmospheric Research
trenbert@ucar.edu, http://www.cgd.ucar.edu/staff/trenbert/

Dr. Kevin E. Trenberth is a Distinguished Senior Scientist in the Climate Analysis Section at the National Center for Atmospheric Research. From New Zealand, he obtained his Sc. D. in meteorology from Massachusetts Institute of Technology. He was a lead author of the 1995, 2001 and 2007 Scientific Assessment of Climate Change reports from the Intergovernmental Panel on Climate Change (IPCC), and shared the 2007 Nobel Peace Prize which went to the IPCC. He served from 1999 to 2006 on the Joint Scientific Committee of the World Climate Research Programme (WCRP), and he chaired the WCRP Observation and Assimilation Panel from 2004 to 2010 and now chairs the Global Energy and Water Exchanges (GEWEX) scientific steering group. He has also served on many national committees. He is a fellow of the American Meteorological Society (AMS), the American Association for Advancement of Science, the American Geophysical Union, and an honorary fellow of the Royal Society of New Zealand. In 2000 he received the Jule G. Charney award from the AMS; in 2003 he was given the NCAR Distinguished Achievement Award; and in 2013 he was awarded the Prince Sultan Bin Abdulaziz International Prize for Water. He edited a 788 page book Climate System Modeling, published in 1992 and has published 500 scientific articles or papers, including 55 books or book chapters, and over 225 refereed journal articles. He has given many invited scientific talks as well as appearing in a number of television, radio programs and newspaper articles. He is listed among the top 20 authors in highest citations in all of geophysics.

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

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

Expeditions Project 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 William Norris Professor and Head 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 also served as the Director of Army High Performance Computing Research Center (AHPCRC) from 1988 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 Assistant 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 data mining, bioinformatics and computational biology, high performance computing, and their applications, and has published more than 50 papers in various peer-reviewed journals and conferences. His research is supported by National Science Foundation and Department of Energy.

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

Arindam Banerjee is an Associate Professor at the Department of Computer & Engineering and a Resident Fellow at the Institute on the Environment at the University of Minnesota, Twin Cities. He received his Ph.D. from the University of Texas at Austin in 2005, where his dissertation was nominated for the best dissertation award. His research interests are in Machine Learning, Data Mining, Convex Analysis and Optimization, and their applications in complex real world problems. He has won several awards including the NSF CAREER award in 2010, the McKnight Land-Grant Professorship at the University of Minnesota, Twin Cities, 2009-2011, and the prestigious IBM PhD fellowship for the academic years 2003-2004 and 2004-2005. His publications have won five best paper awards, including the best paper award at the SIAM International Conference on Data Mining (SDM), 2004, best student paper award at SDM, 2012, and best application paper award at SDM, 2013.

Shyam Boriah – University of Minnesota
sboriah@cs.umn.edu

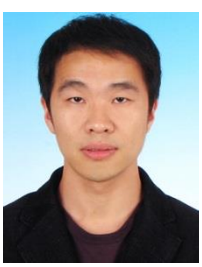
Shyam Boriah is a Research Associate in the Department of Computer Science and Engineering at the University of Minnesota. His research interests are in data mining and machine learning, primarily in spatial, time series and spatio-temporal analysis. He has developed algorithms for global scale analysis in the Earth science and environmental domains. His broader research interests are in developing data analysis techniques and tools to address issues in food supply, climate change and deforestation.

Ruben Buaba – North Carolina Agricultural & Technical State University
rbuaba@ncat.edu

Ruben Buaba is an Adjunct Assistant Professor and Postdoctoral Researcher in the Department of Electrical and Computer Engineering at North Carolina Agricultural and Technical State University. He received his Ph.D. in Electrical Engineering from North Carolina Agricultural and Technical State University in 2012. His research interests include Controls, Data Mining, Image Processing, Machine Learning, Search Algorithms and High Performance and Scalable Computing. He obtained his B.S. in Electrical and Electronic Engineering from Kwame Nkrumah University of Science and Technology, Ghana in 2007. In the summer of 2009, he was an Intern at NOAA/NGDC, Boulder, Colorado, where he investigated a Similarity Search Algorithm for satellite image retrieval application. He has recently developed software called SIBRA (Satellite Image-Based Retrieval Application), which provides an interactive image search platform for image search analyses. His research work is partially supported by the NSF Expedition in Computing. He has published more than 9 papers in peer-reviewed journals and conferences.

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.



Zhengzhang Chen – Northwestern University
zhengzhang.chen@northwestern.edu

Zhengzhang Chen 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 2012, and his research interests include data mining, bioinformatics, graph algorithms, social computing, and machine learning.

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. His doctoral research was part of a 5-year $10M NSF-funded Expeditions in Computing grant to develop novel numerical techniques to study and monitor climate change. As part of the Expeditions team, James developed scalable data mining algorithms to analyze large climate datasets with applications to tropical cylone forecasting and mesoscale ocean eddy monitoring. James' research has been 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.

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

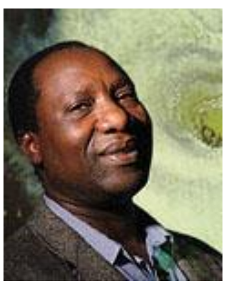
Stefan Liess is a Research Associate in Atmospheric Sciences at the University of Minnesota. He analyzes climate dynamics and climate change with observations, general circulation models, and high-resolution regional models. His specific research interests are intraseasonal variability and predictability 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.

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

Dr. Nagiza F. Samatova is an Associate 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, Bioinformatics, Systems Biology, Data Management, Scientific and High Performance Computing, and Machine Learning. She is the author of over 130 publications in peer-reviewed journals and conference proceedings.



Fredrick 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.

Karsten Steinhaeuser – University of Minnesota
ksteinha@umn.edu

Karsten Steinhaeuser is a Research Associate in the Department of Computer Science and 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 (but not limited to) climate, ecology, and social networks. He is actively involved in shaping an emerging research area called climate informatics, which lies at the intersection of computer science and climate sciences, and his interests are more generally in interdisciplinary research and scientific problems relating to climate change and sustainability. He co-organizes the IEEE ICDM Workshop on Knowledge Discovery from Climate Data and the International Workshop on Climate Informatics, among others, and is engaged in numerous other professional service and community building activities. Karsten earned his PhD in Computer Science and Engineering at 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 and Engineering (2007) and a BS, Summa Cum Laude in Computer Science (2005), both from the University of Notre Dame.

Poster Presentations

Thursday, August 15, 2013
Listed in Alphabetical order of Last Name of Presenter

(This page left intentionally blank)

Presenter : **Saurabh Agrawal Airan, University of Minnesota**
Title: *Graph Based Analysis of Dynamic Teleconnections*

Contributors: Agrawal, Kawale, A Kumar, Ormsby, Rehberger, Liess, Steinbach, Steinhaeuser, Chatterjee, Snyder, V Kumar (UMN), Ganguly (NEU), Samatova, Semazzi (NCSU)

Pressure dipoles are important long distance climate phenomena (teleconnection) characterized by pressure anomalies of opposite polarity appearing at two different locations at the same time. Such dipoles have proven important for understanding and explaining the variability in climate in many regions of the world, e.g., the El Nino climate phenomenon is known to be responsible for precipitation and temperature anomalies at large parts of the globe.

We present a novel graph based approach called shared reciprocal nearest neighbors that considers only reciprocal positive and negative edges in the shared nearest neighbor graph to find dipoles. To show the utility of finding dipoles using our approach, we show that the data driven dynamic climate indices generated from our algorithm generally perform better than static indices formed from the fixed locations used by climate scientists in terms of capturing impact on global temperature and precipitation. Another salient point of this approach is that it can generate a single snapshot picture of all the global dipole connections in a given dataset. This allows a comparison between different data sets, e.g., climate model simulations, which can be very useful for evaluating climate models. Given the importance of teleconnections and the importance of model simulations in understanding the impact of climate and climate change, this methodology has the potential to provide valuable insights into future climate projections. Further, we are able to identify some new connections that appear as significant and we are investigating one such phenomenon near Australia.

Presenter: **Michael Angus and Kanchana Padmanabhan, North Carolina State University**
Title: *Coupled Heterogeneous Association Rule Mining (CHARM): Application toward Inference of Modulatory Climate Relationships*

Contributors: Gonzalez, Tetteh, Angus, Padmanabhan, Pendse, Srinivas, Villanes, Semazzi, Samatova (NCSU), Kumar (UMN)

The complex dynamic climate system often exhibits hierarchical modularity of its organization and function. Scientists have spent decades trying to discover and understand the driving mechanisms behind the western African Sahel summer rainfall variability, mostly via hypothesis-driven and/or first-principles based research. Their work has furthered theory regarding the connections between various climate patterns, but the key relationships are still not fully understood. We present Coupled Heterogeneous Association Rule Mining (CHARM), a computationally efficient methodology that mines higher-order relationships between these subsystems' anomalous temporal phases with respect to their modulating effect on the system’s functional response. We apply this methodology to infer putative pathways/cascades of modulating events and the modulating signs that collectively define the network of pathways for the rainfall anomaly in the Sahel. We apply this data-driven approach to climate science data, aiming to quantify known underlying mechanisms, and/or suggest new mechanisms and interconnections to study. This allows for building a more comprehensive model of the modular relationships influencing Sahelian summer rainfall.

Here, we particularly focus on the influence of the Atlantic El Niño Southern Oscillation signal (EATL) on the local geo-potential height pattern and on the teleconnections between various climate indices and the equatorial Lower Level Westerlies (LLWs), both of which play an important role in modulating rainfall in the Sahel (Nicholson 2009). We emphasize these two findings as the former gives us confidence in our approach, due to the consistency with the climate literature, while the latter highlights the potential for discovering previously unrecognized relationships.

Presenter: **Gonzalo Bello, North Carolina State University**
Title: *Climate Index Discovery using Supervised Clustering: An Application to Seasonal Rainfall Prediction in the Greater Horn of Africa*

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

Climate indices provide a generalized overview of an aspect of the climate system, and can be used to analyze the influence of this aspect of the system on regional weather. In this work, we propose a data-driven methodology for the automated discovery of climate indices, using a supervised approach for clustering spatio-temporal multivariate climate data. Our supervised approach allows us to construct response-specific climate indices that explain a higher percentage of the variability of the response.

We apply this methodology to the discovery of climate indices for seasonal rainfall variability in the Greater Horn of Africa, and use the climate indices discovered to predict seasonal rainfall in the region. Rainfall variability has a profound socio-economic impact on the Greater Horn of Africa, because of its severe effect on agriculture production and hydro-electric power generation. More accurate forecasts of seasonal rainfall would greatly benefit logistical planning efforts in the region, by allowing policymakers to anticipate potential food and power shortages.

Initial experiments were performed to predict September-December seasonal rainfall at 4 synoptic stations (Arusha, Kilimanjaro, Moshi, and Same) located in the North Eastern Highlands region of Tanzania. Preliminary results show potential improvements on the accuracy of the prediction of the September-December rainfall season with respect to state-of-the-art methodologies for data-driven discovery of climate indices.

Presenter: **Mace Blank, University of Minnesota**
Title: *Global Land Cover Change Detection*

Contributors: Khandelwal, Blank, Mithal, Boriah, Kumar (UMN)

A variety of natural and anthropogenic changes including wild fires, deforestation, insect infestation and urbanization are impacting the forests around the world. These changes in land surface potentially undermine the capacity of ecosystems to maintain freshwater and forest resources, and regulate climate and air quality. As a result, monitoring of land resources such as forests has become a major priority of researchers and policymakers around the world. In this work, we focus on two types of land cover changes, namely, fire and urbanization. MODIS sensor (250m spatial resolution and 16 day temporal resolution) onboard Terra and Aqua satellites provide suitable data for global monitoring of fires. Products such as Active Fire Product and Burned Area Product are widely used global products for fire monitoring. But they are known to have errors. In this work, we propose a multi view approach for detecting fire events and characterizing land cover conversion after the fire. Specifically, we aim to integrate these available products with event detection algorithms developed by our group in a way that the final product is more accurate than all individual products. In context of urbanization, spatial resolution of MODIS is not suitable and thus we choose Landsat data for monitoring urbanization. Here, we propose a Hidden Markov Model based scheme which takes time series of imperfect classification maps and generate more accurate classification maps using the temporal context and domain related constraints.

Presenter: **Ruben Buaba, North Carolina Agricultural & Technical University**
Title: *Nearest neighbor search in large high-dimensional dataset*

Contributors: Buaba, Homaifar (NCAT), Kihn (NOAA/NCDC)

Searching large datasets to find similar records has become a big challenge in this era of “big data”. Computational resources are not catching up with the enormous amount of data that are being generated. The interesting fact is that data points are becoming highly dimensional and complex; making the search problem more difficult. A nearest neighbor search may be formulated as follows: Given a set S of n data points in a d-dimensional metric space, "P" , the task is to preprocess these points so that, given any query point q∈"P" the data point nearest to q can be reported quickly. Over the years, scientists have proposed a number of algorithms, which only turned out to provide relatively modest constant factor improvements and suffer as the dimension increases. No known method simultaneously achieves the goals of roughly linear space and logarithmic query time in any fixed dimension greater than two. Consequently, many researchers have become proponents of approximate nearest neighbor search. This gave birth to Locality Sensitive Hashing (LSH), which guarantees the retrieval of nearest neighbors by searching just a fraction of the search space (dataset). We have significantly improved the LSH and made it highly scalable. The improved algorithm outperforms the existing LSH in terms of computational complexity (50% improvement), memory requirement (50% improvement) and query runtime (twice as fast). Experiments conducted on 1.6 million images, chosen from the Defense Meteorological Satellite Program imagery dataset, have affirmed the theoretical findings regarding the improvements of the computational, memory and runtime complexities.

Presenter: **Soumyadeep Chatterjee, University of Minnesota**
Title: *Data-driven approach to improve predictability of Indian summer monsoon*

Contributors: Airan, Chatterjee, Rehberger, Liess, Boriah, Banerjee, Chatterjee, Kumar (UMN)

Indian monsoon has a direct impact on the economies of the south Asian countries and the lives of more than 16 % of total population of the Earth. Improvement in forecasting of Indian monsoon would significantly help these nations prepare for the annual event, particularly extremes.

Despite advances in seasonal climate forecasting using dynamical models, skill in predicting the Indian monsoon by such methods deteriorates with increasing lead time. The primary objective of this study is to build up a data-driven model that has 1) higher prediction power of monsoons using data mining, statistical and machine learning techniques compared to current state-of-the-art and 2) capture the relationships between Indian monsoon and different global climate variables.

The initial phase of this study focuses on analyzing the performance of linear regression models with precipitation during monsoon months (June, July, August, September) for a region of India as a predictand. We further extended the analysis of linear regression models (with k-fold cross-validation) with many more climate variables introduced as possible predictors (with one/two predictors at a time), at different levels of spatial aggregation. To overcome the effects of local factors on the precipitation, subdivisions that are in spatial neighborhood of each other and have similar response to a set of predictors were grouped together which led to significant improvement in the response to the same set of predictors with respect to the individual subdivisions. Also, this highlighted the dominance of different climate variables in different regions of India which further suggests that these regions are influenced by different factors.

As the preliminary results indicate, grouping the regions shows significant improvement in the prediction power even for a simple linear regression model and thus coming up with models that exploit the spatial information from the data is the next natural step. Along similar lines, mining patterns in temporal dimension is another possible extension. For instance, the relationships can vary across El-Nino and La-Nina time regimes and studying these variations would further aid in improvement in the prediction. Extending beyond linear models may also improve predictive power.

Presenter: **Zhengchan Chen, North Carolina State University**
Title: *EventForecaster: A Framework for Forecasting Extreme Events in Complex Dynamic Systems*

Contributors: Chen, Jenkins, Liao, Semazzi, Samatova (NCSU), Choudhary (NW), Agrawal, Kumar (UMN)

In complex dynamic systems, accurate forecasting of extreme events, such as hurricanes, is a highly underdetermined, yet very important problem. Traditional statistical methods simplify this problem by constraining the forecast of extreme events to a specific region or by considering only the linear relations between the response variable and the predictands. In this work, we present a systematic framework EventForecaster for robust forecasting of spatio-temporal extreme events by utilizing the characteristics of extreme event prediction, namely nonlinear relationships in an underdetermined problem space (namely, FORECASTER algorithm) and systems exhibiting hierarchical relationships (namely, EDETECTOR algorithm). FORECASTER supports multivariate spatio-temporal data. And EDETECTOR is not limited to system-subsystem or whole-part paradigm, but works for various hierarchical or modular system structures including linear corrections, associations, etc. By combining FORECASTER with EDETECTOR, we can reach more than 70%–80% accuracy in seasonal forecasting of (a) tropical cyclones in a Northern Hemisphere, (b) hurricanes and landfalling hurricanes in North Atlantic, and (c) North African rainfall.

Presenter: **Debasish Das, Northeastern University / Temple University**
Title: *A Sparse Bayesian Model for Dependence Analysis of Extremes: Climate Applications*

Contributors: Das (NEU & Temple), Ganguly (NEU), Obradovic (Temple)

In many real applications, such as climate, finance and social media among others, we are often interested in extreme events. An important part of modeling extremes is discovery of covariates on which the quantities related to the extremes are dependent, as this may lead to improved understanding and the discovery of new causal drivers of extremes. Despite developments in sparse covariate discovery algorithms, adaptations to extremes can fail because the tail attributes do not follow a Gaussian distribution. In this paper, we proposed a sparse Bayesian framework for discovery of covariates that influences the frequency of extremes based on the Poisson description of extremes frequency and hierarchical Bayesian description of a sparse regression model. We developed an efficient approximation algorithm based on the variational Bayes approach to estimate the distribution over regression coefficients that indicate dependence of extremes on the corresponding covariates. Experiments with synthetic data demonstrate the ability of the approach to accurately characterize dependence structures. Applications to rainfall extremes suggest new insights relevant for improved understanding of hydrological extremes under climate variability and change.

Presenter: **James Faghmous, University of Minnesota**
Title: *Abstracting the Pacific Ocean's Impact on North Atlantic Tropical Cyclones*

Contributors: Faghmous, Le, Liess, Kumar (UMN)

The warming anomalies of sea surface temperatures (SSTs) along the near-equatorial Pacific Ocean (ENSO) have well documented global long-range weather teleconnections from rainfall in southern India to mudslides in the western United States. In this work, we focus on one of ENSO's teleconnections with North Atlantic tropical cyclone (TC) activity. Traditionally, ENSO's impact on Atlantic TC has been abstracted by monitoring the warming of static regions along the equatorial Pacific Ocean. We propose that the spatial distribution of Pacific Ocean warming might provide better predictive insights into ENSO-Atlantic TC activity better than warming anomalies alone. We present a distance-based ENSO index that tracks the location of maximum near-tropical Pacific warming anomaly instead of its absolute warming. Our spatial ENSO index correlates better with seasonal TC activity than standard ENSO indices, especially with increased lead times.

Presenter: **James Faghmous, University of Minnesota**
Title: *Spatio-Temporal Data Mining for Scalable Ocean Eddy Monitoring*

Contributors: Faghmous, Uluyol, Le, Styles, Mithal, Boriah, Kumar (UMN)

Rotating coherent structures of water, known as ocean eddies are the oceanic analog of storms in the atmosphere and a crucial component of ocean dynamics. In addition to dominating the ocean's kinetic energy, eddies play a significant role in the transport of water, salt, heat, and nutrients. Therefore, understanding current and future eddy activity is a central climate challenge to address future sustainability of marine ecosystems. The emergence of sea surface height observations from satellite radar altimeter has recently enabled researchers to track eddies at a global scale. The majority of studies that identify eddies from observational data employ highly parametrized connected component algorithms using expert filtered data, effectively making reproducibility and scalability challenging. In this paper, we improve upon the state of the art connected component eddy monitoring algorithms to track eddies globally. This work makes three main contributions: first, we do not pre-process the data therefore minimizing the risk of wiping out important signals within the data. Second, we employ a physically-consistent convexity requirement on eddies based on theoretical and empirical studies to improve accuracy effectively reducing the computational complexity from quadratic to linear time in the size of each eddy. Finally, we are able to effectively break-up merged eddies, which existing methods cannot accomplish. We compare our results to those of the state-of-the-art method and discuss the impact of our improvements on the difference in results.

Presenter: **Gholamreza Fetanat, North Carolina Agricultural & Technical University**
Title: *Objective Tropical Cyclone Intensity Estimation from Satellite Images using Data Mining Techniques*

Contributors: Fetanat, Homaifar (NCAT), Knapp (NOAA/NCDC)

Tropical cyclones (TCs) are a significant threat to life and property. Developing and improving objective techniques to estimate a TC's intensity remain a challenge. An objective method to estimate tropical cyclone intensity using historical hurricane satellite data (HURSAT) is developed and tested. This new method, referred to as Feature Analogs in Satellite Imagery (FASI), requires a TC's center location to extract azimuthal brightness temperature (BT) profiles from current imagery as well as BT profiles from imagery 6, 12 and 24 hours prior. Instead of using regression techniques, the estimated TC intensity is determined from the ten closest analogs to this TC based on the BT profiles using a K-nearest-neighbor algorithm. The FASI technique was trained and validated using intensity data from aircraft reconnaissance in the North Atlantic Ocean, where the data were restricted to include storms that are over water and south of 45˚N. This subset comprised of 2016 observations from 165 storms during 1988 – 2006. Several tests were implemented to statistically justify the FASI algorithm using n-fold cross-validation. The resulting average mean absolute intensity error was 10.9 kt (50% of estimates are within 10 kt) or 8.4 mb (50% of estimates are within 8 mb); its accuracy is on par with other objective techniques. This approach has the potential to provide global TC intensity estimates that could augment intensity estimates made by other objective techniques. The simplicity, objectivity and consistency of the FASI technique should make it a valuable tool for estimating the intensity of TCs in operational environment.

Presenter: **Hendrix, William, Northwestern University**
Title: *Accelerating Climate Discovery through High Performance Computing*

Contributors: Jin , Hendrix, Patwary, Agrawal, Liao, Choudhary (NWU), Fu, Wang, Banerjee (UMN)

As a data-rich science, many high-impact questions in climate science would take incredible amounts of time to answer using the power of a single compute node. In this poster, we describe some of our recent and ongoing projects and collaborations in applying high performance computing to climate problems.

Several important combinatorial optimization problems, including drought detection, can be formulated as maximum a posteriori (MAP) inference in discrete graphical models. In collaboration with researchers at the University of Minnesota, we implemented a parallel MAP inference algorithm based on Bethe-ADMM. By using an edgewise graph partitioning and a neighborhood identification scheme to facilitate the message passing, we achieved near-linear scaling to thousands of cores.

Many other climate-related problems can be accelerated by decomposing them into independent tasks. For example, we collaborated with researchers from Northeastern University to produce a parallel algorithm for estimating uncertainty in Indian precipitation extremes, reducing the compute time from days to minutes while increasing the number of bootstrap iterations. As another example, we have ongoing work with NC A&T University on a parallel hash-based technique for efficiently identifying similar hurricanes based on satellite imagery. We've also previously developed several parallel data analysis algorithms, including high performance clustering and frequent itemset mining techniques scalable to thousands of compute nodes.

In addition to adopting sound parallel work division strategies, the use of parallel I/O, including MPI I/O and PnetCDF, is key to achieving good parallel performance. Widely-adopted scientific applications, PnetCDF provides a set of APIs for accessing netCDF files in parallel and sustains high I/O bandwidth up to tens of thousands of computer cores. The interface is able to read or write one or more variables in arbitrary patterns and provides both blocking and nonblocking functions calls to meet different application requests.

Presenter: **Mohammad Gorji Sefidmazgi, North Carolina Agricultural & Technical University**
Title: *Temperature regime/trends analyzing during 1950-2010 in North Carolina*

Contributors: Gorji, Homaifar, Sayemuzzaman, Jha, (NCAT) Liess(UMN)

An important characteristic of climatic time series is non-stationarity. It means that their statistical properties are changing with time and the time series cannot be represented uniquely. A common problem in this field is the analysis of high-dimensional time series containing different regimes. The problem of non-stationary time series clustering is defined to find these regimes numerically. This study investigates the regimes/trends based on the average temperature across the state of North Carolina. Historical climatic observations from over 249 weather stations spread across the state during 60 years have been used in the analysis. In this study a non-stationary time series clustering technique based on the Finite Element method (FEM) is used to identify local climatic regimes and the corresponding linear trend for each regime of the time series. Statistical significance of trends for each regime are tested using Mann-Kendall approach. We anticipate that the climate trend of this study will identify region-specific climate variability in North Carolina in different periods of time. Furthermore, we investigated the relationship between the Atlantic Multi-decadal oscillations (AMO), the sunspot cycle and the trends/regimes detected based on our proposed clustering method. It is shown that long-term regimes/trends in NC are mostly due to AMO and solar activities.

Presenter: **Zhe Jiang, University of Minnesota**
Title: *Focal-Test-Based Spatial Decision Tree Learning: A Summary of Results*

Contributors: Jiang, Shekhar, Zhou, Knight, Corcoran (UMN)

Given a raster spatial framework, as well as training and test sets, the spatial decision tree learning (SDTL) problem aims to learn a decision tree model minimizing classification errors as well as salt-and-pepper noise. The SDTL problem is important due to many societal applications such as land cover classification in remote sensing. However, the SDTL problem is challenging due to the spatial autocorrelation of class labels, and the potentially exponential number of candidate trees. Related work is limited due to the use of local-test-based decision nodes, which cannot adequately model spatial autocorrelation during test phase, leading to high salt-and-pepper noise. In contrast, we propose a focal-test-based spatial decision tree (FTSDT) model, where the tree traversal direction for a location is based on not only local but also focal (i.e., neighborhood) properties of the location. Experimental results on real world remote sensing datasets show that the proposed approach reduces salt-and-pepper noise and improves classification accuracy.

Presenter: **Ankush Khandelwal, University of Minnesota**
Title: *A computational approach for globally monitoring surface extent of lakes*

Contributors: Khandelwal, Anderson, Boriah, Kumar (UMN), Vatsavai (ORNL)

Temperature increase and precipitation variations over the last 50 years have resulted in significant changes in many lakes in the world, as shown in their size, morphology, and ecology. Moreover, surface water is one of the most varying variables in hydrological cycles in terms of its spatial and temporal distribution. Thus, there is an urgent need to monitor water bodies across the globe for more accurate hydrological modelling and effective water resource management. MODIS sensor (250m spatial resolution and 16 day temporal resolution) onboard Terra and Aqua satellites has been widely used for both regional and global studies of variation in water extent. In this work, we propose a classification based approach with a training framework that allows efficient training at global scale with high temporal frequency. We aim to incorporate the spectral shift in land cover classes that occur across time steps so that minimum number of classifier can be learned to classify all the time steps. The results have shown that a single classifier which incorporates spectral shift can perform accurately over all the time steps for a given region. For evaluation, water extent maps generated every 16 days were compared with radar altimeter data using correlation as metric.

Presenter: **Evan Kodra, Northeastern University**
Title: *Projections of a widening range of temperature extremes in a warming environment*

Contributors: Kodra, Ganguly (NEU)

Projections with associated uncertainty based on statistical extreme value theory and a bootstrap analysis reveal consistently larger increases in the hottest seasonal temperature extremes compared to the coldest ones, especially further from the tropics. These asymmetric changes in tail probability distributions of temperature appear robust when explored through 14 CMIP5 climate models and three reanalysis datasets. A statistical linear mixed effect model quantifies how that asymmetry depends on the covariates region, season, extreme type, and climate model variability. An assessment of potential physical drivers supports results from the asymmetry.

Presenter: **Devashish Kumar, Northeastern University**
Title: *Projected changes in precipitation extremes with implications for flood hazards resilience over India*

Contributors: Mishra (IIT Gandhinagar), Kumar, Ganguly (NEU)

Flood and flash flood disaster risks over India are exacerbated by a growing population, emerging economy, and non-robust infrastructures. However, an assessment of the non-stationary nature of rainfall extremes, particularly with a view to flood and flash flood hazards resilience, has been lacking. Here we show robust change in the projections of precipitation extremes over India in the 21st century with a metric that relates directly to structural design and floods hazards management. The results from 32 Couple Model Intercomparison Project Phase 5 (CMIP5) earth system models over India suggest on an average a two-fold increase in the intensity and frequency of precipitation extremes in many regions and the daily 100 year precipitation maxima occurring once every 20 year by the end of the century. The median risks are projected to increase by 55.7, 54, and 35% for hydrologic structures designed for 20, 50, and 100 year of service life, respectively.

Presenter: **Chaunté Lacewell, North Carolina Agricultural & Technical University**
Title: *Identification of Cloud Clusters Developing into Tropical Cyclones*

Contributors: Lacewell, Homaifar (NCAT), Lin (NCAT)

Increasing intensity of hurricanes, extreme droughts, floods, and rising sea levels might occur due to rising temperature during the warming phase of the climate change. Especially considering that our planet will reach 9 billion inhabitants by mid-century, the associated social, economic, and environmental impacts are enormous. In this study, we are particularly interested in determining whether cloud clusters (CCs) will develop into a tropical cyclone (TC), which continues to be an unresolved problem. Many studies have been conducted to distinguish between developing and non-developing CCs but forecasters need new techniques using pattern recognition to determine whether a TC will develop from a loosely organized cluster of clouds. Prior studies have attempted to predict tropical cyclogenesis (TCG) using numerical weather prediction models, satellite and radar data. Due to the complexity of cloud patterns, satellite data are used to initialize these dynamic models since TCs form in areas where little or no in situ data are available. Dynamic models still show discrepancies; hence, it is beneficial to use solely satellite data which is fully based on remote sensing of events that have actually occurred. In response, this research investigates whether a cloud cluster (CC) will develop into a TC without using numerical weather prediction models. Global gridded satellite data are used and are readily available when advanced data are not available.

Presenter: **Scott Sellars, University of California – Irvine**
Title: *The Application of Object-Oriented Data Analysis for Developing a Higher Dimensional Knowledge Discovery Dataset: Regional Hydroclimatrology of Western U.S Atmospheric Rivers*

Contributors: Sellars, Nguyen, Chu, Gao, Hsu, Sorooshian (UCI)

Using an object-oriented data analysis approach and satellite derived precipitation data, we investigate fundamental climate phenomena associated with the variability of Atmospheric Rivers (AR) (large plumes of moisture transported from the tropics) that impact California. Understanding the fundamental causes of regional variability in hydroclimatrology and hydrometeorology is one of the most important aspects in water resource planning. Scientists and engineers are in need of methods to organize and understand vast amounts of data, especially data from state-of-the-art, high resolution remote sensing platforms associated with their particular region of interest. This poster will discuss a recently reported object-oriented, global precipitation database approach to understanding regional hydroclimatology, specifically focused on California. This approach analyzes precipitation events as objects, allowing for attributes and statistics to be derived. An "object" can be thought of as an identified item or event with distinct attributes representing the existence of the event. Our database, which contains segmented objects of precipitation allows for high dimensional attributes to be included in an analysis of each segmented precipitation object. Some of these attributes are volume (m3), maximum precipitation intensity (mm/hr), average intensity (mm/hr), duration (hr), average speed (km/hr), El Nino Southern Oscillation (ENSO) and Madden Julian Oscillation (MJO) phase. We currently have 52 attributes or "characteristics", including many other climate phenomena for each object. The precipitation data used for populating the object-oriented and descriptive statistics databases is the Precipitation Estimation from Remotely Sensed Information using Artificial Neural Networks (PERSIANN) .25 degree dataset, which covers from 60N to 60S and from March 1st, 2000 to January 1st, 2011. This poster will report results describing the MJO and ENSO (including other climate phenomena) impact on certain characteristics of ARs including volume, originating location, frequency, duration and change in the seasonality of extreme precipitation. In addition, these results will show the potential usefulness of organizing data into objects for the detection of structure in the statistics of the object population, which enhances interpretability and understanding of climate variability and change.

Presenter: **Vidyashankar Sivakumar, University of Minnesota**
Title: *Combining Climate Model Predictions with Spatiotemporal Smoothing*

Contributors: Das, Subbian, Chatterjee, Johnson, Sivakumar, Banerjee (UMN)

Global Circulation Models (GCM) are mathematical models used for weather forecasting, understanding the climate and projecting climate change. Due to different modeling assumptions there is high variability between future projections of different GCM models. A popular practice is to combine the output from the various GCM models using weighted averaging models to reduce variability and prediction error. In models like superensemble a separate least squares regression model is fit for each location and regression parameters obtained denote the weights of the individual models for each location. But such models have been found to overfit the training data due to absence of any regularization and they also do not constrain the regression parameters in neighboring locations to be smooth. These models also do not incorporate the possibility of changes in the regression parameters over time. In our work we model the problem of combining GCM model outputs as a multiple least squares regression problem with spatial and temporal regularization. We use a graph Laplacian based spatial regularization, which yields smoothly varying spatial coefficients, and total variation based temporal regularization, which encourages the difference in temporal coefficients to be sparse. We use Alternating Direction Method of Multipliers (ADMM) to efficiently solve the optimization problem. Experimental results show that our model performs favorably compared to models such as superensemble and uniform average in terms of accuracy as measured by average RMSE. Further, our models yields smoothly varying spatial coefficients for model combination as measured by the Kendall Tau and Spearman correlation coefficients. Aggregate analysis also suggests that certain general circulation models tend to consistently perform well in certain regions of the world.

Presenter: **Muhammad Sohail, North Carolina Agricultural & Technical University**
Title: *Identifying Hurricane Eye using Local Binary Pattern and Edge Feature Extraction*

Contributors: Sohail, Buaba, Fetanat, Homaifar, Mensah (NCAT), Knapp (NOAA/NCDC)

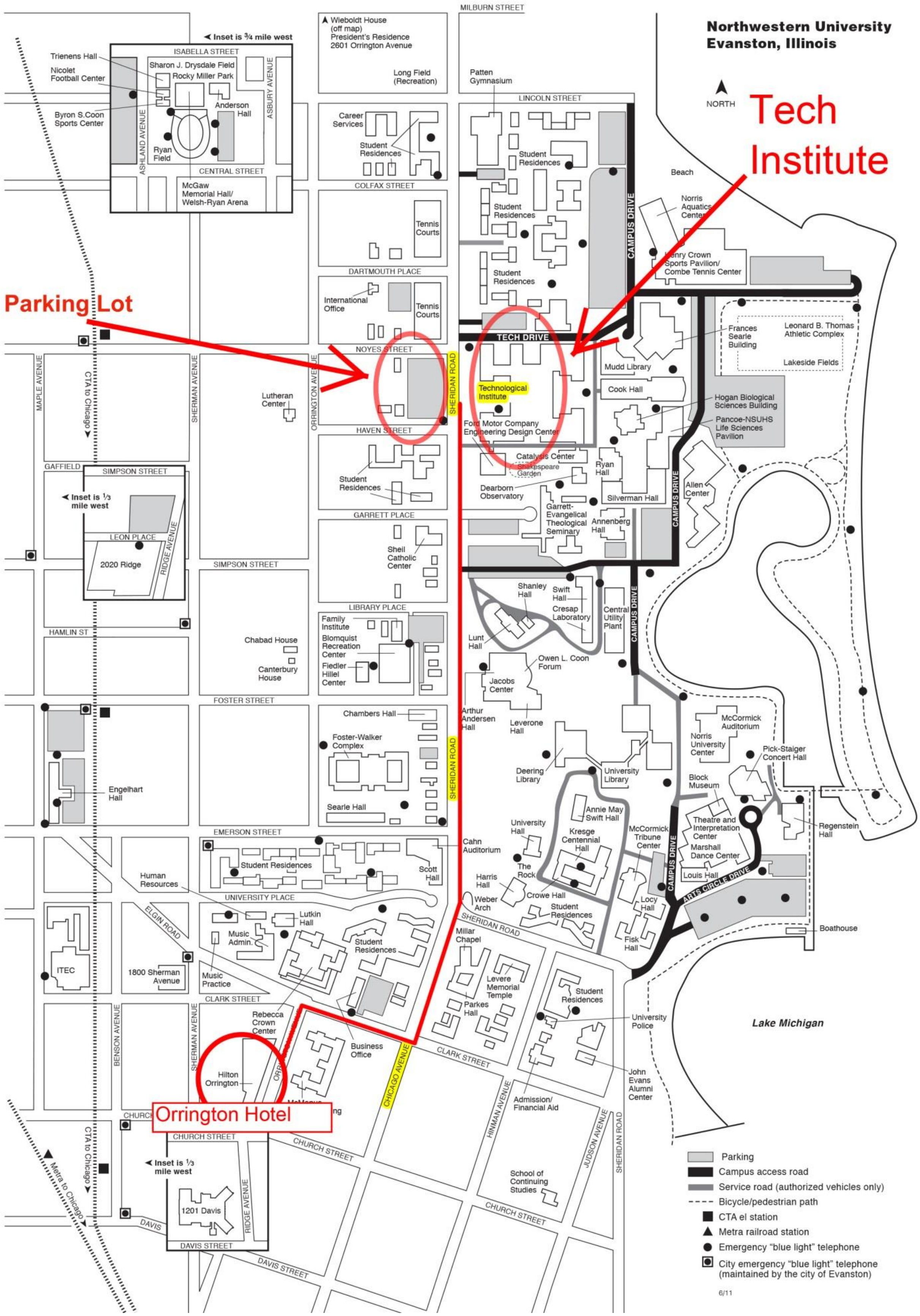
Hurricanes cause massive damage and destruction especially along the coastal areas. One of the best known approaches to track any major hurricane and its predicted path is to map the hurricane eye from image to image, as the hurricane develops. The objective of this research is to identify a developed storm based on its eye recognition and extract orientation-invariant features that can help track the eye as it propagates through the storm. Using brightness temperature (BT) profiles of the imagery, the regeneration of infrared images shows a different color histogram of the eye compared to the rest of the storm. Thus, by using Local Binary Pattern algorithm, the recognition of the hurricane eye can be obtained easily if the eye is well defined. However, for a storm with unclear eye location, the use of Edge Feature Extraction is proposed and investigated. This technique requires identifying a point-of-interest (hurricane eye) and then extracts features based on the edges of the eye wall. These features include orientation-invariant features such as perimeter, area; and the orientation of the eye. Experiments have been conducted on eight storms chosen from the timeline of historical hurricane satellite data (HURSAT), from 2004 to 2008. These storms are found in the North Atlantic, which have major impacts on coastline due to their development into major hurricanes.

Presenter: **Xun Zhou, University of Minnesota**
Title: *Discovering Persistent Change Windows (PCW) in Spatiotemporal Datasets*

Contributors: Zhou, Shekhar, Oliver (UMN)

Given a region S comprised of locations that each have a time series of length |T|, the Persistent Change Windows (PCW) discovery problem aims to find all spatial window and temporal interval pairs <Si, Ti> that exhibit persistent change of attribute values over time. PCW discovery is important for critical societal applications such as detecting desertification, deforestation, and monitoring urban sprawl. The PCW discovery problem is challenging due to the large number of candidate patterns, the lack of monotonicity where sub-regions of a PCW may not show persistent change, the lack of predefined window sizes for the ST windows, and large datasets of detailed resolution and high volume. Previous approaches in ST change footprint discovery have focused on local spatial footprints for persistent change discovery and may not guarantee completeness. In contrast, we propose a space-time window enumeration and pruning (SWEP) approach that considers zonal spatial footprints when sfinding persistent change patterns. We provide theoretical analysis of SWEP's correctness, completeness, and space/time complexity. We also present a case study on vegetation data that demonstrates the usefulness of the proposed approach. Experimental evaluation on synthetic data show that the SWEP approach is orders of magnitude faster than the naive approach.

Wireless access during the workshop:



Northwestern University
Evanston, Illinois
NORTH
Tech Institute
Parking Lot
Orrington Hotel
Technological Institute
Ford Motor Company Engineering Design Center
Hilton Orrington
TECH DRIVE
SHERIDAN ROAD
CHICAGO AVENUE
CAMPUS DRIVE
NOYES STREET
HAVEN STREET
LINCOLN STREET
COLFAX STREET
DARTMOUTH PLACE
GARRETT PLACE
LIBRARY PLACE
FOSTER STREET
EMERSON STREET
UNIVERSITY PLACE
CLARK STREET
CHURCH STREET
DAVIS STREET
MILBURN STREET
ORRINGTON AVENUE
SHERMAN AVENUE
MAPLE AVENUE
BENSON AVENUE
HINMAN AVENUE
JUDSON AVENUE
ELGIN ROAD
ARTS CIRCLE DRIVE
Mudd Library
Cook Hall
Frances Searle Building
Leonard B. Thomas Athletic Complex
Lakeside Fields
Hogan Biological Sciences Building
Pancoe-NSUHS Life Sciences Pavilion
Henry Crown Sports Pavilion/ Combe Tennis Center
Norris Aquatics Center
Beach
Patten Gymnasium
Long Field (Recreation)
Wieboldt House (off map) President's Residence 2601 Orrington Avenue
Student Residences
Career Services
Tennis Courts
International Office
Lutheran Center
Catalysis Center
Shakespeare Garden
Dearborn Observatory
Ryan Hall
Silverman Hall
Allen Center
Garrett-Evangelical Theological Seminary
Annenberg Hall
Sheil Catholic Center
Shanley Hall
Swift Hall
Cresap Laboratory
Central Utility Plant
Lunt Hall
Owen L. Coon Forum
Jacobs Center
Leverone Hall
Arthur Andersen Hall
Family Institute
Blomquist Recreation Center
Fiedler Hillel Center
Chabad House
Canterbury House
Chambers Hall
Foster-Walker Complex
Searle Hall
Engelhart Hall
Deering Library
University Library
McCormick Auditorium
Norris University Center
Pick-Staiger Concert Hall
Block Museum
Theatre and Interpretation Center
Marshall Dance Center
Louis Hall
Regenstein Hall
Annie May Swift Hall
Kresge Centennial Hall
McCormick Tribune Center
University Hall
The Rock
Harris Hall
Crowe Hall
Weber Arch
Locy Hall
Fisk Hall
Boathouse
Cahn Auditorium
Scott Hall
Human Resources
Lutkin Hall
Music Admin.
Music Practice
ITEC
1800 Sherman Avenue
Rebecca Crown Center
Business Office
Millar Chapel
Levere Memorial Temple
Parkes Hall
University Police
John Evans Alumni Center
Admission/ Financial Aid
School of Continuing Studies
Lake Michigan
Inset is ¾ mile west
Trienens Hall
Nicolet Football Center
Byron S.Coon Sports Center
Sharon J. Drysdale Field
Rocky Miller Park
Anderson Hall
Ryan Field
McGaw Memorial Hall/ Welsh-Ryan Arena
ISABELLA STREET
CENTRAL STREET
ASHLAND AVENUE
ASBURY AVENUE
Inset is ⅓ mile west
GAFFIELD
SIMPSON STREET
LEON PLACE
RIDGE AVENUE
2020 Ridge
HAMLIN ST
CTA to Chicago
Metra to Chicago
1201 Davis
Parking
Campus access road
Service road (authorized vehicles only)
Bicycle/pedestrian path
CTA el station
Metra railroad station
Emergency "blue light" telephone
City emergency "blue light" telephone (maintained by the city of Evanston)
6/11