University of Minnesota
Driven to Discover℠

Second Workshop on

Understanding Climate Change from Data

The Annual Meeting of
NSF Expeditions in Computing Award # 1029711

August 6-7, 2012

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

(This page intentionally left blank)

Understanding Climate Change from Data

University of Minnesota – Twin Cities, Minneapolis, MN
3-180 Keller Hall

August 6-7, 2012

Table of Contents

(This page intentionally left blank)

Agenda: Second Workshop on Understanding Climate Change from Data

University of Minnesota - Twin Cities, Minneapolis, MN 3-180 Keller Hall

	Monday Aug. 6, 2012	pg
8:00	Registration & Breakfast	
8:30	Vipin Kumar, University of Minnesota Welcome & Introduction to the NSF Expeditions in Computing on Understanding Climate Change: A Data Driven Approach	37
9:00	NSF Expeditions Research Highlights	
10:00	Q&A	
10:15	Coffee Break	
	Session 1 *Chair: Peter Snyder*	
10:45	David R. Easterling, National Oceanic and Atmospheric Administration *Global Data for Analysis of Climate Extremes*	11
11:05	Lawrence Buja, National Center for Atmospheric Research *Climate 2.0: Usable Climate Science and Services for Society*	12
11:25	Michael Wehner, Lawrence Berkeley National Laboratory *TECA, 13TB, 80,000 processors--- Or: Characterizing extreme weather in a changing climate*	13
11:45	Bruno Sansó, University of California Santa Cruz *Assessing Regional Climate Model Predictions*	14
12:05	Lunch Break	
	Session 2 *Chair: Snigdhansu Chatterjee*	
13:30	Siegfried Schubert, NASA *On the Causes of and Long Term Changes in Eurasian Heat Waves*	15
13:50	Upmanu Lall, Columbia University *Nonlinear Multivariate Projections and long range ENSO predictability*	16
14:10	Steve Easterbrook, University of Toronto *Getting Climate Models and Data into the Open: How hard can it be?*	17
14:30	Ian Foster, Argonne National Laboratory *Big Process for Big Data*	18
14:50	Slobodan Simonovic, University of Western Ontario *System Dynamic Modelling of Interactions within the Society-Biosphere-Climate System*	19
15:10	Coffee Break	

	Session 3 *Chair: Karsten Steinhaeuser*	
15:40	Nitesh Chawla, University of Notre Dame *Collaboratory for Adaptation to Climate Change*	20
16:00	Tom Dietterich, Oregon State University *Machine Learning Methods for Timing of Biological Events*	21
16:20	Claire Monteleoni, George Washington University *Global Climate Model Tracking Using Geospatial Neighborhoods*	22
16:40	Yan Liu, University of Southern California *Sparse Extreme Value Modeling for Climate Change Attribution Analysis*	23
17:00	Closing Remarks	
18:00	Poster Session & Dinner	

Tuesday, Aug. 7, 2012

8:00	Registration & Breakfast	
8:30	Welcome & Recap of August 6th	
	Session 4 *Chair: Shashi Shekhar*	
8:40	Ashok Srivastava, NASA *Scalable Regression for Multimodal Data*	25
9:00	Zoran Obradovic, Temple University *Modeling Multi-Source Remote Sensing Observations of Varying Quality in the Presence of a LargFraction of Missing Values*	26
9:20	J. Ronald Eastman, Clark University *Geotemporal Analysis, Earth System Science and GIS*	27
9:40	Chris Brunsdon, University of Liverpool *Humans as Sensors - Modelling Climate Change with Volunteered Geographical Information*	28
10:30	Coffee Break	
	Session 5 *Chair: Abdollah Homaifar*	
11:00	Praveen Kumar, University of Illinois at Urbana-Champaign *Characterizing Disaster Impact using LIDAR and AVIRIS: 2011 Lower Mississippi Flood Study*	29
11:20	Lian Xie, North Carolina State University *Decipher Complex Climatic Records for Clues of Tropical Cyclone Temporal Variability, Trends and Spatial Pattern*	30

11:40	Zhan Zhang, NOAA *Advancements in Operations and Research on Hurricane Modeling and Ensemble Prediction System at EMC/NOAA*	31
12:00	Lunch Break	
13:15	Panel Discussion: Ana Barros, Bhaskar Chattaraj, Nitesh Chawla, Marvin Geller, Robert Oglesby, Anand Patwardhan, Soroosh Sorooshian	33
15:00	Closing Remarks	

(This page intentionally left blank)

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

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

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)

Invited Speakers, August 6th:

Name: **David R. Easterling** – National Oceanic and Atmospheric Administration

Title: *Global Data for Analysis of Climate Extremes*

Abstract:
Anthropogenic forced climate change is expected to result in changes in both the mean climate state and climate extremes. Documenting changes in extremes, such as drought, heavy precipitation, heat waves or tropical storms requires long-term observations of climate at sufficiently high temporal resolution to identify individual extreme events and is spatially complete enough to use at both global and regional scales. This presentation discusses various data sets that are suitable for examining changes in extremes in the observed record. Issues discussed include problems with data quality, such as data homogeneity and data availability. Data sets discussed include global sets for monthly and daily temperature and precipitation, tropical cyclones, and reanalyses. In addition, a brief discussion of recent research in precipitation extremes is included.

Bio:
Dr. Easterling is currently Chief of the Global Climate Applications Division at NOAA's National Climatic Data Center in Asheville, NC. He received his Ph.D. from the University of North Carolina at Chapel Hill in 1987 and served as an Assistant Professor in the Atmospheric Sciences Program, Department of Geography, Indiana University-Bloomington from 1987 to 1990. In 1990 he moved to the National Climatic Data Center as a research scientist, was appointed Principal Scientist in 1999, and Chief of the Global Climate Applications Division in 2002. He has authored or co-authored more than eighty research articles on climate science in journals such as Science, Nature and the Journal of Climate. Dr. Easterling was a Lead Author on the Nobel Prize winning Intergovernmental Panel on Climate Change (IPCC) Fourth Assessment Report, a Convening Lead Author for the U.S. Climate Change Science Program (CCSP) Synthesis and Assessment Product (SAP) 3.3 on Climate Extremes and was a Contributing Author to the IPCC Second and Third Assessment Reports. He is currently a Lead Author on the IPCC Special Report on Climate Extremes and the IPCC Fifth Assessment Report. Dr. Easterling is a Fellow of the American Meteorological Society and his research interests include the detection of climate change in the observed record, particularly changes in extreme climate events and the assessment of climate model simulations for changes in extreme climate events.

Name: **Lawrence Buja** – National Center for Atmospheric Research
southern@ucar.edu

Title: *Climate 2.0: Usable Climate Science and Services for Society*

Abstract:
As the reality of human-induced climate change is increasingly accepted by policymakers and the public at large, the demand for usable regional-scale information about observed and projected changes and impacts is growing rapidly. The ability to create and provide this information is dependent on improvements in climate models, research frameworks, and analysis tools. This presentation will describe current trends in climate science, climate change modeling and applications. Recent efforts and advancements in climate services across the globe will be discussed.

Bio:
Dr. Lawrence Buja is the Director of NCAR's Climate Science and Applications Program at the National Center for Atmospheric Research (NCAR) in Boulder, Colorado, which carries out interdisciplinary research on social, economic, and political activities related to climate at local, regional and global scales. CSAP addresses impacts, adaptation and vulnerability to climate change by generating scenarios of projected climate change, developing tools and methods for analyzing current and future vulnerability, and conducting integrated analyses of climate change impacts and adaptation.

Dr Buja also serves as the scientific project manager for the Climate Change and Prediction group. This group carried out the climate simulations of the earth's past, present and future climate with NCAR's Community Climate System Model (CCSM) that made up the joint US NSF/DOE submission to the Intergovernmental Panel on Climate Change (IPCC). Together with the CCSM runs carried out on the Japanese Earth Simulator, these CCSM runs were the largest data submission to the IPCC AR4 effort by any modeling center in the world. In addition to carrying out IPCC climate simulations, Lawrence is a contributing author to both the 2001 IPCC Third Assessment Report (AR3) (http://www.grida.no/climate/ipcc_tar/wg1) and the breakthrough 2007 IPCC Fourth Assessment Report (AR4) (http://ipcc-wg1.ucar.edu/wg1/wg1-report.html).

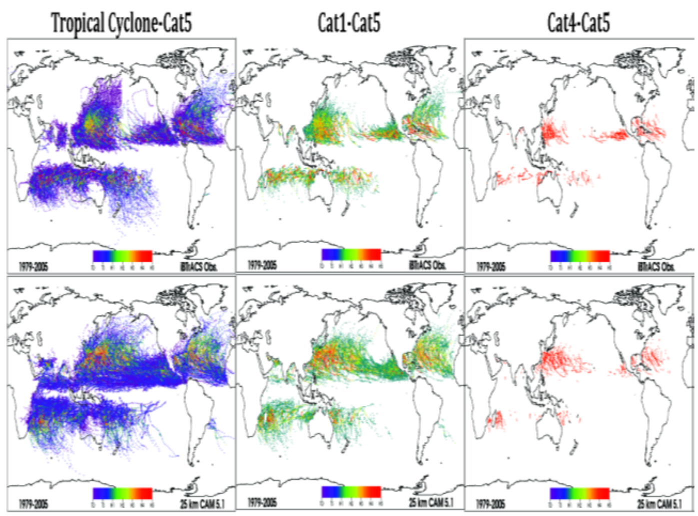
Lawrence also works closely with the World Bank, the InterAmerican Development Bank and other international agencies applying NCAR's climate and regional model expertise to help guide sustainable development investment strategies throughout the developing world.

Name: **Michael Wehner** – Lawrence Berkeley National Laboratory

Title: *TECA, 13TB, 80,000 processors--- Or: Characterizing extreme weather in a changing climate*

Abstract:

Today's highest resolution climate models enable the study of how extreme weather, particularly intense storminess, will change as the overall global climate warms. The analysis challenges posed by this class of problems can be severe involving tens of petabytes of model output and months of computer processor time. We have developed the "Toolkit for Extreme Climate Analysis"--- TECA, to identify extreme storms and other weather events contained in the sub---daily output of high---resolution atmospheric models. We are using this efficient parallel analysis tool to identify tropical and extratropical cyclones, atmospheric rivers and blocking events. We will discuss the techniques involved in event identification and present results from our analyses. Our practical experience in data management and workflow will also be discussed.



Tropical cyclone tracks identified by TECA in a $0.23^{o}x0.31^{o}$ CAM 5.1 simulation (lower row) and IBTraCS observations (upper row). Colors indicate maximum wind speeds on Saffir--- impson scale. Left column: Tropical cyclone strength and larger. Middle column: Hurricane strength and larger. Right column: Intense hurricane strength.

Bio:

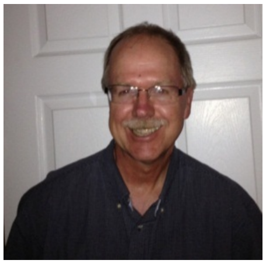
Michael F. Wehner is a staff scientist in the Computational Research Division at the Lawrence Berkeley National Laboratory. Dr. Wehner's current research concerns the behavior of extreme weather events in a changing climate, especially heat waves, intense precipitation, drought and tropical cyclones. Before joining the Berkeley Lab in 2002, Wehner was an analyst at the Lawrence Livermore National Laboratory in the Program for Climate Modeling Diagnosis and Intercomparison. He is the author or co-author of over75 scientific papers. He was also a member of the lead author team for the 2009 White House report, "Global Climate Change Impacts in the United States" and is currently a lead author for both the Fifth Assessment Report of the Intergovernmental Panel on Climate Change and the upcoming 3rd US National Assessment on climate change. Dr.Wehner earned his master's degree and Ph.D. in nuclear engineering from the University of Wisconsin at Madison, and his bachelor's degree in Physics from the University of Delaware.

Name: **Bruno Sansó** – University of California Santa Cruz

Title: *Assessing Regional Climate Model Predictions*

Abstract:
Within the context of the North American Regional Climate Change Assessment Program (NARCCAP) we consider output from the regional climate model RegCM3. This model was implemented by the Paleoclimate and Climate Research Group of the University of California Santa Cruz. Simulations are available for two time periods: current climate conditions, covering 1968 to 2000 and future climate conditions, under the SRES A2 emissions scenario, covering 2038 to 2070. We consider RegCM3 simulations performed using forcings from two different Atmosphere-Ocean General Circulation Models (AOGCM): GFDL and CGCM3. We develop a space-time model for annual summer average daily maximum temperature over a region that covers the southwest of the United States. Our model uses observational records to assess the discrepancies of the two different climate model simulations under present day conditions. Those discrepancies are then propagated into the future to obtain blended forecasts of 21st century climate. The model allows for time-varying spatial heterogeneities, providing local comparisons between the two sets of simulations. Additionally, we estimate the different modes of spatial variability, and use the temporally varying coefficients of the spatial factors for comparisons on a global scale.

Bio:
Bruno Sansó is currently Professor of Statistics and Chair of the Department of Applied Mathematics and Statistics of the University of California Santa Cruz. Professor Sansó obtained his PhD at Universidad Central de Venezuela in 1992. After obtaining the PhD his research activity focused on problems on robust Bayesian inference. More recently he has worked in problems related to model selection, meta-analysis, Bayesian predictive modeling of rainfall in space and time and Bayesian spatio-temporal modeling for environmental and climatological applications. Professor Sansó was part of the Faculty of Universidad Simón Bolívar, Caracas, Venezuela. There he reached the rank of Full Professor at the Department of Scientific Computing and Statistics, of which he was co-founder. In 2001 he joined the newly created Department of Applied Mathematics and Statics at UCSC. He holds active collaborations with researchers from different institutions and has supervised the work of a number of graduate students. One of them won the Savage Award in 2010. Sansó's publications have appeared in the most prestigious statistical journals. One of them won the Annual Prize to the Best Scientific Paper in Engineering by CONICIT, Venezuela, in 2000. Another won the Mitchell Prize awarded in 2009 to a paper presenting an outstanding application of Bayesian methods. Sansó has been Associate Editor of JSPI and Technometrics. He is Editor of Bayesian Analysis. He served as the Treasurer of ISBA. He is elected Member of the ISI and a Fellow of the ASA.



Name: **Siegfried Schubert** – NASA

Title: *On the Causes of and Long Term Changes in Eurasian Heat Waves*

Abstract:
The MERRA reanalysis, other observations, and the GEOS-5 model have been used to diagnose the causes of Eurasian heat waves including the recent extreme events that occurred in Europe during 2003 and in Russia during 2010. The results show that such extreme events are an amplification of natural patterns of atmospheric variability (in this case a particular large-scale atmospheric planetary wave) that develop over the Eurasian continent as a result of internal atmospheric forcing. The amplification occurs when the wave occasionally becomes locked in place for several weeks to months resulting in extreme heat and drying with the location depending on the phase of the upper atmospheric wave. Model experiments suggest that forcing from both the ocean (SST) and land play a role phase-locking the waves. An ensemble of very long GEOS-5 model simulations (spanning the 20th century) forced with observed SST and greenhouse gases show that the model is capable of generating very similar heat waves, and that they have become more extreme in the last thirty years as a result of the overall warming of the Asian continent.

Bio:
Siegfried Schubert is a senior research scientist in NASA's Global Modeling and Assimilation Office (GMAO), where he leads the group on sub-seasonal to decadal climate. He received his Ph. D. in Meteorology from the University of Wisconsin- Madison in 1983. His research interests include climate variability and predictability, droughts, the hydrological cycle, extreme weather and climate events, and reanalysis. Dr. Schubert has authored or co-authored over 100 research articles, and has served as an editor of the Journal of Climate. He is currently a member of the international CLIVAR scientific steering group, and serves on the advisory committee of the Center for Ocean-Land-Atmosphere studies (COLA). As a member of the WCRP drought interest group, he recently helped organize two international workshops on drought, and as the lead of a NOAA drought task force, he is helping to coordinate and facilitate NOAA -funded drought research projects.

Name: **Upmanu Lall –** Columbia University

Title: *Nonlinear Multivariate Projections and long range ENSO predictability*

Abstract:
We demonstrate that a nonlinear decomposition of the thermocline field in the Pacific provides the basis for a 2 year predictability horizon for ENSO using a linear prediction model with the leading nonlinear components identified using maximum variance unfolding. We discuss the identified modes and their relation to different ENSO events and the possible reasons for the long range predictability from different calendar months.

Bio:
My research has been motivated by empirical insights into hydroclimate dynamics through data analysis and theory. It focuses on algorithms for nonlinear dynamical systems and their application to a variety of problems. My background includes hydrology, climate dynamics, nonparametric function estimation, optimization and decision making, and at some point in the distant past numerical modeling. I am also battling water scarcity and policy issues. The systems engineering paradigm of understand, predict, control in a very broad sense captures my approach to problems in somewhat diverse areas.

Name: **Steve Easterbrook –** University of Toronto
http://www.cs.toronto.edu/~sme

Title: *Getting Climate Models and Data into the Open: How hard can it be?*

Abstract:
Open source software and open data are often touted as the holy grail of reproducible computational science. In principle, by making models and data freely available, other scientists can perform their own analysis on the data, and can re-run the code to verify results. The end result is expected to be greater collaboration, wider acceptance of results, and increased trust in the scientific endeavor. In practice, none of this comes easily. In this talk, I will describe a number of challenges that stand in the way of full openness for climate models and data. These include a confusion between repeatability and reproducibility; the challenge of sharing petabyte-sized datasets; the difficulty of capturing full data provenance for highly processed, homogenized datasets; multiple ownership and commercial licensing of meteorological data; the blurring of the distinction between models and observational data; the complexity of description of a model configuration; the challenge of sharing legacy code; the myth of "many eyes" verification; and last, but not least, in a polarized context where scientists are often subject to politically-motivated attacks, the potential for "denial of service" attacks on scientific labs. Having set out these challenges, I will briefly describe the current state of practice for global climate models, and suggest some research goals for improving the situation.

Bio:
Steve Easterbrook is a professor of computer science at the University of Toronto. He received his Ph.D. (1991) in Computing from Imperial College in London (UK), and was a lecturer at the School of Cognitive and Computing Science, University of Sussex from 1990 to 1995. In 1995 he moved to the US to lead the research team at NASA´s Independent Verification and Validation (IV&V) Facility in West Virginia, where he investigated software verification on the Space Shuttle Flight Software, the International Space Station, the Earth Observation System, and several planetary probes. He moved to the University of Toronto in 1999. His research interests range from modelling and analysis of complex systems to the socio-cognitive aspects of team interaction. He has served on the program committees for many conferences and workshops in Requirements Engineering and Software Engineering, and was general chair for RE'01 and program chair for ASE'06. In the summer of 2008, he was a visiting scientist at the UK Met Office Hadley Centre, and in 2010 a visiting scientist at the National Centre for Atmospheric Research in Boulder, Colorado; the Max-Planck-Institute for Meteorology, in Hamburg, and the Institute Pierre Simon Laplace in Paris.

Name: **Ian Foster** – Argonne National Laboratory
foster@anl.gov

Title: *Big Process for Big Data*

Abstract:
We have made much progress over the past decade toward effectively harnessing the collective power of IT resources distributed across the globe. In fields such as high-energy physics, astronomy, and climate, thousands benefit daily from tools that manage and analyze large quantities of data produced and consumed by large collaborative teams.

But we now face a far greater challenge: Exploding data volumes and powerful simulation tools mean that far more—ultimately most?--researchers will soon require capabilities not so different from those used by these big-science teams. How is the general population of researchers and institutions to meet these needs? Must every lab be filled with computers loaded with sophisticated software, and every researcher become an information technology (IT) specialist? Can we possibly afford to equip our labs in this way, and where would we find the experts to operate them?

Consumers and businesses face similar challenges, and industry has responded by moving IT out of homes and offices to so-called cloud providers (e.g., Google, Netflix, Amazon, Salesforce), slashing costs and complexity. I suggest that by similarly moving research IT out of the lab, we can realize comparable economies of scale and reductions in complexity. More importantly, we can free researchers from the burden of managing IT, giving them back their time to focus on research and empowering them to go beyond the scope of what was previously possible.

I describe work we are doing at the Computation Institute to realize this approach, focusing initially on research data lifecycle management. I present promising results obtained to date with the Globus Online system, and suggest a path towards large-scale delivery of these capabilities. I illustrate my discussion with examples from climate science and other disciplines.

Bio:
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.

Name: **Slobodan P. Simonovic –** University of Western Ontario
simonovic@uwo.ca

Title: *System Dynamic Modelling of Interactions within the Society-Biosphere-Climate System*

Abstract:
This lecture presents a tool - the feedback based integrated assessment model ANEMI - that represents the society-biosphere-climate system. This tool provides insight and understanding of complex impacts that climate change has on physical and biological systems and allows for investigation of adaptation strategies through the examination of feedbacks between the climate, economy and the environment. ANEMI is an integrated model that sacrifices resolution for completeness and cannot include the number of physical processes present in Global Climate Models. However, as an integrated model, ANEMI supplies a context for climate change. It integrates nine sectors: climate, carbon cycle, land-use, population, food production, hydrologic cycle, water demand, water quality, and energy-economy. It runs very quickly on personal computers, making multiple simulations and an assessment of model sensitivities possible. Finally, ANEMI improves our understanding of the feedbacks between the socio-economic and climate systems. While it is clear that good governance requires useful and reliable science for decision-making, it is equally clear that the effects of poor planning could be disastrous. Climatic change raises serious economic and social dangers. We must plan well, and good planning requires understanding and foresight. ANEMI model provides support for both. It also provides the Canadian government with a scientifically credible tool, connecting science, governance, economics, and the environment. The policy community should benefit from the results of this research through an enhanced understanding of the interplay between population growth, land use change and climate change impacts on the economy. This enhanced understanding of the impacts should lead to strategies for mitigating the impacts through policy initiatives consistent with Canada's international obligations and domestic objectives. Research in these areas is of critical importance for Canada given the vulnerability of large regions of the country to climate change, and the importance of the oil and gas sector as an element in our economy.

Bio:
Dr. Simonovic is currently a Professor in the Depatment of Civil and Environmental Engineering and the Director of Engineering Studies at the Institute for Catastrophic Loss Reduction . He has over thirty years of research, teaching and consulting experience in water resources systems engineering. He actively works for national and international professional organizations. He has received a number of awards for excellence in teaching, research and outreach. Dr. Simonovic has been invited to present special courses for practicing water resources engineers in many countries. He has published over 350 professional publications and three major textbooks. Dr. Simonovic's primary research interest focuses on the application of systems approach to, and development of the decision support tools for, management of complex water and environmental systems. Most of his work is related to the integration of risk, reliability, uncertainty, simulation and optimization in hydrology and water resources management.

Name: **Nitesh Chawla** – University of Notre Dame
nchawla@nd.edu

Title: *Collaboratory for Adaptation to Climate Change*

Abstract:
Adaptation to climate is a critical component of the climate crisis, motivating an unprecedented coordination and mobilization of data, information, networks and knowledge. The question is not only how to slow or stop climate change, but how to adapt and live with the consequences of climate change. Such adaptation requires an unprecedented mobilization of knowledge about impacts, vulnerabilities, and potential strategies for effective action. Climate change and adaptation discussion go beyond research labs and classrooms, and impacts citizens and community. It thus imposes a need to network people together of different expertise – from computer science to ecology to climate science to policy – dispersed over geography, and share data, information, and knowledge. In this talk, I will present our work on building Collaboratory on Adaptation to Climate Change that is leading the initiative to enable collaboration, analysis, prediction, and discussion on adaptation research and implementation.

Bio:
Nitesh Chawla, PhD is an Associate Professor in the Department of Computer Science and Engineering, and Director of the Interdisciplinary Center for Network Science and Applications (iCeNSA). His research is focused on machine learning, data mining, and network science. He is at the frontier of interdisciplinary applications with innovative work in healthcare informatics, social networks, analytics, and climate/environmental sciences. He is currently leading an NSF Cyber-Enabled Discovery and Innovation (CDI) project on building and studying a virtual organization for adaptation to climate change. He is the recipient of multiple awards for research and teaching innovation including outstanding teacher awards (2007 and 2010), National Academy of Engineers New Faculty Fellowship, and number of best paper awards and nominations. He is the chair of the IEEE CIS Data Mining Technical Committee.

Name: **Tom Dietterich** – Oregon State University
http://web.engr.oregonstate.edu/~tgd/

Title: *Machine Learning Methods for Timing of Biological Events*

Abstract:
The factors controlling the timing of many biological events (e.g., bird migration, moth emergence, fish migration) are poorly understood. This is largely because the timing of the event itself is not observed directly, which makes it difficult to connect the event to possible causal factors. For example, the migration decisions of an individual bird or the exact moment of emergence of an individual moth are not observed with current field methods. Instead, it is common to have observations that consist of counts of the number of organisms active at irregularly-spaced points in time. This talk will discuss a general probabilistic modeling framework, Collective Graphical Models, that provides a way of estimating precise timings from count data. We will describe the methodolgy and show how it can be applied to estimate timings of moth emergence and bird migration. Our long term goal is to understand the biology of the timing of lifecycle events ("phenology") and the role of climate (and the effect of climate change) on the timing of these events.

Bio:
Tom Dietterich is Professor and Director of Intelligent Systems Research at Oregon State University. He is co-PI of the NSF Expedition in Computational Sustainability and co-PI of the NSF BirdCast project, both of which involve the application of machine learning to modeling the distribution and dynamics of species for purposes of conservation. Dietterich is past President of the International Machine Learning Society (the parent organization of the International Conference on Machine Learning) and President-Elect of the Association for the Advancement of Artificial Intelligence (AAAI). He is a Fellow of the ACM, AAAI, and AAAS. His publications are available from his website: http://web.engr.oregonstate.edu/~tgd/

Name: **Claire Monteleoni** – George Washington University

Title: *Global Climate Model Tracking Using Geospatial Neighborhoods*

Abstract:
A key problem in climate science is how to combine the predictions of the multi-model ensemble of global climate models. Recent work in machine learning (Monteleoni et al. 2011) showed the promise of an algorithm for online learning with experts for this task. We extend their Tracking Climate Models (TCM) approach to (1) take into account climate model predictions at higher spatial resolutions and (2) model geospatial neighborhood influence between regions. Our algorithm enables neighborhood influence by modifying the transition dynamics of the Hidden Markov Model from which TCM is derived, allowing the performance of spatial neighbors to influence the temporal switching probabilities for the best climate model at a given location. In experiments on historical data at a variety of spatial resolutions, our algorithm demonstrates improvements over TCM, when tracking global temperature anomalies. This is joint work with Scott McQuade, appearing at AAAI 2012, in the Special Track on Computational Sustainability and AI.

Bio:
Claire Monteleoni is an assistant professor of Computer Science at The George Washington University, which she joined in 2011. Previously, she was research faculty at the Center for Computational Learning Systems, and adjunct faculty in the Department of Computer Science, at Columbia University. She was a postdoc in Computer Science and Engineering at the University of California, San Diego, and completed her PhD and Masters in Computer Science, at MIT. Her research focus is on machine learning algorithms and theory for problems including learning from data streams, learning from raw (unlabeled) data, learning from private data, and Climate Informatics: accelerating discovery in Climate Science with machine learning. Her papers have received several awards. In 2011, she co-founded the International Workshop on Climate Informatics, which she is co-chairing again this year. She serves on the Editorial Board of the Machine Learning Journal, and recently served on the Senior Program Committees of ICML and UAI, 2012.

Name: **Yan Liu** – University of Southern California
yanliu.cs@usc.edu, http://www-bcf.usc.edu/~liu32/

Title: *Sparse Extreme Value Modeling for Climate Change Attribution Analysis*

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

Bio:
Yan Liu is an assistant professor in Computer Science Department at University of Southern California from 2010. Before that, she was a Research Staff Member at IBM T.J. Watson Research Center. She received her M.Sc and Ph.D. degree from Carnegie Mellon University in 2004 and 2006. Her research interest includes developing scalable machine learning and data mining algorithms with applications to climate modeling, social media analysis, computational biology, and business analytics. She has received several awards, including 2007 ACM Dissertation Award Honorable Mention, best application paper award in SDM 2007, and winner of several data mining competitions, including KDD Cup 2007, 2008, 2009 and INFORMS data mining competition 2008. She has published over 50 referred articles and served as a program committee and organizing committee of major machine learning and data mining conferences. Her research work is supported by NSF, DARPA, Yahoo, IBM and ExxonMobil.

(This page intentionally left blank)

Invited Speakers, August 7th:

Name: **Ashok Srivastava** – NASA
ashok.n.srivastava@nasa.gov, https://c3.nasa.gov/dashlink/members/5/

Title: *Scalable Regression for Multimodal Data*

Abstract:
Regression problems on massive data sets are ubiquitous in many application domains including the Internet, earth and space sciences, and finances. In many cases, regression algorithms such as linear regression or neural networks attempt to fit the target variable as a function of the input variables without regard to the underlying joint distribution of the variables. As a result, these global models are not sensitive to variations in the local structure of the input space. Several algorithms, including the mixture of experts model, classification and regression trees (CART), and others have been developed, motivated by the fact that a variability in the local distribution of inputs may be reflective of a significant change in the target variable. While these methods can handle the non-stationarity in the relationships to varying degrees, they are often not scalable and, therefore, not used in large scale data mining applications. In this paper we develop Block-GP, a Gaussian Process regression framework for multimodal data, that can be an order of magnitude more scalable than existing state-of-the-art nonlinear regression algorithms. The framework builds local Gaussian Processes on semantically meaningful partitions of the data and provides higher prediction accuracy than a single global model with very high confidence. The method relies on approximating the covariance matrix of the entire input space by smaller covariance matrices that can be modeled independently, and can therefore be parallelized for faster execution. Theoretical analysis and empirical studies on various synthetic and real data sets show high accuracy and scalability of Block-GP compared to existing nonlinear regression techniques.

Bio:
Ashok N. Srivastava, Ph.D. is the Principal Scientist for Data Sciences at NASA. He currently leads the System-Wide Safety and Assurance Technology (SSAT) Project and was the Principal Investigator of the Integrated Vehicle Health Management (IVHM) project at NASA. He is also the lead of the Data Sciences Group at NASA Ames Research Center. He started DASHLink, a social networking site for people in systems health management, data mining, and related fields. In 2012, he was chosen and the Editor-in-Chief of the AIAA Journal of Aerospace Computing, Information, and Communication. Dr. Srivastava is the author of many research articles in data mining, machine learning, and text mining, and has edited a book on Text Mining: Classification, Clustering, and Applications (with Mehran Sahami, 2009). He is currently editing two more books: Advances in Machine Learning and Data Mining for Astronomy (with Kamal Ali, Michael Way, and Jeff Scargle) and Data Mining in Systems Health Management (with Jiawei Han). Dr. Srivastava has given seminars at numerous international conferences. He has a broad range of business experience including serving as Senior Consultant at IBM and Senior Director at Blue Martini Software. In these roles, he led engagements with numerous Fortune Global 500 companies including Bank of America, Chrysler Corporation, Saks 5th Avenue, Sprint, Chevron, and LG Semiconductor. He has won numerous awards including the IEEE Computer Society Technical Achievement Award for "pioneering work in Intelligent Information Systems," the NASA Exceptional Achievement Medal for contributions to state-of-the-art data mining and analysis, the NASA Distinguished Performance Award, several NASA Group Achievement Awards, the IBM Golden Circle Award, and the Department of Education Merit Fellowship.

Name: **Zoran Obradovic** – Temple University
www.dabi.temple.edu/~zoran

Title: *Modeling Multi-Source Remote Sensing Observations of Varying Quality in the Presence of a Large Fraction of Missing Values*

Abstract:
We will first describe our method for information fusion and learning from remote sensing data of varying quality based on Continuous Conditional Random Fields. When applied to aerosol optical depth (AOD) retrieval using data of varying quality, this method successfully exploited spatio-temporal properties of observations and outperformed existing statistical and deterministic AOD retrieval models. Next, we will discuss our method for learning from spatio-temporal data with a large fraction of missing values. This method outperformed alternatives when imputing up to 80% of nonrandom missing values in aerosol observations. Using together imputed and observed data we have developed a predictor of aerosol optical depth that was much more accurate than predictors based on alternative imputation methods when tested rigorously over the entire continental US. Finally, we will show how to completely avoid the data imputation step when learning form partial observations by taking into account the uncertainty of each instance due to the missing values. Our method (to be published at AAAI 2012 conference) was shown to outperform the alternatives when there is a large fraction of missing values in data (more than 25%).

Bio:
Zoran Obradovic is professor of Computer and Information Sciences and the director of the Center for Data Analytics and Biomedical Informatics at Temple University in Philadelphia. He has published about 250 articles addressing data mining challenges in health informatics, climate and ecological management and other domains. Obradovic is an editorial board member at seven journals and serves as the executive editor at the journal on Statistical Analysis and Data Mining, which is the official publication of the American Statistical Association (ASA). He was the program chair at six and track chair at seven data mining conferences and is the general co-chair for the SIAM Data Mining (SDM) conference in year 2013 and 2014.

Name: **J. Ronald Eastman** – Clark University

Title: *Geotemporal Analysis, Earth System Science and GIS*

Abstract:
It has become increasingly apparent that the earth system has entered a highly dynamic and transitional phase. With a goal of monitoring and understanding these changes a wide variety of earth observing systems have been put in place providing a wealth of environmental image time series. However, the development of tools to analyze these data has not kept pace. While the Geographic Information Science community has been very active in the development of spatial analysis procedures, the character of geotemporal data and the needs of earth system/climate scientists are very different from those of GIS. GIS is primarily synthetic in character whereas geotemporal analysis is truly analytic. GIS is layer oriented whereas geotemporal analysis is multidimensional, with surprising implications for analytical operations which depend upon a decomposition of variance. Earth System Science is particularly focused on interactions in coupled systems. Thus it is not uncommon to want to analyze data with five dimensions or more. With geotemporal data, noise is substantial and can at times be larger than the signal. Further, while the objects of GIS are fixed in space, the objects of geotemporal data move and evolve. These concepts are illustrated through an exploration of a new suite of software tools known as the Earth Trends Modeler. The Earth Trends Modeler is a geotemporal system that provides elements for seasonal and inter-annual trend analysis, series decomposition based on Principal Components, Multichannel Singular Spectrum Analysis and Empirical Orthogonal Teleconnection analysis, Fourier and Wavelet spectral analysis and lagged linear modeling. The application of these tools is illustrated using a range of multi-decadal earth observation time series.

Bio:
J. Ronald Eastman is Professor of Geography in the Graduate School of Geography at Clark University. He is also the Director of Clark Labs – a research unit focused on geographic/earth system information technology to address the needs of effective decision making for social and environmental security and sustainable development. Dr. Eastman is best known as the developer and chief architect of the IDRISI, Land Change Modeler and Earth Trends Modeler software systems.

Name: **Chris Brunsdon** – University of Liverpool
christopher.brunsdon@liverpool.ac.uk

Title: *Humans as Sensors - Modelling Climate Change with Volunteered Geographical Information*

Abstract:
There is a long tradition of volunteers collecting and reporting different types of information about the environment we live in. One example is the data set considered here, which consists of Volunteered Geographical Information (VGI) and data provided by expert researchers monitoring the first bloom dates of lilacs from 1956 to 2003 is used to investigate changes in the onset of the North American spring. In one sense, this can be regarded as data provided by sensors - where the sensors were human beings. It is argued that care must be taken when analysing data of this kind, with particular focus on the issues of lack of experimental design, and of Simpson's paradox. Approaches used to overcome this issue make use of random coefficient modelling, and bootstrapping approaches. It is also argued that some issues experienced here may be used for the analysis of hardware-based sensors, where issues such as breakdowns incremental installation may present similar problems in terms of data modelling and analysis. Once the suggested methods have been employed, a gradual advance in the onset of spring is suggested by the results of the analysis. This talk will outline the 'story of the data', the methods used for analysis, and the outcome of the analysis.

Bio:
Chris Brunsdon is currently Professor of Human geography at the University of Liverpool. He was previously Professor of Geographical Information at the University of Leicester from 2005 until 2011. He received a BSc in Mathematics from the University of Durham in 1984 where he was awarded a Horsfall Scholarship, an MSc. (with distinction) in Medical Statistics from the University of Newcastle upon Tyne in 1985, and a PhD in Geography from the same university in 1990. He is a Fellow of the Royal geographical Society. His interests include the methodologies underlying spatial statistical analysis and geographical information systems, and their application in a number of subject areas, in particular the modelling of crime patterns, house prices and health. More recently, these interests have extended to the analysis of climate data.

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

Title: *Characterizing Disaster Impact using LIDAR and AVIRIS: 2011 Lower Mississippi Flood Study*

Abstract:
The mega-flood in the lower Mississippi River in Spring 2011 created an opportunity to study the impact of large-scale disaster on the landscape. Extensive data were collected during and after the event, which include post-flood Lidar mapping of the landscape, and AVIRIS and MASTER imaging spectroscopy data using NASA ER2 aircraft. Pre-flood lidar has been made available by US Army Corps of Engineers (Memphis) through an MOU. Analyses are performed to understand the geomorphic and geochemical impact of large scale flooding on the landscape using the AVIRIS and MASTER data in conjunction with Lidar and other ancillary data such as flow and soil sample analyses. Our hypothesis is that the geochemical signatures of the flood impact are strongly correlated with the geomorphic disturbance induced on the landscape due to the high flows and associated erosion-deposition patterns. The research results focus on the characterization of the patterns of geomorphic impact for interpretation of geochemical signatures, and investigation of biogeochemical impacts. At O'Bryan's ridge (south of Wyatt, Missouri) there was significant scouring immediately following the flood as the rapidly flowing water encountered an extra 4 ft (1.2 m) of depth. Several hundred hectares of crop land were eroded into deep gullies. O'Bryan's Ridge was created by the relic of historical meandering of the Mississippi and the scouring during this flood were seaded by existing gullies. Difference between pre- and post-flood lidar provide us with spatial patterns of erosion and deposition across the floodplain. Preprocessing of AVIRIS and MASTER data is performed involving: orthorectification, atmospheric corrections using ATCOR4 software, and radiometric normalisation. Subsequently, the spectral signatures of AVIRIS and MASTER are correlated with laboratory analysis of soil grab samples which were obtained coincident with ER2 flights. The laboratory analyses identify soil particle size distribution and chemical composition. We are studying soil chemistry using the AVIRIS data by correlating with laboratory chemical analysis of soil samples and spectral matching techniques using the Tetracorder and Specpr technologies developed by USGS. Initial results will be presented at the meeting.

Bio:
Praveen Kumar is the Lovell Professor of Civil and Environmental Engineering. Dr. Kumar is currently 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 improve our understanding of hydrologic processes over a range of space and time scales with particular emphasis on understanding and modeling multiple scale non-linear interactions among sub-processes.

Name: **Lian Xie** – North Carolina State University

Title: *Decipher Complex Climatic Records for Clues of Tropical Cyclone Temporal Variability, Trends and Spatial Pattern*

Abstract:
Data doesn't cheat, but the same dataset can lead to drastically different and even opposite conclusions as a result of incompleteness of the data and different perspectives from the observer. In this presentation, examples will be given to illustrate how opposite conclusions on the variability and trends of the climate in recent decades may be derived from the same climatic dataset. Such an example highlights the need for more comprehensive statistical tools for data analysis. A method, known as Hilbert-Huang Transform (HHT), which can be used to extract essential climatic signals at various frequencies, is introduced and its applications in tropical cyclone study is illustrated. We will then provide an example of the application of statistical procedures in analyzing two-dimensional tropical cyclone track data and the development of prediction models for seasonal tropical cyclone activity.

Bio:
Dr. Xie is currently a professor in meteorology and physical oceanography at North Carolina State University. He received his B.S. and M.S. degrees in meteorology from Nanjing Institute of Meteorology, China in 1982 and 1985, respectively, and Ph.D. in Meteorology and Physical Oceanography from the University of Miami, Florida in 1992. He joined North Carolina State University faculty in 1992. Since then, he has directed or co-directed 35 externally funded projects with a total project funding exceeding $20 million on topics including air-sea interactions, modeling and predictions of the regional, particularly coastal, atmospheric and oceanic processes. He was a member of the U.S. Weather Program Prospectus Development Team on Landfalling Hurricanes and on Coastal Meteorology, and received two awards from the National Weather Service for helping to improve the forecasts of storm surge. In August 2008, he led an effort to provide real-time hourly surface wind forecasts for the sailing competition of the Olympic Game in Qingdao, China, and received a "Certificate of Appreciation Award" by Qingdao Meteorological Bureau. His has published over 100 peer-reviewed journal articles and book chapters. He has been frequently invited as guest speaker or lecturer at international workshops including the "International Tropical Cyclone Disaster Reduction Training Workshop" organized by the World Meteorological Organization of the United Nations in March 2007 and the "Asia-Pacific Regional Training Course on Ocean Dynamics and Climate" organized by UNESCO/International Ocean Commission in 2011. He was named China's prestigious "Chang Jiang (Cheung Kung) Scholar" by the Educational Ministry of China in 2007.

Name: **Zhan Zhang** – Environmental Modeling Center, NOAA
Zhan.zhang@noaa.gov

Title: *Advancements in Operations and Research on Hurricane Modeling and Ensemble Prediction System at EMC/NOAA*

Abstract:
The atmosphere-ocean coupled NCEP/EMC operational Hurricane Weather Research and Forecast (HWRF) modeling system has been providing numerical guidance on tropical cyclone tracks and intensity to the hurricane forecasters at NHC for Atlantic and Eastern Pacific basins since 2007. Through a major collaborative effort supported by NOAA's Hurricane Forecast Improvement Project (HFIP), for the first time, a high-resolution hurricane model operating at cloud-permitting 3km resolution was implemented into NCEP operational system in 2012. Retrospective testing of the high-resolution HWRF system showed significant improvement in track, intensity and structure forecasts for 2010-2011 hurricane seasons compared to previous versions of the operational HWRF. A challenging problem in hurricane modeling is to reduce uncertainties in hurricane vortex initialization and model physics to further improve track and intensity forecasts. Ensemble prediction system (EPS) has proved to be effective in improving the forecast skills of numerical models. The key to a successful hurricane ensemble forecast is to generate a set of ensemble that effectively represents the uncertainties in hurricane model prediction system. This talk reviews the ongoing operational and research efforts at NCEP/EMC in advancing the operational HWRF modeling system. Results from the retrospective and real-time high resolution HWRF forecasts will be presented along with ongoing efforts on various methods for ensemble generation and post-processing techniques. Multi-model ensembles, multi-physics ensembles, Kernel Density Estimation, cluster analysis, and assessment of forecast skills from these ensembles will be discussed.

Bio:
Zhan Zhang is a research scientist at Environmental Modeling Center (EMC) at National Centers for Environmental Prediction (NCEP). Zhang received his BS and MS degrees in Meteorology from Beijing University, P.R. China, and the Ph. D. degree in Tropical Meteorology from the Florida State University in 1997. He has worked at several major operational weather forecasting centers, including National Meteorological Center (NMC) of China, European Center for Medium-range Weather Forecasts (ECMWF), and Environmental Modeling Center at NOAA. His research interests include hurricane modeling, hurricane ensemble prediction system, and ensemble post-processing method. He has developed some innovative methods of analyzing huge data sets generated by hurricane ensembles, including EOF based perturbation method which improved hurricane track and intensity forecasts. He co-authored a research paper in Science on multi-model superensemble. He is actively involved in generating effective and efficient ensembles for operational hurricane model system, multi-model/multi-physics ensembles, and forecasting forecast skills using hurricane ensembles. He is also involved in regional and vortex scale data assimilation using hybrid approach of Grid Statistical Interpolation (GSI) and Ensemble Kalman Filter (EnKF).

(This page intentionally left blank)

Panel Discussion: Big Data for Climate Science

Moderators: **Arindam Banerjee & Peter Snyder** – University of Minnesota

The climate sciences have become increasingly important in addressing the impacts of climate change on Earth's inhabitants. Significant progress has been made over the last three decades in understanding the physical processes contributing to climate change, attributing the role of greenhouse gases and feedbacks to planetary warming, and in mitigating additional warming of the planet. However, over the last decade new challenges fueled by societal demands have emerged that test the limits of traditional climate science methodologies. The inclusion of an increasing amount of data from satellites, ground observations, and models has made the problem more challenging and new tools are needed to process and interpret this data. Computational big data approaches that can handle large spatio-temporal data have shown promise both in the detection of new and important climate behavior as well as in the generation of hypotheses that the climate science community can then investigate further with physics-based methodologies.

The goals of the panel discussion are:

(i) To identify bridges between physical understanding and data-driven analysis of the climate system.

(ii) To identify methods, models, techniques which may be useful in the analysis of climate data and the understanding of climate change.

(iii) To identify new areas where data mining research needs to advance to address the challenges posed by climate data analysis problems.

(iv) To provide methods for successfully answering policy questions related to impacts of climate change.

The panelists include some of the eminent researchers and experts working on different aspects related to climate change at the intersection of computational sciences, statistics, climate sciences, and climate change impacts: Ana Barros, Bhaskar Chattaraj, Nitesh Chawla, Marvin Geller, Robert Oglesby, Anand Patwardhan, and Soroosh Sorooshian.

<u>Panelists</u>

Name: **Ana Barros** – Duke University
ousf@duke.edu

Bio:
Ana Barros is a Professor of Engineering at Duke University. Dr. Barros was in the engineering faculty at the University of Porto, Penn State University, and Harvard University before joining Duke University in 2004. Through model development, field and laboratory experiments, and data analysis, her goal is to facilitate and promote bridges across environmental science and engineering, physics, applied mathematics, and information sciences. Her primary research interests are in Hydrology, Hydrometeorology and Environmental Physics with a focus on water-cycle processes in the coupled land-atmosphere-biosphere system particularly in regions of complex terrain, the study of multiscale interface phenomena in complex environments across the Earth Sciences, remote sensing of the environment (precipitation, clouds, soil moisture, and vegetation), climate predictability and risk assessment of natural hazards. She received a Ph.D. from the University of Washington in 1993. She also received an MSc in Environmental Science from OHSU, and an MSc in Ocean Engineering and a Diploma in Civil Engineering with double major in Structures and Hydraulics

from UPorto. Dr. Barros has been actively involved in NRC committees with the Space Studies Board and Board in Atmospheric Sciences and Climate. She is a Senior Member of the IEEE and a Fellow of the American Meteorological Society. Dr. Barros currently serves as Chief Editor of the Journal of Hydrometeorology.

Name: **Nitesh Chawla** – University of Notre Dame
Please see bio on page 20.

Name: **Bhaskar Chattraj** – Tigerrisk Partners
bchattaraj@tigerrisk.com

Bio:
Bhaskar began his career in the insurance industry in 1998 at AIR Worldwide Corporation, where he was the Manager of the Core Products Quality Group. In that position he was responsible for statistical analysis and testing of all catastrophe models and underlying industry exposures. Additionally, Bhaskar played an integral role in the development of a proprietary catastrophe software engine and was involved in special client risk management projects such as providing analytical expertise in the design of catastrophe bonds.
Prior to joining TigerRisk Partners in 2009, Bhaskar was the Assistant Vice President in the Analytics department at John B. Collins Associates, Inc. In that capacity he was responsible for building analytical tools for catastrophe risk analysis, creating tools and processes for optimizing portfolios of catastrophe-prone exposures including depopulations from Citizens Property Insurance Corporation in Florida, reinsurance and ILW pricing, and creating growth and start-up scenarios for several companies. Bhaskar currently leads the Catastrophe Risk Modeling Team at TigerRisk Partners.
Bhaskar earned both a Bachelor's degree and Master's degree in Statistics from Indian Statistical Institute and a Master's degree in Economics from Boston University.

Name: **Marvin Geller** – Stony Brook University

Bio:
Marvin A. Geller is a Professor in the School of Marine and Atmospheric Sciences at Stony Brook University. His research career has spanned more than 4 decades, and his research has extended from climate influences of soil moisture to the behavior of the ionospheric E-region. He has authored well over 100 refereed publications in leading journals on many topics, but over his career his principal specialty has been on atmospheric dynamics, and more specifically on wave motions in the atmosphere, subjects in which he continues to be active. In recent years, he has been working on parameterizations of sub-grid scale atmospheric gravity waves in climate

models, among other things. He is a Fellow of both the American Meteorological Society and the American Geophysical Union, and has received several other honors, among which are NASA's Distinguished Public Service medal and COSPAR's International Cooperation Medal. He has supervised 18 PhD students to completion to-date. He did his undergraduate studies in Applied Mathematics and his PhD studies in Meteorology, both at MIT.

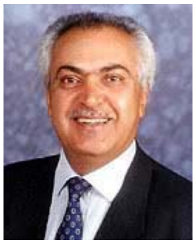
Name: **Robert Oglesby** – University of Nebraska, Lincoln

Bio: Robert 'Bob' Oglesby is a Professor of Climate Modeling at the University of Nebraska, Lincoln, with joint appointments in the Department of Earth and Atmospheric Sciences and the School of Natural Resources. Bob received his BS in Physical Geography from the University of California, Davis in 1985 and his PhD in Geophysical Fluid Dynamics from Yale University in 1990. Prior to arriving at UNL in 2006, he was a senior scientist for 5 years at NASA and prior to that spent 10 years on the faculty of Purdue University. Bob's research interests include the causes of drought, the impact of deforestation on climate, and key mechanisms of climate change, both past and future. He has authored or co-authored over 100 refereed journal papers and book chapters on these subjects. Bob is also currently involved with in-country training in the development and use of high-resolution climate change models for vulnerability and impacts studies in Central America and Asia.

Name: **Anand Patwardhan** – Indian Institute of Technology Bombay

Bio:
Anand Patwardhan is Professor at SJMSOM and served as the Executive Director of the Technology Information, Forecasting and Assessment Council, an autonomous organization in the Ministry of Science & Technology. He holds adjunct faculty positions at Carnegie Mellon University and at the School of Public Policy, Georgia Tech.

Anand has worked extensively with industry, government and international organizations as a consultant and an advisor. He was a lead author of the third & fourth assessment reports of the Intergovernmental Panel on Climate Change and the Millennium Ecosystem Assessment. He is a member of the Scientific and Technical Advisory Panel (STAP) of the Global Environment Facility (GEF); of the Scientific Steering Committee of the Global Carbon Project and of the Expert Committee on Climate Change of the Government of India.



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

Bio:
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; 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.

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.

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

Bio:
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.

Name: **Ankit Agrawal**– Northwestern University
ankitag@eecs.northwestern.edu

Bio:
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.

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

Bio:
Arindam Banerjee is an Associate Professor at the Department of Computer & and Engineering and a Resident Fellow at the Institute on the Environment at the University of Minnesota, Twin Cities. 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, Information Theory, 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, the J. T. Oden Faculty Research Fellowship from the Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, 2006, and the prestigious IBM PhD fellowship for the academic years 2003-2004 and 2004-2005. He has also won several awards for his publications, including the Best Paper Award at the SIAM International Conference on Data Mining (SDM), 2004, the Best Research Paper Award under University Cooperative Society Research Excellence Awards, University of Texas at Austin, 2005, and the Best of SIAM Data Mining (SDM) Award at the SIAM International Conference on Data Mining, 2007.

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

Bio:
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.

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

Bio:
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.

Name: **Alok Choudhary** – Northwestern University
choudhar@eecs.northwestern.edu

Bio:
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.

Name: **Jonathan Foley** – University of Minnesota
jfoley@umn.edu

Bio:
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.

Name: **Auroop Ganguly** – Northeastern University
a.ganguly@neu.edu

Bio:
Auroop R. Ganguly is an associate professor of civil and environmental engineering who heads Northeastern's Sustainability and Data Sciences Laboratory. His recent research interests at ORNL have focused on climate extremes, uncertainty and impacts, as well as hydrometeorology and surface water hydrology, in addition to transportation and infrastructure security, remote and *in-situ* sensor data analysis, and the supply chain. His methodological interests include nonlinear modeling, time series and spatial statistics, extreme value theory, data mining, complex model evaluation, risk analysis, uncertainty quantification, and decision sciences. Following his graduate studies, Prof. Ganguly worked on prediction and uncertainty problems in science and business for 12 years in a DOE National Laboratory, academia, and in the private sector. He has published in the *Proceedings of the National Academy of Sciences, Physical Review E., Transportation Systems,* and the *Journal of the Operational Research* Society, as well as edited a book entitled, Knowledge *Discovery from Sensor Data*. Prof. Ganguly earned a Ph.D. in Civil and Environmental Engineering from the Massachusetts Institute of Technology in 2002, an M.S. in Civil Engineering from the University of Toledo in 1997, and a B.Tech. With Honors in Civil Engineering from the Indian Institute of Technology , India, in 1993. Ganguly is a member of the American Geophysical Union, American Meteorological Society, Sigma Xi, Institute of Electrical and Electronics Engineers, and the American Society of Civil Engineers.

Name: **William Hendrix** – Northwestern University
whendrix@northwestern.edu

Bio:
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.

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

Bio:
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.

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

Bio:
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.

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

Bio:
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.

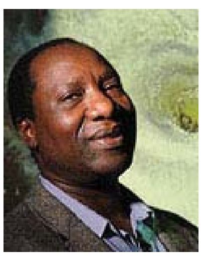
Name: **Stefan Liess** – University of Minnesota
liess@umn.edu

Bio:
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.

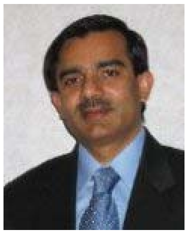
Name: **Nagiza Samatova** – North Carolina State University
samatova@csc.ncsu.edu

Bio:
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



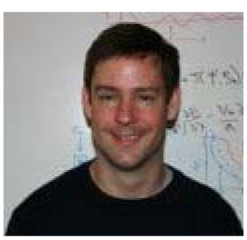
Name: **Fredrick Semazzi** – North Carolina State University
fred_semazzi@ncsu.edu

Bio:
Dr. Semazzi has served in several senior positions at 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; & Professor at North Carolina State University, Raleigh, North Carolina, 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; 2009-2012). 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 on twelve research grants, with total funding of more than $25 million.



Name: **Shashi Shekhar** – University of Minnesota
shekhar@cs.umn.edu

Bio:
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 260+ 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 program co-chair for the Intl. Conference on Geographic Information Science (2012). Earlier, he served on multiple National Research Council 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 co-chair for the Intl. Symposium on Spatial and Temporal Databases (2011) and ACM 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). More details are available from http://www.cs.umn.edu/~shekhar.



Name: **Peter Snyder** – University of Minnesota
pksnyder@umn.edu

Bio:
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.

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

Bio:
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.

Name: **Karsten Steinhaeuser** – University of Minnesota
ksteinha@umn.edu

Bio:
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

Monday, August 6, 6-8 PM
Listed in Alphabetical order of Last Name of Presenter

Presenter: **Gonzalo Bello, NCSU**

Title: *The Influence of the Eastern Africa Climatic Conditions in Modulating Atlantic Hurricane Variability*

Contributors: Bello, Harlalka, Waniha, Samatova, Semazzi (NCSU)

Previous studies have identified NAO, AO, Niño, Atlantic Dipole, AMO and other well-known climate indices as major sources of variability of Atlantic Hurricane activity.

This study was, in part, motivated by the recent discovery of an important source of the African Easterly Waves (AEWs) over the Ethiopian Highlands in close vicinity to the Indian Ocean. Yet none of the known sources of variability of Atlantic Hurricane activity is attributed to the Indian Ocean and the adjacent land mass. Among many different atmospheric variables, near-surface relative humidity has been shown to be one of the best predictors of Hurricane seasonal activity in the Pacific Ocean, but it has not been effectively exploited for Atlantic tropical cyclones. For this reason, in our study we adopted relative humidity at surface level as our primary atmospheric variable for the investigation.

The poster presents correlation and causality analysis as evidence of the important influence of the Indo-Pacific region SSTs (source region) on the African Easterly Waves genesis region over the Ethiopian Highlands, which in turn have profound influence on the Atlantic Hurricane activity. A new index based on relative humidity over Eastern Africa outperforms all the other known major climate indices in terms of its unique contribution in explaining the variability of Atlantic Hurricane activity. It is apparent from this study that comprehensive understanding of the variability of the Atlantic Hurricanes should include consideration of climate variability over Eastern Africa and the role of the Indo-Pacific Ocean.

Presenter: **Ruben Buaba, NCAT**

Title: *Fast Locality Sensitive Hashing Algorithm For Approximate Nearest Neighbor Search: A Practical Data Mining Approach*

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

We present a new Locality Sensitive Hashing (LSH) scheme to build a data structure for the Defense Meteorological Satellite Program (DMSP) imagery dataset. The LSH is an index-based data structure that allows spatial item retrieval over a large dataset. The data structure is to find similar images to a given query image and should do this in sub-linear search time. Searching a large dataset to find similar images to a query image using the Linear Search (LS) is computationally exhaustive, especially, if the images have high dimensionality. This is due to the fact that LS iterates through the entire dataset to find the similar images to a given query image. In the existing LSH scheme, each data sample is randomly projected onto k-dimensional subspace. The k then determines the number of hash tables (L) that should be created. We theoretically demonstrated that the existing LSH scheme drives up the computational cost, the memory requirement and the query runtime. We therefore propose a new LSH scheme in which the number of tables (L) rather determines the choice of the k-dimensional subspace. Experiments conducted on 1.6 million DMSP satellite images have shown that our proposed scheme saves more than 50% on memory requirement; cuts the computational cost by more than 50% and improves the query runtime by more than a factor of 2.

Presenter: **Soumyadeep Chatterjee, UMN**
Title: *Hierarchical Sparse Regression Models: Climate Applications*

Contributors: Soumyadeep Chatterjee, Banerjee, Snigdhanshu Chatterjee (UMN), Ganguly (NEU)

Understanding the influence of ocean atmospheric variables on land variables, such as rainfall, is one of the key challenges in climate science. Improved understanding and advancement in predictive modeling of precipitation events have high value for policymakers in taking future policy decisions. Although physical models of the atmosphere provide partial understanding of processes, such as that of rainfall formation, high complexity and nonlinearity of the atmospheric system demands novel advancements through statistical methods to provide insights into predictive modeling. In this work, we model the prediction of precipitation as a regression problem. We consider temporal measurements of atmospheric variables as covariates. We utilize a recently proposed sparse regression model, called Sparse Group Lasso (SGL), for predictive learning. SGL belongs to a general class of hierarchical tree-structured norm regularized M-estimators. We provide statistical consistency guarantees for this general class of sparse models, which are competitive with the best-known guarantees for sparse regression models. Our experiments demonstrate the utility of using sparse models for predictive modeling, as they are efficient to optimize, significantly control model complexity and are empirically found to be robust in covariate selection.

Presenter: **Debasish Das, NEU**
Title: *Data-driven characterization and prediction of extremes*

Contributors: Das, Kodra, Ganguly (NEU), Kumar, Chatterjee, Banerjee (UMN), Obradovic (Temple)

Understanding and predicting precipitation extremes is important due to the deep science challenges as well as the widespread impacts on flood hazards and water resources decisions. Unfortunately, processes pertaining to precipitation are among the least well understood and therefore precipitation is one of the climate variables that are not predicted well by the physics-based climate models. However, it has been observed that precipitation extremes tend to have a dependence on atmospheric variables ranging from temperature, humidity and precipitable water, to updraft velocity and horizontal wind components. We propose three different approaches towards understanding precipitation extremes based on three different sources of information. First, we exploit the ancillary climate variables (or covariates) that are either observed or predicted well from the physics-based models for information content about the precipitation extremes. We introduced spatially penalized L1-regularised regression to simultaneously identify the variables that influence precipitation extremes at each individual location and discover a neighborhood within which these influence is active.
In our second approach we propose to address the non-linear relationships between precipitation extremes and the covariates by taking into account the physics knowledge about the dominant processes influencing the complex phenomenon of extreme precipitations. Going forward, we plan to use the indicator variables of dominant processes, which are non-linear functions of the ancillary climate variables, to characterize and predict precipitation extremes. In our third approach we plan to estimate the spatial and temporal dependence structure among extremes and exploit this dependence structure to predict future occurrences of extremes. However normal correlation measures may not be

appropriate here since they do not capture the co-occurrence pattern in the extremes. Instead, we may have to use tail dependence measures. Finally, we plan to integrate all three approaches for a better descriptive and predictive model for extremes.

Presenter: **Lindsey Dietz, UMN**
Title: *Modeling the Relationship of Wind Speed and Pressure in Atlantic Tropical Cyclones*

Contributors: Dietz, Chatterjee (UMN)

Prior research has suggested that the relationship between minimum pressure and maximum wind speed in a tropic cyclone can be estimated by the physical model: $\max WindSpeed = \delta(\mu - \min Pressure)^{\xi}$ where $\delta = \sqrt{200}$, $\mu = 1013$, and $\xi = 0.5$. This research examines the strengths and shortcomings of this model and tests the physical parameter estimates via least squares and maximum likelihood estimation. Additionally, we investigated modeling each variable by extreme value distributions and used general linear modeling to incorporate these distributions.

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

Contributors: Faghmous, Le, Liess, Kumar (UMN), Mesquita (BCCR), Ganguly (NEU), Semazzi (NCSU)

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 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, UMN**
Title: *Spatio-Temporal Data Mining for Scalable Ocean Eddy Monitoring*

Contributors: Faghmous, Styles, Mithal, Boriah, Liess, Kumar (UMN), Vikebø (NIMR), Mesquita (BCCR)

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, NCAT**
Title: *Tropical cyclone intensity estimation using temporal analysis of satellite data*

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

Tropical cyclones (TCs) are becoming an increasing threat to life and property. Developing an automated technique to estimate TC intensity and to overcome the existing errors in estimation is still a challenge. Dvorak technique is the state-of-the-art method that has been used over three decades for estimating the intensity of a tropical cyclone. We have developed and tested an automated method for estimating TC intensity based on just the existing historical data. A case study using North Atlantic Hurricane Satellite data from 1978-2006 is considered. Our intensity estimation algorithm has two parts: temporal constraints and image analysis. However, we focus on the temporal constraints. Temporal information provides a priori estimates of the intensity of the storm. The temporal analysis uses the age of the cyclone, 6, 12 and 24 hours prior intensities as predictors of the expected intensity. The 10 closest analogs (determined by a K-nearest-neighbor algorithm) are averaged to estimate the intensity. Several validation tests are conducted to statistically justify the proposed algorithm using K-Fold Cross-Validation. The resulting average root mean squared error (RMSE) of our algorithm is approximately 4.6 knots compared to 11.7 knots from the Dvorak technique on the same dataset. To analyze the effect of noise, a Gaussian noise of zero mean with 5 knot and 6 hour standard deviations are considered for the prior intensities and duration respectively. The results indicate that the average RMSE is approximately 8.2 knots compared to 11.7 knots from the Dvorak technique. Overall, 30% to 55% improvement has been achieved compared to Dvorak technique.

Presenter: **Ilias Fountalis, GaTECH**
Title: *A network-based method for the study of climate teleconnections*

Contributors: Fountalis, Dovrolis, Bracco (GaTECH)

Network analysis provides a powerful, but only marginally explored framework to validate climate models. It refers to a set of tools, algorithms and metrics used to study complex non-linear dynamical systems. Its main premise is that the underlying topology of a system has a strong impact on its dynamics and evolution. Here we present a fast, scalable cutting edge toolbox to infer networks from climate systems. Our goal is to uncover relations in the climate system that might not be captured by more traditional techniques while at the same time explaining known climate phenomena in terms of network structure or its metrics.
Our approach is also novel in terms of the network definition. We identify areas, the nodes of our network, which are clusters of cells highly homogenous in respect to the underlying climate variable. Links between these areas capture teleconnections in the climate system. Our network is robust to noise naturally existing in such spatio-temporal data sets , to the choice of the correlation measure used to access the strength of the teleconnections as well as to the choice of the reanalysis data set. We analyze the network structure of both reanalysis data sets and CMIP 5 model outputs. We quantify how each model reproduces major teleconnections and identify common or specific errors in comparing model outputs and observations. We therefore are able to discuss sources of uncertainties in climate models, clustering models with qualitatively different topological characteristics as distinct climate scenarios.

Presenter: **Qiang Fu, UMN**
Title: *Drought Detection on Large Scale Precipitation Datasets: A Parallel Algorithm*

Contributors: Fu, Wang, Banerjee, Liess, Snyder (UMN)

Droughts are one of the most damaging climate-related hazards. Due to the importance of understanding droughts, we consider the problem of their detection based on gridded datasets of precipitation and formulate it as one of the finding the most likely configuration of a Markov Random Field. Motivated by the fact that gridded precipitation datasets are usually of large scale with at least millions of variables, we propose an efficient parallel drought detection algorithm. We apply this algorithm to solve a drought detection problem based on the Climate Research Unit precipitation dataset with more than 7 million variables. The empirical results show that the algorithm successfully identifies the major droughts of the twentieth century in different regions of the world. We also implement the parallel algorithm using Open MPI and the experimental results demonstrate that the parallel algorithm scales nicely with the number of cores used. When we use ten computing cores, the drought detection problem with more than 7 million variables can be solved in about 15 minutes.

Presenter: **Mohamed Gebril, NCAT**
Title: *Multi-Objective Optimization for Predictive Modeling Selection*

Contributors: Sohail, Gebril, Homaifar (NCAT)

Support vector machines (SVMs) are widely used in pattern recognition applications. Finding the optimal model selection for real-world problems is vital for prediction tasks. Most real-world problems involve multiple objectives therefore optimizing those objectives simultaneously is very crucial. Evolutionary approaches have been successful in solving these Meta heuristic problems. In this work, we propose a multi-objective optimization (MOO) method based on meta-modeling prediction using non-dominant sorting genetic algorithm (NSGAII). The NSGAII is a type of MOO method in which each objective function of support vector regression is represented optimally to find the trade-off hyper-plane between objective variables. Since there is always a need for a good kernel function, carefully chosen trade-off parameter that controls the trade off between allowing training errors that enforce rigid margins, and less number of support vectors reduces the computational complexity that produces a final pareto optimal front. The optimization of these functions is essential. The epsilon-intensive loss function controls the training error that includes a distance measure. Pareto optimal fronts are then used by the decision maker to select the preferred optimal model. We discuss a way to select additional experimental data for revising objective functions and exploring correlation among them. Finally, we illustrate the effectiveness of the proposed method through several climate datasets of different scales and attributes to measure the mean absolute error (MAE) and the root mean squared error (RMSE).

Presenter: **Mohamed Gebril, NCAT**
Title: *Indexing Satellite Images via Learning*

Contributors: Gebril, Homaifar (NCAT)

The ability to search for a specific region of interest in a vast collection of images has become a growing problem. To overcome the lack of accuracy and speed for large-scale object recognition, different combinations of features with careful selection of attributes, measures and semi-supervised classification modules were tested and provided much better performance than using single feature extraction method. The hybrid system of Classification/Searching methods is capable of handling multi-classes' case with comparable computational complexity both in training and at run time, and has a good retrieval performance in practice. We developed a Multi-Level Indexing (MLI) framework that can effectively index imagery based on its feature's attributes to reduce the sample size significantly for the Image Matching (IM) comparisons for retrieval. MLI can effectively tackle the drawback of parameters tuning such as the number of hash tables of Locality Sensitive Hashing and the computational cost of SVM (support vector machines). Different combinations of features with careful selection of attributes, measures and classification modules provided much better performance. In order to avoid over-fitting, we proposed a probability max-margin using probabilistic latent semantic analysis (pLSA) to address optimization problem. We showed an average improvement in precision of 18.7% was achieved using MLI.

Presenter: **Nikai Gibson, NCAT**
Title: *Visualizing Improved Accuracy in Automated Eddy Detection and Tracking*

Contributors: Gibson (NCAT), Styles, Faghmous (UMN)

Vortical eddies account for a large majority of the mesoscale variability in sea surface height of the world's oceans. The state-of-the-art algorithm for detecting eddies in global ocean data first identifies closed contours in sea surface height anomaly snapshots, then stitches these features together between consecutive time frames to form persistent eddy tracks. However, the extremely dynamic nature of eddies injects significant inaccuracy into the results generated by the current method. In the spatial domain, this is often the result of a failure to accurately discern distinct eddies when they occur in tightly packed groups. In stitching over time, the algorithm is limited by an inability to account for the splitting or merging of objects. We present concrete examples demonstrating the challenges inherent to the separate spatial and temporal components of this approach. We demonstrate through case studies how our work makes significant improvement in discerning discrete eddies within a single time frame, and make the argument that re-framing the problem to allow for eddy merging and splitting would greatly improve accuracy in future stitching procedures.

Presenter: **Mohammad Gorji-Sefidmazgi, NCAT**
Title: *Non-Stationary Time Series Clustering for Meteorological Systems*

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

Due to development of computational and measurement capabilities in meteorology and climate research, amount and precision of raw data in the form of time series has increased in recent years. Many of these time series have transitions between different local regimes. The aim of non-stationary time series clustering is to identify these persistent regimes and also construct a dynamical model for each regime. In this poster, we apply Finite Elements Method (FEM) for clustering of meteorological time series. The FEM-K-trend method, where K is the number of clusters, is a newly developed method in literature. In contrast to FEM methods, most of the non-stationary time series clustering methods assume local stationarity of time series. On the other hand, FEM method can find clusters even when there is time trend in the regime without any explicit assumption about probabilistic models (like Gaussian, etc.). Here, instead of finding local stationary regimes, the concept of model distance functional is defined and a convex linear combination of these functionals is used to present the dynamical model of the time series. In this study, we present the FEM-K-trend method to find the regimes in time series which contains information of temperature from different points in North Atlantic and Europe in 5 years. This time series is multi-dimensional and there is a need to find local regimes and linear time trends in each regime.

Presenter: **William Hendrix, NW**
Title: *High performance data clustering*

Contributors: Agrawal, Choudhary, Hendrix, Liao, Patwary, Rangel (NW)

Clustering is a fundamental data mining technique for identifying natural groupings or structure within a dataset. We have developed several high performance clustering techniques as part of an effort to provide a library of high performance implementations of standard data mining kernels. In particular, we have created highly scalable implementations of density-based and hierarchical clustering algorithms, both of which exhibit speedups of thousands on thousands of MPI processes. Additionally, we have developed AGORAS, a clustering algorithm related to k-medoids based on data sampling. AGORAS is unique in that its runtime is dependent on the distribution of the data, but its performance is independent of the full data size. We believe that these clustering algorithms have wide applicability across a number of problems, both in the climate domain and beyond.

Presenter: **Megan Heyman, UMN**
Title: *Statistical Dimension Reduction Analysis*

Contributors: Heyman, Chatterjee (UMN)

The central goal of climate analysis techniques like EOF and SVD is dimension reduction. Finding the correct number of significant relationships is of utmost importance for correct interpretation. Traditional EOF analysis (used, for example, in the computation of Arctic Oscillation) is conducted assuming independent observations with linear relationships. Dimension reduction techniques (linear and non-linear) such as Ridge regression are examined in terms of the statistical sampling distribution of parameters. Theoretical and simulation results are presented.

Presenter: **Zhe Jiang, UMN**
Title: *Learning Spatial Decision Tree for Geographical Classification: A Summary of Results*

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

Given learning samples from a spatial raster dataset, the geographical classification problem aims to learn a decision tree classifier that minimizes classification errors as well as salt-n-pepper noise. The problem is important in many applications, such as land cover classification in remote sensing and lesion classification in medical diagnosis. However, the problem is challenging due to spatial autocorrelation. Existing decision tree learning algorithms, i.e. ID3, C4.5, CART, produce a lot of salt-n-pepper noise in classification results, due to their assumption that data items are drawn independently from identical distributions. In contrast, we propose a spatial decision tree learning algorithm, which incorporates spatial autocorrelation effect by a spatial information gain (SIG) measure and neighborhood wise tree node test. The proposed approach is evaluated in a case study on a remote sensing dataset from Chanhassen, MN. Case study results show that the proposed approach outperforms the traditional approach in not only reducing salt-n-pepper noise but also improving classification accuracy.

Presenter: **Jaya Kawale, UMN**
Title: *Graph Based Analysis of Dynamic Teleconnections*

Contributors: Kawale, A Kumar, Ormsby, 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 dipole interconnections on the globe in a given dataset and thus makes it possible to study changes in dipole interactions and movements. We further show that a modified framework can help us identify delayed relationships in any given dataset and show that our approach is able to identify MJO like behavior. We also present a novel algorithm to test the significance of the teleconnection patterns in any given dataset so as to remove some of the spuriously correlated regions. Using our framework, we are able to identify most of the known dipoles as significant. Further, we are able to identify some new connections that appear as significant and we are investigating one such phenomenon near Australia.

Presenter: **Evan Kodra, NEU**
Title*: Exploring the future of United States fresh water availability: uncertainty from climate and population change*

Contributors: Kodra, Kumar, Ganguly (NEU) Parish, Sorokine, Branstetter, Erickson III, Singh (ORNL) Steinhaeuser (UMN)

The future of fresh water availability is recognized as a global challenge. It will be critical to understand uncertainties in regional changes for decision makers in sectors such as water resources management, energy generation, civil and international conflict, and food security. Here we explore two key components of uncertainty that need to be characterized as fully as possible, namely, changes in population and climate. Using a suite of previous (CMIP3) and current (CMIP5) generation global climate models as well as two population change storylines, we find spatial variability in projected United States water availability and large uncertainties exemplified both by model disagreement as well as assumptions regarding future population dynamics. A comprehensive approach for characterizing and quantifying uncertainty in future fresh water availability from these two sources is needed.

Presenter: **Arjun Kumar, UMN**
Title: *A New Teleconnection : The Australian Southern Oscillation*

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

A possibly new teleconnection has been discovered off the east coast of Australia in the region around Tasman sea and Southern Ocean. Found in pressure anomalies using a novel graph based approach called shared reciprocal nearest neighbors, this dipole appears in reanalysis datasets such as NCEP, JRA, ERA and MERRA. In contrast, the HadSLP2 observations do not show the new dipole, because they do not include observations in the Tasman Sea. Tests are performed in order to understand the uniqueness of the dipole and its relationship to existing well known phenomena. The dipole index is correlated with known dipole indices such as the SO, AAO with which it shares a marginally high correlation of less than 0.4 and other northern teleconnections with which it is shown to have a poor relationship. We limit further analysis with only the AAO and SO indices as these are spatially close, have a higher correlation with the new index and tend to influence it in one or more seasons. Seasonal analysis is done to look at the variation in strength as well as its influence on other variables such as TAS (Temperature at Surface), OLR (Outgoing Longwave Radiation), Precipitation etc. We also look at composite maps and do significance tests to determine the significant regions in these maps. We also determine regions that are influenced by the new dipole index alone and are not influenced by other dipoles namely the SO and AAO by looking at difference maps. Our tests have shown that we may indeed be looking at a new phenomenon and further tests are being conducted to confirm that.

Presenter: **Devashish Kumar, NEU**
Title: *Regional and seasonal mean precipitation and temperature intercomparison across CMIP3 and CMIP5 climate model ensembles*

Contributors: Kumar, Kodra, Ganguly (NEU)

We examine the hypothesis that the new generation of climate models, the Climate Modeling Intercomparison Project version 5 (CMIP5) suite of models improves over the previous generation, CMIP3 suite. The bases of the hypothesis are that higher grid resolutions, more sophisticated physics, or comprehensive earth system model components, may enhance finer scale projections. The ensemble of models is evaluated in terms of their ability to simulate precipitation and near-surface air temperature at regional and seasonal scales. The metrics considered for the evaluation include past performance skills, quantified here through bias maps, as well as model agreement, assessed through inter-model comparisons. The bias maps use the Global Precipitation Climatology Project (GPCP) for precipitation and NCEP-II for temperature as reference data. Other metrics considered in the study are meridional profiles, multi-model bounds, and mean square errors or differences. Historical performance and future model agreements are based on multimodel ensemble statistics as well as pairwise comparisons of model versions in CMIP3 and CMIP5. To compare model projections in the future, greenhouse gas emissions scenarios and concentration pathways in CMIP3 and CMIP5 are selected to enable effective comparisons of projections across the model versions. While the models and scenarios across CMIP3 and CMIP5 may not always be directly comparable, our premise is that projections still need to be evaluated across versions to understand the science, improve predictions, and inform adaptations. Here we employ 11 carefully selected model-pairs, multi-model ensembles, and emissions scenarios across the two model generations. For future projections, of particular interest are regions and seasons where the sign of the change differs within or across model generations, as well as any changes in multi-model ensembles. Preliminary results suggest that newer generation global circulation models do not appear to improve significantly either in terms of skills (past performance) or consensus (model agreement). However, in certain cases, the possibility of solidifying important insights may be suggested. In addition, regional and seasonal differences appear to persist between model pairs across generations, for example, over regions such as South America and parts of Africa and Asia.

Presenter: **Rachindra Mawalagedara, UNL**
Title: *The Climatic Effects of Deforestation in South and Southeast Asia*

Contributors: Mawalagedara, Oglesby (UNL)

Deforestation can have strong effects on climate by modifying the surface albedo, evapotranspiration and exchange of latent and sensible heat between the surface and the atmosphere, thus altering the spatial and temporal patterns of temperature and precipitation. In order to understand the impacts of deforestation on the climate, Weather Research and Forecasting model (WRF) was used along with forcing from NCEP/NCAR Reanalysis data (NNRP). The study focused on South Asia, Southeast Asia and Sri Lanka, all regions where the Asian monsoon plays an important role in determining the regional climate and deforestation is currently or potentially a major issue. A control run as well as two idealized runs were carried out and analyzed for each of the three years: 1988, 1991 and 1993. The two idealized runs consisted of a completely deforested run where all the land-use categories other than inland water were replaced with grassland and a completely forested run where evergreen broadleaf forest was used. These changes, while extreme, represent the maximum possible range of changes due to deforestation. Simulations show that the climatic conditions over land regions become warmer and drier in response to deforestation with precipitation, evapotranspiration and cloud cover all showing a decrease. Analysis of the moisture budget shows that amounts of precipitation over South Asia and Sri Lanka are much greater than local evapotranspiration. This demonstrates the importance of the contribution from an external moisture source (e.g., the Indian Ocean) to the regional precipitation. All the changes seen in the simulations are present in both annual and monsoon season values, suggesting that the impacts due to deforestation have the ability to override even the strongest monsoon signal.

Presenter: **James Middleton, NCAT**
Title: *Understanding Relationships between Fire Season Severity and Sea Surface Temperature Anomalies*

Contributors: Middleton (NCAT), Blank, Karpatne, Steinhaeuser, Boriah, Steinbach, Kumar (UMN)

Finding relationships between climate drivers (e.g. El Niño Index) and natural disturbance events (both oceanic and land cover based) is of prime concern, especially for understanding diverse phenomena of Earth’s complex systems. Such drivers might induce non-linear influences on target events with varying temporal lags and spatial coverage in different geographic and climatic regions, making the problem complex and challenging. In a recent work by Chen et al., relationships between Fire Season Severity (FSS) in South America and Sea Surface Temperature (SST) Anomalies in the Atlantic and Pacific Ocean were studied from 2001 to 2009 and utilized for predicting FSS at future instances. Specifically, they used Oceanic Niño Index (ONI) and Atlantic Multidecadal Oscillation (AMO) as climate indices responsible for affecting FSS. Their scheme comprised of transforming the driver time series at multiple time lags and measuring its Pearson's Correlation with an annually aggregated FSS time series observed at a particular spatial location. We extend and explore the findings in their work by implementing a Graphic User Interface (GUI) for visualizing the intermediate transformation steps involved and the obtained correlation values in greater detail, at varying spatial resolution and temporal scales. We further present randomization experiments to statistically analyze the significance of the obtained correlations results. Future work would focus on extending and improving the underlying machinery for finding relationships, which would honor data characteristics and pattern semantics, such as spatial coherence, temporal consistence, and inherent count data characteristics of FSS. Incorporating other climate indices and capturing their multivariate influence behavior on FSS at larger spatial and temporal scales should also be pursued.

Presenter: **Abhishek Nandy, UMN**
Title: *Model Selection: An approach based on Bias Variance tradeoff*

Contributors: Nandy, Chatterjee (UMN)

In this poster we will focus on devising a generic algorithm for the selection of models including but are not limited to regression model selection, covariate selection, random and mixed effect models, bandwidth selection or non parametric and semi parametric curve estimation. The problem of model selection is a well studied problem in both the Frequentist and Bayesian paradigms working reasonably well in regression setup. Our aim is to extend the notion of model selection from regression models to a wider class of models that satisfy some regularity conditions. Also the other aim of this study is to conceptualize rigorously the ideas of model variance and model bias to be able to use them as criterion for selecting models based on given data. Another aspect of this problem is that sufficient emphasis has been laid on the applications of this novel methodology on climate data especially problems related to model selection in extreme-value modeling, with application to precipitation.

Presenter: **Neeti Neeti, Clark**
Title: *Characterizing implications of two-dimensional space-time orientations for Principal Components Analysis of geographic time series: a case study of sea surface height in the equatorial Pacific*

Contributors: Neeti, Farrar, Eastman (Clark)

Availability of various remotely sensed spatio-temporal datasets offers opportunities to develop and analyze sophisticated space-time models for studying climate change and their impacts. Converting these datasets into knowledge systems requires both sophisticated methods and good understanding of the datasets. These datasets can be described as a cube with latitude, longitude and time as three dimensions. The variability within a dataset is commonly analyzed using Principal Components Analysis (PCA). This research introduces new approaches for PCA to investigate space-time variability in an image time series by using the concept of tensors for decomposing the cube. Using the concept of tensors, an image time series can be analyzed using six different orientations by grouping the basic elements of the cube called voxels along different dimensions. Voxels grouped along the columns or rows of the cube result in profiles. Voxels grouped across different planes result in slices. The traditional S- and T-mode PCA are thus the profile- and slice-modes across time and space respectively. This research introduces two profile-modes and two slice-modes across and along longitude and latitude respectively. The study, carried out using 7-day anomalies of the AVISO Sea Surface Height product for 1997-1999 in the tropical Pacific region filtered for Tropical Instability Waves (TIW), shows that profile-mode across longitude and slice-mode along longitude were able to capture the propagation characteristics of TIW. The other two orientation modes can detect dominant latitudinal locations for TIW.

Presenter: **Zachary Pearson, NCAT**
Title: *Predicting Rainfall Distribution Using Quantile Regression*

Contributors: Pearson (NCAT), Chatterjee (UMN)

Extreme rainfalls are believed to be happening around the world since the beginning of time. It has high impacts on human population through flooding, loss of electricity, mudslides, water damage to buildings and cars. We are doing research to predict rainfall distributions and thus infer the probability of occurrence of rainfall extremes. We will be doing our studies over the United States of America. We will be using 7 different variables as covariates and formulate the problem as a regression problem. The variables we consider are air temperature at 5 different levels, geopotential height, updraft velocity at 5 different levels, precipitable water, specific humidity at 5 different levels, surface pressure at 5 different levels, and horizontal wind speeds. Since wind has both direction and speed, we consider two components, uwind and vwind, and consider these components as different covariates. In order to create our model, we will use regression analysis. This will help us get a good line of fit for the data. We will implement quantile regression in order to predict different regimes of rainfall events, such as normal and extreme rainfall. It will separate out data into different percentiles to define the regimes, and we will fit a different regression model for each different quantile. Accurate prediction of percentiles of rainfall is an important first step in order to do a two-stage predictive modeling, prediction of rainfall regimes, followed by prediction of the actual rainfall distribution.

Presenter: **Mark Rogers, UC-Berkeley**
Title: *Dynamic Tensor Systems for Weather Prediction*

Contributors: Rogers, Li (UCB)

Given the temperature traces of a spatial grid, how to find patterns in the data and predict the future trend? In our work, we propose a novel model for spatial temporal time series, dynamic tensor systems (DTS). Our model can capture the spatial correlations among temperature sequences from multiple locations as well as temporal dynamics. In addition, our model can detect correlations among multiple sources of sensors, including wind, temperature, humidity, and precipitation. Our proposed DTS models the hidden dynamics in the data sequence as a series of latent tensor states that evolve over time according to a Markov chain. The transition of the latent state, as well as the linear mapping from latent states to observed tensor data, e.g., temperature and wind measurements, are modeled using tensor linear-Gaussian distributions. We have developed algorithms to estimate the latent factors and predict the next few time ticks of weather data. We performed experiments on both simulated weather datasets and a real tropical sea surface temperature dataset. Our preliminary evaluation shows promising results in predicting temperatures for the next few time ticks.

Presenter: **Scott Sellars, UC-Irvine**
Title: *Object Based Methods and Machine Learning Techniques to Remote Sensing Precipitation Data for Knowledge Discovery*

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

A new object oriented precipitation database is developed and investigated in terms of the precipitation object characteristics and different time varying climate phenomena that impact regional hydroclimate by using a Gaussian Mixture Model (GMM). Instead of employing the traditional data analysis approaches, which look at information at each data point within a data set, this poster will demonstrate the application of a connectivity algorithm developed by Dr. Wei Chu (CHRS Researcher), to a high resolution near global precipitation dataset. This connectivity algorithm segments precipitation events into four dimensional objects (latitude, longitude, time, and intensity) and stores the information in a database. The precipitation data used for populating the object oriented database is the Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks (PERSIANN) .25 degree dataset, which ranges from 60N to 60S and from March 1st, 2000 to September 30th, 2009. A descriptive statistics algorithm is also developed and applied to all objects contained within the database. These descriptive statistics will be stored in a Nxd data matrix, where N is the number of objects and d is the number of dimensions (or features/characteristics) and will be used as input to the GMM. Preliminary results will be presented describing the precipitation characteristic structure of the precipitation objects by using the unsupervised learning techniques (e.g. GMM), which can learn the intrinsic structure contained in the Nxd descriptive statistic database. This structure will be compared with different spatio-temporal climate phenomena for gaining new knowledge between changes in regional precipitation and climate.

Presenter: **Rahni Sumler, NCAT**
Title: *Exploring Extreme Precipitation Events using Spatial-Temporal Hot-Spot Analysis.*

Contributors: Sumler (NCAT), Zhou (UMN)

Climate extreme events are rare events, defined by some to occur much less than 5% of the time. However these events, such as extreme precipitations, are believe to occur more often in the environment of global warming. Not only did these events cost many lives, the change in frequency and intensity changes the definition of flood, drought, and storms. Analyzing the precipitation data to understand change pattern of extremes is thus important to societal applications such as infrastructure design and policy making. However, the change trend and rate of precipitation extremes are not globally uniform. They are various in different places and in different time, which makes it hard to capture the regional pattern.
Existing work on precipitation extreme data analysis have not sufficiently explored spatiotemporal(ST) data analysis techniques, such ST hot-spot detection, and thus may not accurately capture all the regional patterns. In contrast, we propose to apply spatial-temporal hot spot detection techniques, such as the Local Indicator of Spatial Association (LISA) and Scan Statistics to find significant changes of precipitation extremes at regional scales. The analysis is carried out on the CRU precipitation time series data with 100 years' length. The results will also be visualized in forms of maps and time series plots to assist validation by domain experts.

Presenter: **Pascal Waniha, NCSU**
Title: *Influence of the North Atlantic Oscillation on the Sub-Saharan African Seasonal Climates Reveals Multiple Pathways*

Contributors: Tetteh, Pendse, Gonzalez, Waniha, Semazzi, Samatova (NCSU), Kumar (UMN)

Previously established North Atlantic Oscillation (NAO) forcing over northern Africa suggested a homogeneous regional-scale climate response, obscuring its actual impacts and pathways at different spatial scales. Empirical evidence is hereby provided to show that multi-year NAO variability impacts the Sahel Region, Western Sahel, and Eastern Sahel boreal summer climates differently. A combined multi-scale, multi-phenomena annual cycle impact analysis of global ocean surface temperature and tropospheric features exclusively capture intraseasonal-to-subseasonal NAO variability from winter to mid-spring. This detection is one of the primary sources of variability in the summer climates significantly attributed to different temporal variability unique to each spatial scale, hence separating the dynamics. The corresponding general circulation patterns reveal spatially coherent, but distinct upper level flows that teleconnect to the Himalayas anticyclonic variability. Mechanisms associated with anomalously wet and dry conditions vary along preferential pathways, as the NAO evolves through diverse spatial scales and interacts with distinct oceanic-atmospheric modes.

Presenter: **Pascal Waniha, NCSU**
Title: *West African Seasonal Climate Variability and Predictability*

Contributors: Tetteh, Chen, Waniha, Semazzi, Samatova (NCSU)

Satisfactory understanding of the variability of the West African Seasonal Climate and its predictability remains an elusive problem in climate science. A multi-scale, multi-phenomena methodology, that takes into account, ocean and atmospheric conditions has been developed to investigate the Western Sahel seasonal (July-September; JAS) using data for 1950-2008. The primary objective of this study is to understand the relationships between the West African Seasonal Climate and the extra tropical North Atlantic Oscillation (NAO) which is investigated through combined consideration of three time scales, namely, monthly (intraseasonal), bimonthly (subseasonal), and seasonal (trimester).
The poster will present results from our phenomenological model, which catalogues the best set of predictors from an initial pool of 60 for investigating the predictability of the region's climate. Multi-year surface relative humidity (RH) ensemble forecasts for April 1987-2010 over Ghana using just the monthly timescale predictors, up to a 12-month lead time based on one, five, and ten-year moving window forecasts, and using the FORECASTER Model will also be presented.
The model achieves high accuracy (0.6 - 0.8), with the five and ten-year forecasts generally being more skillful than the one year. Results from this study (model) may be applied in the management and planning of important social-economic sectors such as health and agriculture. Furthermore, the diagnostic tools developed in this study have important applicability to the validation of IPCC climate change projection models.

Presenter: **Sen Yuan, UMN**
Title: *The Asymptotics of M-estimators by Minimizing a Sequence of Convex Functions with Applications in Extreme Climates*

Contributors: Yuan, Chatterjee (UMN)

The climate extremes such as hurricane, extreme precipitation are of great concerns in climate science. The data analysis methodologies mainly depend on generalized extreme value distribution and its related models. Instead of considering the maximum value of the sample, the top k largest values always contain more information about the tail behavior of the distribution. Our studies on M-estimators can be regarded as a much more general case which deals with the choice for k. We focus on a general frame work which can fit in the scenarios of extreme values if we specify our model to be a certain kind. By solving a sequence of convex functions, we obtain a sequence of M-estimators that eventually converge to different distributions according to difference choices of rate of one tuning parameter. Therefore, the rate of that tuning parameter reflects the ideas on how to the top k extreme values with a right asymptotic distribution.

Presenter: **Xun Zhou, UMN**
Title: *Discovering Interesting Sub-paths in Spatiotemporal Datasets*

Contributors: Zhou, Shekhar, Mohan, Liess, Snyder (UMN)

Given a spatiotemporal (ST) dataset and a path in its embedding spatiotemporal framework, the goal of the problem is to to identify all interesting sub-paths defined by an interest measure. Sub-path discovery is of fundamental importance for understanding climate changes, agriculture, and many other applications. However, this problem is computationally challenging due to the massive volume of data, the varying length of sub-paths and non-monotonicity of interestingness throughout a sub-path. Previous approaches find interesting unit sub-paths (e.g., unit time interval) or interesting points. By contrast, we recently propose a Sub-path Enumeration and Pruning (SEP) approach that finds collections of long interesting sub-paths. Two case studies using climate change datasets show that SEP can find long interesting sub-paths which represent abrupt climate change. We provide theoretical analyses of correctness, completeness and computational complexity of the SEP approach. We also provide experimental evaluation of two traversal strategies for enumerating and pruning candidate sub-paths. This research adds to our recent work by offering a new computational strategy, namely, the sub-path enumeration and pruning with pruning border approach (SEPPER). Analytical results show that our new SEPPER approach is correct and complete. Result from new experimental evaluation with both synthetic and real data show that our new approach outperforms the SEP approach in our previous work. We also present a new case study with eco-climate dataset that verifies the effectiveness of the approaches.

Wireless access during the workshop:

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

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