Climate Program Transition Workshop
May 14-16, 2018

SPEAKER TITLES/ABSTRACTS

Brian Blanton
RENCI

"Introduction and Overview of the SAMSI CLIM Risk and Coastal Hazards Working Group"

This short introduction to the SAMSI CLIM Risk and Coastal Hazards Working Group session will provide an overview of the topics discussed over the year, the diversity of the group's expertise and interests, and the research activities undertaken by several of the group's members. Details of three activities are presented in more detail in this session.

Won Chang
University of Cincinnati

"Ice Sheet Model Calibration using Zero-Inflated Continuous Spatial Data"

Rapid changes in Earth's cryosphere caused by human activity can lead to significant environmental impacts. Computer models provide a useful tool for understanding the behavior and projecting the future of Arctic and Antarctic ice sheets. However, these models are typically subject to large parametric uncertainties due to poorly constrained model input parameters that govern the behavior of simulated ice sheets. Computer model calibration provides a formal statistical framework to reduce and quantify the uncertainty due to such parameters. Calibration of ice sheet models is often challenging because the relevant model output and observational data take the form of zero-inflated continuous spatial data; handling of such data type is one of the open problems in the field of spatial statistics. In this work, we will introduce potential ideas to overcome the computational and inferential challenges by combining the logistics principal component analysis (LPCA) and likelihood based principal component analysis. I will demonstrate the fidelity of our approach using the example of calibrating PSU-3D ice model based on Bedmap2 dataset.

Ansu Chatterjee
University of Minnesota

“Extremes and Networks: a review of activities”

We present a summary of the activities carried out by various members of the “Extremes and Networks" working sub-group. These include studies on different connectivity metrics on networks and their applicability on data on extremes, case studies on networks of precipitation extremes for southern USA and for India, and resampling methods for extremes.

Asim Dey
University of Texas, Dallas

“Modeling Weather-induced House Insurance Risks”

Insurance industry is one of the most vulnerable sectors to climate change. Assessment of the future number of claims and incurred losses is critical for disaster preparedness and risk management. In this project, we study the effect of precipitation on joint dynamics of weather-induced home insurance claims and losses. We discuss new statistical and machine learning approaches to forecasting future climate-induced claim dynamics, while accounting for nonlinear multivariate dependence structure and quantifying for the associated modeling and data input uncertainties. We illustrate our methodology by application to attribution analysis and forecasting of weather-induced home insurance claims in Canada.

Robert Erhardt
Wake Forest University

“The Nexus of Climate Data, Insurance, and Adaptive Capacity”

This workshop will facilitate a national, interdisciplinary scientific research discussion on modeling and managing climate change risks between three different but related research communities: the climate modeling and data community, statisticians, and researchers within the insurance and reinsurance industries. This workshop will be aimed at active scientific researchers in each community, with the goal of sharing perspectives, methods, and databases; defining gaps of existing research; and formation of actionable research paths which can only be addressed by teams which are new, collaborative and interdisciplinary. It will be held November 8 and 9, 2018, at The Collider in Asheville, NC.

David Gagne
NCAR

Talk 1 Title: “The Climate Informatics Working Group: An Overview”

The climate informatics working group investigated many different avenues in which machine learning and other data science tools can be applied to climate science. The centerpiece of the group's work is a review paper on deep learning for the geosciences that highlights how deep learning has been applied to climate science problems in recent years and some potential applications and challenges going forward. The talk will also discuss some of the other joint work between the Climate Informatics Working Group and other subgroups.

Talk 2 Title: “A Generative Adversarial Network Stochastic Parameterization of the Lorenz '96 Model”

Jointly Presented with Aneesh Subramanian

Generative adversarial networks (GANs) are a form of deep learning designed to generate realistic samples from any multidimensional data distribution through the process of training two neural networks to optimize against each other. GANs have been used successfully to generate realistic synthetic images of faces and scenes and have also shown promise for generating spatial fields. In this project we developed a GAN that can generate stochastic realizations of the subgrid state of the Lorenz '96 dynamical model and compared the GAN with other stochastic statistical parameterizations on weather and climate timescales. Stochastic parameterizations combine a model of the mean state of a process with a noise-generating model to account for process variability and to sample from a wider portion of the feature space. The GAN combined external random processes with internal stochasticity in the form of dropout to generate a full set of subgrid values that closely approximated the true feature space. A Lorenz '96 ensemble successfully ran with a GAN parameterization in both short "weather" runs and longer "climate" runs. We examine the spread-skill relationship of the GAN parameterizations at weather timescales and how well the GAN approximates the true model's climatological distributions overall and within certain regimes. We also examine the sensitivity of GAN performance to certain parameter choices.

Azar Ghahari
University of Texas

Climate change and agriculture are intrinsically interrelated, and agricultural risk management is considered to be the fastest growing line of business in property and casualty insurance companies in North America. Challenges with modeling complex weather and climate dynamics bring to the fore statistical problems of analyzing massive, multi-resolution, multi-source data with a nonstationary space-time structure, nonlinear relationships between weather events and crop yields, and the respective actuarial implications due to imprecise risk estimation.

Modern machine learning procedures offer multiple operational benefits and are now proved to deliver a highly competitive performance in a variety of applications, their potential in actuarial sciences and, particularly, agricultural insurance, remains largely untapped. In this project, our goal is to introduce a modern deep learning methodology into assessment of climate-induced risks in agriculture and to evaluate its potential to deliver a stable and superior performance for various data inputs. We present a pilot study of deep learning algorithms using historical crop yields, weather station-based records, and gridded weather reanalysis data for Manitoba, Canada. Our results show that deep learning can attain a higher prediction accuracy, based on benchmarking its performance against more conventional approaches, especially in multi-scale, heterogeneous data environments of agricultural risk management.

Co-authors: N. Newlands, Y.R. Gel. and V. Lyubchich

Keywords: Artificial intelligence, agricultural risk, crop insurance, climate change, deep learning, extreme weather event

Donata Giglio
University of California, San Diego

“Estimating Oxygen in the Southern Ocean using Argo Temperature and Salinity”

An Argo based estimate of Oxygen (O_2) at 150 m is presented for the Southern Ocean (SO) from T/S, O_2 Argo profiles collected during 2008-2012. The method is based on supervised machine learning, i.e. Random Forest (RF) regression, and provides an estimate for O_2 on gridded Argo T/S fields. Results show that the Southern Ocean State Estimate (SOSE) and the World Ocean Atlas 2013 climatology may overestimate annual mean O_2 in the SO, both on a global and basin scale. A large regional bias is found east of Argentina, where high O_2 values in the Argo based estimate are closer to the coast compared to other products. SOSE may also underestimate the annual cycle of O_2. Regions where the RF method does not perform well (e.g. eastern boundaries) are identified comparing the actual SOSE O_2 fields to the RF estimate from model profiles co-located with observations. The RF based method presented here has the potential to improve our understanding of O_2 annual mean fields and variability from available (sparse) O_2 measurements. Also, it may guide the design of future enhancements to the current array of O_2 profiling floats, and prove effective for other biogeochemical variables (e.g. nutrients and carbon).

Yawen Guan
SAMSI

“Metrics for Evaluating Sea Ice Models"

Sea ice models governed by physical equations have been used to simulate the state of the ice including features such as ice thickness, concentration, and motion. Recent satellite observations with high spatio-temporal resolution have also provided unique opportunities to examine ice motion and deformation. These multiple disparate data sources prompted the research questions in our working group: How do we evaluate the skill of models in simulating ice features? How do we identify numerical model parameter space that produces realistic state of the ice? I will discuss an ongoing project that explores some potential methods for validating sea ice models.

Jonathan Hobbs
JPL

“Incorporating Spatial Dependence in Remote Sensing Inverse Problems”

The remote sensing working group has investigated methodology for atmospheric remote sensing retrievals, which are mathematical and computational procedures for inferring the state of the atmosphere from remote sensing observations. Satellite data with fine spatial and temporal resolution present opportunities to combine information across satellite pixels using spatio-temporal statistical modeling. We present examples of this approach at the process level of a hierarchical model, with a nonlinear radiative transfer model incorporated into the likelihood. In this framework, we assess the impact of various statistical properties on the relative performance of a multi-pixel retrieval strategy versus an operational one-at-a-time approach. The prospect of adopting the approach is illustrated in the context of estimating atmospheric carbon dioxide concentration with data from NASA's Orbiting Carbon Observatory-2 (OCO-2).

Huang Huang
SAMSI

“Inference on the Future Climate States from Multiple Ensembles using Bayesian Hierarchical Models”

Climate scientists have been developing a lot of climate models for variables of interest, like temperature, pressure, based on physical dynamics. Due to different techniques in implementing the climate models and the uncertainty in the climate system, the variable values are not identical from different model outputs. Scientists have been working on sensible statistical models to combine different climate model outputs, and most of them assume the exchangeability of all the climate model outputs, which, however, may have similar origins or share common components that lead to model dependence. In this work, we present a Bayesian hierarchical model to account for the model dependence, which gives a better inference for the underlying process for the variable of interest. In addition, we use the spatial Gaussian random field to allow for spatial correlation in the modeling, offering us a sensible map of the inferred future climate states.

Whitney Huang
SAMSI

"Estimating Extreme Strom Surge Levels: A Statistical Perspective"

Storm surge is an abnormal rise of water, largely induced by the strong winds of a hurricane, that could cause tremendous damage in coastal areas. Therefore, it is critically important to estimate the surge levels especially those extreme ones. However, the estimation of surge levels poses an unique statistical challenge due to the rareness of hurricanes in space and time. To overcome this difficulty, the join probability modeling of hurricane characteristics combined with hydrodynamic simulations is currently the recommended method by the Federal Emergency Management Agency (FEMA) for calculating the extreme surges in terms of 10-, 50-, 100-, and 500-year return levels.

In this talk, I will present the FEMA's approach from a statistical perspective starting from the estimation of the distributions of hurricane characteristics to the design and analysis of the hydrodynamic simulations. I will highlight the challenges and how we might improve the current practice in terms of estimation and uncertainty qualification.

Raphael Huser
KAUST

"Climate Extremes and Max-stable Processes"

Extreme environmental events such as droughts, floods and heat-waves take place in space and time, and it is necessary to take this into account when evaluating their risks and estimating their probabilities. During this lecture, I will review recent work on this topic, focusing on max-stable processes, which are the natural extension of univariate extreme-value models for more complex spatial phenomena. The most widely-used max-stable models, the Brown–Resnick and extremal-t processes will be described, followed by a brief overview of other types of models and of approaches to inference based on high threshold exceedances, with an emphasis on likelihood-based methods. The ideas will be illustrated by applications to heavy rainfall close to Jeddah and to high temperatures in the region of Madrid.

Emily Kang
University of Cincinnati

"Statistical Emulation with Dimension Reduction for Complex Physical Forward Models"

The retrieval algorithms in remote sensing generally involve complex physical forward models that are nonlinear and computationally expensive to evaluate. Statistical emulation provides an alternative with cheap computation and can be used to calibrate model parameters and to improve computational efficiency of the retrieval algorithms. We introduce a framework of combining dimension reduction of input and output spaces and Gaussian process emulation technique. The functional principal component analysis (FPCA) is chosen to reduce to the output space of thousands of dimensions by orders of magnitude. In addition, instead of making restrictive assumptions regarding the correlation structure of the high-dimensional input space, we identity and exploit the most important directions of this space and thus construct a Gaussian process emulator with feasible computation. We will present preliminary results obtained from applying our method to OCO-2 data, and discuss how our framework can be generalized in distributed systems. This is joint work with Jon Hobbs, Alex Konomi, Pulong Ma, and Anirban Mondal, and Joon Jin Song.

Ben Lee
Pennsylvania State University

“A Fast Particle-Based Approach for Computer Model Calibration”

Complex computer models play a prominent role in climate science, particularly in projecting future climate. These models have key parameters that need to be inferred ("calibrated") based on observational data. We describe a sequential Monte Carlo method that is well suited for calibration problems for which standard Markov chain Monte Carlo methods and model emulation approaches are computationally burdensome. The motivating scientific problem for our work is a computer model for projecting the future of the Antarctic ice sheet.

Nathan Lenssen
Columbia University

“Tunable Testbed for Detection and Attribution Methods”

The field of detection and attribution has been growing for a couple of decades and has recently seen a increase in the quantity and sophistication of methods. The difficulty of comparing these methods has motivated the design of a simulation testbed. Such a testbed is difficult to develop as it involves generating spatiotemporal fields with complex and flexible covariance structures that do not inherently favor any of the methods to be tested. The presented testbed has the ability to generate a wide class of isotropic and non-isotropic correlation matrices to simulate the climate variability. The forcing response fields are tunable, spatially correlated fields with adjustable signal-to-noise ratios. The flexibility of the simulation method allows us to replicate a variety of climate model-like output in a testable setting. In addition to the methods used in the testbed, we present synthetic data for a few simulated climate scenarios and demonstrate the intended use of the testbed by comparing the performance of ordinary least squares (OLS) and total least squares (TLS) on the synthetic data. This is joint work with Dorit Hammering, Alexis Hannart, and the SAMSI working group on Detection and Attribution.

Fei Lu
Johns Hopkins University

“Bayesian Approach to Climate Reconstructions using Stochastic Energy Balance Models”

Spatio-temporally resolved reconstructions of past climate are important in understanding the response of the climate system to changes in external forcings. Unfortunately, they are subject to large uncertainties and complex proxy-climate structures, and a physically reasonable interpolation between the sparse proxy observations is difficult.

We propose a Bayesian hierarchical framework for temperature reconstructions of the last deglaciation (from about 21ky BP to about 6ky BP) which combines a stochastic energy balance model (SEBM) with an observation operator for proxy data containing spatio-temporally integrated information on the temperature evolution. The SEBM is motivated by the non-linear model (Fanning and Weaver, 1996), but with an additional stochastic term accounting for unresolved processes.

Particle MCMC (Andrieu et al. 2010) is used for inference. It combines sequential Monte Carlo methods with MCMC techniques and exploits the forward structure of the SEBM to create MCMC proposals for the high dimensional, non-Gaussian posterior. We present initial results from perfect model tests. In the future, we plan to perform pseudo-proxy experiments with the CCSM3 TraCE21K simulation of the last deglaciation and apply the framework to a state of the art proxy synthesis.

Authors: Fei Lu, Johns Hopkins, Nils Weitzel, University of Bonn, Adam Monahan, University of Victoria

Adam Monahan
University of Victoria

“Stochastic Parameterization of Subgrid-Scale Air-Sea Fluxes”

Fluxes of mass, energy, and momentum between the atmosphere and ocean are parameterized in climate models as dependent on the near-surface wind speed. While what is desired is the flux spatially averaged across a model grid box, the wind information available to compute the flux is the grid box-mean vector wind. In general, the mean of the flux will not equal the flux estimated from the magnitude of the mean vector wind - and the difference between these two quantities is not necessarily a deterministic function of the resolved variables.

To compare the true and resolved fluxes, we have coarse-grained a high-resolution atmospheric simulation over the tropical Indian Ocean and Western Pacific. We will present a characterization of the stochastic dependence of the true fluxes on the resolved variables, both in terms of their local marginal structure and their space/time dependence. Finally, we will discuss implications of these results for the development of stochastic parameterizations in weather and climate models.

Sohini Raha
North Carolina State University

“On the Probability Distributions of the Intensity and Duration of Heatwaves”

Heatwaves are affecting the human mortality and the global climate adversely. There are many existing ad hoc definitions of heatwaves in literature. Here, we are identifying a notion of intensity and duration of the possible heatwaves and finding a proper definition using their distributions. We are modelling these two factors and attempting to find the predictive ability of our model. We have also obtained some preliminary results on obtaining the probability distribution of duration and intensity for a sequence of stationary random variables. We are exploring these concepts on a heat data of Atlanta from the year 1991-2012. The next goals are to relax the stationary assumption and obtaining more general results.

Brook Russell
Clemson University

"Investigating Precipitation Extremes in the US Gulf Coast through the use of a Multivariate Spatial Hierarchical Model"

Over a seven day period in August 2017 Hurricane Harvey brought extreme levels of rainfall to the Houston area, resulting in catastrophic flooding that caused loss of human life and damage to personal property and public infrastructure. In the wake of this event, there is growing interest in understanding the degree to which this event was unusual and estimating the probability of experiencing a similar event in other locations. Additionally, we investigate the degree to which the sea surface temperature in the Gulf of Mexico is associated with extreme precipitation in the US Gulf Coast. This talk addresses these issues through the development of an extreme value model.

We assume that the annual maximum precipitation values at Gulf Coast locations approximately follow the Generalized Extreme Value (GEV) distribution. Because the observed precipitation record in this region is relatively short, we borrow strength across spatial locations to improve GEV parameter estimates. We model the GEV parameters at US Gulf Coast locations using a multivariate spatial hierarchical model; for inference, a two-stage approach is utilized. Spatial interpolation is used to estimate GEV parameters at unobserved locations, allowing us to characterize precipitation extremes throughout the region. Analysis indicates that Harvey was highly unusual as a seven-day event, and that GoM SST seems to be more strongly linked to extreme precipitation in the Western part of the region.

Christian Sampson
SAMSI

"Understanding Sea Ice Data for Data Assimilation"

Sea ice is a porous composite of ice and brine with a variety of microstructures that control a broad range of transport phenomena through it. These transport phenomena play a strong role in regulating the thermodynamics and optical properties of the ice at both small and large scales. As a result, sea ice microstructure is intimately tied to many of the most important parameters in numerical sea ice models used for climatological and operational forecasting. These parameters are also important for understanding remote sensing data for sea ice. Data assimilation schemes serve as powerful tools to improve model forecasts, quantify their uncertainty, and estimate model parameters through the combination of both model and observational information. In this talk I will discuss some recent work quantifying the differences between two of the most common crystallographic types of sea ice and how understanding these differences can be used to improve both models and observations of sea ice.

Deborah Sulsky
University of New Mexico

"Overview of Sea-Ice Modeling and Statistical Challenges"

Accurate quantification and simulation of the relative contributions of the thermodynamics and dynamics to the ice thickness distribution are crucial for understanding the behavior and the vulnerability of the polar ice cover in a warming climate. To improve our understanding and projection of the future behavior of the Arctic sea-ice system a constitutive model has been developed that is based on elasticity combined with a cohesive crack law that predicts the initiation, orientation and opening of leads.

Simulations of the Arctic show deformation, motion, lead openings, effective thickness, and ice compactness. Corresponding satellite observations can be used to validate and calibrate the model. The challenges associated with developing metrics for validation and calibration will be described. The additional challenges when cracks are present will also be discussed.

Surya Tokdar
Duke University

"Semiparametric Models for Extremes"

I will review the main ideas and research directions that have emerged through our discussions in the Semiparametric Extremes subgroup.

Erik Van Vleck
University of Kansas

"Projected Data Assimilation"

In this talk we present a framework for splitting data assimilation problems based upon the model dynamics. This is motivated by assimilation in the unstable subspace (AUS) and center manifold and inertial manifold techniques in dynamical systems. Recent efforts based upon the development of particle filters projected into the unstable subspace will be highlighted.

Sooin Yun
University of Illinois

“Comparing Two Spatio-temporal Fields with Application to Climate Data”

Comparing two spatio-temporal random fields is often of interest in climatology. Motivated by evaluating the synthetic climate from climate models and assessing the skill of climate field reconstructions, we proposed two approaches for the field comparison with one focusing on the large scale comparison and the other on regional scale. I will first present the large scale comparison developed based on the functional data analysis, how it is used to evaluate the climate field reconstruction, and how the results are interpreted and understood. Then I will present a new adaptive multiple testing procedure that enables the regional comparison. This new approach incorporates the spatial dependence to boost power by modeling the mixing probability as a smooth function over space. Non-parametric, semi-parametric, and parametric approaches combined with EM-algorithm are carried out to estimate the mixing probability function. I will show the regional comparison of mean and covariance function between two climate model outputs.0

Zhengyuan Zhu
Iowa State University

“Statistical Approaches for Un-Mixing Problem and Application to Satellite Remote Sensing Data”

Remote sensing data from satellite with high temporal resolution typically have lower spatial resolution, with one pixel often spanning over a square kilometer. The signal recorded by such satellite at a pixel is typically a mixture of reflectance from different types of land covers within the pixel, resulting in a mixed pixel. In this talk we introduce a couple of parametric and non-parametric statistical approaches to deal with the un-mixing problem which integrate information from multiple sources, and present some preliminary results applying the methodology to data from the SMOS (Soil Moisture and Ocean Salinity) mission and the OCO-2 (Orbiting Carbon Observatory 2) mission, which motivated this research.