samsi
NSF·Duke·NCSU·UNC

Climate Program Opening Workshop
August 21-25, 2017

SPEAKER TITLES/ABSTRACTS

Michal Branicki
University of Edinburgh

“An Information-Theoretic Framework for Improving Multi-Model Predictions & Data Assimilation Techniques”

Multi Model Ensemble (MME) predictions are a popular ad-hoc technique for improving predictions of high-dimensional, multi-scale dynamical systems. The heuristic idea behind MME framework is simple: given a collection of models, one considers predictions obtained through the convex superposition of the individual probabilistic forecasts in the hope of mitigating model error. However, it is not obvious if this is a viable strategy and which models should be included in the MME forecast in order to achieve the best predictive performance. I will present an information-theoretic approach to this problem which allows for deriving a sufficient condition for improving dynamical predictions within the MME framework; moreover, this formulation gives rise to systematic and practical guidelines for optimising data assimilation techniques which are based on multi-model ensembles. Time permitting, the role and validity of â€˜fluctuation-dissipationâ€™ arguments for improving imperfect predictions of externally perturbed non-autonomous systems - with possible applications to climate change considerations - will also be addressed.

Chris Bretherton
University of Washington

“Developing Stochastic Parameterizations of Subgrid Variability of Clouds and Turbulence using High-Resolution Simulations”

Climate model parameterizations of cumulus convection and other clouds that form due to small-scale turbulent eddies are a leading source of uncertainty in predicting the sensitivity of global warming to greenhouse gas increases. Even though we can write down equations governing the physics of cloud formation and fluid motion, these cloud-forming eddies are not resolved by the grid of a climate model, so the subgrid covariability of cloud processes and turbulence must be parameterized. Many approaches are used, all involving numerous subjective assumptions. Even when optimized to match present-day climate, these approaches produce a broad range of predictions about how clouds will change in a future climate.

High resolution models which explicitly simulate the clouds and turbulence on a very fine computational grid more realistically simulate cloud formation compared to observations. But it has proved challenging to translate this skill into better climate model parameterizations. We will present one naturally stochastic approach for this using a computationally expensive approach called ‘superparameterization’ and then we will lay out a vision for how machine learning could be used to do this translation, which amounts to a form of stochastic coarse-graining. Developing the statistical and computational methods to realize this vision is a good challenge for this SAMSI year.

Alberto Carrassi
Nansen Environmental and Remote Sensing Center

“Issues in Ensemble Prediction and Data Assimilation Using a Lagrangian Model of Sea-Ice”

In the first part of the talk, we will present a sensitivity analysis of a novel sea ice model. *neXtSIM* is a continuous Lagrangian numerical model that uses an elastobrittle rheology to simulate the ice response to external forces. The response of the model is evaluated in terms of simulated ice drift distances from its initial position and from the mean position of the ensemble. The simulated ice drift is decomposed into advective and diffusive parts that are characterized separately both spatially and temporally and compared to what is obtained with a *free-drift* model, i.e. when the ice rheology does not play any role. Overall the large-scale response of *neXtSIM* is correlated to the ice thickness and the wind velocity fields while the *free-drift* model response is mostly correlated to the wind velocity pattern only. The seasonal variability of the model sensitivity shows the role of the ice compactness and rheology at both local and Arctic scales. Indeed, the ice drift simulated by *neXtSIM* in summer is close to the *free-drift* model, while the more compact and solid ice pack is showing a significantly different mechanical and drift behavior in winter. In contrast of the *free-drift* model, *neXtSIM* reproduces the sea ice Lagrangian diffusion regimes as found from observed trajectories. The forecast capability of *neXtSIM* is also evaluated using a large set of real buoy’s trajectories. We found that *neXtSIM* performs better in simulating sea ice drift, both in terms of forecast error and as a tool to assist search-and-rescue operations. Adaptive meshes, as the one used in *neXtSIM*, are used to model a wide variety of physical phenomena. Some of these models, in particular those of sea ice movement, use a remeshing process to remove and insert mesh points at various points in their evolution. This represents a challenge in developing compatible data assimilation schemes, as the dimension of the state space we wish to estimate can change over time when these remeshings occur.

In the second part of the talk, we highlight the challenges that such a modeling framework represents for data assimilation setup. We then describe a remeshing scheme for an adaptive mesh in one dimension. The development of advanced data assimilation methods that are appropriate for such a moving and remeshed grid is presented. Finally we discuss the extension of these techniques to two-dimensional models, like *neXtSIM*.

Howard Chang
Emory University

“Projecting Health Impacts of Climate Change: Embracing an Uncertain Future”

Global climate change affects human health most notably by increasing the frequency and intensity of dangerous heat waves, wildfires and hurricanes. In addition to extreme weather events, climate change can also lead to a myriad of persistent environmental changes that impact public health. Health impact assessment refers to the analytic framework for evaluating how a policy or program affects population health. It is frequently applied in climate and public health research to quantify future health and economic burdens attributable to various consequences of climate change.

Performing health impact assessment entails the integration of various data. For projecting future climate-related health impacts, analyses require three sources of information: (1) health effects of environmental exposures, (2) projections of future exposures, and (3) distributions of exposures and effects in the future population. Each information source is subject to uncertainty because of data availability and assumptions made for the future. Climate research is highly interdisciplinary, bringing together tremendous amount of data, theory, and modeling efforts to provide timely knowledge for one of the most pressing issues of our time. Statistical modeling techniques and probabilistic reasoning can plan an important role in ensuring these findings are informative, accurate, and reproducible.

This presentation will discuss recent development in statistical methods for quantifying health impacts of climate change, as well as related open problems in environmental epidemiology and exposure assessment.

Michael Wehner
Lawrence Berkeley National Laboratory

"Computational and Mathematical Challenges in Climate Modeling"

We present a survey of computational and applied mathematical techniques that have the potential to contribute to the next generation of high-fidelity, multi-scale climate simulations. Examples of the climate science problems that can be investigated with more depth with these computational improvements include the capture of remote forcings of localized hydrological extreme events, an accurate representation of cloud features over a range of spatial and temporal scales, and parallel, large ensembles of simulations to more effectively explore model sensitivities and uncertainties.

Numerical techniques, such as adaptive mesh refinement, implicit time integration, and separate treatment of fast physical time scales are enabling improved accuracy and fidelity in simulation of dynamics and allowing more complete representations of climate features at the global scale. At the same time, partnerships with computer science teams have focused on taking advantage of evolving computer architectures such as many-core processors and GPUs. As a result, approaches which were previously considered prohibitively costly have become both more efficient and scalable. In combination, progress in these three critical areas is poised to transform climate modeling in the coming decades.

Noel Cressie
University of Wollongong

“A Bird's-Eye View of Statistics for Remote Sensing Data”

Remote-sensing data offer unprecedented opportunities to address Earth-system-science challenges, such as understanding the relationship between the atmosphere and Earth's surface using physics, chemistry, biology, mathematics, and computing. Statistical methods have often been seen as a hybrid of the latter two, so that a lot of attention has been given to computing estimates but far less to quantifying the uncertainty of the estimates. In my "bird's-eye view," I shall give a way to look at the problem using conditional probability models and three states of knowledge. Examples will be given of analyzing remotely sensed data of a leading greenhouse gas, carbon dioxide.

Daniel Crichton
NASA/JPL

"Software Architecture Considerations in the Analysis of Highly Distributed Data and Computational Analysis"

Climate Science presents several data intensive challenges that are the intersection of software architecture and data science. This includes developing approaches for scaling the analysis of highly distributed data across institutional and system boundaries. JPL has been developing approaches for quantitatively evaluating software architectures to consider different topologies in the deployment of computing capabilities and methodologies in order to support the analysis of distributed climate data. This talk will cover those approaches and also needed research in new methodologies as remote sensing and climate model output data continue to increase in their size and distribution.

Imme Ebert-Uphoff
Colorado State University

"Methods for Causality Analysis in Climate Science"

Knowledge of cause-effect relationships is central to the field of climate science, supporting mechanistic understanding, observational sampling strategies, experimental design, model development and model prediction. While the major causal connections in our planet's climate system are already known, there is still potential for new discoveries in some areas. The purpose of this talk is to make this community familiar with a variety of available tools to discover potential cause-effect relationships from observed or simulation data. Some of these tools are already in use in climate science, others are just emerging in recent years. None of them are miracle solutions, but many can provide important pieces of information to climate scientists. An important way to use such methods is to generate *cause-effect hypotheses* that climate experts can then study further. In this talk we will (1) introduce key concepts important for causal analysis; (2) discuss some methods based on the concepts of *Granger* causality and *Pearl* causality; (3) point out some strengths and limitations of these approaches; and (4) illustrate such methods using a few real-world examples from climate science.

Kristie Ebi
University of Washington

"Climate and Health"

Climate change could have far-reaching consequences for human health across the 21st century. At the same time, development choices will alter underlying vulnerability to these risks, affecting the magnitude and pattern of impacts. The current and projected human health risks of climate change are diverse and wide-ranging, potentially altering the burden of any health outcome sensitive to weather or climate. Climate variability and change can affect morbidity and mortality from extreme weather and climate events, and from changes in air quality arising from changing concentrations of ozone, particulate matter, or aeroallergens. Altering weather patterns and sea level rise also may facilitate changes in the geographic range, seasonality, and incidence of selected infectious diseases in some regions, such as malaria moving into highland areas in parts of sub-Saharan Africa. Changes in water availability and agricultural productivity could affect undernutrition, particularly in parts of Asia and Africa. These risks are not independent, but will interact in complex ways with risks in other sectors. Policies and programs need to explicitly take climate change into account to facilitate sustainable and resilient societies that effectively prepare for, manage, and recover from climate-related hazards.

Montse Fuentes
Virginia Commonwealth University

"A Multivariate Dynamic Spatial Factor Model for Speciated Pollutants and Adverse Birth Outcomes"

Evidence suggests that exposure to elevated concentrations of air pollution during pregnancy may increase risks of birth defects and other adverse birth outcomes. While current regulations put limits on total PM2.5 concentrations, there are many speciated pollutants within this size class that likely have distinct effects on perinatal health. However, due to correlations between these speciated pollutants it can be difficult to decipher their effects in a model for birth outcomes. To combat this difficulty we develop a new multivariate spatio-temporal Bayesian model for speciated particulate matter using dynamic spatial factors. These spatial factors can then be interpolated to the pregnant women's homes to be used in a birth outcomes model. The model for birth outcomes allows the impact of pollutants to vary across different weeks of the pregnancy in order to identify susceptible periods. The proposed innovative methodology is implemented using pollutant monitoring data from the Environmental Protection Agency and birth records from the National Birth Defect Prevention Study.

Work in collaboration with Kimberly Kaufeld, Brian Reich, Amy Herring, Gary Shaw and Maria Terres.

Andrew Gettelman
National Center for Atmospheric Research

"A Cloud of Numbers: Representing Physical Processes in the Earth System with Mathematics"

Physical processes in the earth system are modeled with mathematical representations called parameterizations. This talk will describe some of the conceptual approaches and mathematics used do describe physical parameterizations focusing on cloud parameterizations. This includes tracing physical laws to discrete representations in coarse scale models. Clouds illustrate several of the complexities and techniques common to many physical parameterizations. This includes the problem of different scales, sub-grid scale variability. Discussions of mathematical methods for dealing with the sub-grid scale will be discussed. In-exactness or indeterminate problems for both weather and climate will be discussed, including the problems of indeterminate parameterizations, and inexact initial conditions. Different mathematical methods, including the use of stochastic methods, will be described and discussed, with examples from contemporary earth system models.

Dorit Hammerling
NCAR

"Compression and Conditional Emulation of Climate Model Output"

Numerical climate model simulations runs at high spatial and temporal resolutions generate massive quantities of data. As our computing capabilities continue to increase, storing all of the generated data is becoming a bottleneck, and thus is it important to develop methods for representing the full datasets by smaller compressed versions. We propose a statistical compression and decompression algorithm based on storing a set of summary statistics as well as a statistical model describing the conditional distribution of the full dataset given the summary statistics. The statistical model can be used to generate realizations representing the full dataset, along with characterizations of the uncertainties in the generated data. Thus, the methods are capable of both compression and conditional emulation of the climate models. Considerable attention is paid to accurately modeling the original dataset, particularly with regard to the inherent spatial nonstationarity in global temperature fields, and to determining the statistics to be stored, so that the variation in the original data can be closely captured, while allowing for fast decompression and conditional emulation on modest computers.

Murali Haran
Pennsylvania State University

“Some Statistical Challenges in Studying the West Antarctic Ice Sheet”

The melting of the West Antarctic ice sheet (WAIS) is likely to cause a significant rise in sea levels. Studying the present state of WAIS and predicting its future behavior involves the use of computer models of ice sheet dynamics as well as observational data. I will outline general statistical challenges posed by these scientific questions and data sets.

This discussion is based on joint work with Yawen Guan (Penn State/SAMSI), Won Chang (U. of Cincinnati), Patrick Applegate, David Pollard (Penn State)

Charles Jackson
University of Texas

“Ice Sheet Contribution to Sea Level Rise”

This problem represents an interesting opportunity for scientists and statisticians to collaborate since the problem is too big for either community. The science is not well established, although fairly sophisticated ice flow models exist. They are even becoming relevant to explain some of the complexity seen in observational data. At the same time, the complex phenomena we see in observations may not be particularly relevant to assessing the risks of significant increases in sea level rise over the near future. The talk will review what we have learned about this problem through the PISCEES SciDAC project. This problem is rich with challenges and opportunities, particularly for realigning how our two communities engage each other. The talk will review the computational, scientific, and mathematical "reality checks" that might stop any reasonable person from considering this topic further. I then will point out how each of these challenges could be mitigated if these different perspectives were better integrated.

Matthias Katzfuss
Texas A&M University

“A General Framework for Vecchia Approximations of Gaussian Processes”

Gaussian processes (GPs) are commonly used as models for functions, time series, and spatial fields, but they are, in general, computationally infeasible for large datasets. Focusing on the typical setting of observations from a GP containing a nugget or noise term, we propose a generalization of Vecchia's approximation as a framework for GP approximations. We show that our general Vecchia approach contains many popular existing GP approximations as special cases, allowing a comparison of the different approaches within a unified framework. Representing the models by directed acyclic graphs, we determine the sparsity of the matrices necessary for inference, which leads to new insights regarding the computational properties. Based on these results, we propose a novel sparse general Vecchia approximation, which ensures computational feasibility for large datasets but can lead to tremendous improvements in approximation accuracy over Vecchia's original approach. We provide theoretical results and conduct numerical comparisons. We conclude with guidelines for the use of Vecchia approximations. This project is joint work with Joseph Guinness (NCSU).

Vipin Kumar
University of Minnesota

“Big Data in Climate: Opportunities and Challenges for Machine Learning”

The climate and earth sciences have recently undergone a rapid transformation from a data-poor to a data-rich environment. In particular, massive amount of data about Earth and its environment is now continuously being generated by a large number of Earth observing satellites as well as physics-based earth system models running on large-scale computational platforms. These massive and information-rich datasets offer huge potential for understanding how the Earth's climate and ecosystem have been changing and how they are being impacted by humans actions. This talk will discuss various challenges involved in analyzing these massive data sets as well as opportunities they present for both advancing machine learning as well as the science of climate change in the context of monitoring the state of the tropical forests and surface water on a global scale.

Research funded by the NSF Expeditions in Computing Program and NASA

Kenneth Kunkel
North Carolina State University

“Understanding the Physical Causes of Observed Trends in Extreme Precipitation: How Can Statistics Help?”

Numerous studies have found an average increase in extreme precipitation for both the U.S. and Northern Hemisphere mid-latitude land areas, consistent with the expectations arising from the observed increase in greenhouse gas concentrations (now more than 40% above pre-industrial levels). However, there are important regional variations in these trends that are not fully explained. These trend studies are typically based on direct analyses of observational station data. Such analyses confront multiple challenges, such as incomplete data and uneven spatial coverage of stations. Central scientific questions related to this general finding are: Are there changes in weather system phenomenology that are contributing to this observed increase? What is the contribution of increases in atmospheric water vapor? There are also questions related to application of potential future changes in planning. Because of the rarity (by definition) of extreme events, trends are mostly found only when aggregating over space. When would we expect to see a signal at the local level? What are the uncertainties surrounding future changes and their potential incorporation into future design? Further development of statistical/mathematical methods, or innovative application of existing methods, is desirable to aid scientists in exploring these central scientific questions. This talk will describe characteristics of the observation record and the issues surrounding the above questions.

Mr. Prabhat
Lawrence Berkeley National Laboratory

“Deep Learning for Extreme Weather Detection”

Extreme weather events pose great potential risk on ecosystem, infrastructure and human health. Analyzing extreme weather in the observed record (satellite, reanalysis products) and characterizing changes in extremes in simulations of future climate regimes is an important task. Thus far, extreme weather events have been typically specified by the community through hand-coded, multi-variate threshold conditions. Such criteria are usually subjective, and often there is little agreement in the community on the specific algorithm that should be used. We propose the use of a different approach: machine learning (and in particular deep learning) for solving this important problem. If human experts can provide spatio-temporal patches of a climate dataset, and associated labels, we can turn to a machine learning system to learn the underlying feature representation. The â€˜trainedâ€™ ML system can then be applied to novel datasets, thereby automating the pattern detection step. Summary statistics, such as location, intensity and frequency of such events can be easily computed as a post-process.

We will report compelling results from our investigations of Deep Learning for the tasks of classifying tropical cyclones, atmospheric rivers and weather front events. For all of these events, we observe 90-99% classification accuracy. We will also report on progress in localizing such events: namely drawing a bounding box (of the correct size and scale) around the weather pattern of interest. Both tasks currently utilize multi-layer convolutional networks in conjunction with hyper-parameter optimization. We utilize HPC systems at NERSC to perform the optimization across multiple nodes, and utilize highly-tuned libraries to utilize multiple cores on a single node. We will conclude with thoughts on the frontier of Deep Learning and the role of humans (vis-a-vis AI) in the scientific discovery process.

Leonard Smith
London School of Economics

“Science, Simulation and Insight: Developing Confidence when Extrapolating Complicated Complex Systems”

Scientific approaches to challenges in climate-like systems differ fundamentally from similar looking challenges in weather-like systems. In weather-like systems one repeatedly attempts to answer the same (or rather similar) questions leading to repeated, truly out-of-sample outcomes; the life-time of a model is long compared to the lead-time of each forecast, allowing a large and growing out-of-sample forecast-outcome archive. In climate-like systems a model's lifetime is a small fraction of the lead-times of most interest, often exceeding the professional lifetime of a researcher. Quantitative out-of-sample forecast-outcomes are very small if not empty. That said broad predictions of unique phenomena and qualitative change may abound, if often without a specific time-line. How is a decision maker, or a scientist, to develop a rational level of confidence in proposed "predictions" in climate-like systems? How might scientists communicate the implications of "as good as it gets" science, while laying bare often severe limitations of today's "best available" simulations? How does one report (?or avoid?) impressive novel mathematical manipulations in model-land that can be expected to tell us nothing about the real world? Can one maintain a focus on communicating scientific outputs which are adequate for purpose without being drawn either into Machiavellian distractions or into the oversell of model-land results for real world application? What challenges do statisticians, applied mathematicians and others face when adapting traditional aids (experimental design, data analysis, interpretation of simulation ...) to climate-like systems? Can multidisciplinary teams effectively evaluate to the relevance of solid "in principle" analysis techniques which can only be applied to today's state-of-the-art models? Approaches to the development and long term maintenance of confidence are discussed, from delivery options and internal terminology, through tests of internal consistency, to confessed deep ignorance and the communication both of confidence and of uncertainty.

Deborah Sulsky
University of New Mexico

“Modeling Arctic Sea Ice”

As high-resolution simulations become increasingly possible and favored, questions are being raised about isotropic constitutive models for sea ice that are based on averaging material behavior over 100 km scales. At finer resolutions, it may not be appropriate to average over concentrated deformations which occur in leads and ridges since small regions do not contain sufficient numbers of these features at arbitrary orientations to support the assumption of isotropy. An elastic-decohesive constitutive model for pack ice has been developed that explicitly accounts for leads. The constitutive model is based on elasticity combined with a cohesive crack law that predicts the initiation, orientation and opening of leads.

This presentation shows results of the elastic-decohesive constitutive model applied to sea ice in the Arctic. We show simulation runs for lead openings, effective thickness and ice compactness. Ice compactness and thickness simulations are then compared to observations to validate the model. To this end we develop performance metrics that provide quantitative measures for the comparison. These metrics are easy to read, interpret and implement and show that the model performs reasonably well during winter months but less so during summer months.

Jason West
University of North Carolina

“The Effects of Climate Change on Human Health through Changes in Air Quality”

Climate change affects human health in several different ways, and one important effect is through changes in air pollution. Here I will discuss the state of science currently on how climate change affects air pollution, and the resulting effects on human health, drawing from the broad literature and highlighting studies from my lab. In particular, I will discuss: 1) The effects of air pollution on health globally, 2) How climate change affects air pollution, 3) The effects of climate change on air pollution and health globally, and 4) The co-benefits of greenhouse gas mitigation for air pollution and health, globally and in the US.