samsi

NSF Duke NCSU UNC

MUMS Agent-Based Modeling and Uncertainty Quantification Workshop March 11-12, 2019

SPEAKER TITLES/ABSTRACTS

Robert Axtell
George Mason University and Santa Fe Institute

"Reduced-Scales and Alternative Representations for Large/Full-Scale Agent-based Models"

We have calibrated an ultra-large scale model of the U.S. economy involving 120 million employee agents who interact to form some 6 million firms. We use firm-level micro-data on the entire private sector. These data feature many extreme properties, e.g., Pareto-distributed firm sizes (by multiple measures), heavy-tailed growth rate/fluctuations (e.g., log-Laplace distributions that become lognormal over long time scales), labor productivity approximately Lévy-distributed with a power law high productivity tail, and so on. This model is VERY expensive to evaluate and we have worked hard to parallelize it on custom hardware (100s of cores, TB of RAM, GPU, etc.). However, an alternative to make the model more usable is to create reduced-scale/form models. Reduced-scale models maintain the economic interpretability of the model but require rescaling all the output measures, of which there are nearly 50 in this case, with rescaling being heterogeneous across these measures and non-trivial with all the heavy-tails in the data. Alternatively, reduced-form models may maintain direct commensurability of output to data but make behavioral/policy interpretation difficult. This talk will map some of the tradeoffs between model intelligibility, usability, and scalability.

David Banks
SAMSI

"Statistical Challenges in Agent-Based Modeling"

This talk is an overview of several species of agent-based models. The goal is to identify common aspects, and to frame questions about how statistical inference on such models might be done.

Bobby Gramacy
Virginia Tech

"Replication or Exploration? Sequential Design for Stochastic Simulation Experiments"

We investigate the merits of replication, and provide methods that search for optimal designs (including replicates), in the context of noisy computer simulation experiments. We first show that replication offers the potential to be beneficial from both design and computational perspectives, in the context of Gaussian process surrogate modeling. We then develop a look ahead based sequential design scheme that can determine if a new run should be at an existing input location (i.e., replicate) or at a new one (explore). When paired with a newly developed heteroskedastic Gaussian process model, our dynamic design scheme facilitates learning of signal and noise relationships which can vary throughout the input space. We show that it does so efficiently, on both computational and statistical grounds. In addition to illustrative synthetic examples, we demonstrate performance on three challenging real-data simulation experiments from Bayes factor hyperprior sensitivity to inventory management and epidemiology.

Dave Higdon
Virginia Tech

"Bayesian Model Calibration and Prediction Applied to (Stochastic) Epidemic Simulations"

Agent-based models (ABMs) use rules at the individual level to simulate a social, or social-technical system, producing structured behavior when viewed at an aggregated level. For example, dynamic network simulation models commonly evolve a very large collection of agents interacting over a network that evolves with time. Such models are often used simulate epidemics or transportation, typically producing random trajectories, even when model parameters and initial conditions are identical. This introduces a number of challenges in designing ensembles of model runs for sensitivity analysis and computer model calibration. This talk will go through a case study of a recent epidemic, seeking to forecast the epidemic's behavior given initial administrative information. This talk discusses two different approaches for combining observations with this model for estimation and prediction. The first uses methods from traditional computer model calibration; the second approach - still under development - uses a sequential Monte Carlo approach.

Mevin Hooten
Colorado State University

"First-order Statistical Emulators"

Statistical emulators have been critical for providing inference based on complex differential equation models, computer models, and agent-based models. By approximating true model outputs while accounting for uncertainty in the approximation, emulators offer a way to fit statistical models to real data much more efficiently than when using the original model. In contrast to more classical second-order statistical emulators (i.e., based on covariance structure), I present first-order statistical emulation ideas that may be easier for practitioners to implement, can be used in a regularization setting, and can be used to fit agent-based models (ABMs). In particular, some ABMs in ecology present a unique set of challenges for emulation. I summarize these challenges for emulating ecological ABMs and propose possible remedies.

Junxiu Liu
Tufts University

"Cost-Effectiveness of the U.S. Federal Restaurant Menu Calorie Labeling Law for Improving Diet and Health: A Microsimulation Modeling Study"

Objectives Excess caloric intake is linked to weight gain, obesity and related diseases including type 2 diabetes and cardiovascular disease (CVD). Obesity incidence has been on the rise, with almost 2 of 3 people being overweight or obese in the US. In 2018, the US federal government passed a law mandating the labeling of calories on all menu items across chain restaurants, as a strategy to support informed consumer choice and reduce caloric intake. Yet, potential health and economic impacts of this policy remain unclear.

Methods We used a validated microsimulation model (CVD-PREDICT) to estimate reductions in CVD events, diabetes cases, gains in quality-adjusted life-years (QALYs), costs, and cost-effectiveness of two policy scenarios: (1) implementation of the federal menu calorie labelling (*menu calorie label*), and (2) further accounting for corresponding industry reformulation (*menu calorie label + reformulation*). The model utilized nationally representative demographic and dietary data from NHANES 2009-2016; policy effects on consumer intake and BMI-disease effects from published meta-analyses; and policy effects on industry reformulation, policy costs (policy administration, industry compliance and reformulation) and health-related costs (formal and informal healthcare costs, productivity costs) from established sources. We conservatively modeled change in calories to change in weight using an established dynamic weight-change model. Findings were evaluated over 10 years and lifetime from a healthcare and societal perspective. Costs were inflated to constant 2018 USD, and costs and QALYs were discounted at 3% annually. We performed probabilistic analyses and a range of one-way sensitivity and subgroup analyses to assess the robustness of our findings.

Results Sample statistics were shown (Table). American adults (35+) consume ~21% calorie from restaurants (Figure) that would be reduced by 2% due to this law at the population level. Government administration costs were estimated at 11.6$M, industry compliance costs at 652$M, and industry reformulation costs at 9.2$B. Findings for all other analyses will be presented at the meeting.
Conclusions These findings will provide much needed evidence on the health and economic impacts of the US menu calorie labeling law.

John Schuler
George Mason University

"Nonparametric Estimation of General Equilibrium Price Vectors"

Agent-based economic modeling often requires the determination of an initial equilibrium price vector. Calculating this directly requires algorithms of exponential computational complexity. It is known that a partial equilibrium price can be estimated using a median of trades. This paper explores the possibility of a multivariate generalization of this technique using depth functions.

Laura Schultz
George Mason University

"Practical Bayesian Optimization for Agent Based Transportation Simulators"

Simulators play a major role in analyzing multi-modal transportation networks. As complexity of simulators increases, development of calibration procedures is becoming an increasingly challenging task. Current calibration procedures often rely on heuristics, rules of thumb and sometimes on brute-force search. In this talk we consider a statistical framework for calibration that relies on Bayesian optimization. Bayesian optimization treats the simulator as a sample from a Gaussian process (GP). Tractability and sample efficiency of Gaussian processes enable computationally efficient algorithms for calibration problems. We show how the choice of prior and inference algorithm effect the outcome of our optimization procedure. We develop dimensionality reduction techniques that allow for our optimization techniques to be applicable for real-life

problems. We develop a distributed, Gaussian Process Bayesian regression and active learning models. We demonstrate those to calibrate ground transportation simulation models.