

Monday, May 13

8:00am
Registration and Continental Breakfast

9:00am
Welcome

9:15am
Our Earthrise: Awakening the World to Complexity
Adam Bly, *Founder and CEO, Startup in Stealth*

Adam Bly recently led data at Spotify and is currently building a new purpose-driven data and AI infrastructure startup. He previously founded the data analytics company Seed Scientific, which was acquired by Spotify in 2015. He was a Visiting Senior Fellow in Science, Technology & Society at Harvard Kennedy School and has lectured at Harvard, MIT, Peking University, The National Academy of Sciences, The Royal Society, NASA, MoMA, and the World Economic Forum. He has advised governments, national political campaigns, and international organizations on science and data policy. Adam previously founded the popular science magazine, Seed (2001-2010), and is the editor of Science is Culture: Conversations at the New Intersection of Science & Society. He served as Vice Chair of the Global Agenda Council on Complex Systems and currently serves on the National Leadership Council of the Society for Science & the Public. Adam was named a Young Global Leader by the World Economic Forum and is the recipient of the Golden Jubilee Medal from Queen Elizabeth II. He was born in Montreal and currently lives in New York City.

10:15am
Break

10:45am
Smart Cities Panel Discussion

Enabling a Systems of Systems in Smart Cities Through AI

Panel Moderator: *Ms. Karen Lightman (Carnegie Mellon University)*

Members: *Santiago Garces (City of Pittsburgh), Ben Levine (MetroLab Network), Steve Smith (Carnegie Mellon University)*

Cities are a complex system of systems with interweaving dependencies and connections. Today's computing technologies combined with advanced AI and modeling techniques are helping us define these connections in unprecedented ways, allowing for a new sense of clarity and transparency to policy and decision making.

This session, presented in a moderated panel format, will explore real-world use cases of applying technology to improve decision making in a wide array of applications and implications. We will discuss how computer visioning algorithms are increasing public mobility and safety by helping to predict landslides and helping understand the way public spaces are used in order to better plan downtown events. We will discuss how a combination of advanced analytics and edge computing techniques are helping improve intersections by measuring air quality and capturing near-miss accident data. Finally, we'll explore how predictive modeling and machine learning techniques are being used to determine fire risk, helping city officials target inspections and prevent infrastructure loss.

These pilot projects, enabled by Metro21: Smart Cities Institute and deployed in the city of Pittsburgh, combine modeling techniques and technology to impact safety, efficiency, urban planning, and more,

and allow municipal stakeholders to make more informed policy and management decisions than ever before.

12:00pm
Lunch

1:30pm

Confessions of a Lifelong Modeler, and a View from NSF

Jim Kurose, Assistant Director, National Science Foundation

Dr. Jim Kurose is an Assistant Director of the National Science Foundation, where he leads the Directorate for Computer and Information Science and Engineering (CISE). With an annual budget of nearly \$1B, CISE's mission is to uphold the nation's leadership in scientific discovery and engineering innovation through its support of fundamental research in computer and information science and engineering, transformative advances in cyberinfrastructure, and preparation of a diverse computing-capable workforce. Jim also co-chairs the Networking and Information Technology Research and Development (NITRD) Program, the Subcommittee on Machine Learning and AI, and the Subcommittee on Open Science of the National Science and Technology Council (NSTC), facilitating the coordination of these R&D efforts across Federal agencies. Recently, Jim also served as the Assistant Director for Artificial Intelligence in the US Office of Science and Technology Policy.

Jim is on leave from the University of Massachusetts, Amherst, where he is Distinguished University Professor of Computer Science with research interests in computer network protocols and architecture, network measurement, sensor networks, multimedia communication, and modeling and performance evaluation. He considers himself "a lifelong modeler" of computer networks and their protocols, as well as other complex distributed systems. He is a Fellow of the ACM and IEEE, and has received a number of research and education honors, including the ACM Sigcomm Lifetime Achievement Award, the IEEE Infocom Award, and the IEEE Computer Society Taylor Booth Education Medal. With Keith Ross, he coauthored the textbook, *Computer Networking: A Top-Down Approach*. He holds a Bachelor of Arts degree in physics from Wesleyan University, and a Ph.D. in computer science from Columbia University.

2:30pm
Break

3:00pm

AI-assisted Model Construction - Session I

This is one of two sessions devoted to semi-autonomous construction of models. The other is on Wednesday morning. One day, machines will read text, ingest data, process images, scan social media, consolidate electronic health records and integrate these and other sources into causal, mechanistic models. This session is a sneak peak at the future of semi-automated modeling.

Eidos, INDRA, & Delphi: From Free Text to Executable Causal Models

Dr. Rebecca Sharp (University of Arizona), Dr. Adarsh Pyarelal (University of Arizona), Dr. Benjamin Gyori (Harvard Medical School), Mr. Keith Alcock (University of Arizona), Dr. Egoitz Laparra (University of Arizona), Dr. Marco A. Valenzuela-Escárcega (University of Arizona), Dr. Ajay Nagesh (University of Arizona), Mr. Vikas Yadav (University of Arizona), Dr. John Bachman (Harvard Medical School), Mr. Zheng Tang (University of Arizona), Ms. Heather Lent (University of Arizona), Ms. Fan Luo (University of Arizona), Mr. Mithun Paul (University of Arizona), Dr. Steven Bethard (University of Arizona), Dr. Kobus Barnard (University of Arizona), Dr. Clayton Morrison (University of Arizona), Dr. Mihai Surdeanu (University of Arizona)

An Interpretable Modelling Pipeline for Observational, Multi-modal Biomedical Data

Mr. Vineet Raghu (University of Pittsburgh), Mr. Xiaoyu Ge (University of Pittsburgh), Prof. Panayiotis Benos (University of Pittsburgh), Prof. Panos Chrysanthis (University of Pittsburgh)

4:00pm

Lighting Talks and Posters - Session 1

This session begins with five-minute synopses of posters. The poster session follows. To ensure that poster presenters have the opportunity to see other posters, half of the posters will be presented today, half tomorrow.

Reproducibility Score for Computational Artifacts

Mr. Rakan Alseghayer (University of Pittsburgh)

Experiments for computationally-driven science are often difficult to reproduce due to underspecified parameters, limited metadata, and/or complex software. In computer science, Artifact Evaluation (AE) has emerged to incentivize authors to provide documentation, metadata, and packaging to make experiments more reproducible. AE often relies on third-party peer review to check the consistency of experimental artifacts with claims. In this paper, we take a preliminary step to examine whether machine learning can be used to construct models that predict the reproducibility of results. In addition, these models generate reproducibility scores as vectors that qualify and explain the reproducible results. These reproducibility scores can also facilitate the finding and reusability of the computational artifact that produced the results. Using data from manual peer review, we show that such models indeed hold promise for automatic artifact evaluation

Knowledge-driven Model Assembly and Execution

Dr. Alfredo Gabaldon (GE Research)

We report on a system under development that uses knowledge graphs--knowledge bases that use graphs as the main data structure---of scientific knowledge to assemble scientific models in response to users' requests to compute the value of a set of "output" variables from a set of given "input" variables. A user may also upload observational data and request the system to assemble models as hypotheses that explain the observed data.

On Modeling Adaptive Governance For Resilience in Social-Ecological Systems

Dr. Mehran Homayounfar (University of Florida)

Adaptive governance is often mentioned when discussing management of complex coupled human-natural systems or social-ecological systems facing rapid social and environmental changes. In this study, we aim at deriving governance policies that maximize resilience of the coupled natural-human systems. Our analysis is built on a stylized dynamical model that mathematically operationalizes a widely used conceptual framework, which links social components, natural resource, and infrastructure in social ecological systems. We numerically solve the Hamilton-Jacobi-Bellman (HJB) equation to determine the policies—in the form of taxation and investment in public infrastructure—that maximize a quantitative metric of the system's resilience, which is based on earlier analysis of the model. In particular, our analysis examines how the governance policies should vary in response to changes in biophysical and social factors to "optimize" the system for resilience. The results and methodology reported here should enable researchers to investigate how the optimal policies vary under different scenarios of resource availability and other incentives for the system's social agents.

Stochastic Progression Forecasting for Alzheimer's and Parkinson's Diseases

Ms. Xian Teng (University of Pittsburgh)

Forecasting patients' disease progressions with rich longitudinal clinical data has attracted much attention in recent years due to its potential application in healthcare. Existing works often tackle it by formulating a regression or classification problem that produces a single prediction in a deterministic way, which fails to deal with complex uncertain scenarios such as progression uncertainty (i.e., multiple possible trajectories) and data uncertainty (i.e., incomplete and imprecise observations). We move beyond the deterministic formulation to ask a more challenging question: What is the distribution of a patient's possible health states at a future time? For this purpose, we propose a novel deep generative model, along with an associated neural network architecture to solve this problem. Our method contains internal stochastic components that can tolerate departures of observed data from patients' true health states, and more importantly, is able to produce a comprehensive estimate of future disease progression possibilities. Based on two public datasets related to Alzheimer's

disease and Parkinson's disease, we demonstrate that our method achieves robust and superior performance than deterministic baseline approaches.

Ensemble Management under Concept Drift via Reinforcement Learning

Dr. Paul Ardis (GE Research)

Concept drifts in regression and classification problems have been shown to frustrate static modeling approaches. We show that reinforcement learning can be effectively applied to manage ensembles of models, providing needed modeling dynamism. This approach is demonstrated against industrial data to showcase potential applications in gas turbine and other complex system control domains.

Resilience Engineering for Climate Extremes and an Interconnected Vulnerable Earth (RECEIVE)

Nishant Yadav (Northeastern University)

We propose the use of resilience engineering, coupled with established resilience frameworks and risk analysis, to address climate hazards and extreme events. We aim to take an interdisciplinary approach to inform preparedness and recovery of interconnected infrastructure systems in the case of cascading failures through improved forecasting of climate hazards. With the use of big data and deep learning, we can improve the resolution of climate events to better inform smaller regions and their associated decision makers. We seek to create solutions in which hybrid approaches are taken to incorporate physics, data science, and private/public partnerships.

Using Agent-Based Modeling to Analyze Enforcement Alternatives in Spectrum Sharing Scenarios

Pedro Bustamante (University of Pittsburgh)

As radio spectrum sharing matures, one of the main challenges becomes finding adequate governance systems, and appropriate enforcement mechanisms. Historically, these processes were assigned to a central entity (in most cases a governmental agency). Nevertheless, existing literature on Common Pool Resources (CPRs) shows that other governance mechanisms are possible. In this work, we study an alternative governance approach for self-enforcement: a "government-less" environment (i.e., a distributed enforcement approach). In this "narchy" situation (i.e., lack of formal government intervention), norms, rules, and enforcement mechanisms are solely the product of repeated interactions among the intervening agents. For this reason, we have selected Agent-Based Modeling (ABM) as our primary tool. ABM allows us to represent the complex framework for spectrum sharing of the 1695-1710MHz band as a set of interacting agents in a defined environment. Further, we study how from the simple interaction of agents and their rules a long-term system can emerge based on a self-governance situation.

Are You There? Identifying Unavailability in Mobile Messaging

Pranut Jain (University of Pittsburgh)

Delays in response to mobile messages can cause negative emotions in message senders and can affect an individual's social relationships. Recipients, too, feel a pressure to respond even during inopportune moments. A messaging assistant which could respond with relevant contextual information on behalf of individuals while they are unavailable might reduce the pressure to respond immediately and help put the sender at ease. By modelling attentiveness to messaging, we aim to (1) predict instances when a user is not able to attend to an incoming message within reasonable time and (2) identify what contextual factors can explain the user's attentiveness---or lack thereof---to messaging. In this work, we investigate two approaches to modelling attentiveness: a general approach in which data from a group of users is combined to form a single model for all users; and a personalized approach, in which an individual model is created for each user. Evaluating both models, we observed that on average, with just seven days of training data, the personalized model can outperform the generalized model in terms of both accuracy and F-measure for predicting inattentiveness. Further, we observed that in majority of cases, the messaging patterns identified by the attentiveness models varied widely across users. For example, the top feature in the generalized model appeared in the top five features for only 41% of the individual personalized models.

Stochastic simulations of hybrid models with DiSH

Khaled Sayed (University of Pittsburgh)

In this work, we present DiSH (Discrete, Stochastic, Heterogeneous simulator), a stochastic simulator which is capable of simulating networks with multi-valued elements in deterministic or stochastic modes. We focus on order of update and thus incorporate information about timing, taking into account that different processes are not synchronized and certain changes occur slower than others. Another feature of DiSH is the use of grouped rules to model multi-valued elements and delays. The DiSH-Sim is publicly available and is being used to validate discrete cancer microenvironment and infectious disease models as well as economic and environmental models making it a very useful tool for discrete and logical modeling.

HPV Vaccine Impact Modelling: One-dose vs Two-Dose

Liza Hadley (London School of Hygiene & Tropical Medicine)

The Human Papilloma Virus (HPV) causes cervical cancer and is the fourth most common cancer in women, with an estimated 311,000 deaths each year (Gavi data). Using statistical modelling, we explore the impact and value of one-dose versus two-dose vaccination. This is critical research, as a one-dose campaign could resolve the considerable vaccine shortage seen in 2017, and reduce logistical constraints to vaccine delivery. Countries most at risk will then be able to introduce the HPV vaccine into their national programmes.

The presented methodology is also of interest – we use outputs from dynamic models to inform a static multi-cohort model. With support from Marc Brisson at Laval University and Yoon Choi at Public Health England, we estimate the indirect impact of vaccination (i.e. herd effects). This is incorporated into our existing model of cervical cancer incidence, and used to compare the predicted reduction in incidence between a one-dose and two-dose vaccination campaign.

This ongoing research will also look at the cost-effectiveness of each campaign, and then be used to directly inform public health decision-making.

A Mechanistic Model for How Temperature Affects Transmission of Lymphatic Filariasis by *Aedes Polynesiensis*

Sindiso Nyathi (Stanford University)

Lymphatic Filariasis (LF) is a neglected tropical disease caused by nematode parasites (primarily *Wucheria bancrofti*) and transmitted by a wide range of mosquito vectors. The global burden of LF is 120 million people, with about 886 million people at risk, primarily in tropical areas. The main vector of LF in Pacific Island nations is *Aedes polynesiensis*, which also transmits dengue virus and Ross River virus. Here, we fit thermal responses to previously published data and build a mechanistic model that quantifies how these non-linear, temperature-dependent mosquito and parasite traits affect the basic reproductive ratio, R_0 . This model can predict the influence of changing climate and seasonality on the geographic distribution, incidence and burden of LF.

6:00pm

Graduate Student Session

This session will feature a round-table conversation between grad students and a diverse group of individuals from government, industry and academia. The intention of the discussion is to expose students to a range of career opportunities. We will also emphasize other topics that the grad students have identified as most helpful.

Tuesday, May 14

8:00am

Registration and Continental Breakfast

9:00am

Social Models and Simulation

This session focuses on human social dynamics and agent-based models. It is partly about use cases, partly about simulation as a way to evaluate models.

FRED Web: an online platform for creating, running and analyzing agent-based models

Dr. David Sinclair (University of Pittsburgh), Dr. Mary Krauland (University of Pittsburgh), Mr. David Galloway (University of Pittsburgh), Mr. Michael Lann (University of Pittsburgh), Mr. Robert Frankeny (University of Pittsburgh), Dr. Mark Roberts (University of Pittsburgh), Dr. Donald Burke (University of Pittsburgh), Dr. John Grefenstette (University of Pittsburgh)

Enabling Model Complexity Through an Improved Workflow

Mrs. Christine Allen (Institute for Health Metrics and Evaluation, University of Washington), Mr. James Collins (Institute for Health Metrics and Evaluation, University of Washington), Mr. Zane Rankin (Institute for Health Metrics and Evaluation, University of Washington), Mr. Derrick Tsoi (Institute for Health Metrics and Evaluation, University of Washington), Ms. Kate Wilson (Institute for Health Metrics and Evaluation, University of Washington), Ms. Kelly Compton (Institute for Health Metrics and Evaluation, University of Washington), Dr. Abraham Flaxman (Institute for Health Metrics and Evaluation, University of Washington)

Modeling Gender Inequity in Household Decision-making

Ms. Allegra Cohen (University of Florida), Prof. Gregory Kiker (University of Florida)

10:15am

Break

10:45am

What our Modeling Assistants Know About the World

Our machine collaborators should have the benefit of extant data resources, ontologies, and semi-structured data about the world, and should be capable of inferring ontological information from data resources.

The Scientific Variables Ontology: A Blueprint for Custom Manual and Automated Creation and Alignment of Machine-Interpretable Qualitative and Quantitative Variable Concepts

Dr. Maria Stoica (University of Colorado, Boulder), Dr. Scott Peckham (University of Colorado, Boulder)

Project Tycho 2.0, An open data repository to accelerate development and testing of dynamic models in population biology

Dr. Wilbert Van Panhuis (University of Pittsburgh)

Parsing, Representing and Transforming Units of Measure

Mr. Basel Shbita (University of Southern California), Mr. Arunkumar Rajendran (University of Southern California), Dr. Jay Pujara (University of Southern California), Prof. Craig Knoblock (University of Southern California)

12:00pm

Lunch

1:30pm

Language and Human-machine Modeling

Much of the world's knowledge is written down in natural language, and much science is done collaboratively by humans who communicate in language. Can machines discuss models with people?

Human-Machine Collaboration for World Modeling

Dr. James Allen (Institute for Human and Machine Cognition), Dr. Choh Man Teng (Institute for Human and Machine Cognition)

Interpreting causal expressions with gradable adjectives to assemble dynamics models

Dr. Adarsh Pyarelal (University of Arizona), Dr. Rebecca Sharp (University of Arizona), Dr. Clayton Morrison (University of Arizona), Dr. Kobus Barnard (University of Arizona)

2:30pm

Break

3:00pm

Roundtable – The Future of the Modeling the World's Systems Conference

This is the second annual MWS conference. Shall we keep going? Who are “we”? Are we creating/serving a community? How would we measure efficacy? How big do we want to become? Shall we form a professional society? What should the conference format be? What did we like most / least about this version? The conference was created to build a community of many kinds of stakeholders around the problem of modeling and managing the world's complicated, interacting systems. Assuming that's still the goal, and we've made a good start, what should we be doing to improve?

4:00pm

Lighting Talks and Posters - Session 2

This session begins with five-minute synopses of posters. The poster session follows. To ensure that poster presenters have the opportunity to see other posters, half of the posters were presented yesterday, half today.

contact : An R Package for Processing High-Resolution Spatiotemporal Data into Contact and Social Networks

Trevor S. Farthing (North Carolina State University)

Point data obtained from real-time location systems are increasingly being used to quantify inter-animal, and animal-environment interactions. We have developed a package containing functions for 1.) "cleaning" real-time-location data via filtration, 2.) spatial and temporal data interpolation, and 3.) point- and polygon-based network creation. We provide package users with the tools to overcome or mitigate effects of limitations associated with common methods for deriving contact networks from real-time-location points, such as: 1.) the inability of point-based distance thresholds to identify where contacts occur on animals' bodies, 2.) uncertainty in contact occurrence caused by real-time-location-system accuracy, and 3.) unknown drivers of contact. Additionally, contact functions are not limited to describing inter-animal contacts, as they can also identify and quantify contacts between individuals and fixed areas (e.g., home ranges, waterbodies, buildings, etc.). We hope that by providing this resource for facilitating epidemiological, ecological, and ethological research, we can vastly improve network-model realism and researchers' ability to draw inferences from real-time location data.

Mathematical modelling to study the horizontal transfer of antimicrobial resistance genes in bacteria: current state of the field and recommendations

Quentin J. Leclerc (London School of Hygiene & Tropical Medicine)

Antimicrobial resistance (AMR) is one of the greatest public health challenges we are currently facing. To develop effective interventions against this, it is essential to understand the processes behind the spread of AMR. These are partly dependent on the dynamics of horizontal transfer of resistance genes between bacteria, which can occur by conjugation (direct contact), transformation (uptake from the environment) or transduction (mediated by bacteriophages). Mathematical modelling is a powerful tool to investigate the dynamics of AMR, however its application to study the horizontal transfer of AMR genes is currently unclear. In this systematic review, we searched for mathematical modelling studies which focused on horizontal transfer of AMR genes. We compared their aims and methods

using a list of predetermined criteria, and utilized our results to assess the current state of this research field. Of the 26 studies we identified, most focused on the transfer of single genes by conjugation in *Escherichia coli* in culture, and its impact on the bacterial evolutionary dynamics. Our findings highlight the existence of an important research gap on the dynamics of transformation and transduction, and the overall public health implications of horizontal transfer of AMR genes. To further develop this field and improve our ability to control AMR, it is essential that we clarify the structural complexity required to study the dynamics of horizontal gene transfer, which will require cooperation between microbiologists and modellers.

Analysis of Composition of Microbiomes with Bias Correction (ANCOM-BC)

Huang Lin (University of Pittsburgh)

Increasingly, researchers are finding associations between human diseases and human microbiome. Therefore, determination of differentially abundant taxa across groups, while controlling false discovery rate, is an important problem. Unlike commonly seen genomic data, the microbiome data are compositional data with excess zeros. Consequently, it is not appropriate to apply standard methods when identifying differentially abundant taxa. Recently a method called "Analysis of Composition of Microbiomes (ANCOM)" was proposed by Mandal et al. (2015) to address this problem which can be computationally intensive. In this paper, we introduce a bias-correction methodology, referred to as "Analysis of Composition of Microbiomes with Bias Correction (ANCOM-BC)", that is computationally more efficient than ANCOM while retaining all its important features. Thus, similar to ANCOM, it controls the FDR and maintains a high power. However, unlike ANCOM, ANCOM-BC also provides p-value associated with each individual taxon and can be applied to multi-group comparisons.

Data-Driven Decision Support by Coupling Nanotech, Public Health Risk Models, and Indigenous Knowledge in Rural Colombia

V.L. Morgan (University of Florida)

Mercury is a dangerous neurotoxin that can cause a plethora of health effects at levels as low as 6 ppb in drinking water. Effects are often heightened in Artisanal Small-Scale Gold Mining (ASGM) communities as this economic activity takes place only in low-income regions where governments have insufficient regulatory frameworks and/or lack the capacity to control informal extractive industries, which makes the goal of phasing-out usage of mercury extremely challenging. The goal of this study is to assess the public health risk and develop a participatory monitoring and management plan of ASGM workers and the communities surrounding them in La Toma, Colombia. The research used sensors coupled with health risk assessments in order to identify areas and populations with higher and lower risk from drinking water exposure. The information from the sensor-model platform was used to create a decision risk analysis framework for mitigating mercury threats from ASGM. Mercury was analyzed using a suite of different analytical tools (fluorescent nanoparticles, electrochemical sensors, standard, and analytical methods). A simplified hazard quotient risk calculation tool was used to determine the health risk of different community members based on water ingestion. After calculating the risk to be significantly higher than the safety threshold recognized by the World Bank (0.2), a preliminary decision analysis framework was created using stochastic multi-attribute acceptability analysis. On-going research will explore economic, public health, public acceptability, and ecological factors related to the smart sensor-risk system to rank mitigation alternatives for the two communities. All of these tools will be used to design an Agent-Based Model (ABM) in order to predict how the communities will react to different management plans and scenarios to create peace and prosperity for the communities in Colombia.

S100A9 Inhibitor Tasquinimod: A novel strategy to inhibit small cell lung cancer progression and metastasis

Salha Sassi (The Ohio State University (OSU); Benghazi University, Libya)

Small cell lung cancer (SCLC) treatment is a major clinical challenge at present as it is highly refractory to available drugs. The MDSCs/macrophages are known to help SCLC develop resistant to available therapies. S100A9 (Migration inhibitory factor-related protein 14 (MRP14)) is an EF-hand calcium-binding protein that has been involved in cell migration, invasion, proliferation, and tumor metastasis in various type of cancers, however not much is known about its role in SCLC. In this study, we found that S100A9 protein is highly up-regulated in various types of pulmonary neuroendocrine carcinomas (NEC) patient tissues compared to normal using tissue microarrays. We also observed that SCLC patients with higher S100A9 expression have significantly increased

numbers of macrophage in the stroma. We have also shown that pre-treatment of the cells with S100A9 inhibitor (Tasquinimod) suppressed in-vitro cell migration, invasion, and colony formation. In addition, we analyzed the efficacy of S100A9 inhibitor against SCLC using in vivo mouse models. S100A9 inhibitor significantly reduces tumor growth and metastasis in SCLC in xenograft mouse models. We further observed that S100A9 inhibitor suppressed myeloid-derived suppressor cells (MDSC) populations and TAMs of the M2-polarized phenotype in SCLC. Moreover, we found myeloid cells sequestered from tumors of treated mice expressed were M1 type as they showed higher levels of inducible nitric oxide synthase (iNOS), and lower levels of arginase-1. Molecular analysis revealed that Tasquinimod decreases expression of IL6, IL10, and TGF- β 1 in the cancer cells which helps inhibit macrophage activation to TAMs. Reduced proliferation and vascularization were observed in the tumors obtained from animals treated with S100A9 inhibitor. We also observed S100A9 inhibitor suppressed osteolytic bone formation in ex-vivo resorption assay. Overall, our studies, for the first time, show that Tasquinimod that targets S100A9 signaling could be used as a novel therapeutic strategy against SCLC.

The multifaceted relationship between extreme weather events, displacement and conflict: New insights from Somalia

Lisa Thalheimer (University of Oxford)

Extreme weather and migratory events have been topics of great interest for decades. More recently, a debate has emerged whether the human impact of climate change can lead to armed conflict, and how conflict and extreme climate events interact in inducing large-scale displacement. This paper explores the relationship between conflict and displacement in the context of droughts and floods across Somalia, a country in the East Africa region, in which the population has historically been using migration as coping strategy for the effects of recurring climatic extremes and socio-economic uncertainties. Since 2015, Somalia has been affected by a humanitarian crisis, paired with on-going conflict, which subsequently reduced resilience of its population. Applying panel econometric methods to monthly within-country migration observations from 18 regions together with spatio-temporal conflict and climate data, this paper quantifies the impacts of conflict and extreme weather on within-country displacement over the period of 2016 to 2018. Empirical evidence suggests significant interaction effects between conflict and extreme weather events on migration, where pre-existing conflict conditions act as accelerators of climate-induced displacement.

Using dynamic models to explore disparities in smoking prevalence and secondhand smoke exposure

Andrea R. Titus (University of Michigan)

Smoking prevalence and secondhand smoke (SHS) exposure patterns in the United States are characterized by persistent disparities across sociodemographic groups. Simulation modeling can be used to project future trends and explore the potential impact of interventions on tobacco-related health outcomes and equity.

We used dynamic models to simulate the prevalence of smoking and SHS exposure for four population groups: non-Hispanic Black (NHB) males, NHB females, non-Hispanic White (NHW) males, and NHW females between 2000 and 2050. We first implemented a series of difference equations to simulate the number of adults (>18 years) and the number of current smokers in the U.S. for each population group, age, and year. The model was calibrated using demographic and mortality data specific to each population group and an optimization algorithm to estimate group-specific smoking cessation parameters. We then ran regression models to empirically estimate the association between smoking prevalence and secondhand smoke exposure in each group using a nationally representative dataset. Using this association in combination with demographic data, we developed a second set of simulation models projecting SHS exposure prevalence. Finally, we explored the potential impact of interventions to 1) increase smoking cessation and 2) decrease the likelihood of contact between nonsmokers and active smokers. We compared results in terms of relative and absolute reductions in SHS exposure by race/ethnicity and sex.

Steps Towards Modeling Hepatitis E Virus (HEV) Infections in the Host

Kush Kumar Yadav (Ohio State University)

HEV is a major health concern in both developing and industrialized countries causing 56,600 deaths from acute viral hepatitis worldwide. HEV genotypes (g), g-1 and g-2 are strictly limited to human

infections and g-3 and g-4 are zoonotic, infecting animals (pigs, rabbits and deer) and humans. Typically, HEV causes less than 1% mortality in otherwise healthy individuals but the mortality rate for pregnant women can increase by up to 30%. Different factors been proposed for HEV mortality during pregnancy but none have been shown to be the definitive mechanism to enhanced mortality. To date, there has been no defined animal model system available for the study of HEV (human infecting strains) which has led to limited knowledge in its pathogenicity and virulence in pregnant women. This is further complicated by the lack of animal model system and difficulty to culture the virus. ORF4 is known to enhance g-1 replication. ORF4 is expressed only in genotype-1 HEV. The genotype 3 HEV strain infecting pigs does not cause pregnancy mortality and does not encode ORF4. HEV RNA (P1 and P6) was transfected in to ORF4 expressing and ORF4 absent (empty) huh7 cell lines. Despite not naturally encoding ORF4, replication of g-3 HEV was enhanced by the presence of g-1 ORF4 protein. This result suggests that the function of ORF4 protein from genotype-1 HEV is transferrable and capable of enhancing, at least, the replication of g-3 HEV. These findings direct us towards the development of a g-1 / g-3 chimeric virus which could help in the development of the first pig model system for evaluating the role of ORF4 during HEV infection while pregnant.

Parsing, Representing and Transforming Units of Measure

Basel Shbita (University of Southern California)

Data-intensive models have become critical to understanding the world. In order to reuse or combine datasets to support modeling, scientists must select, understand, and align them manually, a laborious process that requires understanding different domains and formats. To assist the modeling process, we present an unsupervised approach that identifies units in source data and provides a corresponding semantic representation. Then, we provide a method that enables scientists to perform data transformations, such as unit conversions, which are commonly necessary in modeling world systems. Our preliminary results demonstrate that our method can be used to automatically capture and transform units over spreadsheets achieving an F1-score of 0.48 in unit detection and parsing, and an accuracy of 62% in the semantic representation and transformation.

Comprehensive Reanalysis of DNA Methylation Array Data

Sean K. Maden (Ohio State University)

DNA methylation (DNAm), specifically the covalent addition of a methyl group to a cytosine in a cytosine-guanine dinucleotide, is the best studied epigenetic mark. Aberrant DNAm is studied as a biomarker of risk and progression of diseases, including cancers. DNAm is also leveraged to model biological aging and describe aberrant patterns of biological aging that may predict disease risk and patient outcome. Patterns of methylation can be studied across many published and publicly available to improve the predictive power of fitted DNAm models, uncover novel regulatory dynamics, and improve understanding of how epigenetic dynamics impact biological processes.

We obtained DNAm array data and metadata from over 35k and 2.5k samples run on the Illumina HM450k and EPIC DNAm array platforms, respectively, from the Gene Expression Omnibus. These DNAm array platforms are the most prevalent among GEO studies, where HM450k assays over 480k CpG loci genome-wide, while the newer EPIC platform assays over 850k loci. To automate harmonization of sample metadata across studies, we applied a machine learning pipeline to maps sample metadata to curated, hierarchical ontology terms with natural language processing. With this large-scale DNAm array compilation, we conducted comprehensive sample- and CpG probe-wise, cross-study analyses.

We observed that published DNAm samples, including tissues, primary cells, and cell lines, derive primarily from cancer, blood, and brain. We further observed detection p-values for CpG probes reflect probe signal magnitudes, but not dynamic range of measured DNAm, indicating commonly applied detection p-value filters can remove informative probes. Efforts to manually validate mapped ontology terms and investigate sample type-specific methylation dynamics are ongoing. We will publish our cross-study DNAm data compilation as an open-access resource for the greater DNAm research community.

An Interpretable Modelling Pipeline for Observational, Multi-modal Biomedical Data

Vineet Raghun (University of Pittsburgh)

The abundance of measurable, personalized data has the potential to transform clinical decision making. Next generation sequencing technologies allow researchers to measure genes, proteins, and

metabolites at the cellular level in individuals. Wearable and imaging data enable monitoring of clinical phenotypes and environmental factors. Since interventions can rarely be performed on human subjects, much of these data sources are observational. For modeling systems to have impact, causal knowledge must be inferred from these multi-scale observational data sources. In this paper, we propose an interpretable, causal, graphical modelling paradigm for multi-modal, observational data. Our paradigm builds upon existing domain knowledge to improve accuracy and interpretability, enabling knowledge discovery

Automated assembly of models: applications in biology and world systems

Yasmine Ahmed (University of Pittsburgh)

In biology, model creation is highly dependent on human intervention. It requires reading hundreds of papers to extract useful information. Moreover, the amount of biological data is constantly growing, further exacerbating the issues of data inconsistency and fragmentation. Therefore, automating the process of building new and extending existing models is critical for consistent, comprehensive, and robust studies of biological systems. Here, we propose a method to automatically and efficiently assemble and incorporate the information extracted from literature into models through the process of model extension. This is achieved by integrating clustering, with simulation and formal analysis to allow for an automated iterative process of assembling, testing and selecting the most relevant models.

Specifically, clustering identifies connected groups of newly extracted interactions from literature, while simulation and formal analysis select those clusters of elements that provide the best model extension towards satisfying pre-defined system properties. To test our approach, we compared the model that we obtained using our automated extension method with the previously published manually extended T cell model. Besides automatically and rapidly reconstructing the manually extended model, our approach can provide multiple viable extended model versions. As such, it replaces a large number of tedious, or even impractical, manual experiments, and guides alternative interventions into real biological systems. Moreover, our model extension automation methodology is not limited to biology, as it is applicable to many other systems. To this end, our future work includes applications to large systems that include social, economic, political, environmental and agricultural components.

Collective Conflict Prediction using Big Data Analytics

Chrism Watson Ross (University of New Mexico)

How do we detect and predict collective conflict in real time? Globalization drives complex socioeconomic and political breakdowns which traditional techniques and data sources lack sufficient precision and power to analyze and mitigate, as our deeply interconnected world provides greater opportunity for collective organization, rumor mongering, incitement by fabricated stories, and propaganda dissemination. Triggers- a tweet gone viral, a troll swarm painting a target, emergence of a new terrorist organization, or new alliances driving political conflict- all exhibit near-real-time digital signatures. We aim to develop a probabilistic conflict forecasting system capable of assimilating real-time observations, in the spirit of those used for weather forecasting, to assess current unstable regimes and forecast potential conflict. Utilizing multiple machine learning linguistic models we have analyzed subreddit threads on reddit, and have applied several topic modelling approaches to a multilingual data set of N tweets. We hypothesize that with regression, graph theoretical, and temporal analysis we will learn to infer topic network structure and dynamic characteristics correlated with real-time conflict, both online and on the ground.

5:30pm

Demos

The demo session includes presentations from:

Probabilistic Relational Agent-based Models (PRAM)

Paul R. Cohen (University of Pittsburgh), Tomek Loboda (University of Pittsburgh)

PRAM is a probabilistic agent-based modeling framework that unifies compartmental models, probabilistic models and agent-based models. It also assembles agent-based simulations largely autonomously. We will demonstrate PRAM assembling models from data about Pittsburgh.

MOIRE: Modeling and Inference from REading

Pascale Proulx (Uncharted Software) Fahd Husain (Uncharted Software) Benjamin M. Gyori (Harvard Medical School) Adarsh Pyarelal (University of Arizona)

MOIRE is a super-system for large-scale causal inference jointly developed by Uncharted Software, the University of Arizona, and the Harvard Medical School for DARPA's World Modelers program. The components of this super-system constitute an end-to-end causal modeling pipeline, involving (i) the knowledge assembly of causal fragments read from external sources, (ii) the exploration and correction of this knowledge base of causal statements, (iii) the iterative assembly and parametrization of a causal graph, and (iv) interactive interventions with the resulting computational causal model. The specific components discussed will be INDRA, a library for causal knowledge assembly from raw sources; CauseMos, a human-machine interface for graph exploration and model assembly; Delphi, the computational back-end engine for probabilistic causal modeling; and CauseWorks, a planning interface for intervention modeling. Current challenges in modeling dynamic complex systems will be addressed, along with sketches of future directions for causal research.

Vivarium: Enabling Model Complexity Through an Improved Workflow

James Collins (University of Washington) Christine Allen, Researcher (University of Washington)

Using data from the Global Burden of Disease study, the Institute for Health Metrics and Evaluation's Simulation Science team created a framework and modeling process to enable development of modular healthcare intervention packages, variable groupings of customizable intervention microsimulations. This allows for more timely and efficient model development, and provides the potential for greater model complexity. The team accomplished this by expanding upon their existing simulation framework, then identifying three specific process improvement goals: a method for specifying intervention packages; a modeling process that increases productivity; and a documentation strategy that facilitates transparency. They met those goals by continuing development of their existing tool, developing a clearly defined and iterative model workflow, and integrating within that workflow a clear set of templated phase outputs.

Project Tycho 2.0

Dr. Wilbert Van Panhuis (University of Pittsburgh)

Dynamic models that represent a real-world scenario, such as a measles epidemic in location A and year Y, can be used to inform policy making in the real-world, for example about a vaccination program. Models representing real-world scenarios require observational datasets, but such datasets are often difficult to get. We created a global repository named Project Tycho for standardized datasets. The Project Tycho repository has 4100 registered users and our datasets have been used to create 47 new scientific works, including 17 modeling papers, demonstrating the value of standardized data repositories to accelerate model development and testing.

Semantic Program Analysis for Scientific Model Augmentation

Dr. James Fairbanks (Georgia Tech Research Institute)

SemanticModels.jl is a system that leverages techniques from static and dynamic program analysis to mine the code and text associated with scientific models in order to improve, compare, and combine models via the iterative extraction and inclusion of latent semantic information. We provide a set of code introspection methods that allow scientists to extract information about how a program's methods, variables, and data types interact to implement concepts from the scientific domain. This information in turn can be used to inform a variety of type reflection methods. By framing our metamodeling objectives as metaprogramming tasks, SemanticModels.jl enables writing programs that generate models. To this end, we present a category theory-based framework for defining metamodeling tasks and extracting semantic information from model implementations, and show how SemanticModels.jl can be used to augment scientific workflows in the epidemiological domain.

ModelingToolkit.jl

Dr. Chris Rackauckas (Massachusetts Institute of Technology)

Domain-specific languages (DSLs) enable scientific programmers to describe their models in a high-level format while still allowing for internal optimizations. However, while the homoiconicity of Julia has encouraged the proliferation of DSLs, code reuse between DSLs has been stymied due to the lack of a common intermediate representation for performing inspection, symbolic transformation, and

compilation of scientific models. In this paper, we present ModelingToolkit.jl, a Julia library and intermediate representation for scientific DSLs. We discuss the internal representations of symbolic expressions and systems, chosen to implicitly enforce logical invariants. We showcase how ModelingToolkit is being used in the pharmacometric modeling DSL PuMaS.jl, and describe future improvements to incorporate this system into other Julia packages.

Lightweight Approaches to Needs Estimation

Ben Watkins (Kimetrica,) Michelle Archuleta (Kimetrica)

Policy end-decision makers are tasked with understanding a complex world ecosystem and making strategic and tactical decisions that will inherently define the success or failure of their mission. Oftentimes these decisions are made in isolation without understanding the drivers of influence in the ecosystem, without leveraging relevant and rich datasets, and without predictive analytical approaches that can provide intuition.

Our aim is to showcase the power of lightweight machine-learning models that coupled together provide a framework for understanding, interpreting and predicting a complex world ecosystem. In particular, we will focus on understanding a needs surface this may be a malnutrition needs surface such as the number of malnutrition cases in a particular geographical area or a caloric deficit needs surface that provides a geographic representation of the estimated daily caloric intake for a population. Our approach is to provide a modeling framework that is easily adaptable such that new models could be plugged in or swapped out. Graphical representations of models, variable CAGs, scenario definitions and superset dashboards are designed to provide intuition into a complex world ecosystem model. Our end goal is to harness rich datasets, build out complex machine-learning and expert models, while delivering a product that communicates simplicity and intuition that can be delivered to a policy decision-maker.

The Ecological Forecasting Initiative

Heather J. Lynch (Stony Brook University) and Michael C. Dietze (Boston University)

Near-term ecological forecasting is an emerging interdisciplinary research area focused on accelerating environmental research and making it more relevant to society. Over the last few years, a growing network of ecologists have been building a community of practice around these ideas, leveraging existing networks in ecology, conservation biology, and resource management and actively recruiting expertise in computing (e.g., data science, cyberinfrastructure) and the social sciences (e.g., decision support, stakeholder engagement) critical for long-term sustainability. The inaugural Ecosystem Forecasting Initiative Conference (EFI2019) is occurring at the same time at the Modeling the World's Systems Conference at AAAS Headquarters. Beyond this conference, the EFI is establishing persistent working groups in six areas (Organizational Partnerships, Decision Science, Public Engagement, Training & Diversity, Theory/Synthesis, Methods & Tools, Cyberinfrastructure) and have launched a website (<https://ecoforecast.org>) containing information about emerging field of ecosystem forecasting, ongoing projects, and a community bulletin board. We encourage anyone interested to get involved and look forward to discussing our efforts and how others can get engaged with the process

Wednesday, May 15

7:30am

Registration and Continental Breakfast

8:30am

AI-assisted Modeling - Session 2

This session continues the theme of semi-automated model construction, introduced on Monday.

ModelingToolkit.jl: An Intermediate Representation for Scientific Domain-Specific Languages

Mr. Harrison Grodin (Carnegie Mellon University), Mr. Yingbo Ma (University of California Irvine), Dr. Chris Rackauckas (Massachusetts Institute of Technology)

Semantic Program Analysis for Scientific Model Augmentation

Mrs. Christine Herlihy (Georgia Tech Research Institute), Mr. Kun Cao (Georgia Institute of Technology), Mr. Sreenath Reparti (Georgia Institute of Technology), Dr. Erica Briscoe (Georgia Tech Research Institute), Dr. James Fairbanks (Georgia Tech Research Institute)

Machine-Assisted Extraction of Formal Semantics from Domain Specific Semi-Formal Diagrams

Dr. Eric Davis, Mr. Alec Theriault, Dr. Eddy Westbrook, Mr. Ryan Wright (Galois, Inc.)

AutoMATES: Automated Model Assembly from Text, Equations, and Software

Dr. Adarsh Pyarelal, Dr. Marco A. Valenzuela-Escárcega, Dr. Rebecca Sharp, Mr. Paul Hein, Mr. Jon Stephens, Mr. Pratik Bhandari, Mr. HeuiChan Lim, Dr. Saumya Debray, Dr. Clayton Morrison (University of Arizona)

10:15am

Break

10:30am

Toward a Modeling Stack

Chair: Bruce Childers, *University of Pittsburgh*

Panel Members: *Joshua Elliot (DARPA), Eric Davis, James Fairbanks (GTRI), John Bachman (Harvard Medical School), Adarsh Pyarelal (Arizona)*

This is a panel discussion about the possibility of building a community resource, a software environment that might be called the Modeling Stack. The conference introduced us to a variety of technologies to support semi-autonomous modeling. Could they be unified in a single environment that provides ontologies, data resources, workflow systems, compilation of models from multiple sources, collaboration tools, curation tools and more? What would it take to build a unified and unifying environment for modeling? This session will feature panelists who have made progress toward unified modeling environments.

11:45am

Wrap-up