

Statement of Significance: *Mining relationships among variables in large datasets from complex systems*

Facilitated by advances in high-performance computing and the improved availability of fine-scale socio-economic data, social science researchers have developed new computational modeling methods to explore the dynamics and consequences of human interactions. These methods include agent-based models (ABMs), which create virtual worlds of social agents and their interaction environments in computer code. Adoption of new methods has been motivated by a desire to explore research questions that can not be investigated using traditional equilibrium-based and statistical modeling approaches—questions related to how interactions between heterogeneous, adaptive social actors at a fine scale collectively shape emergent social outcomes at aggregate scales in complex socio-ecological systems. The non-equilibrium, computational structure of ABMs allows the creation of models with a sufficiently high level of detail to explore such questions.

Members of our research team have used ABMs to explore dynamic feedbacks between human resource use and the biophysical environment to develop improved policies for sustainable resource use. Parker and Jin (University of Waterloo, Canada) model interactions between neighborhood design, residential land markets, landscape management and ecosystem service generation, transportation behavior, and traffic patterns. Polhill and Dawson (University of Dundee and the James Hutton Institute, Scotland) model relationships between agricultural land use and management, landscape structure, and biodiversity. Filatova and Voinov (University of Twente, the Netherlands) models how evolution of flood risk perception may lead to abrupt structural changes in coastal cities (mass outmigration, collapse of the housing market, and exposure of the poorest households to the highest risk levels). Barton (Arizona State University, US) models the interactions of land-use and landscape, social dimensions of climate change, and collective action for managing common pool resources. Several team members (Parker, Polhill, and Barton) have been early adopters of these models and leaders in the development of methods, priorities, and infrastructure to support them.

In the last decade, ABMs modelers have made substantial progress to identify compelling research questions, design models to investigate them, and develop methods to bring real-world data into models at both the creation and evaluation stage. Since the ABMs we design have stochastic elements and many potential parameter combinations, multiple model runs that sweep parameters are conducted, creating large quantities of computationally generated, hyper-dimensional, “big data” from which we hope to extract answers to social science research questions. Yet, we lack appropriate methods to mine, analyze, and synthesis large-scale model output data in order to answer our questions. Traditional analysis methods for mapping relationships between input parameters and output data—in both real-world and computational data—are designed for data that are linear, continuous, and normally distributed. However, data from models of complex socio-ecological systems are non-linear, discontinuous, and power-law distributed.

Our proposed project seeks to address this challenge by developing, applying and disseminating an integrated environment for visualization and analysis of data generated by complex systems. The project builds on existing analysis algorithms from each of our projects: prototype web-based data management, analysis, and communication tools; and the NSF-supported CoMSES Net Computational Modeling Library (CML). Each research team will contribute the output data from their computational modeling efforts, as well as existing and newly developed analysis algorithms, to the CML. Each team will also publish the results of their data analysis, comparing application of the algorithms between their generated and real-world validation data. On the CoMSES Net platform, a new suite of collaborative, open-source tools will be developed that will allow any users to 1) post their model output, as well as enter metadata describing the structure of their output data 2) visualize and analyze their output data using our new tools 3) comment on and share their analysis 4) conduct comparative and meta-analysis, drawing on output data from other projects.

Our project builds on the expertise and established collaborative relationships of a team of senior and emerging junior researchers. Our analysis tools and platform will provide an automated means of discovering relationships among variables offer an opportunity to uncover new theories about how social systems work, test the realism of the simulation against knowledge from empirical systems, and propose new parameter subspaces to explore. The collaborative tools will improve communication and reduce barriers to entry to new users.