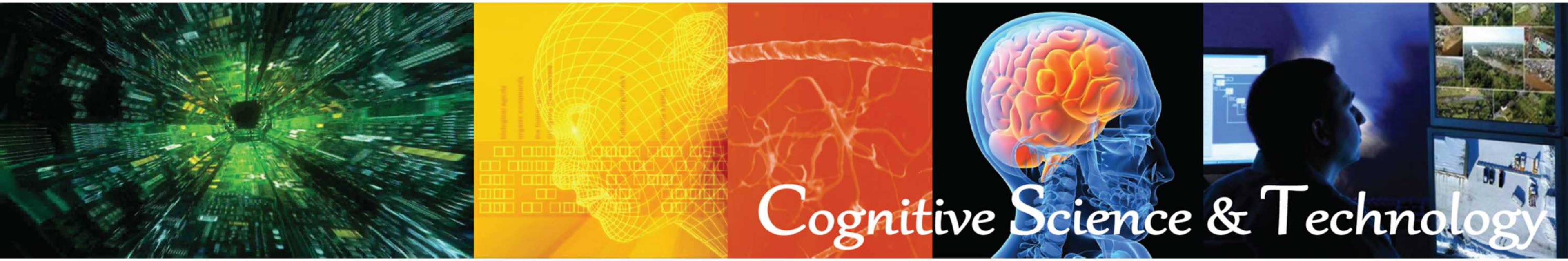
SAND2018-7559C



A Multi-tiered Complexity Metric

Asmeret Naugle, Kiran Lakkaraju, **Stephen Verzi**,
Laura Swiler, Vicente Romero and Michael Bernard

ICCS 2018 July 22-27, 2018

This research was developed with funding from the Defense Advanced Research Projects Agency (DARPA).
The views, opinions and/or findings expressed are those of the author and should not be interpreted as representing the official views or policies of the Department of Defense or the U.S. Government.

Sandia National Laboratories is a multimission laboratory managed and operated by National Technology and Engineering Solutions of Sandia, LLC, a wholly owned subsidiary of Honeywell International, Inc., for the U.S. Department of Energy's National Nuclear Security Administration under contract DE-NA0003525.

Outline

- Why is complexity important in social simulation?
- Considerations in measuring complexity
- Multi-tiered complexity metric
- Example
- Next Steps

Complexity & Social Simulation

- Are we using the right type of simulation to represent a system?
  - Simple enough to understand and analyze
  - Complex enough to represent
- Does complexity of model structure (ground truth) relate to complexity of behaviors observed (outputs)?
- Does real-world social complexity give insight into explaining, predicting, and prescribing systems?

Measuring Complexity

- ***Challenges***
  - Many definitions of complexity – how do we capture what is important?
  - How to avoid the temptation of focusing on easy measurements (e.g., number of actors represented)
- ***Considerations***
  - Want to capture complexity of the actors, environments, interactions, and outputs of a simulation
  - Adding more actors or links between actors could increase complexity but not always
  - Entropy (Shannon) establishes information content but does not distinguish uncertainty from randomness
  - Available actor/group decisions and/or behaviors if they are realized in output and affect other decisions and/or behaviors

A Multi-Tiered Complexity Metric

1. **System intricacy**
   - *How complicated is the system?*
   - Captures information about variables and the potential for interaction among variables
2. **Information theoretic complexity**
   - *What is the information content and uncertainty in the system?*
   - Considers the simulation states, and requires simulation runs to compute
3. **Behavioral complexity**
   - *How do interactions and behaviors of actors in the system affect complexity?*
   - Inspired by efforts in the social sciences to quantify complexity of real social systems

Tier 1: System Intricacy
How complicated is the system?

Metric: **Cyclomatic Complexity**

- Captures the interconnectedness of a graph
- Calculated using ground truth (graphical representation of the causal structure driving the simulation)

$$M = E - N + 2P$$

$M =$ cyclomatic complexity
$N =$ nodes in the graph
$E =$ edges in the graph
$P =$ connected components

Tier 2: Information Theoretic Complexity
Information content & uncertainty

Metric: **Forecast complexity**

- Measure of the minimum amount of information needed for optimal prediction
- Part of the time-series, X^-, is used to predict the rest, X^+, such that $X = (X^-, X^+)$, for (model) f in M (set of models)

$$C = \min_{f \in M} H\big(f(X^-)\big)$$

Tier 3: Behavioral Complexity
Actor interactions & behaviors

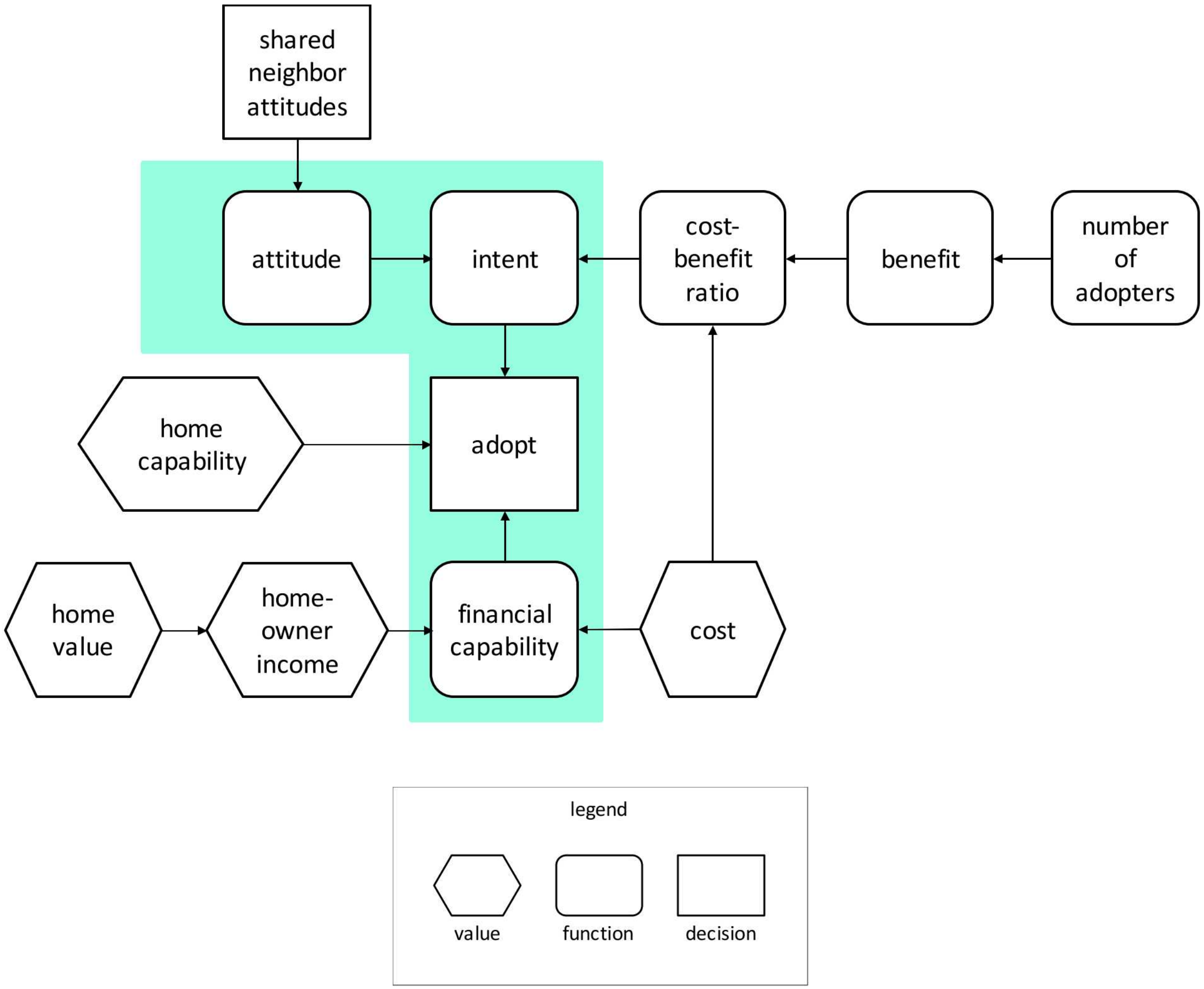
Metric: **Number of Differentiated Relationships**

- Used in the social sciences to measure complexity of animal groups
- Generalizable across simulations, since it allows for the definition of relationship to be tailored to the simulation

Example: The PV World Model

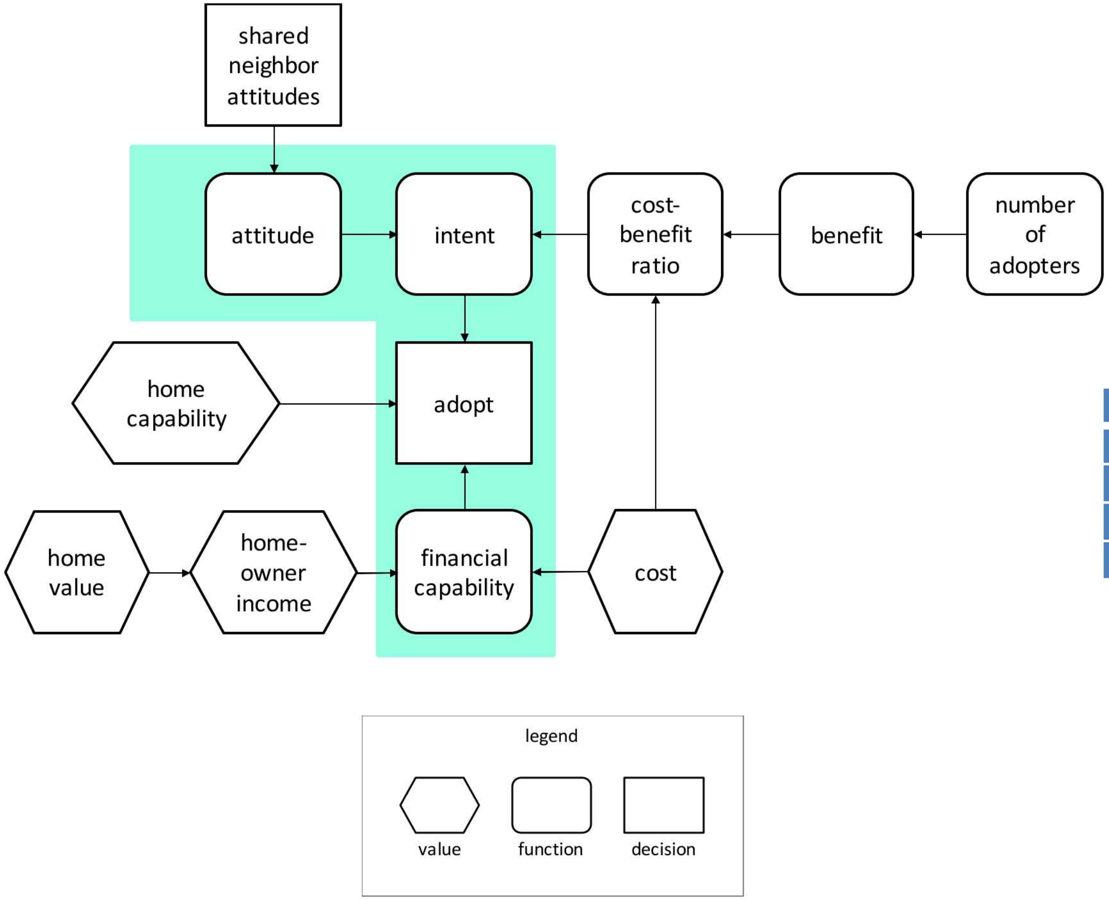
- Created by Sandia staff as part of an internal complex systems class
- Agent based structure
  - Agents represent households deciding whether or not to adopt PV solar
- How will the diffusion of rooftop solar be affected by attitudes, social networks, installation costs and adoption incentives?

PV World Model: Ground Truth



- If two parameters have set values and are perfectly correlated, the link between them is not included
- Includes all direct influences on actor states (in green color)
- Includes all interactions between actors
- Relationships are represented as simply as possible
- Also contains the remaining influences

PV World: Cyclomatic Complexity
System Intricacy



$$M = E - N + 2P$$

M = cyclomatic complexity

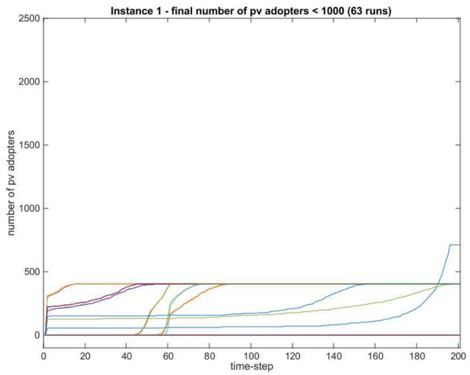
N = nodes in the graph
E = edges in the graph
P = connected components

	System Intricacy
Edges (E)	12
Nodes (N)	12
Connected Components (P)	1
Cyclomatic Complexity	12-12+1*2=2

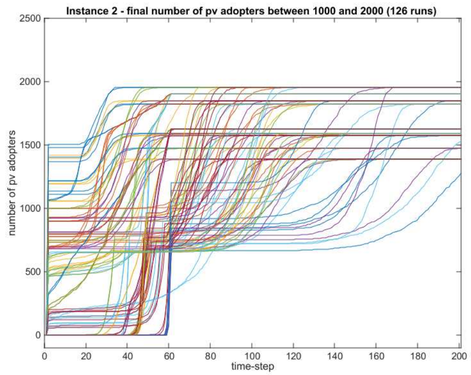
System Intricacy
= E – N + 2P
= 12 - 12 + 2
*= **2***

PV World Model – 3 Instances

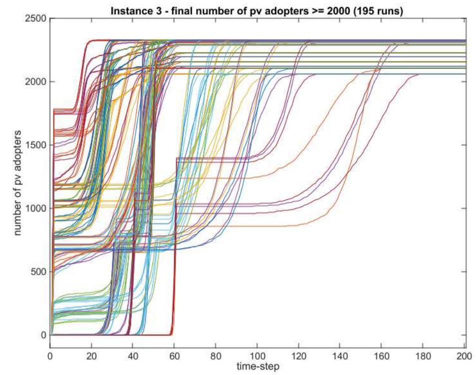
< 1000

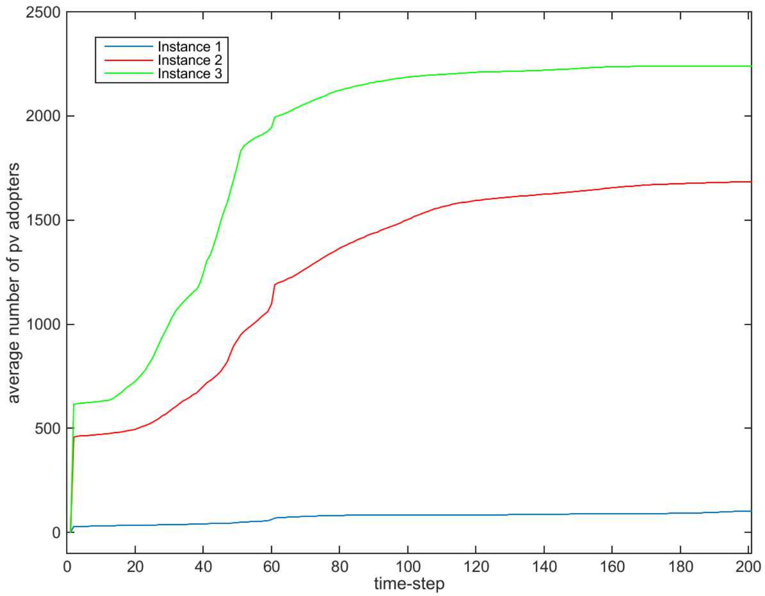


between 1000 and 2000



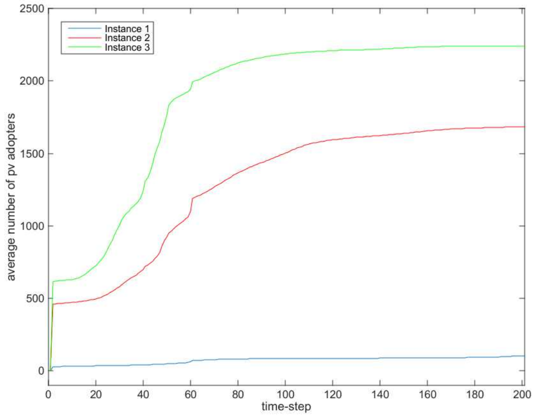
>= 2000



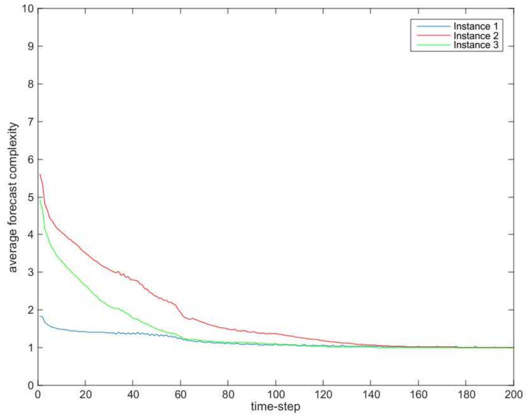


PV World: Forecasting Complexity
Information Theoretic Complexity

- Calculated forecast complexity using a compression algorithm



Number of adopters



Forecast complexity

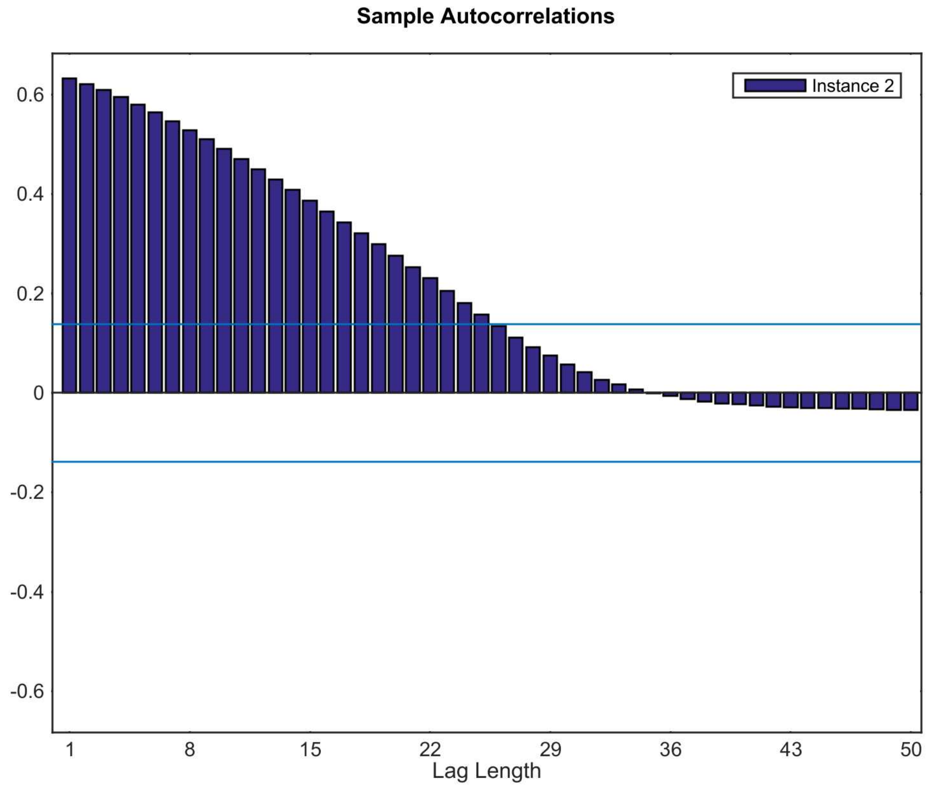
Instance 1: <1000 adopters
Instance 2: 1000—2000
Instance 3: >2000

	Instance 1	Instance 2	Instance 3
Mean FC	1.16	1.81	1.44
Min FC	1.00	1.00	1.00
Max FC	1.83	5.61	4.93

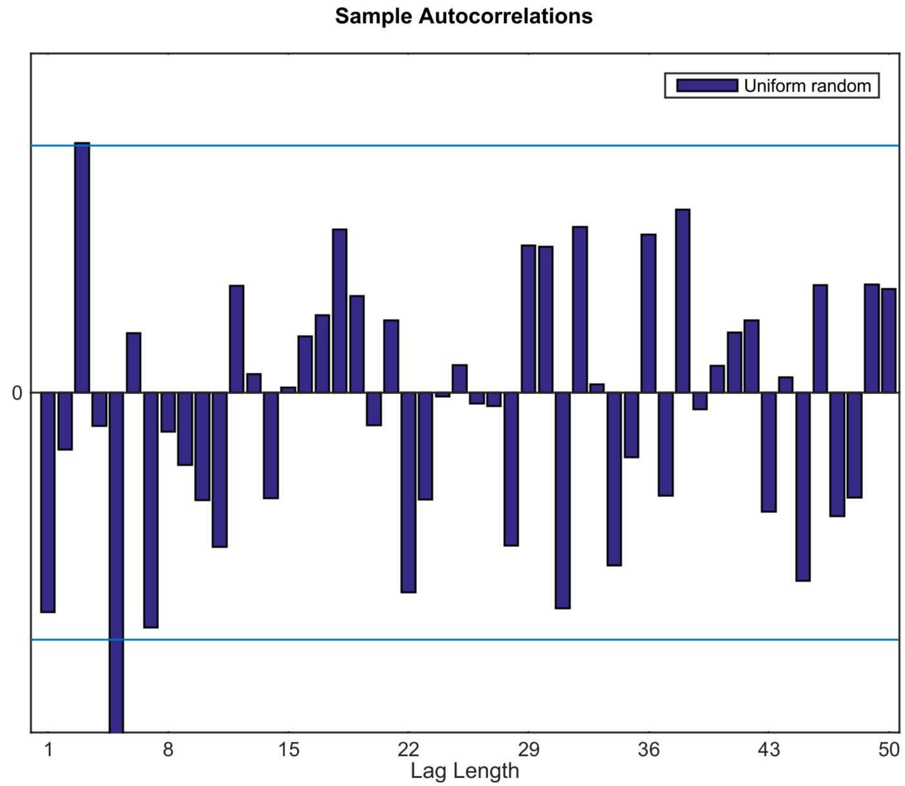
The overall forecast complexity is the weighted (by number of runs per instance) average forecast complexity over these instances.

*Forecast complexity = **1.52***

Example Time-series – Instance 2

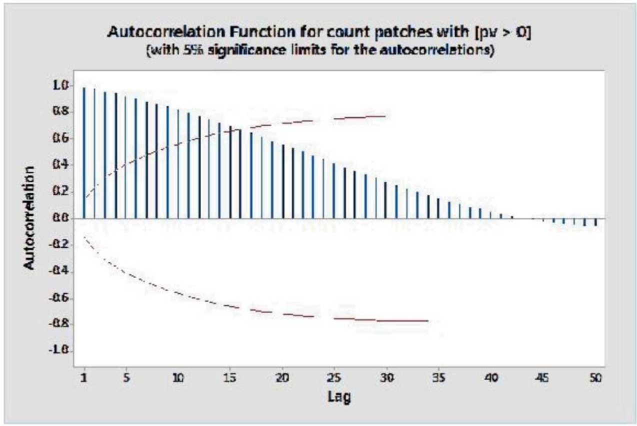


Example Time-series – Uniform random

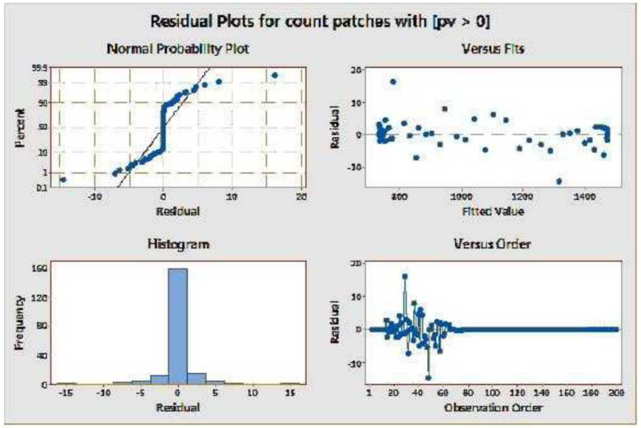


PV World Model – Auto-regressive Model (Instance 2)

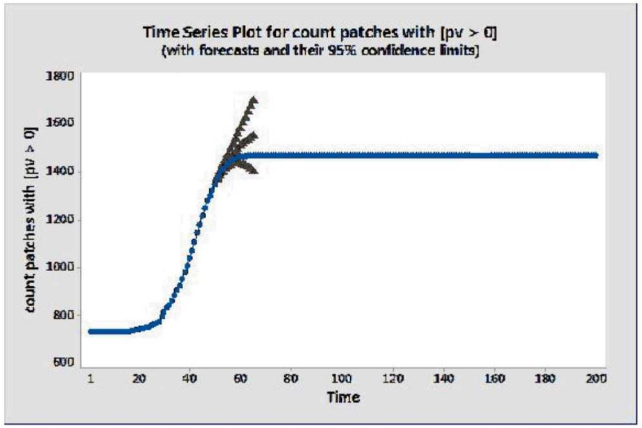
Autocorrelation Function
(55% significance limits)



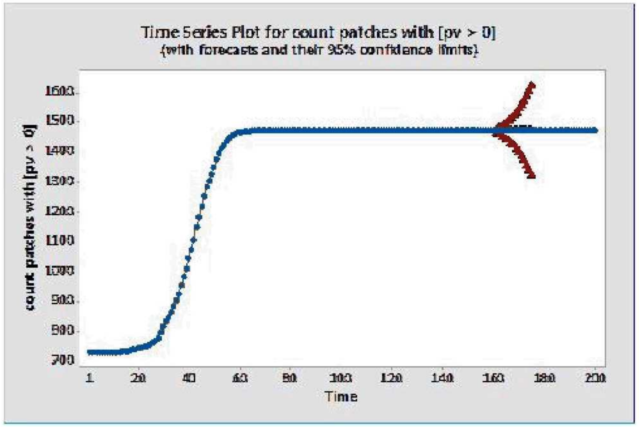
Auto-regressive Model
Residual Plots



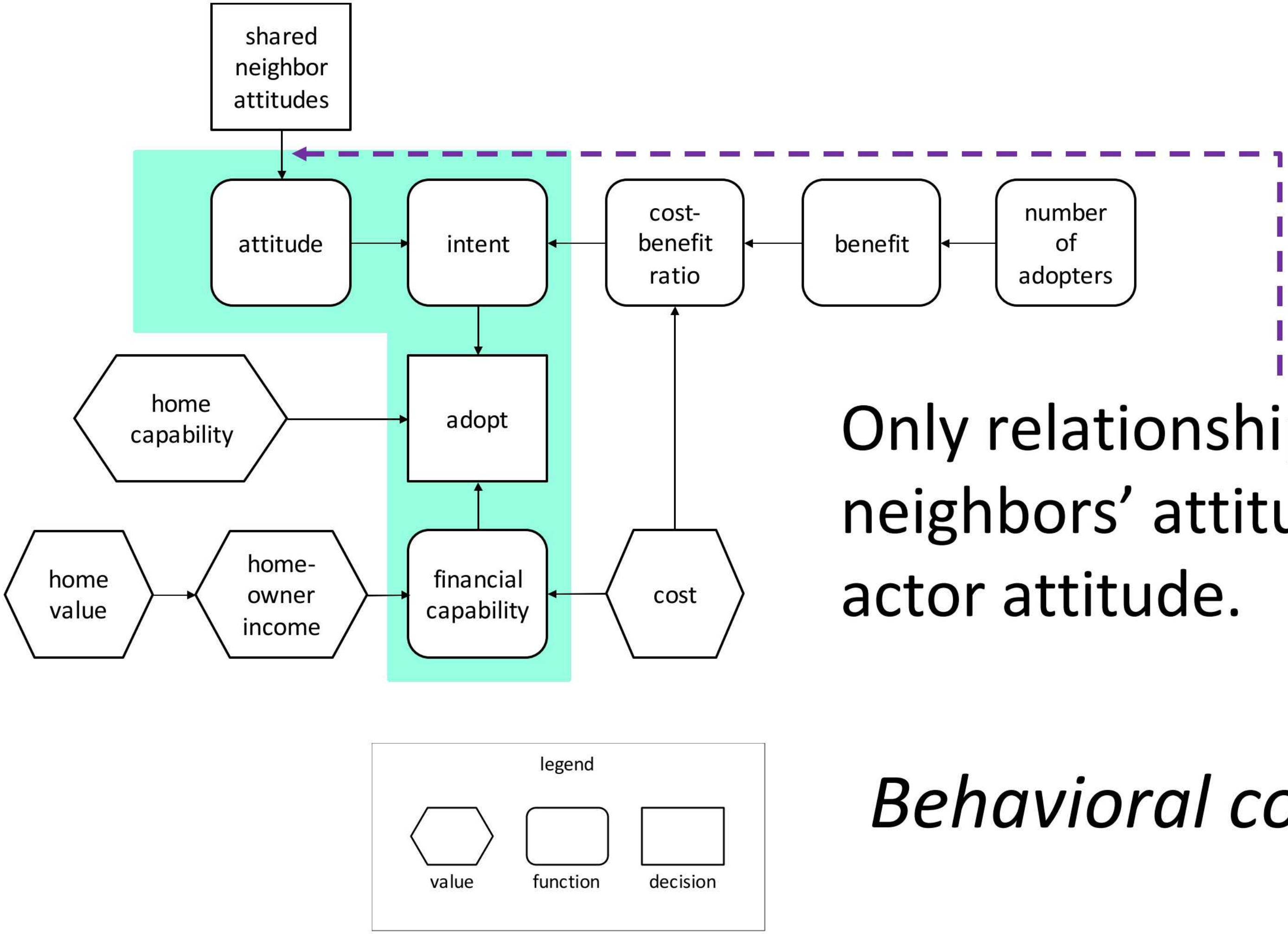
Prediction at t=50



Prediction at t=160



PV World: Differentiated Relationships
Behavioral Complexity



Only relationship in the model: neighbors' attitudes affect actor attitude.

*Behavioral complexity = **1***

Next Steps

- Align multi-tiered complexity metric to simulation dimensions
  - With or without ground truth
  - With or without assumption of social/group interactions (including hierarchical group formation)
- Accounting for feedback loops in ground truth

Thank You

Questions?

Backup

Extended Abstract

Measuring complexity in real-world systems and in computational models of these systems is an ongoing area of research. This research concerns modeled social systems of multiple actors interacting in complex scenarios over lengthy periods of time. Existing theoretical measures such as Kolmogorov complexity or algorithmic information complexity (AIC) are mathematically precise, but they cannot be computed directly as it is extremely difficult to determine the minimal representation (such as an algorithm or code) for any system/model output data. In this research, we explore the suitability of a multi-tiered approach to measure complexity, incorporating three interacting but complementary aspects of complexity: (1) system intricacy, (2) information theoretic complexity, and (3) behavioral complexity.

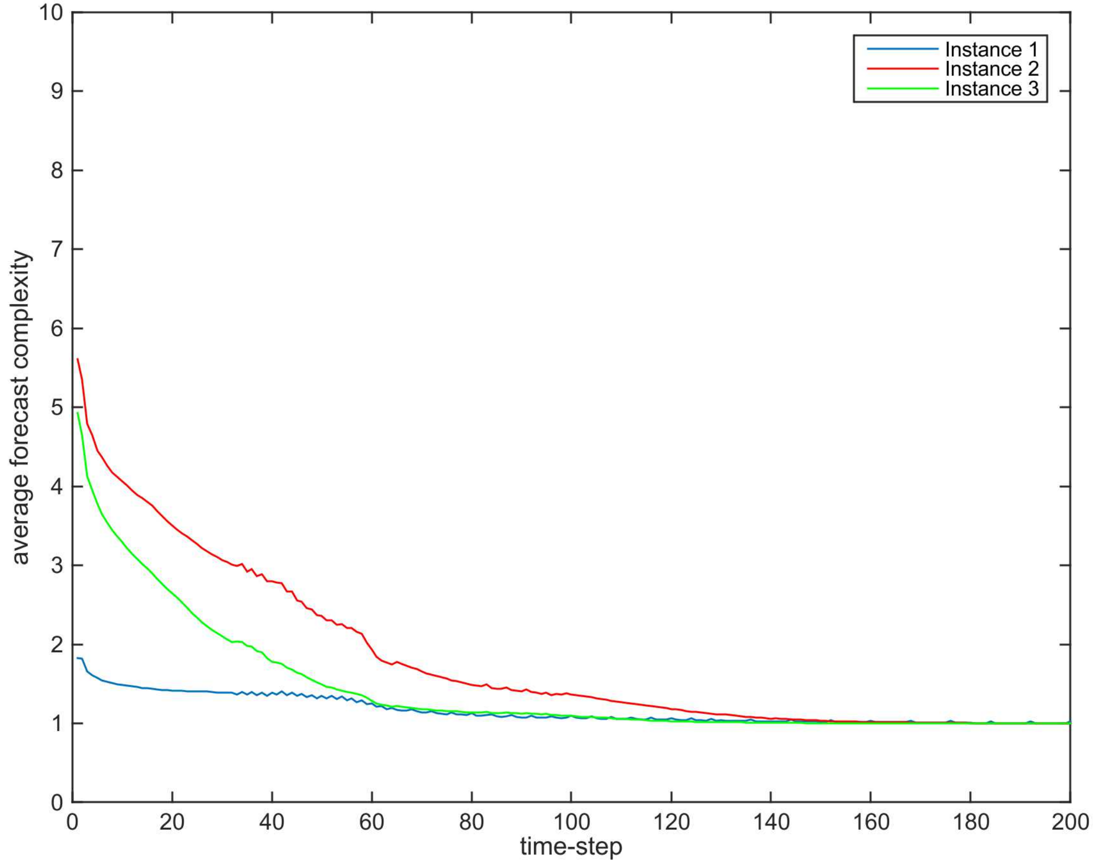
System intricacy is intimately tied to the structures, processes and parameters that determine system dynamics, which makes general measures difficult to identify. One approach is to consider the underlying software implementation and evaluate its complexity. Measures that relate to the parameters and structure of the system are the most relevant as they address the system rather than the implementation. System intricacy measures are based on formal descriptions of the causal structure of the system. Metrics include (but are not limited to): number of actors, number of characteristics of actors, number of potential behaviors per actor, number of types of actors, number of interactions between variables in the causal structure; and assuming a graph representation of the causal structure: the number of spanning trees of the graph, cyclomatic complexity, space of locations in which the actors move, and the number of parameters external to the actor that influence actor behavior.

Information theoretic complexity measures information content as well as uncertainty in information content. These measures are data driven and focus on the empirical data that is the output of the simulation. Information theoretic measures of complexity have been studied in the research community including (but not limited to): entropy or information content, mutual information, forecasting complexity, Kullback-Leibler divergence, and normalized compression distance. Each of these are included in the information theoretic tier of complexity metrics.

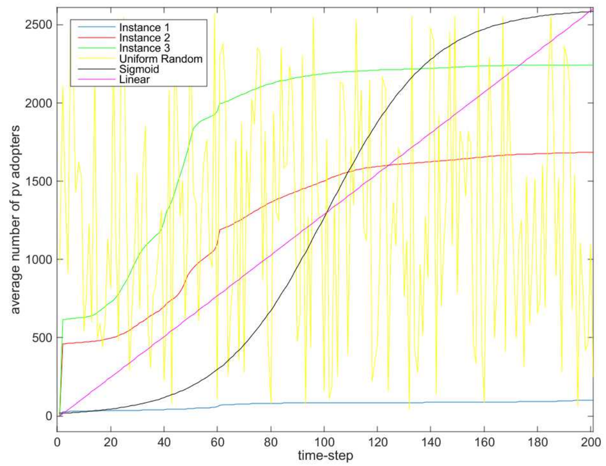
The social sciences have investigated the concept of "social complexity", which is sometimes associated with understanding how a society is governed. Several measures have been proposed to capture this concept, but they often count the existence of specific constructs from the real world, such as the authority of leaders, territorial control, tax extraction, etc. The primary focus of the behavioral complexity metric tier is behaviors and interactions of actors. We leverage definitions from relevant social science literature, including (but not limited to): the number of differentiated relationships; the number of interactions an actor has with other actors; the number of behaviors an actor executes in their lifetime; the number of unique behaviors an actor executes in their lifetime; and the range of change in an actor's characteristics throughout its lifetime. Actors interact with other actors by sending messages, modifying a joint shared state (such as an environment), or executing joint behavior. We extract the interactions between actors and create a social network where the nodes represent actors and the edges represent significant interaction.

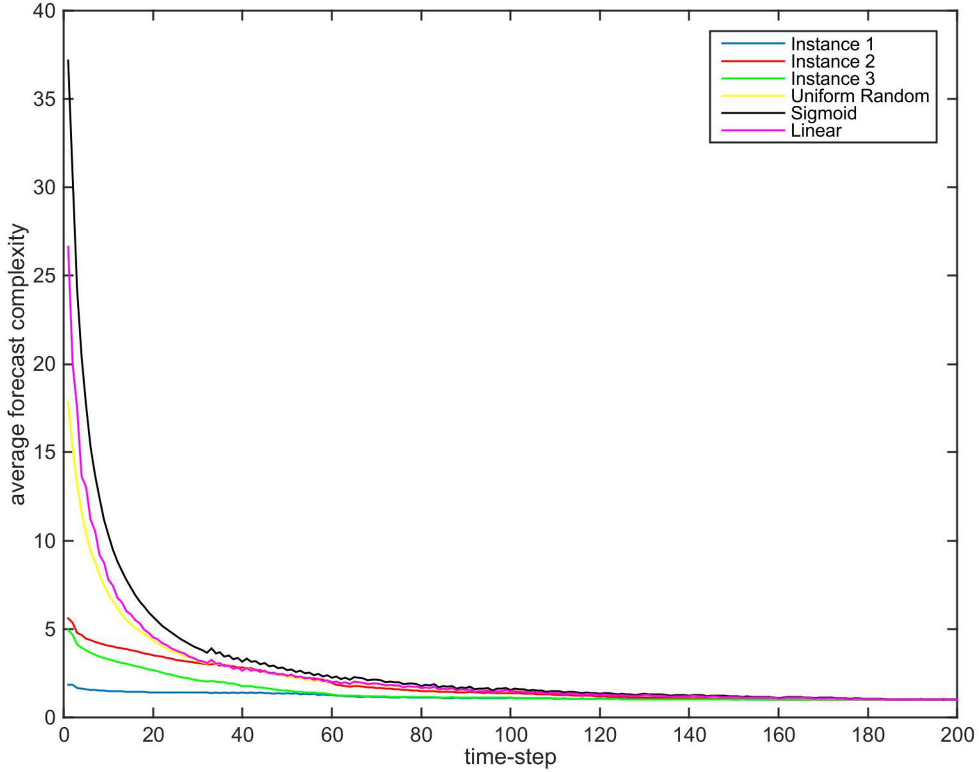
We demonstrate our multi-tiered complexity metric using output from a simple agent-based adoption model as well as other simple (non-complex) and degenerate (i.e., noise-only) time-series.

PV World Model – Forecast Complexity

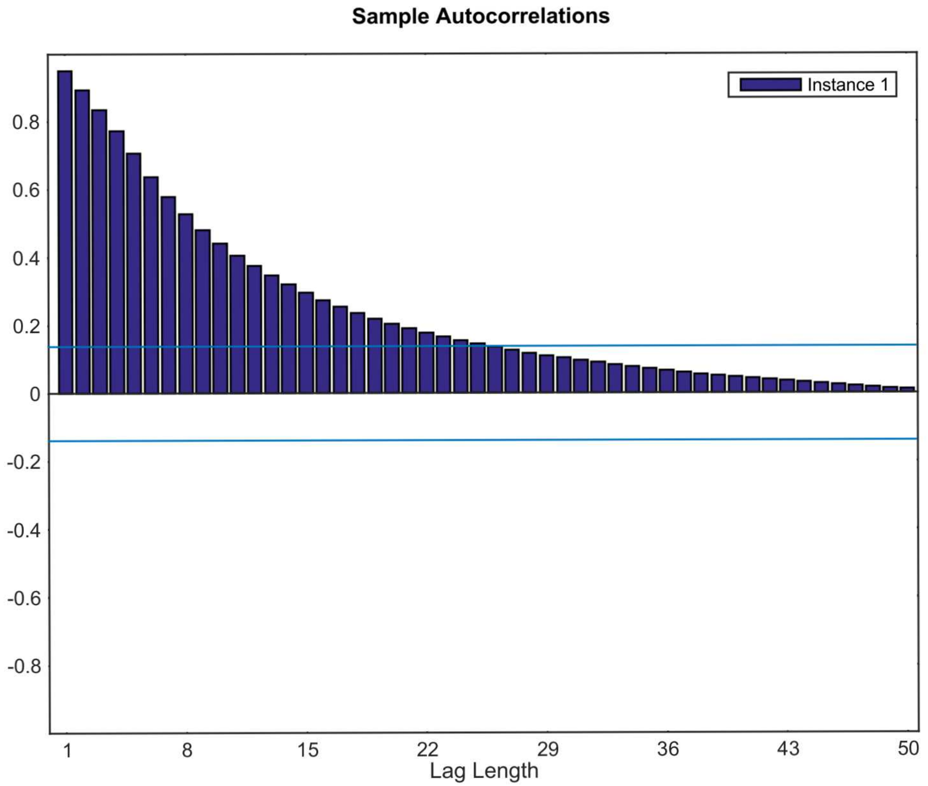


Forecast Complexity

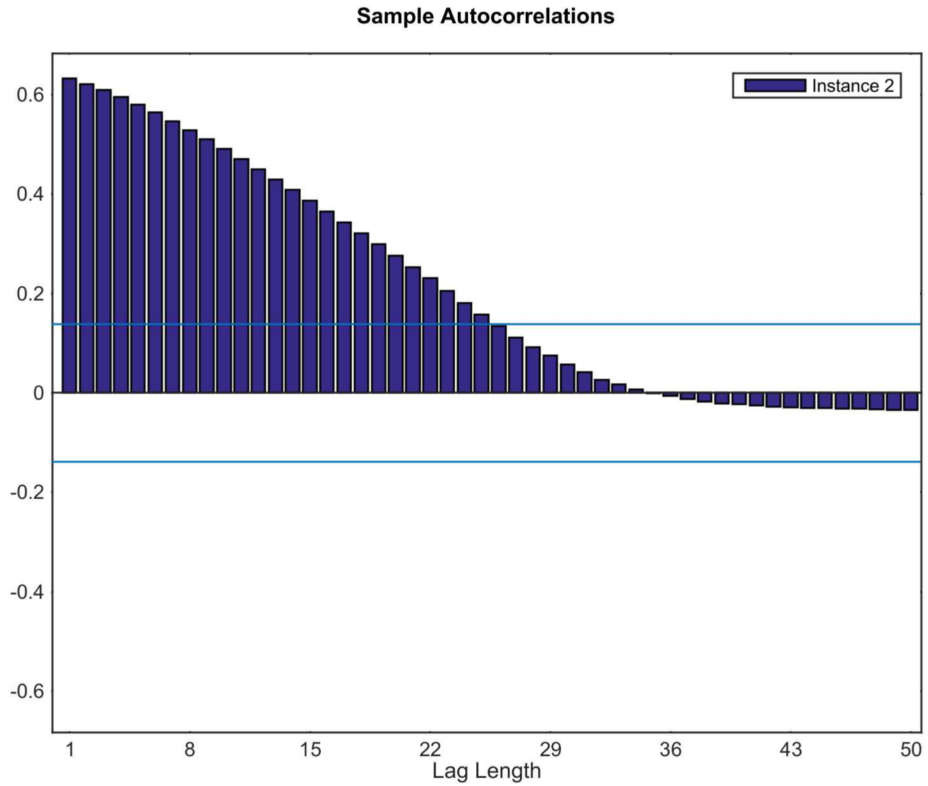




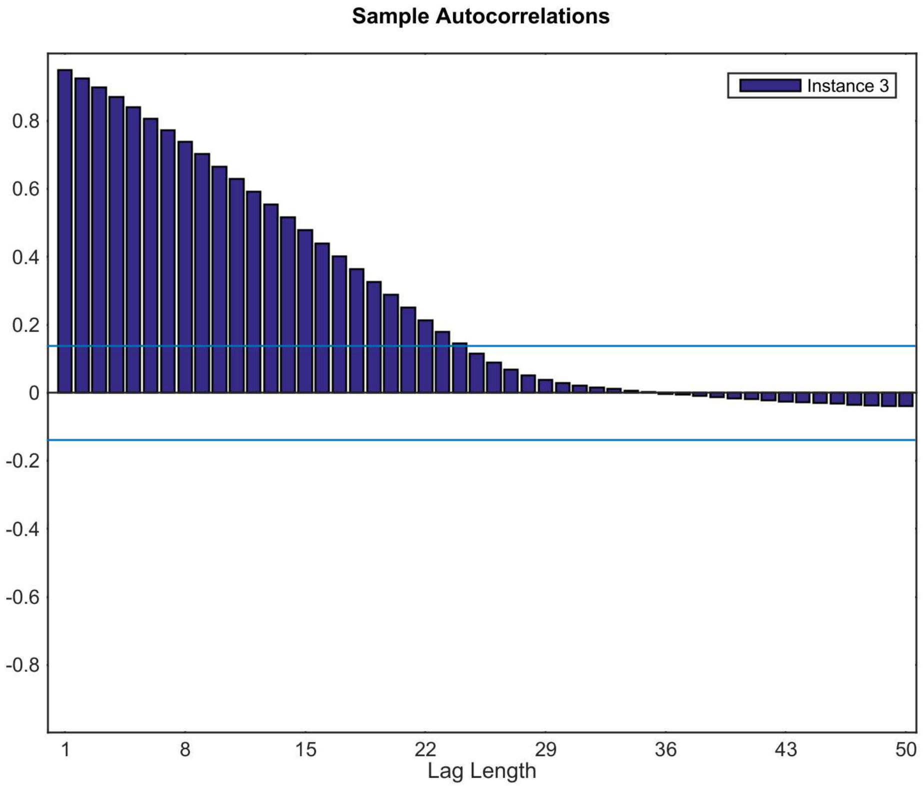
Example Time-series – Instance 1



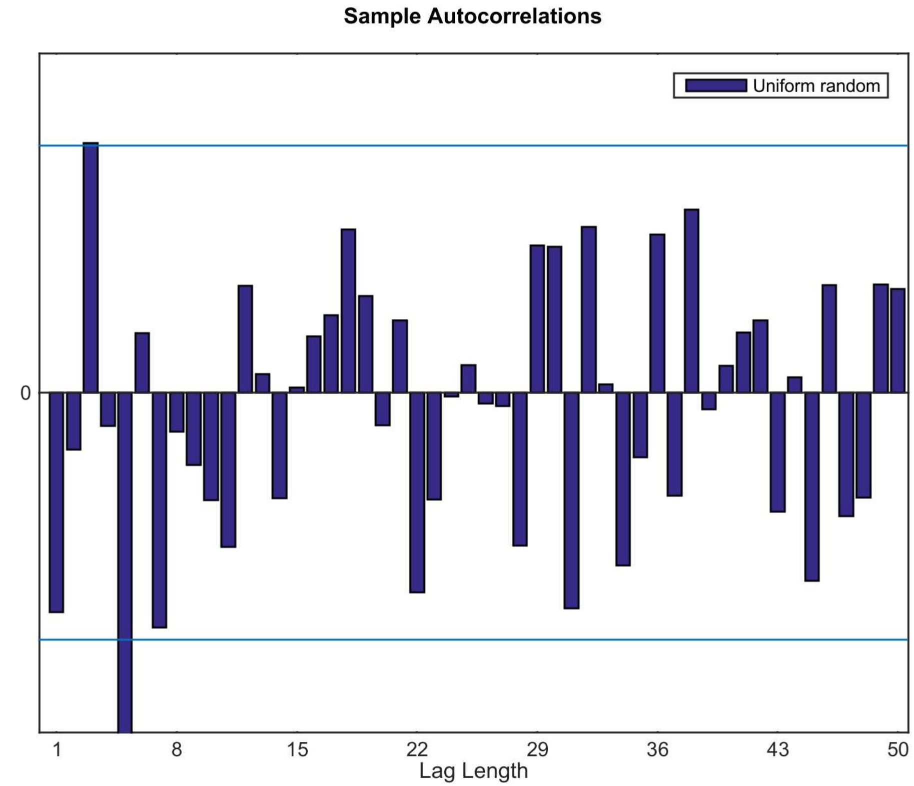
Example Time-series – Instance 2



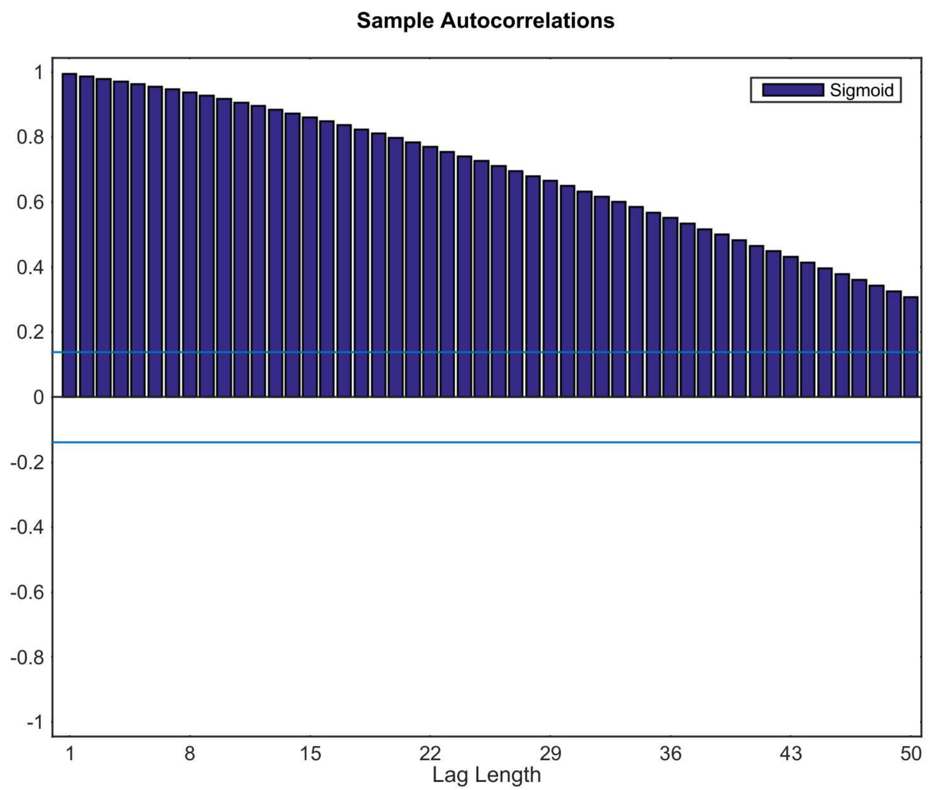
Example Time-series – Instance 3



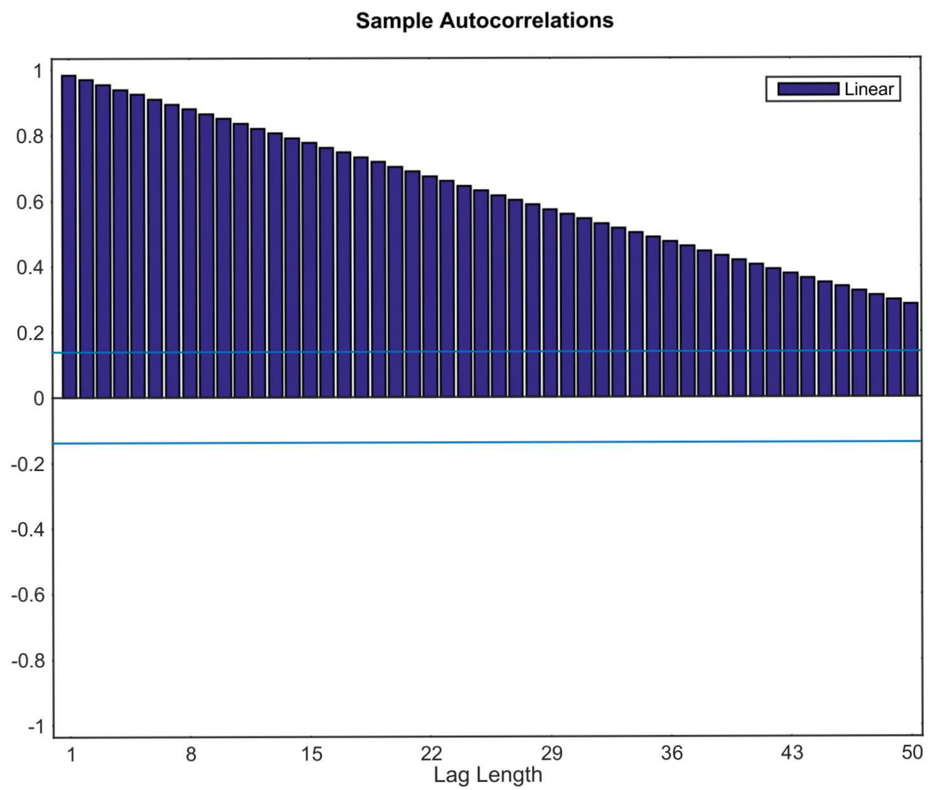
Example Time-series – Uniform random



Example Time-series – Sigmoid



Example Time-series – Linear

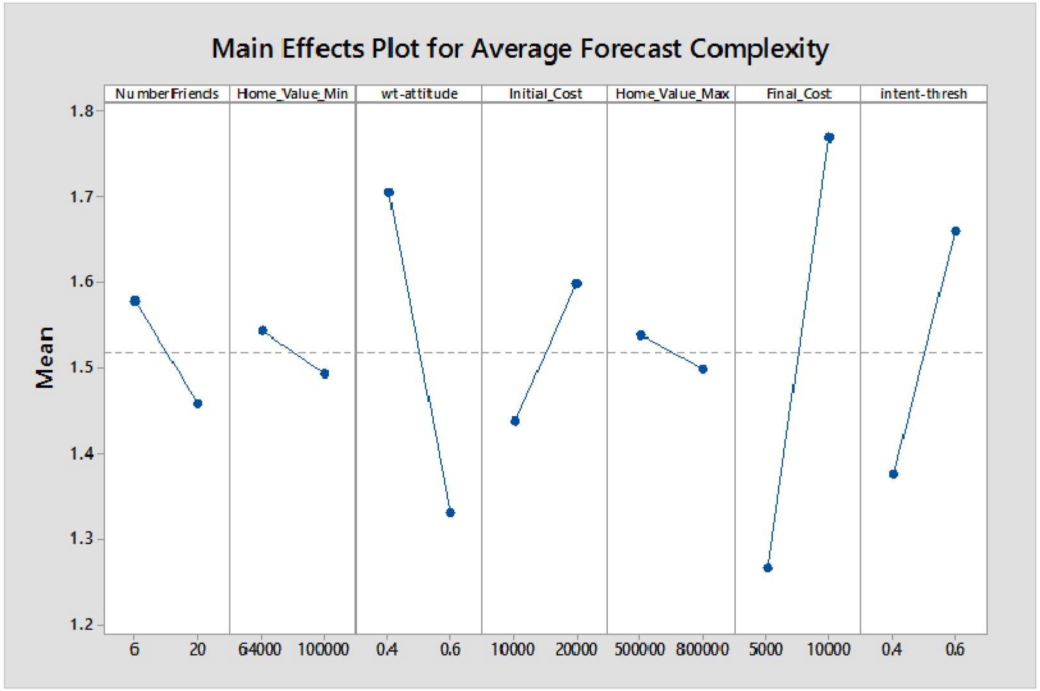


Measuring Flexibility

- ***Definition:*** Ability to manage and manipulate simulation complexity
- ***Significance:*** Complexity should be tunable
- ***Considerations and Potential Metrics***
  - Using complexity metrics
  - Flexibility metrics:
    - Percentage of simulation parameters that significantly increase the complexity of the simulation
      - Significance computed through p-value testing, using a t-test, on multiple runs of the simulation, each with a fixed parameter set
      - Measures the number of parameters involved in complexity of model
    - Range of potential complexity
      - Calculated by varying the parameters identified as significantly increasing the complexity of the simulation.
      - Facilitates ranking based on the size of ranges of complexity.

Example

PV World Model: Flexibility



Main effects: which input parameters have significant effects on the mean responses as we vary the inputs across their respective ranges

Parameter	P-Value
NumberFriends	0.05
Home_Value_Min	0.40
wt-attitude	**0**
Initial_Cost	**0.01**
Home_Value_Max	0.52
Final_Cost	**0**
intent-thresh	**0**

α = 0.05

Percentage of parameters that significantly increase complexity of the simulation = ***57%***

Definition

Ground Truth for Simulations

- **Ground truth:** graphical representation of the causal structure driving the simulation
  - Relationships in the ground truth should be represented as simply as possible
  - Perfectly correlated simulation variables – merged into the same node

Terminology

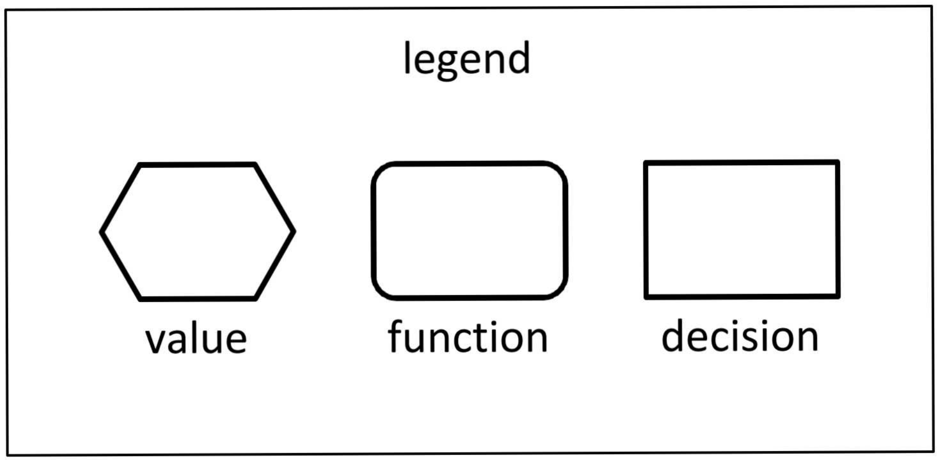
Actors: entities in the simulations

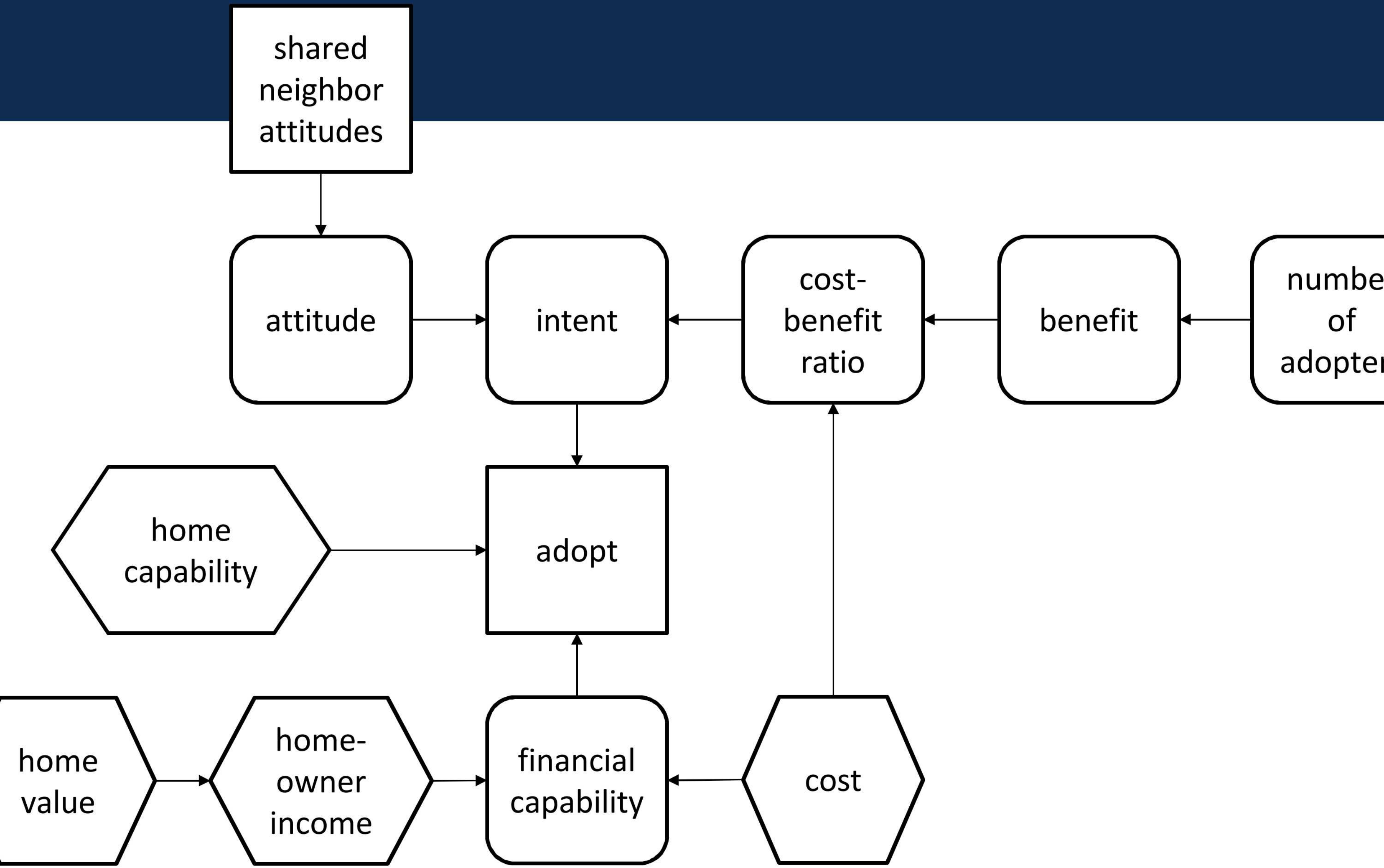
- **Characteristics:** not updated by the simulation
- **States:** updated by the simulation

Behaviors: potential actions that actors can take

Parameters: variables that can influence the simulation.

Simulation variables: data and information about the simulation


legend
value
function
decision


shared neighbor attitudes
attitude
intent
cost-benefit ratio
benefit
numbe of adopte
home capability
adopt
home value
home-owner income
financial capability
cost