Verification and Validation of Scientific and Economic Simulations

Ryan Kennedy, Xiaorong Xiang, Gregory Madey
Computer Science and Engineering
University of Notre Dame

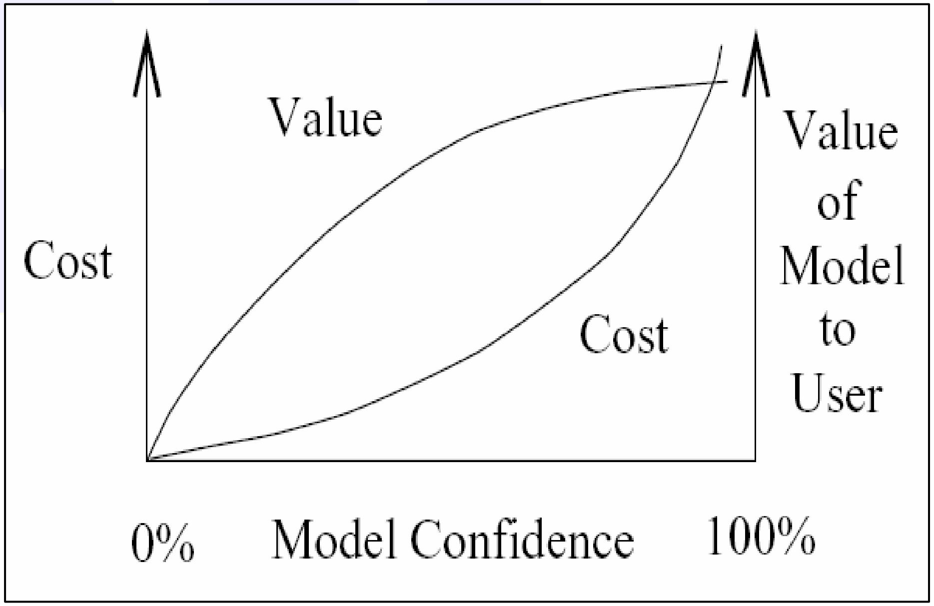
Thomas Cosimano
Department of Finance
University of Notre Dame

Overview

- Introduction
  - Concepts of Verification and Validation
  - Research Objectives and Methods
- Case Studies
  - An Agent-based Scientific Model
  - An Equation-based Economic Model
- Conclusion
- Future Work

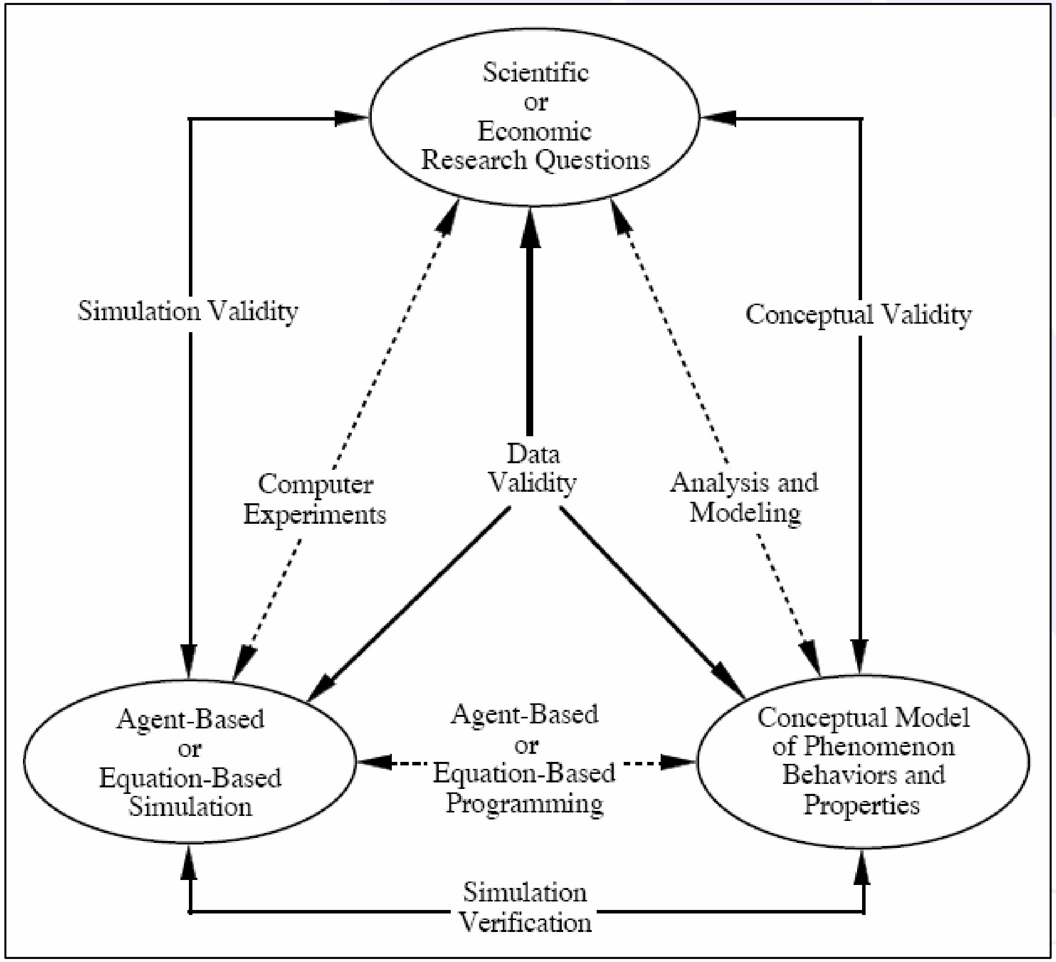
Model Verification & Validation (V & V)

- V & V
  - Verification:
    - *get model right*
  - Validation:
    - *get right model*
- The cost and value influence confidence of model
- Want to utilize V & V for optimal cost-effectiveness



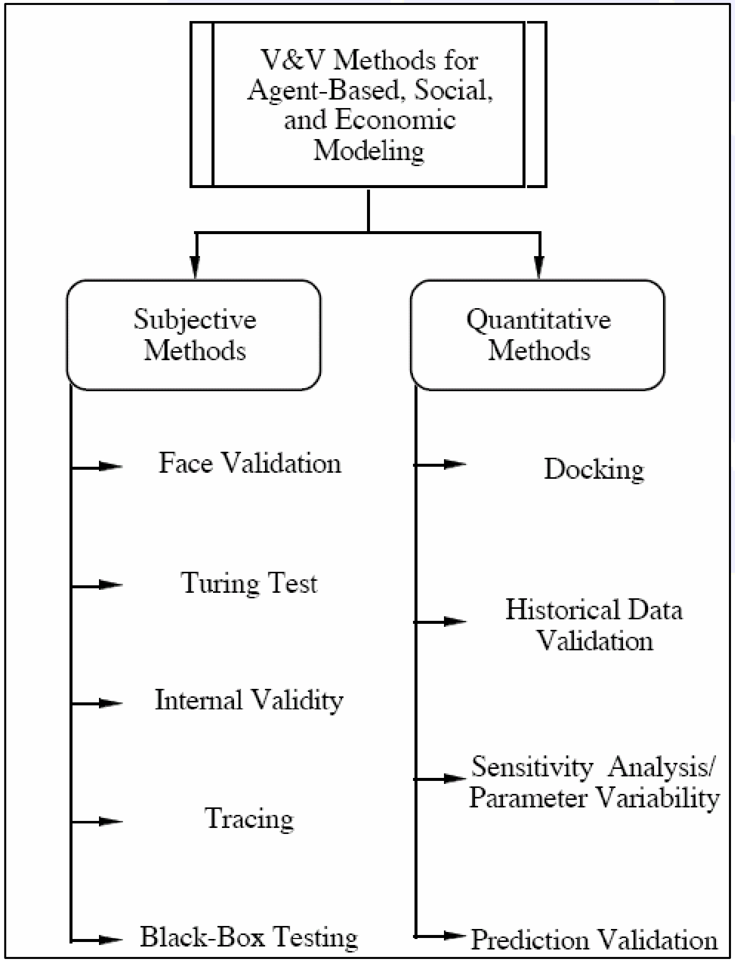
*Adapted from Sargent: "Verification and Validation of Simulation Models"

Verification and Validation Process



*Adapted from Sargent: "Verification and Validation of Simulation Models" and Huang: "Agent-Based Scientific Simulation"

Applicable Verification and Validation Methods



*Balci: "Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice" lists more than 75 Methods

V & V: Subjective Analysis

- Examples of V & V Techniques
  - Face Validity
    - Animation
    - Graphical Representation
  - Turing Test
  - Internal Validity
  - Tracing
  - Black-Box Testing

V & V: Quantitative Analysis

- Examples of V & V Techniques
  - Model-to-Model Comparison (Docking)
  - Historical Data Validation
  - Sensitivity Analysis/Parameter Variability
  - Prediction Validation

What and How

- Research objective
  - Perform V & V on distinct models and identify the more cost-effective techniques
- How
  - Two very different projects as case studies
  - Evaluate and adapt the formalized V & V techniques in industrial and system engineering

Case Study 1: An Agent-based Scientific Model

- NSF funded interdisciplinary project
  - Understanding the evolution and heterogeneous structure of Natural Organic Matter (NOM)
  - E-science example
  - Chemists, biologists, ecologists, and computer scientists
- Agent-based stochastic model
- Web-based simulation model

Case Study 1: NOM

- What is NOM?
  - Heterogeneous mixture of molecules in terrestrial and aquatic ecosystems
- Why study NOM?
  - Plays a crucial role in the evolution of soils, the transport of pollutants, and the global carbon cycle
  - Understanding NOM helps us better understand natural ecosystems

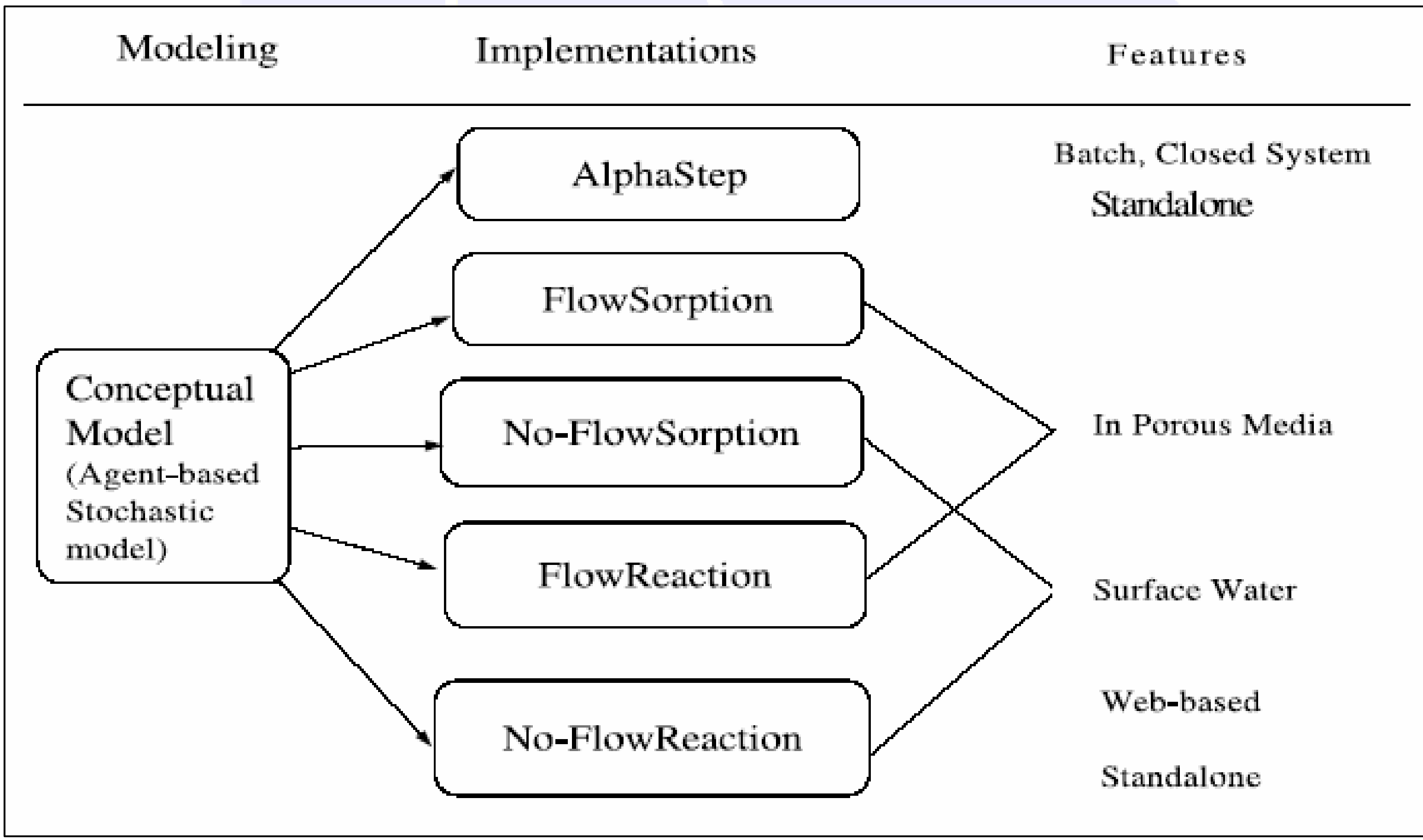
Case Study 1: The Conceptual Model I

- Agents
  - A large number of molecules
    - Heterogeneous properties
      - Elemental composition
      - Molecular weight
      - Characteristic functional groups
  - Behaviors
    - Transport through soil pores (spatial mobility)
    - Chemical reactions: first order and second order
    - Sorption

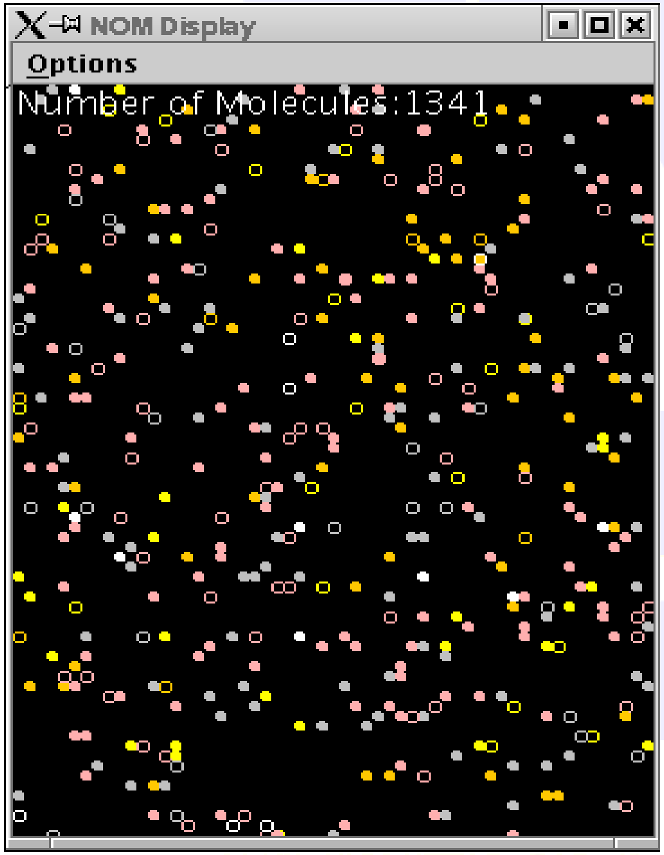
Case Study 1: The Conceptual Model II

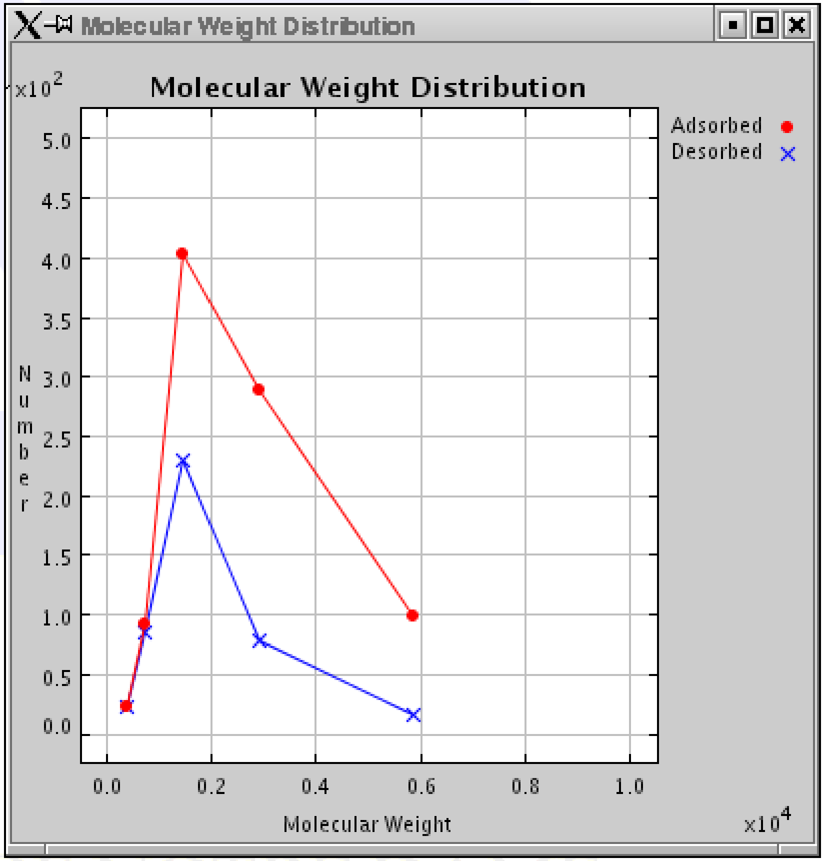
- Stochastic Model
  - Individual behaviors and interactions are stochastically determined by:
    - Internal attributes
      - Molecular structure
      - State (adsorbed, desorbed, reacted, etc.)
    - External conditions
      - Environment (pH, light intensity, etc.)
      - Proximity to other molecules
    - Length of time step, Δt
- Space
  - 2D Grid Structure
- Emergent properties
  - Distribution of molecular properties over time

Case Study 1: Implementations

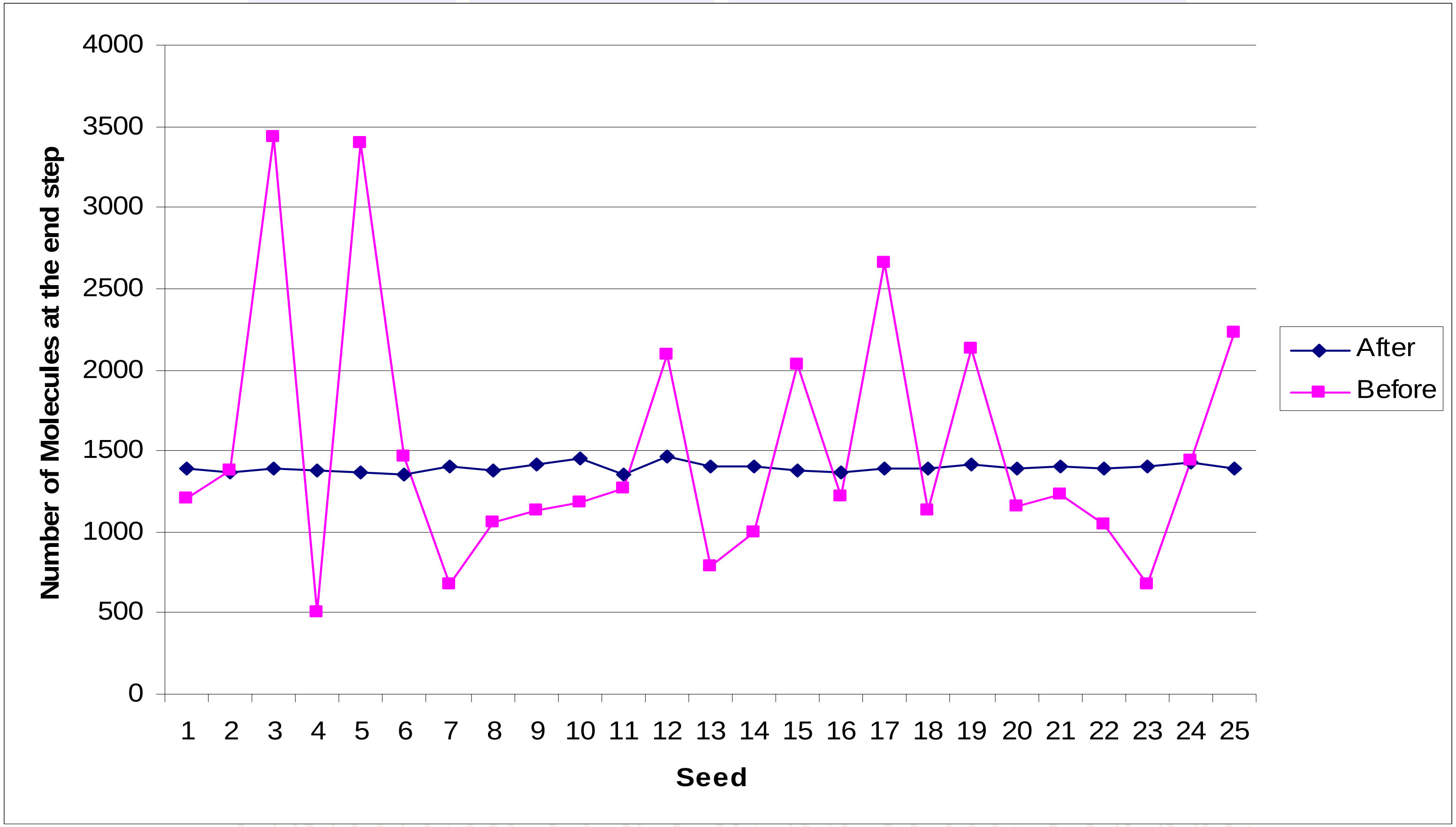


Case Study 1: Face Validity

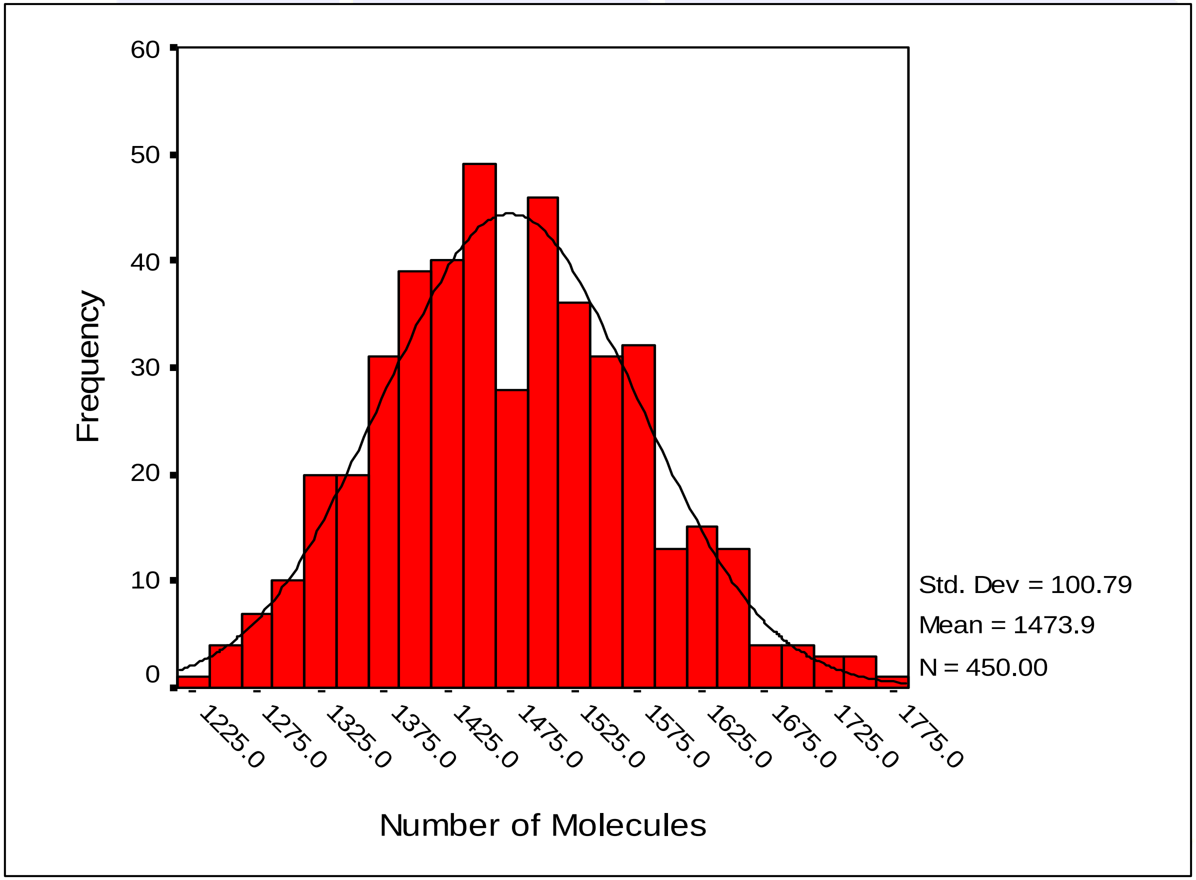




Case Study 1: Internal Validity I



Case Study 1:
Internal Validity II



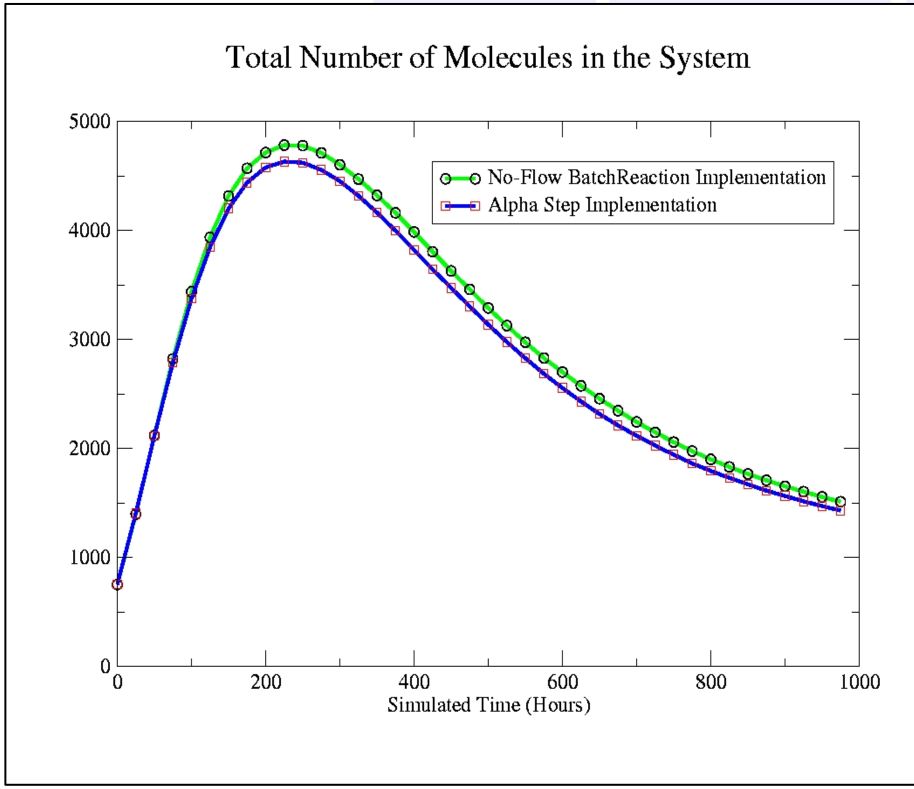
Case Study 1: Model-to-Model Comparison I

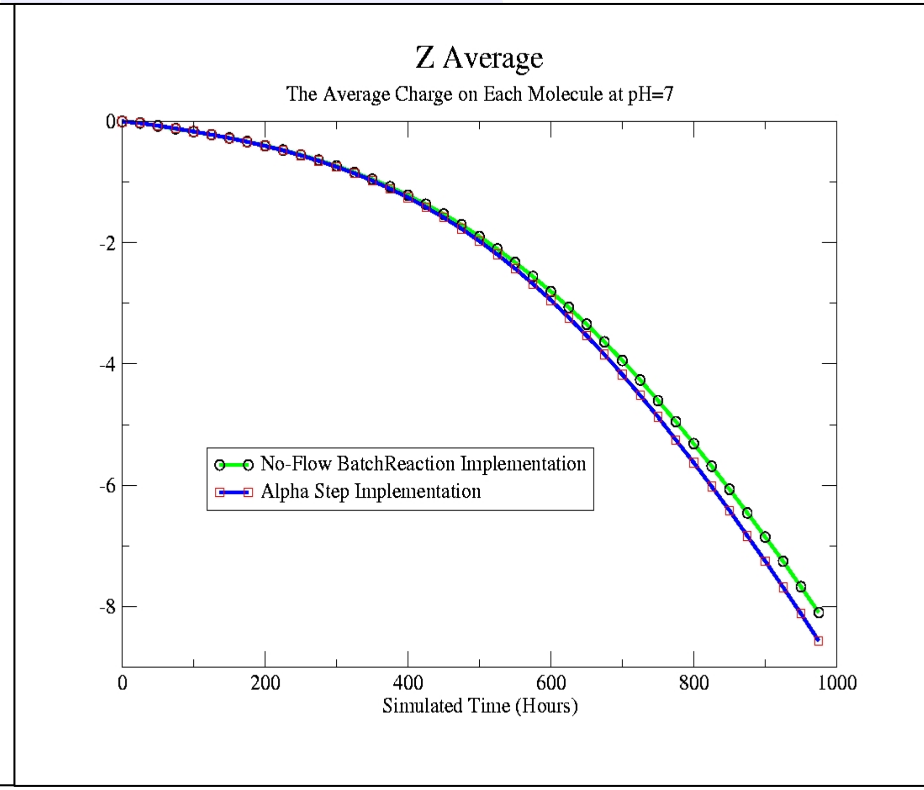
- Compare the model with validated one
- Compare the model with non-validated one
- Different implementations
  - Different programming languages
  - Different packages
- Different modeling approaches
  - Agent-based approach vs. Equation-based approach
- Powerful method for ABS

Case Study 1: Model-to-Model Comparison II

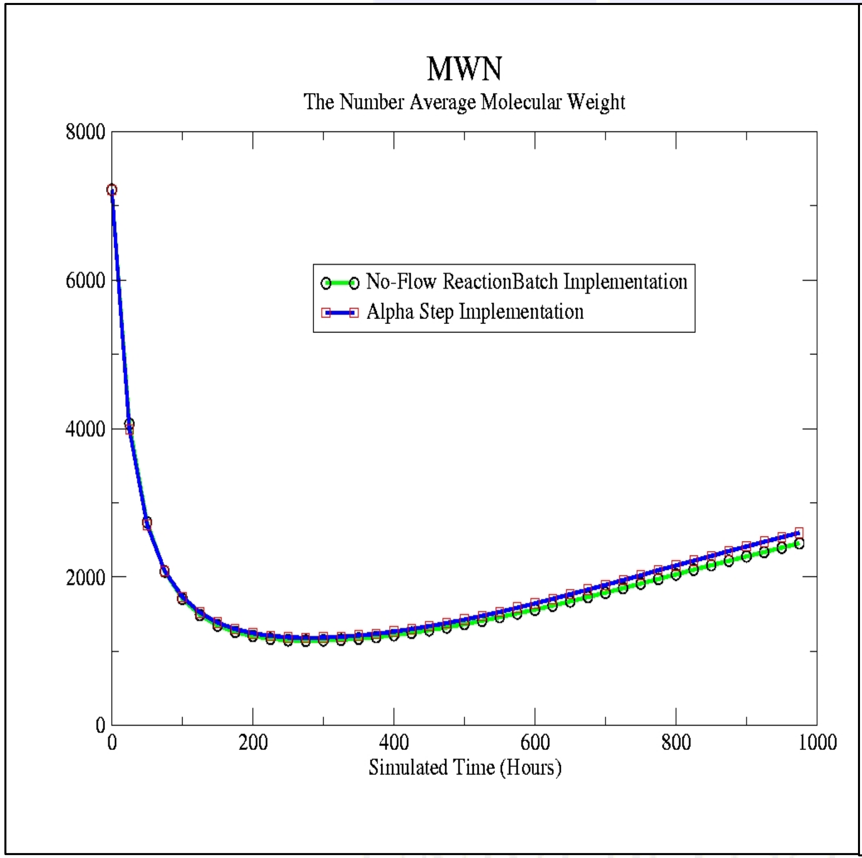
Features	Alpha Step	No-flow Reaction
Developing Group	University of New Mexico, Department of Chemistry	University of Notre Dame, Computer Science and Engineering
Programming language	Pascal	Java (Sun JDK 1.4.2)
Platforms	Delphi 6, Windows	Red hat Linux cluster
Running mode	Standalone	Web based, standalone
Simulation package	None	Swarm, Repast libraries
Animation	None	Yes
Spatial representation	None	2D grid
Second order reaction	Random pick one from list	Choose the nearest neighbor
First order with split	Add to list	Find empty cell nearby

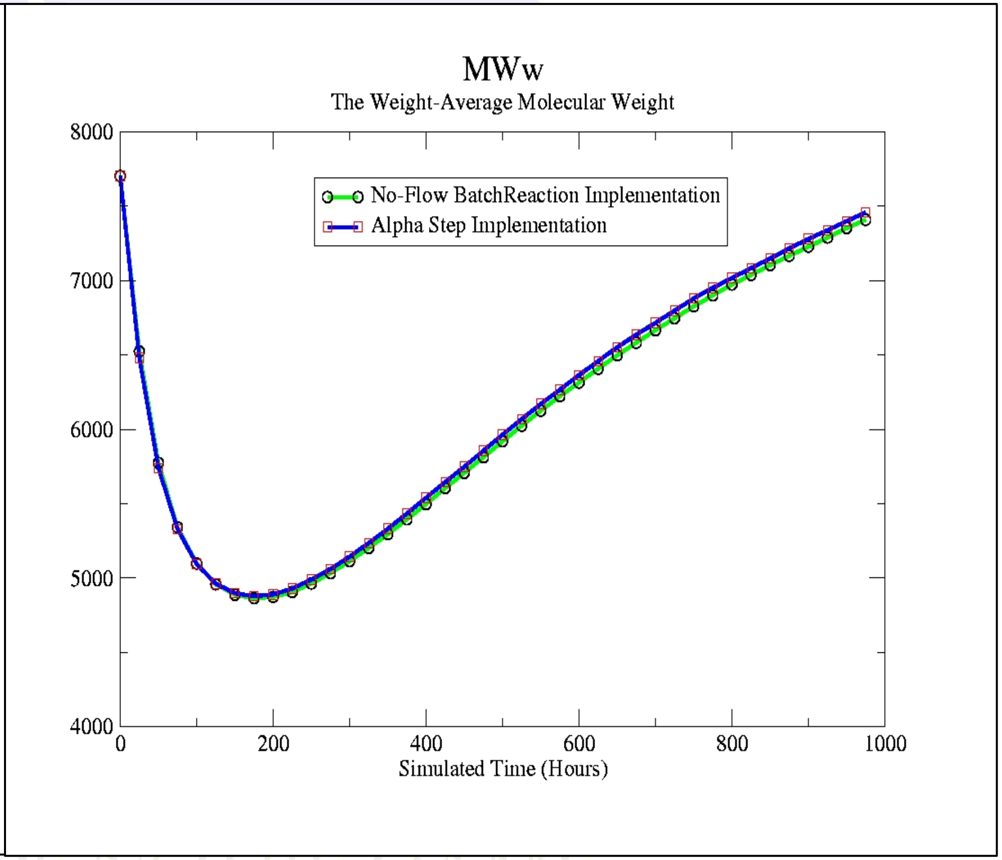
Case Study 1: Model-to-Model Comparison III



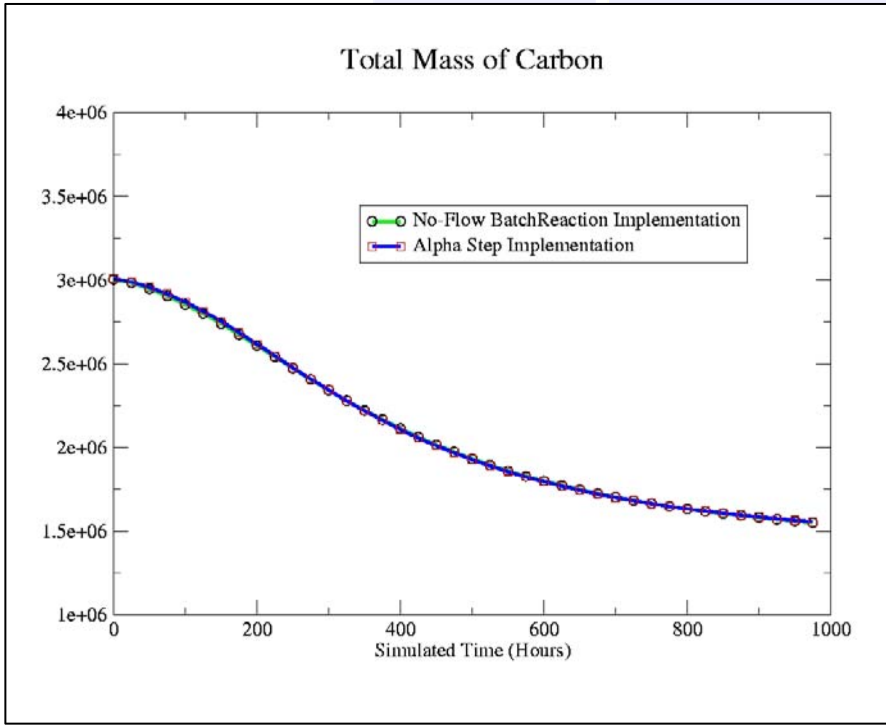


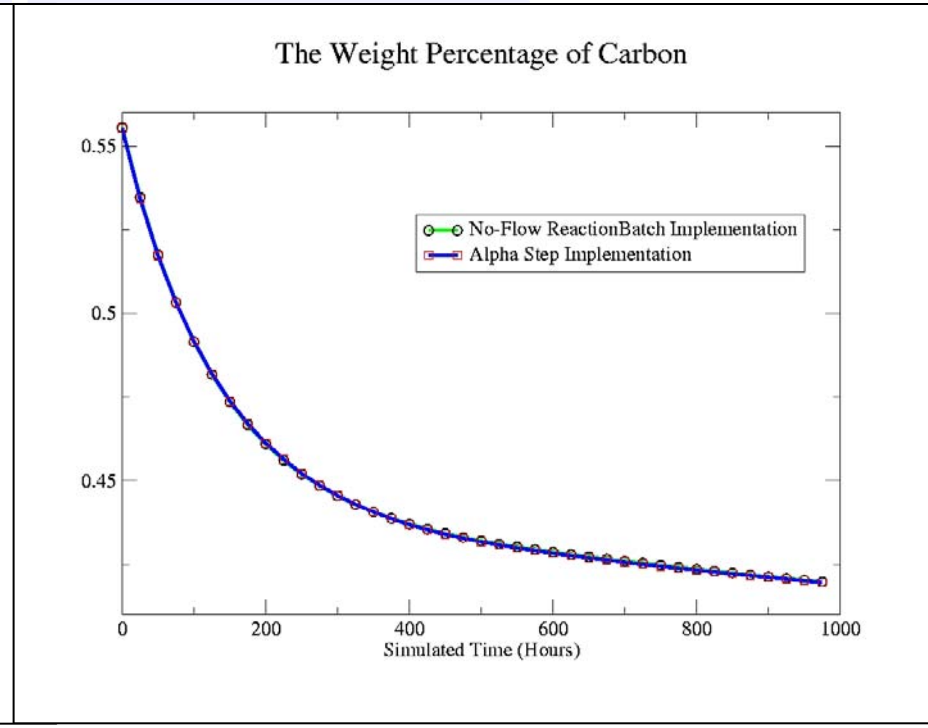
Case Study 1: Model-to-Model Comparison IV





Case Study 1: Model-to-Model Comparison V





Case Study 2: An Economic Model

- Interdisciplinary project
  - Initially written in Matlab within Department of Finance
  - Converted to C++ by Computer Scientists
  - Equation-based system
  - Concerned with identifying ideal economic variables, such as debt, money growth, and tax rate

Case Study 2: The Conceptual Model

- Equation-based system
- Nonlinear projection methods used to solve Ramsey problems in a stochastic money economy
- Goal is to generate the best social welfare for a given economy
- Motivation

$$\widehat{\mu}_{t+1}(\theta_t, g_t, b) = \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_g} b_{ij} \Psi_{ij}(\theta_t, g_t),$$

$$\widehat{\tau}_t(\theta_t, g_t, d) = \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_g} d_{ij} \Omega_{ij}(\theta_t, g_t),$$

$$\widehat{H}_t(\theta_t, g_t, q) = \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_g} q_{ij} \Phi_{ij}(\theta_t, g_t),$$

$$\widehat{\lambda}_{gt}(\theta_t, g_t, v) = \sum_{i=1}^{n_\theta} \sum_{j=1}^{n_g} v_{ij} \Gamma_{ij}(\theta_t, g_t).$$

Case Study 2: Face Verification

	LaGrange Multiplier	Labor	Money Growth	Tax Rate	Cash Good	Credit Good
Matlab	0.138	0.309	-0.009	0.188	0.486	0.621
C++	0.138	0.309	-0.009	0.188	0.486	0.621
Steady State	0.138	0.309	-0.009	0.188	0.485	0.620

Case Study 2: Tracing

- Matlab:

it 44, af 3.7496e-08, rc 0, timer 11.1, l 0.1382704496, m -0.0092286139, t 0.1881024991, h 0.3093668925
cc1 0.4861695543, cc2 0.6212795130, rl 1.0092221442
it 45, af 2.64653e-08, rc 0, timer 11.0, l 0.1382704643, m -0.0092286175, t 0.1881024947, h 0.3093668931
cc1 0.4861695553, cc2 0.6212795120, rl 1.0092221442

- C++:

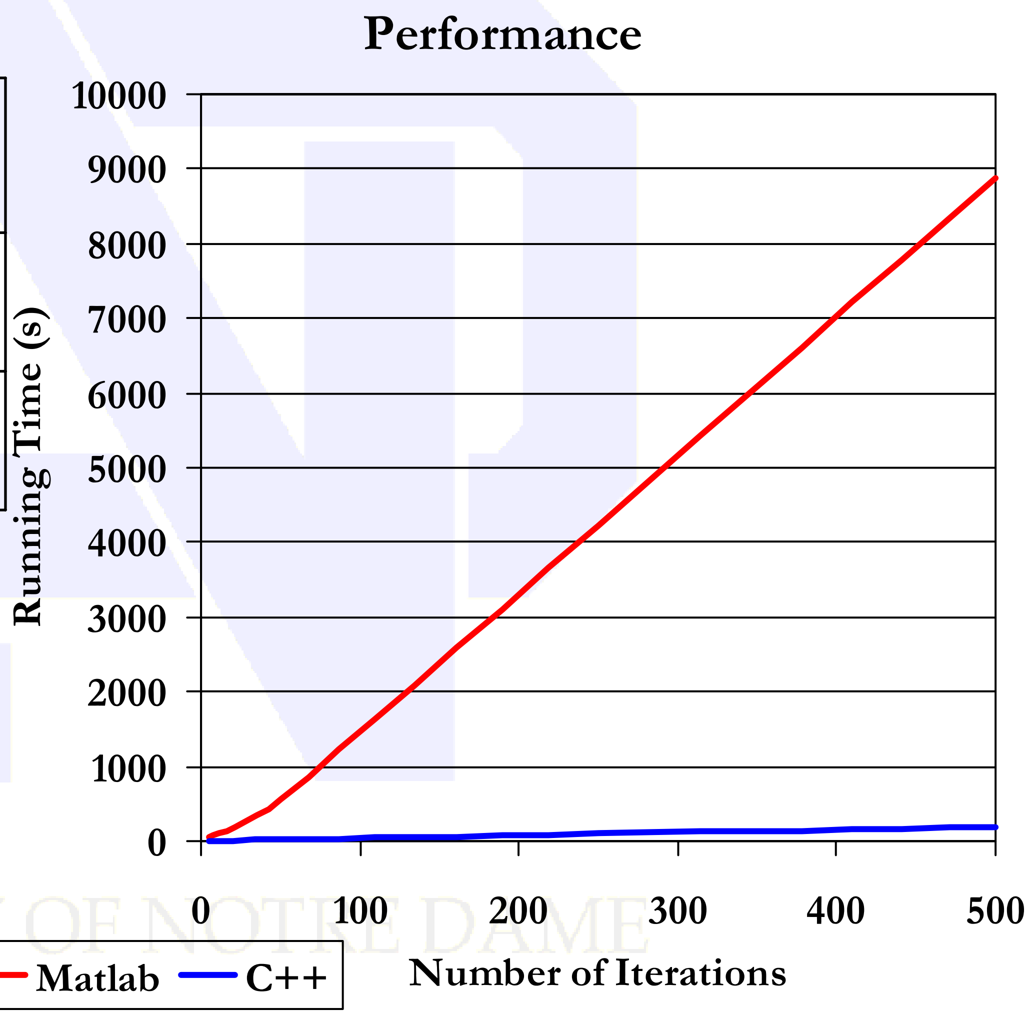
it: 44 af: 0.00144839 rc: 0 l: 0.138359 m: -0.00936025 t: 0.188252 h: 0.309338
cc1: 0.486205 cc2: 0.621244 rl: -0.65888
it: 45 af: 0.00144784 rc: 0 l: 0.138401 m: -0.00937062 t: 0.188239 h: 0.30934
cc1: 0.486208 cc2: 0.621241 rl: -0.665511

Case Study 2: Implementation Characteristics

Features	Matlab	C++
Developing Group	University of Notre Dame, Department of Finance	University of Notre Dame, Computer Science and Engineering
Language	High-Level	Lower-Level
Compiler	Interpreted	GNU Compiler
Good For	Prototyping	Speed
Platforms	Linux, Windows	Linux
Running mode	Standalone	Standalone
Packages	LAPACK, etc…	STL
Variables	Implicit	Declared

Case Study 2: Performance

	5 Iterations	50 Iterations	500 Iterations
Matlab	58 s	568 s	8872 s
C++	2 s	17 s	176 s



Summary & Conclusion

- Applied V & V to distinct case studies to increase model confidence
- Some techniques are more cost-effective

For our models:

	Agent-based	Equation-based
Face Validation/Verification	Very Good	Very Good
Turing Test	Very Good	Good
Internal Validity	Very Good	n/a
Tracing	Fair	Excellent
Black-Box Testing	Good	Good
Model-to-Model Comparison	Very Good	Very Good
Historical Data Verification	Very Good	Very Good
Sensitivity Analysis	Good	Good
Prediction Validation	Good	Fair

Future Work

- Collect and evaluate more statistical data
- Compare simulation results against empirical data
- More stringent and formalized V & V
- Perform more statistical tests

Questions or Comments?

UNIVERSITY OF NOTRE DAME