Introduction to Scientific Modeling

Stephanie Forrest
Dept. of Computer Science
Univ. of New Mexico
Albuquerque, NM

August, 2011
http://cs.unm.edu/~forrest
forrest@cs.unm.edu
Introduction

• The three legs of science
  – Experiment
  – Mathematical theory
  – Simulation and modeling
• Tools for analyzing data
• Methods for discovering new knowledge (the 3\textsuperscript{rd} leg)
• Understanding nature as an information-processing system
What is a model?

• A hypothetical description, often based on an analogy, used in analyzing something.
• A representation of something in which we are interested.
• The representation is reduced---what to throw away?
• Or, a model can explain how something works (mechanistic).
• Examples?
Giovanni Alfonso Borelli (1608-1679)

- Father of biomechanics and an early modeler
- Related animals to machines and used mathematics to prove his theories
- Likened the action of the heart to that of a piston and reasoned that arteries must be elastic
Models are expressed in different ways

- Verbal descriptions, e.g.,
  - The “invisible hand” in economics
  - Clonal selection theory in immunology
  - Bohr’s model of the atom
- Pictures
- Mathematical equations
- Computer programs (model vs. simulation)
Examples?
Examples

- A genetic algorithm is a model of Darwinian evolution
- Stock market simulation
- Crash test dummies (model humans)
- Social network models of social relationships
- The carbon cycle (a compartment model: air, ocean, terrestrial ecology, fossil fuels)
Example: Limits to growth model

Meadows et al.

• 1972 computer model:
  – 5 variables: population, industrialization, pollution, food production, resource depletion

• Goal: Not to make specific predictions, but to explore how exponential growth interacts with finite resources. Because the size of resources is not known, only general behavior can be explored.

• This process of determining behavior modes is "prediction" only in the most limited sense of the word. ... These graphs are not exact predictions of the values of the variables at any particular year in the future. They are indications of the system's behavioral tendencies only.
How do we use models?

• Making predictions (conventional use of models):
  – Quantitative (most analytical models). Often not mechanistic. Examples?
  – Qualitative. E.g., critical parameters, regions of stability or instability. Finding boundary conditions. Examples?
  – Validation: Accurate predictions.

• Existence proofs:
  – Demonstrating that something is possible
  – E.g., von Neumann’s self-reproducing automaton.
  – Related to Gedanken experiments, e.g., Shrodinger’s Cat, Maxwell’s Demon
  – Validation: It works.

• As metaphors:
  – Stimulate thinking, suggest ideas.
  – E.g., neural computation (thinking as a kind of electrical circuit). Validation?
How do we use models? Cont.

• Building intuitions about how complex systems work (exploratory):
  – Examples: Flight simulators, Sim City.
  – Study how patterns of behavior, how resources flow, how co-operation arises, arms races.
  – Sensitivity analysis.
  – Discovering lever points for intervention, e.g., vaccines
  – Validation? (Trust, allows a pilot to land safely).

“The contemplation in natural science of a wider domain than the actual leads to a far better understanding of the actual. “ A.S. Eddington. (first test of Einstein's Theory of Relativity experimentally)
Not everyone agrees

• Def. “Modeling is the application of methods to analyze complex, real-world problems in order to make predictions about what might happen with various actions.” [Shiflet and Shiflet, 2006]

• “Models are metaphors that explain the world we don’t understand in terms of worlds we do. They are merely analogies, provide partial insight, stand on someone else’s feet. Theories stand on their own feet, and rely on no analogies.” [Emanuel Derman, 2012]
How do we evaluate a model?

• Parsimony and simplicity
  – Occam’s Razor (select the competing hypothesis that makes the fewest new assumptions, when the hypotheses are equal in other respects)

• Accuracy of predictions
  – $R^2$ and other statistical tests

• Dynamical model works as claimed:
  – Run it. E.g., patented devices.

• Cogency and relevance of ideas that they produce.

• Falsifiability.

• Consistency---formalize notion of model as a *homomorphic map*. 
Common Modeling Assumptions

• Homogeneity (all agents are identical / stateless)
• Equilibrium (no or very simple dynamics)
• Random mixing
• No feedback (learning)
• Deterministic
• No connection between micro and macro phenomena

• Models with these assumptions can produce some interesting features, e.g., tipping points ($R_0$).
Features of Complex Systems

- Heterogeneous agents
- Non-equilibrium (non-linear dynamics)
- Contact structure (networks, nonrandom mixing)
- Learning / Feedback (agents can change behavior)
- Stochastic behavior (interesting behavior in the tails)
- Emergence (multi-scale phenomena)
Modeling Complex Systems is Difficult

• Closed form solutions rarely exist:
  – Features from previous slide

• Detailed simulations are problematic:
  – Can never hope to get all the details correct.
  – Because systems are nonlinear, small errors can have large consequences.

• Evolution is key:
  – Basic components change over time.
  – Individual variants matter (hard to do theory).
Modeling Complex Systems is Difficult cont.

- Discreteness (e.g., time, state spaces, and internal variable values).
  - Techniques developed to study nonlinear systems are not always directly applicable.
- Spatial heterogeneity.
- Classical ODE (ordinary differential equation) assumptions:
  - Well stirred” (each particle-particle interaction is equally likely).
  - Infinite-sized populations.
  - Spatial homogeneity.
Classes of Scientific Models

- Continuous vs. Discrete
  - E.g., Differential equation vs. Cellular automaton
- Deterministic vs. Probabilistic
  - Dynamical system vs. Markov chain
  - Cellular automaton vs. genetic algorithm
- Spatial vs. nonspatial
- Data-driven vs. theory-driven
  - Bayesian networks vs. expert system
Aggregate Models
Differential Equations

• Represent how a process changes through time as a differential (difference) equation:
  – Time is continuous (discrete)
  – Model components are continuous (density)
  – Deterministic
  – Nonspatial (in simplest case)
• Describes the global behavior of a system
• Averages out individual differences (stateless)
• Assumes infinite-sized populations of model components
  – E.g., assume all possible genotypes always present in population.
• Easier to do theory and make quantitative predictions.
• Examples:
  – Maxwell’s equations
  – Mackey-Glass systems
  – Lotka-Volterra systems

\[
\frac{dx}{dt} = x(\alpha - \beta y) \\
\frac{dy}{dt} = -y(\gamma - \delta x)
\]
Agent-based Models (ABM)
Computational / Individual-based / Particle

• A computational artifact that captures essential components and interactions (i.e. a computer program).

• Encodes a theory about relevant mechanisms:
  – Want relevant behavior to arise spontaneously as a consequence of the mechanisms. The mechanisms give rise to macro-properties without being built in from the beginning.
  – This is a very different kind of explanation than simply predicting what will happen next.
  – Example: Cooperation emerges from Iterated Prisoner’s Dilemma model.
  – Simulation as a basic tool.
  – Observe distribution of outcomes.

• Study the behavior of the artifact, using theory and simulation:
  – To understand its intrinsic properties, and wrt modeled system.
Examples

• Cellular automata
• Genetic algorithms
• Digital immune systems
• Sugarscape
• Prisoner’s Dilemma Tournament
Agent-based models have limitations

- Correspondence problem:
  - What does each primitive component in the model correspond to in the real system?
- What questions can they answer? (qualitative predictions, critical regions of parameter space)
- How to interpret results?
  - Can’t look at a single run.
  - Many contingent behaviors, macro-statistics don’t tell the entire story.
- Scaling issues (e.g., time, error rates, population sizes).
- The mechanistic theory encoded by the model cannot always be stated cleanly.
Agent-based Models: Advantages
(taken from Axelrod 2005)

• Can address problems that are fundamental to many disciplines:
  – Path dependency
  – Effects of adaptive versus rational behavior
  – Effects of network structure
  – Cooperation among egoists
  – Diffusion of innovation
  – Tradeoff between exploitation and exploration
  – Generalism vs. specialization

• Facilitate interdisciplinary collaboration:
  – “A prosthesis for interaction”

• Useful tool when closed form mathematical analyses are intractable.
  – E.g., the evolution of sex

• Can reveal unity across disciplines.

• Can be a “hard sell”:
  – Realism vs. clarity