CS626 Data Analysis and Simulation

Instructor: Peter Kemper

R 104A, phone 221-3462, email:kemper@cs.wm.edu

Today:

Verification and Validation of Simulation Models

Reference: Law/Kelton, Simulation Modeling and Analysis, Ch 5.

Sargent, Verification and Validation of Simulation Models, Tutorial, 2010 Winter Simulation conference.

Verification and Validation

Validation:

"substantiation that a computerized model within its domain of applicability possesses a satisfactory range of accuracy consistent with the intended application of the model"

Verification:

"ensuring that the computer program of the computerized model and its implementation are correct"

Sargent's WSC Tutorial 2010, cites Schlesinger 79

"Verify (debug) the computer program." Law's WSC 09 Tutorial

Accreditation:

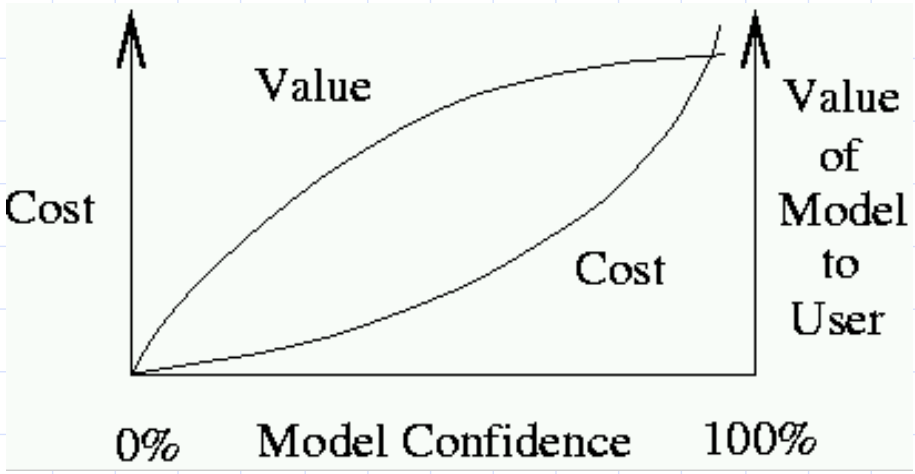
DoD: "official certification that a model, simulation, or federation of models and simulations and its associated data are acceptable for use for a specific application."

Credibility:

"developing in users the confidence they require in order to use a model and in the information derived from that model."

Confidence

- Validity with respect to objective
- Often too costly & time consuming to determine that a model is absolutely valid over complete domain of intended applicability.



Remarks from Law/Kelton

- Model valid => simulation results support decision making similar to experiments with real system
- Complexity of validation process depends on complexity of system and whether a version of the system exists.
- Simulation model is always an approximation, there is no absolute validity.
- Model should be developed for a particular purpose.

Remarks from Law/Kelton

- Performance measures for validation must include those used for decision making.
- Validation should take place during model development.
- Credibility:
  - Needs user understanding & agreement with model's assumptions
  - Needs demonstration that model has been validated & verified.
  - Needs user's ownership of & involvement with the project
  - Profits from reputation of the model developers

How to find an appropriate level of detail?

- A model is an abstraction / a simplification of a real system.
  - Selection of details depends on objective / purpose of a model!
  - Model is not supposed to represent complete knowledge of a topic.
  - Include only details that are necessary to match relevant behavior ... or to support credibility.
- Examples for possible simplifications:
  - Entities in system need not match 1-1 with entities in a model.
  - Unique entities in system may need not be unique in a model.
- Level of detail
  - Level of detail should be consistent with the type of data available.
  - Time and budget constraints influence a possible level of detail.
  - Subject matter experts (SME) and sensitivity analysis can give guidance and what impacts measures of interest!
  - If the number of factor is large, use a "coarse" model to identify most relevant ones.

Output Analysis vs Validation (Law/Kelton)

- Difference between validation and output analysis: Measure of interest

- System: mean μ_S
- Model: mean μ_M
- Estimate from simulating the model: μ_E

Error in μ_E is difference $|\mu_E - \mu_S| = |\mu_E - \mu_M + \mu_M - \mu_S|$

$\leq |\mu_E - \mu_M| + |\mu_M - \mu_S|$

(triangle inequality)

1st part: Focus in output analysis

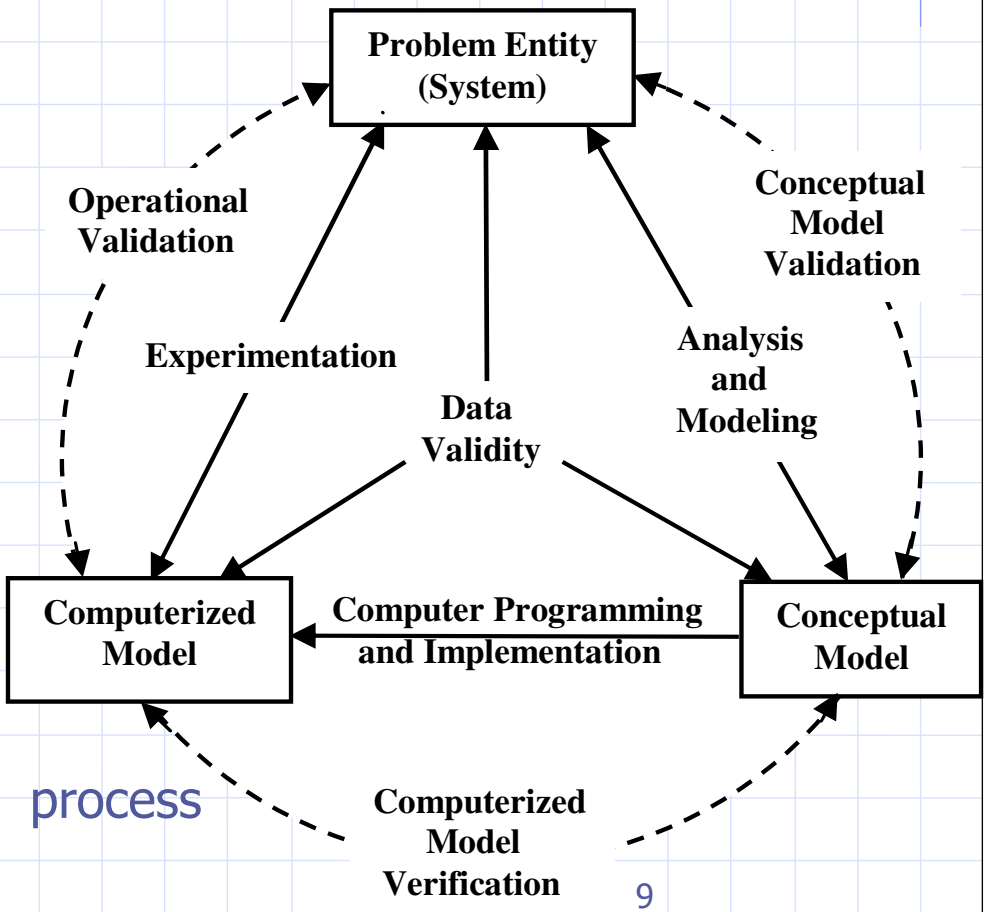
2nd part: Focus in validation

Who shall decide if model is valid?

- Four basic approaches (Sargent)
  - The model development team
    - subjective decision based on tests and evaluations during development
  - Users of the model with members of development team
    - Focus moves to users, also aids in model credibility
    - in particular if development team is small
  - 3rd party
    - independent of both, developers and users
    - in particular for large scale models that involve several teams
    - 3rd party needs thorough understanding of purpose
    - 2 variants
      - concurrently: can be perform complete VV effort
      - after development: may focus on VV work done by development team
  - Scoring model
    - rarely used in practice
    - subjective assignment of scores/weights to categories
    - model is valid if overall & category scores greater than threshold

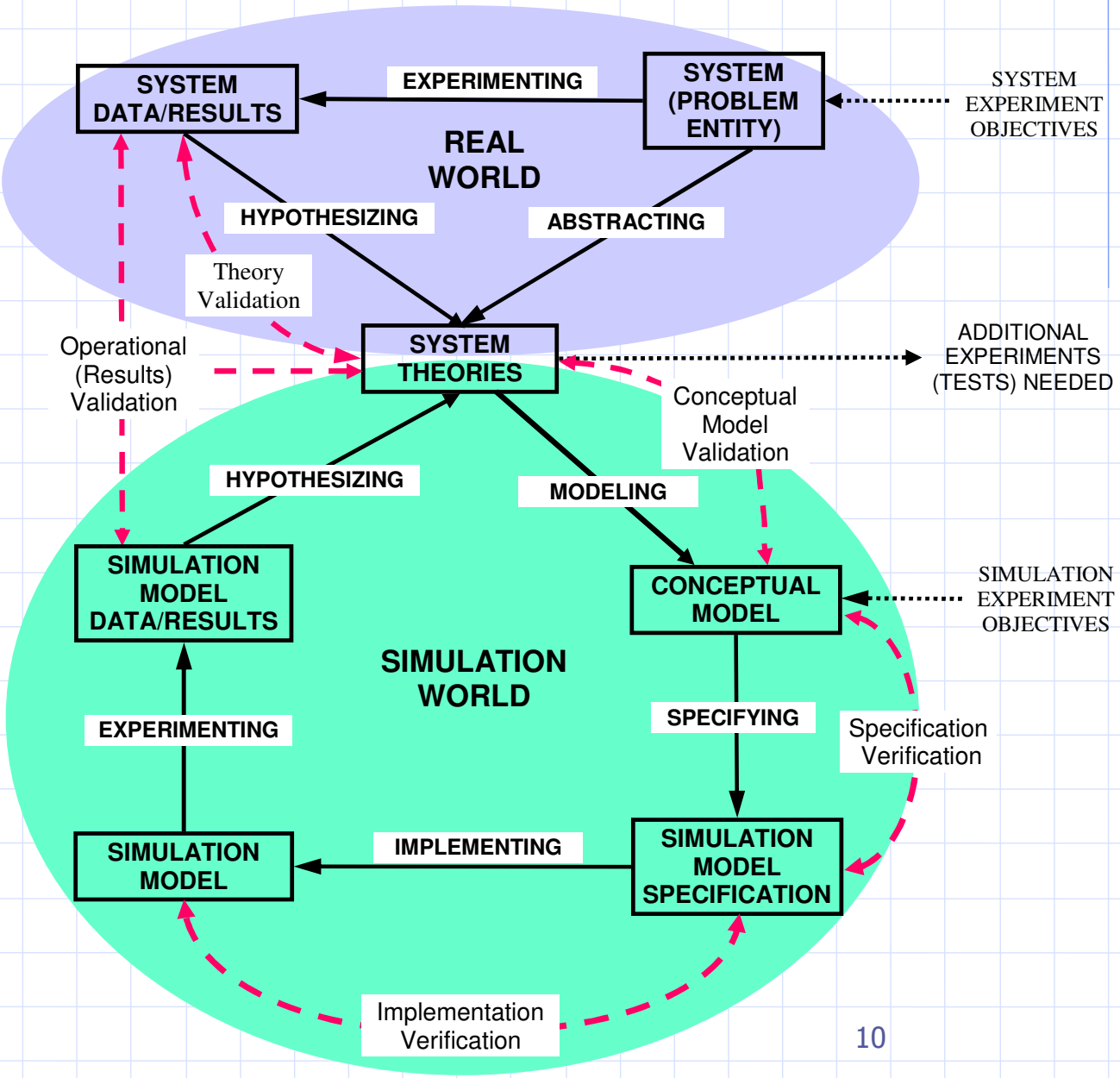
Variant 1: Simplified Version of Modeling Approach

- Conceptual model validation
  - determine that theories & assumptions are correct
  - model representation "reasonable" for intended purpose
- Computerized model verification
  - assure correct implementation
- Operational validation
  - model's output behavior has sufficient accuracy
- Data validity
  - ensure that the data necessary for model building, evaluation, testing, and experimenting are adequate & correct.
- Iterative process
  - also reflects underlying learning process



Variant 2: Real World and Simulation World

- Stresses objectives
- More detailed but conceptually similar
- Iterative process



Aside:

From Stephen Hawking

"Any physical theory is always provisional, in the sense that it is only a hypothesis: you can never prove it.

No matter how many times the results of experiments agree with some theory, you can never be sure that the next time the result will not contradict the theory.

On the other hand you can disprove a theory by finding a single observation that disagrees with the predictions of the theory.

As philosopher of science Karl Popper has emphasized, a good theory is characterized by the fact that it makes a number of predictions that could in principle be disproved or falsified by observation.

Each time new experiments are observed to agree with the predictions the theory survives, and our confidence in it is increased; but if ever a new observation is found to disagree, we have to abandon or modify the theory."

from S. Hawking, A brief history of time, the universe in a nutshell.

Validation Techniques

- Does it match with own/SME's expectations?
  - Animation
  - Operational Graphics observe performance measures during simulation run
  - Face validity
  - Turing test
- Does it match with existing knowledge?
  - Comparison to other models
  - Historical data validation
  - Predictive Validation to compare model predictions with field tests/ system data
  - Degenerate Tests

Validation Techniques

- Sanity checks
  - Degenerate Tests
  - Event validity (relative to real events)
  - Extreme condition tests
  - Traces to follow individual entities
- Historical methods
  - Rationalism (assumptions true/false)
  - Empiricism
  - Positive economics (predicts future correctly)
- Variability
  - Internal validity to determine amount of internal variability with several replication runs
  - Parameter Variability - Sensitivity Analysis

Data validity

- Data needed for
  - building a conceptual model
  - validating a model
  - performing experiments with a validated model
- Valid data necessary for overall approach
  - GIGO: Garbage in - Garbage out
- Sargent: "Unfortunately, there is not much that can be done to ensure that the data are correct."
  - in a scientific sense, i.e., there is no particular procedure to follow other than to carefully
    - collect and maintain data
    - test collected data using techniques such as internal consistency checks
    - screening for outliers and determining if outliers are correct or not

Conceptual model validation

- Conceptual model validation: determining that
  - the theories and assumptions underlying the conceptual model are correct
  - the model's representation of the problem and the models structure, login and mathematical and causal relationships are "reasonable for the intended purpose of the model.
- How to do this?
  - Testing using mathematical analysis and statistical methods, e.g.
    Assumptions: linearity, independence of data, Poisson arrivals
    Methods: fitting to data, MLE parameter estimation, graphical analysis
  - Evaluating individual components and the way those are composed into an overall model by e.g.
    - face validity: experts examine flowchart, graphical model, set of equations
    - traces: tracking entities in each submodel and overall model.

Computerized model verification

- Special case of verification in software engineering
- If a simulation framework is used
  - evaluate if framework works correctly
  - test random number generation
  - model-specific
    - existing functionality/libraries are used correctly
    - conceptual model is completely and correctly encoded in modeling notation of the employed framework
- Means
  - structured walk through
  - traces
  - testing, i.e., simulation is executed and dynamic behavior is checked against a given set of criteria,
    - internal consistency checks (assertions)
    - input-output relationships
    - recalculate estimates for mean and variance of input probability distributions

More concrete:

- Mobius: Compositional model description
- Evaluate individual atomic models in isolation
  - analog: unit testing
  - assign minimal values to state variables and see if dynamic behavior is as expected by fine grained measurements and trace data
  - evaluate qualitative behavior before taking quantitative behavior (specific distributions, performance measurements)
  - evaluate simple special cases that allow for redundant types of analysis (numerical solution of Markov chains vs simulation)
  - compose large models from well-understood atomic models of limited complexity (analog: class and method design)
- Traviando:
  - Perform a trace analysis on isolated submodels / overall model
    - use report generation functionality
    - check ranges of values for state variables
    - occurrence of activities and their state transformations

Operational validity

	Observable System	Non-observable System
Subjective Approach	• Comparison Using Graphical Displays • Explore Model Behavior	• Explore Model Behavior • Comparison to Other Models
Objective Approach	• Comparison Using Statistical Tests and Procedures	• Comparison to Other Models Using Statistical Tests

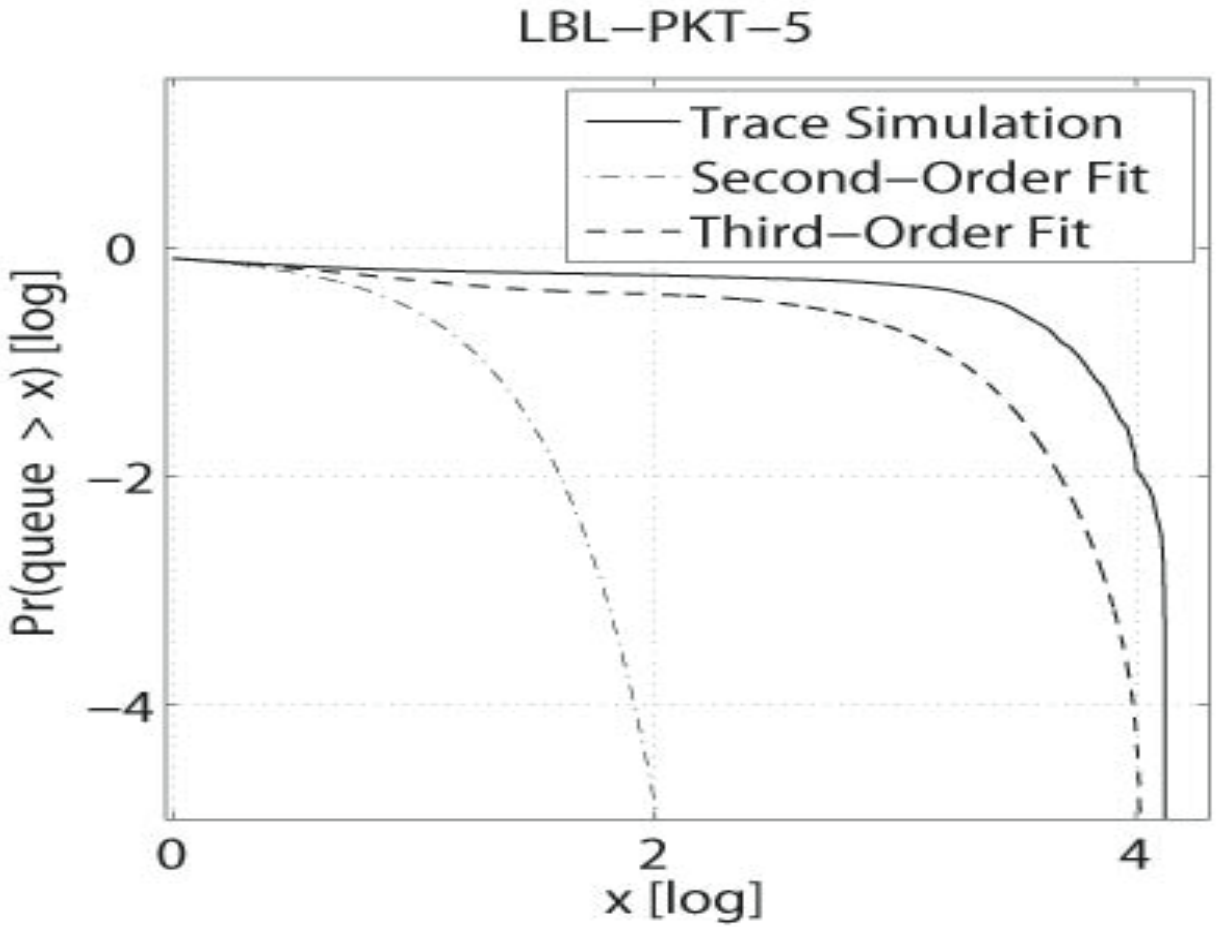
- "Determine whether the simulation model's output behavior has the accuracy required for the model's intended purpose over the domain of the model's intended applicability."
- May reveal deficits in conceptual model as well as in its implementation.

Operational Validity

- Explore Model Behavior
  - Directions of behavior
  - Reasonable / precise magnitudes
  - Parameter variability-sensitivity analysis
    - Statistical approaches: Metamodeling, design of experiments
- Comparisons of Output Behavior (System vs Model)
  - Most effective: trace driven simulation
    - feed measurement data into simulation to closely follow real behavior
  - Use graphs to make subjective decisions
    - Histograms, Box plots, Scatter plots
    - Useful in model development process to evaluate level of detail and accuracy, for face validity checks by subject matter experts, and in Turing tests
  - Use confidence intervals and/or hypothesis tests to make an "objective" decision
    - Problems: underlying assumptions (independence, normality) and/or insufficient system data

Trace driven simulation

- Idea: feed measurement data into simulation model
- Example from MAP fitting work by Casale, Smirni et al.



Documentation of VV effort

- Critical to build credibility, justify confidence
- Detailed documentation on specifics of tests etc
- Separate tables for data validity, conceptual model validity, computer model verification, operational validity

Category/Item	Technique(s) Used	Justification for Technique Used	Reference to Supporting Report	Result/ Conclusion	Confidence In Result
• Theories • Assumptions • Model representation	• Face validity • Historical • Accepted approach • Derived from empirical data • Theoretical derivation				

Low
Medium
High

Strengths
Weaknesses

Overall evaluation for Computer Model Verification	Overall Conclusion	Justification for Conclusion	Confidence In Conclusion

Recommended Procedure

- Agreement between user & developer on VV approach (prior to model development)
- Specification of required amount of accuracy
- Test assumptions & theories underlying simulation model
- In each iteration
  - perform face validity on the conceptual model
  - explore simulation model's behavior using the computerized model
  - make comparisons (if possible) between simulation model and system behavior for at least a few sets of experimental conditions (at least in the last iteration)
- Develop validation documentation for inclusion in the simulation model documentation
- If model is to be used of a period of time, develop a schedule for periodic review of the model's validity.