STEVENS
INSTITUTE of TECHNOLOGY
THE INNOVATION UNIVERSITY®

Verification, Validation, Accreditation

SYS-611: Simulation and Modeling

Paul T. Grogan, Ph.D.
Assistant Professor
School of Systems and Enterprises

Agenda

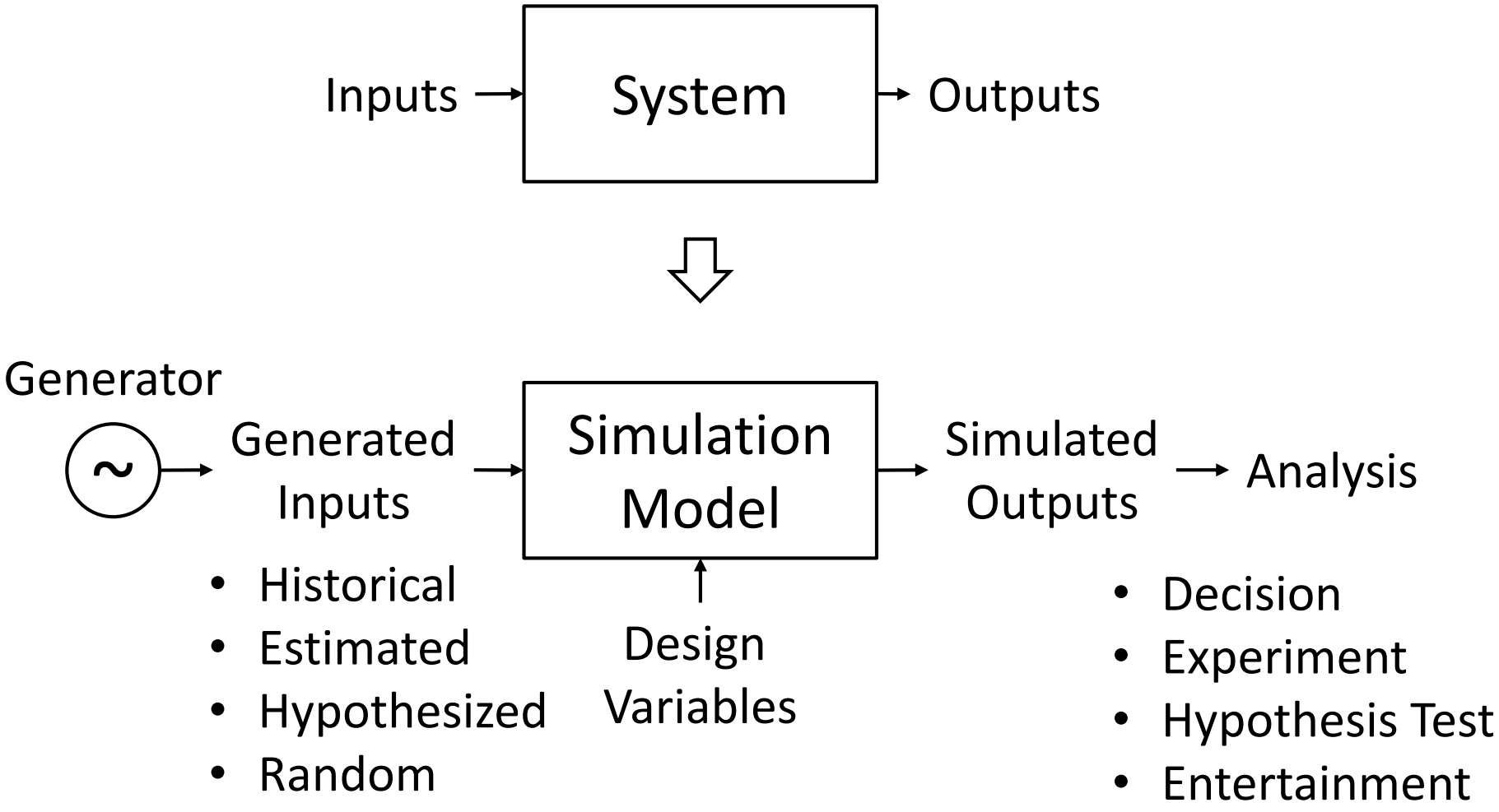
1. Validating Simulation Models
2. Accreditation and Credibility Assessment

Required: J.V. Farr, “Validation, Verification, and Accreditation,” Ch. 9 in *Simulation of Complex Systems and Enterprises*, 2007.

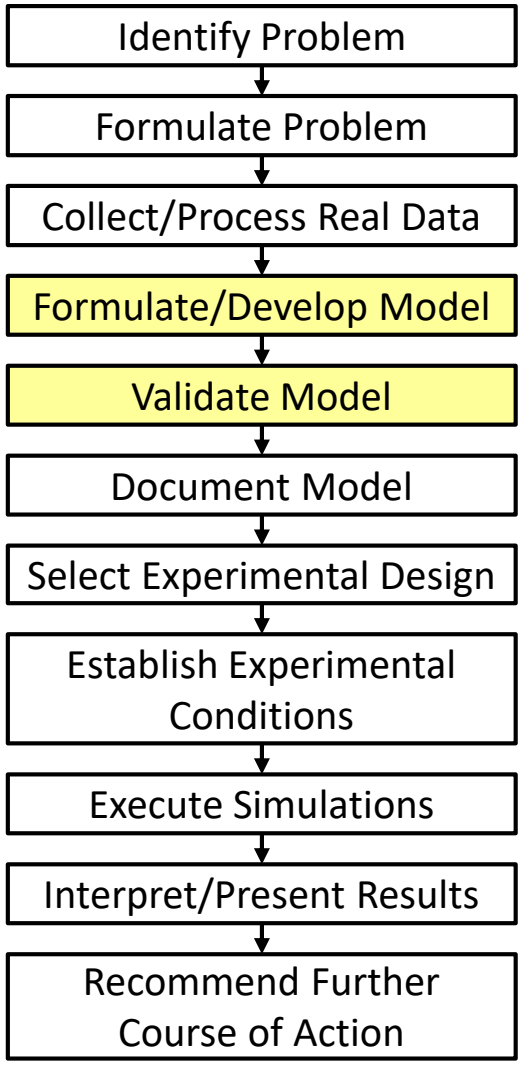
M. Babula et al., “NASA Standard for Models and Simulations: Credibility Assessment Scale,” *AIAA Aerospace Sciences Meeting*, Orlando, FL, January 2009.

Validating Simulation Models

System Modeling & Simulation



Modeling & Simulation Process



4. Formulate/Develop Model

- Verify simulation produces expected results

5. Validate Model

- Compare model output under known conditions to real data

A. Maria, "Introduction to Modeling and Simulation," *Proceedings of the 1997 Winter Simulation Conference*, 1997.

Model Verification & Validation

- **Verification** determines if a model implementation accurately represents the intended description and specifications
  - **Did we build the model right? (Correctness)**

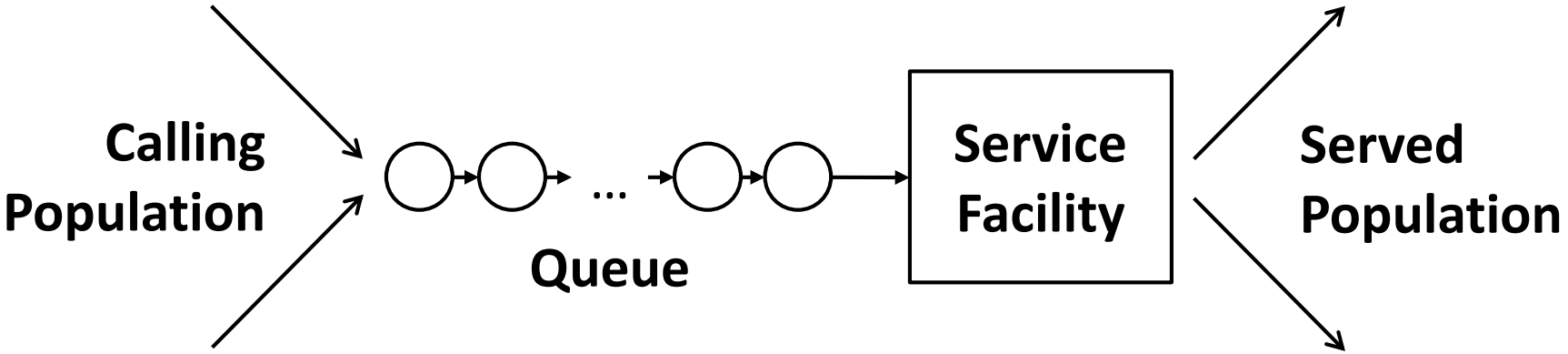
- **Validation** determines the degree to which a model accurately represents the real-world system for its intended uses
  - **Did we build the right model? (Usefulness)**

Validation for Term Project

- **Overall model structure**
  - Does the model match the real world system?
  - Always requires assumptions and limitations
- **Inputs (process generators)**
  - What is the distribution shape and parameters?
  - How do samples compare to observed data?
- **Outputs**
  - How do results compare to observed data?
  - How do results compare to known/existing cases?

Model Structure Validation

- Start with conceptual model
  - Statement of contents and internal representations
  - Combine user's and developer's concepts
  - Recognize assumptions and limitations



“Everything should be made as simple as possible, but not simpler.”

Input Validation

- Use existing sources of data (beware errors)
- Collect new data or observations
  - Subject matter expert (SME) inquiry
  - Follow guidelines for plausible distributions
    - Normal: *some* natural phenomena
    - Triangular: limited knowledge of min, max, and mode
    - Lognormal: positive skew
    - Uniform: no knowledge beside min and max
  - Create custom process generators

Visual Tests for Input Verification

1. Collect the data (as much as practical)
2. Verify and sort the data
3. Select histogram bin size(s)
4. Plot empirical PMF/PDF (histogram) and CDF
5. Calculate descriptive statistics
6. Plot theoretical CDF from descriptive statistics
7. Visually compare CDFs for “good fit”

Visual Tests for Input Verification

Collect Data

Descriptive Statistics

"Thick Pencil"
Min, Max, Mean

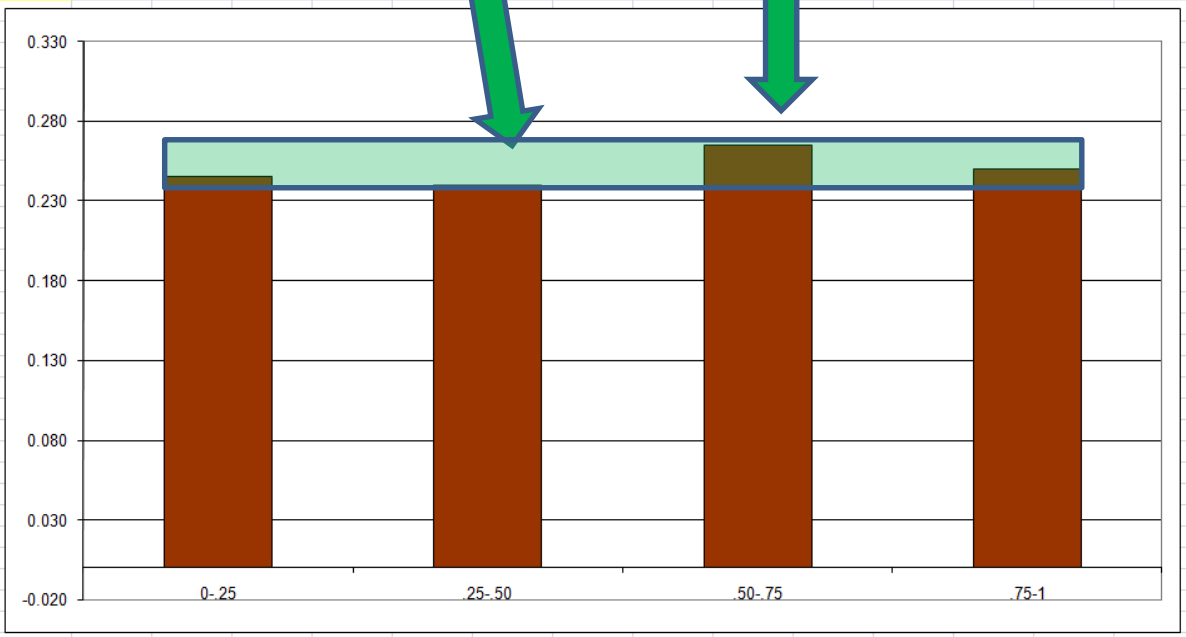
Empirical PMF

Trial	Random Number
1	0.560435835
2	0.64745055
3	0.158375479
4	0.146065348
5	0.784429942
6	0.870522875
7	0.778237876
8	0.449857247
9	0.208313661
10	0.408385148
11	0.095599473
12	0.009019265
13	0.308099374
14	0.85868463
15	0.389672463
16	0.710150339
17	0.573343847
18	0.587740453
19	0.5586827
20	0.955292011
21	0.051735045
22	0.799006779
23	0.607200162
24	0.968496861
25	0.456629602
26	0.315643803
27	0.635396597
28	0.034858052
29	0.437856317
30	0.500020633
31	0.960199288

F9 = Recalculate

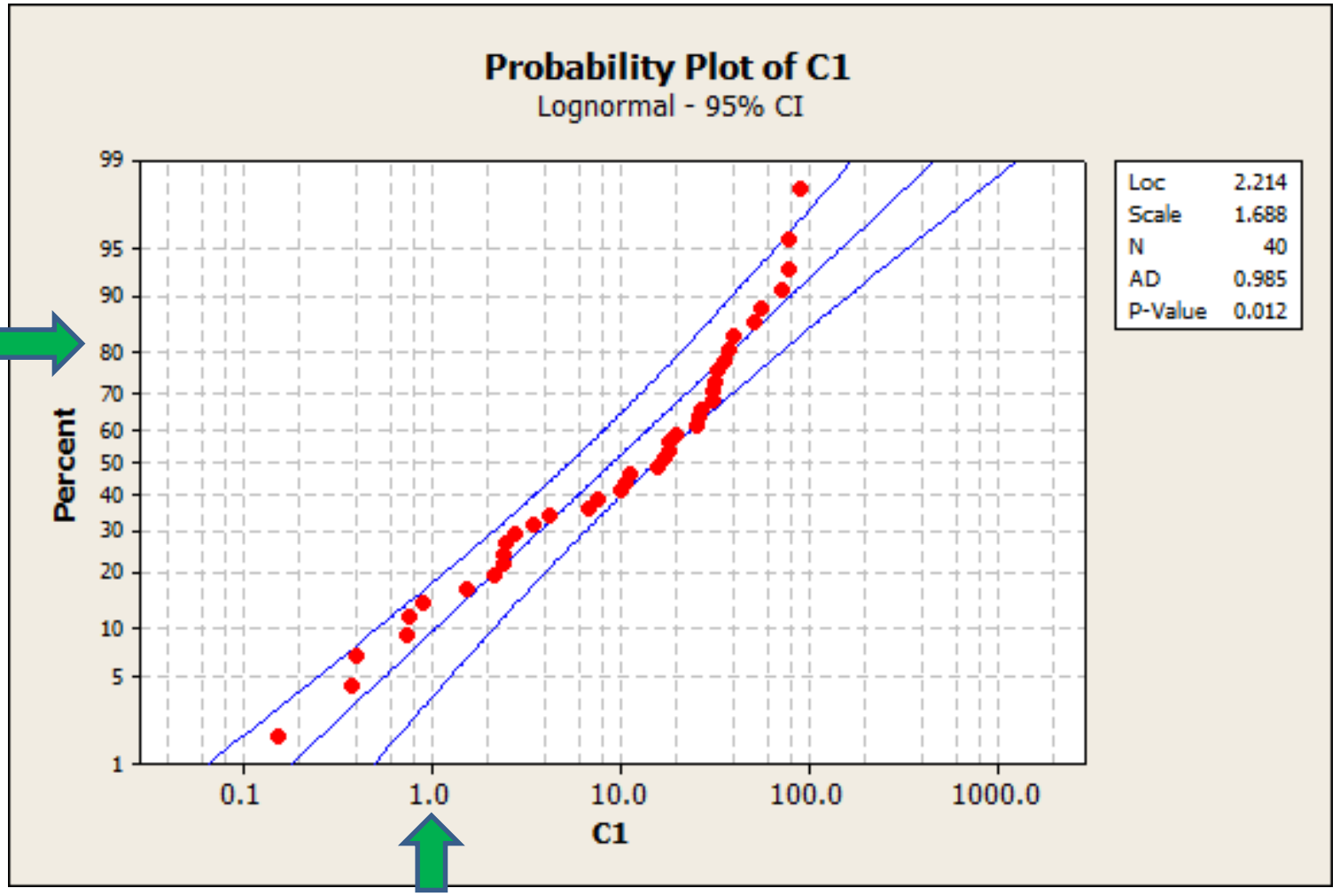
mean =	0.5027
min =	0.0003
max =	0.9978

Fraction	
0-.25	0.245
.25-.50	0.240
.50-.75	0.265
.75-1	0.250



Histogram Bin Size

Visual Tests: Probability Plot



Empirical CDF

Theoretical CDF

Also see Q-Q plot

Statistical Tests for Input Verification

- **Chi-square** Test: Continuous or Discrete
- **Kolmogorov-Smirnov** (K-S) Test: Continuous
- **Anderson-Darling** Test: Continuous

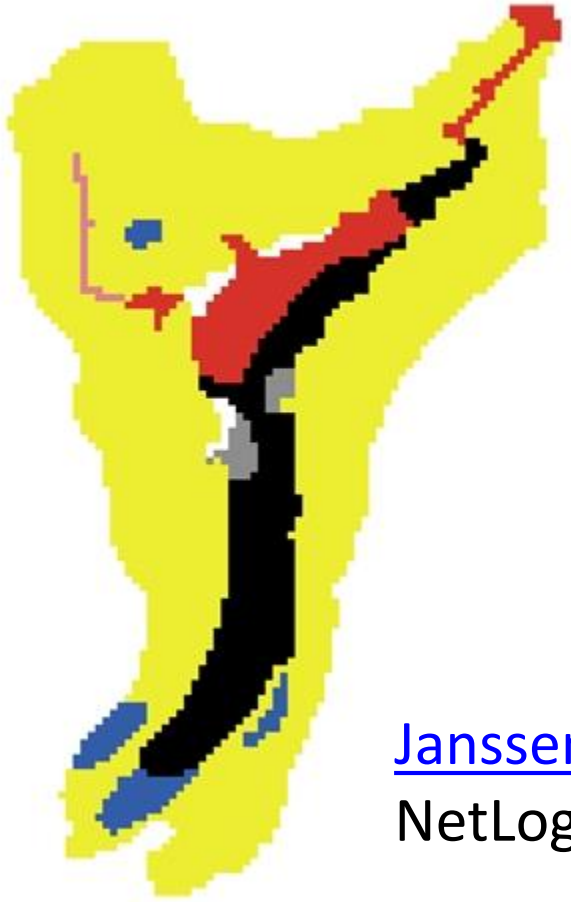
- H_0: Data follow a specified distribution
- H_1: Data do not follow the specified distribution
- Failure to reject H_0 provides evidence the data could come from the specified distribution

Output Validation

- Use existing sources of data (beware errors)
- Collect new data or observations
  - Subject matter expert (SME) inquiry
  - Perform experiments to collect new data
- Compare to known or established baselines
- Explain discrepancies
  - Non-fatal errors: decreases model accuracy
  - Fatal errors: invalidates the model purpose

Artificial Anasazi Model

- ABM of Anasazi people
  - 800-1350 C.E.
  - Long House Valley, AZ
- Compare simulated to historical data
- Environmental factors alone insufficient to explain collapse

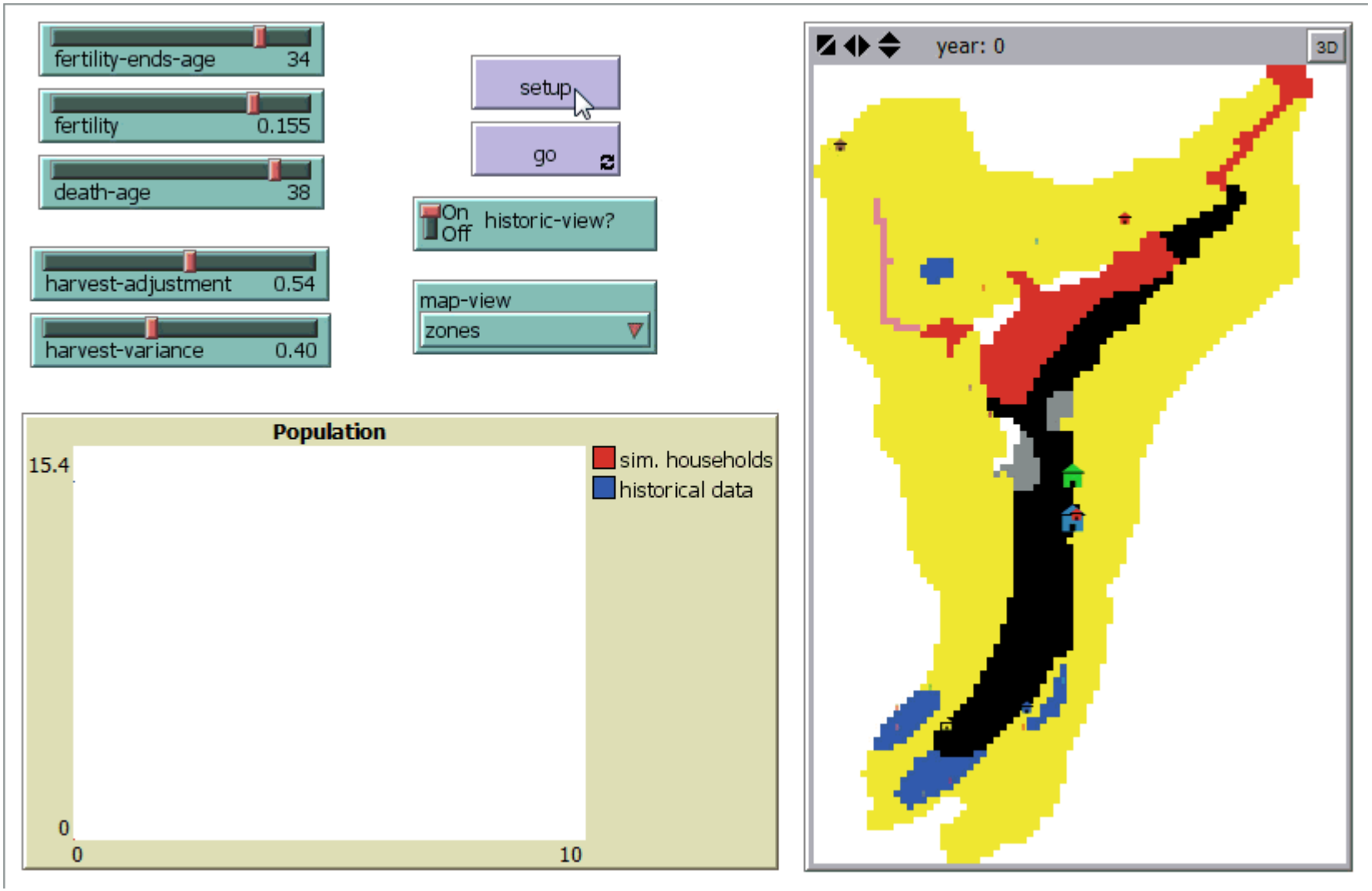


Black: General Valley Floor
Red: North Valley Floor
White: Mid and North Dunes
Gray: Midvalley Floor
Yellow: Nonarable Uplands
Blue: Arable Uplands
Pink: Kinbiko Canyon

Janssen (2009) and NetLogo Model

Anasazi Model

Stonedahl and Wilensky (2010).
NetLogo Artificial Anasazi model.



Simulation Validation

	Structure	Inputs	Outputs
Monte Carlo	Derived state equation	Process generators	Expected value
Discrete Event	State variables, State transition function	Process generators	State trajectories, Expected value
Agent-based	Environment, State variables, State transition function (agent update)	Initial conditions, Process generators	State trajectories

Accreditation and Credibility Assessment

Role of Accreditation

- **Verification** determines the correctness of a model implementation based on its specification. ***Did we build it correctly?***

- **Validation** determines the degree to which a model accurately represents the real world *for its intended uses*. ***Did we build the right thing?***

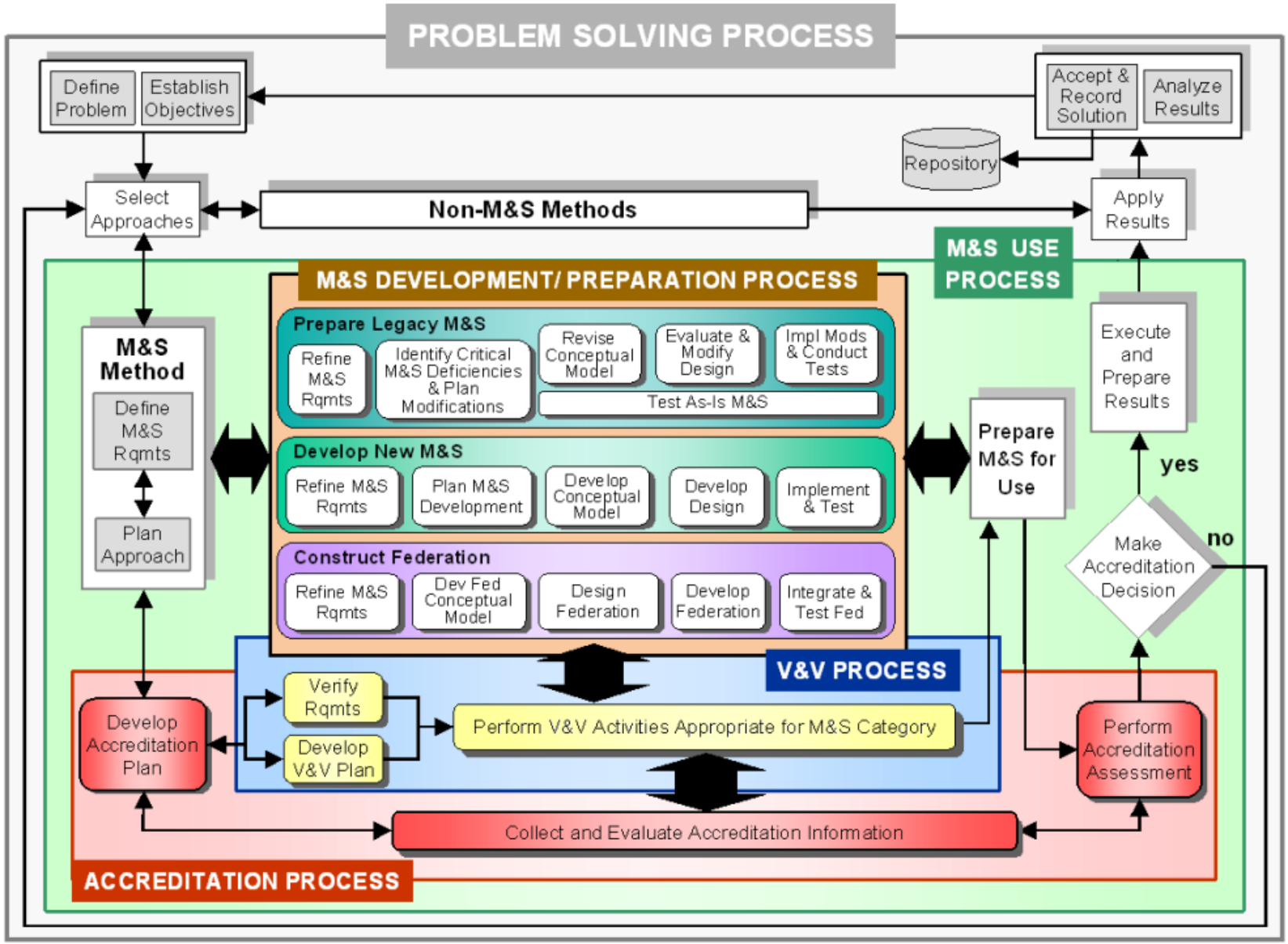
- **Accreditation** officially certifies a model to be acceptable for a specific purpose.

DoD VV&A Standard

- **MIL-STD-3022**: Documentation of Verification, Validation, and Accreditation (VV&A) for Models and Simulations
- Establishes templates for four key products:
  - Accreditation Plan
  - V&V Plan
  - V&V Report
  - Accreditation Report

Accreditation Plan Outline (MIL-STD-3022 Appendix A)

- Problem Statement
- Requirements and Acceptability Criteria
- Assumptions, Capabilities, Limitations, Risks
- Accreditation Methodology
- Accreditation Issues and Mitigation Plans
- Key Participants
- Accreditation Timeline and Resources



MSCO VV&A Recommended Practices Guide http://www.msco.mil/vva_rpg.html

VV&A Roles & Responsibilities

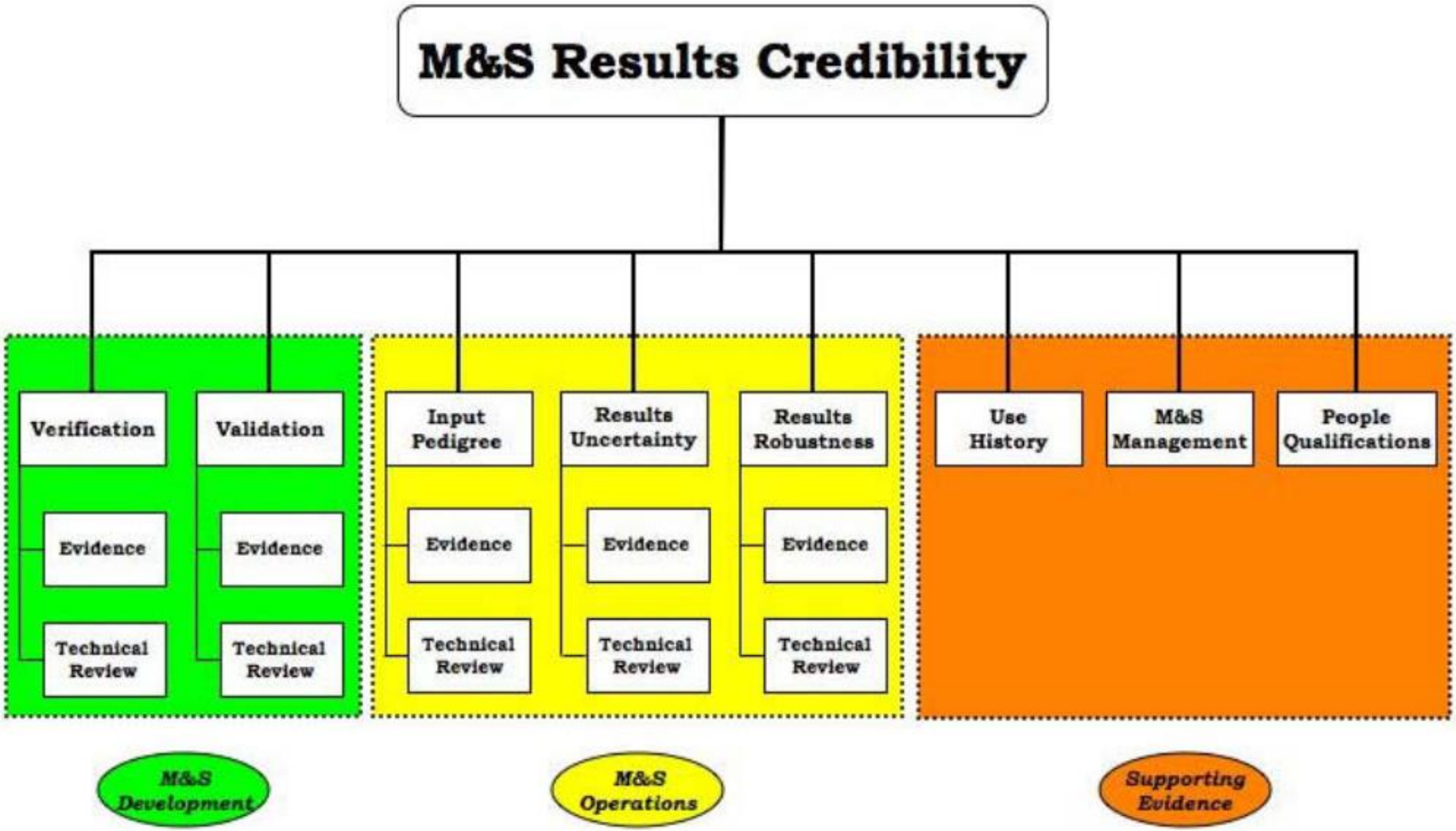
	M&S User	Accreditation Authority	Accreditation Agent	M&S Proponent	V&V Agent	M&S Developer	SME
Define M&S Requirements	Lead/ Approve	Review	Assist	Review	Assist	Assist	Assist
Acceptability Criteria	Assist	Approve	Lead	Review	Assist	Assist	Assist
Accreditation Plan	Review	Approve	Lead	Monitor	Review	Review	Assist
V&V Plan	Review	Monitor	Review	Approve	Lead	Review	Assist
V&V Implementation	Review	Monitor	Monitor	Approve	Lead	Assist	Assist
V&V Data	Review	Monitor	Monitor	Approve	Lead	Assist	Assist
V&V Report	Review	Monitor	Monitor	Approve	Perform		Assist
Accreditation Package				Approve	Perform	Assist	
Accreditation Assessment	Monitor	Monitor	Lead				Assist
Accreditation Report	Review	Approve	Perform				Assist
Accreditation Decision	Review	Perform	Assist				
VV&A Archive & M&S Catalog	Perform	Perform		Perform			
Accredit For Reuse	Review	Approve	Lead	Assist	Assist	Assist	Assist
Accreditation Status	Perform						

Need for Credibility Assessment

- Outcomes of Columbia Accident Investigation Board in 2004 called for NASA standard for models and simulations
- Credibility assessment formalized by 2006
- **NASA STD 7009** published in 2009

Columbia Disaster (2003: Liberman/AP)

NASA Model Credibility Scale



Technical vs. Evidence Review

- Technical reviews weighted by:
  - Informal internal peer review
  - Formal internal peer review
  - External peer review
  - External peer review with independent evaluation
- Evidence reviews conducted on factor-specific scales for some factors

M&S Development Verification

Were the models implemented correctly?

What was the numerical error/uncertainty?

1. Conceptual and mathematical models verified
2. Unit and regression testing of key features
3. Formal numerical error estimation, significant unit test coverage
4. Numerical errors small for all important features in test suite

M&S Development Validation

How well did the M&S results and the referent data compare?

1. Conceptual and mathematical models agree with simple referents
2. Results agree with experimental data or on unit problems
3. Results agree with experimental data for problems of interest
4. Results agree with real-world data

M&S Operations Input Pedigree

How confident are we of the current input data?

1. Input data traceable to informal documentation
2. Input data traceable to formal documentation
3. Input data agree with experimental data for problems of interest
4. Input data agree with real-world data

M&S Operations Results Uncertainty

What is the uncertainty in the current M&S results?

1. Qualitative estimates
2. Quantitative deterministic analysis or expert opinion
3. Quantitative non-deterministic analysis
4. Quantitative non-deterministic and numerical analysis

M&S Operations Results Robustness

How thoroughly are the sensitivities of the current M&S results known?

1. Estimated by analogy to similar problems
2. Sensitivity known for a few parameters
3. Sensitivity known for many parameters
4. Sensitivity known for most parameters; key sensitivities identified

Supporting Evidence Use History

Have the current M&S been used successfully before?

1. Passes simple tests
2. Used previously for critical decisions
3. Previous predictions were later validated by mission data
4. De-facto standard

Supporting Evidence Management

How well managed were the M&S processes?

1. Managed process with roles and responsibilities
2. Established process with documentation
3. Predictable process to measure repeatability of results
4. Continual process improvement with feedback

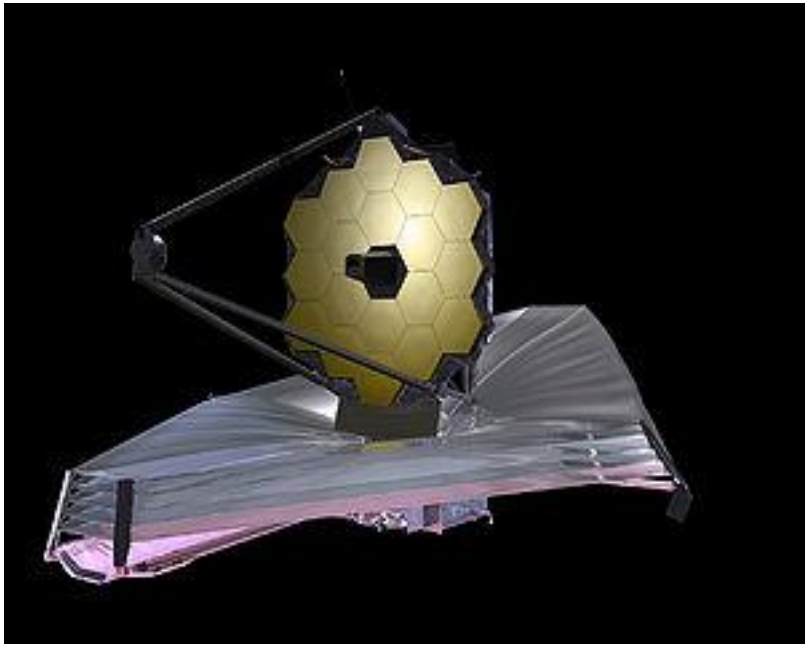
Supporting Evidence Qualifications

How qualified were the personnel?

1. Engineering or science degree
2. Formal M&S training and experience
3. Advanced degree or extensive M&S training and practice
4. Extensive experience developing and using particular M&S

James Webb Space Telescope

- Large infrared space telescope managed by NASA Goddard
- Model of deployed dynamics for vibrations
  - Disturbances from reaction wheels
  - Finite Element Model
  - Commercial code



Credibility Assessment Results

- Technical review yielded all 2s and 3s
  - Credible results for a model that cannot be compared to real-world data (yet!)
  - Evidence review of verification limited by commercial tools (1)