SCIENTIFIC TEST & ANALYSIS TECHNIQUES
CENTER OF EXCELLENCE

FINAL

Model Validation Levels
for Model Authority Quantification

Kyle Provost
Kyle.Provost@HII-TSD.com
25 Apr 2023

Importance of Validation in Modeling & Simulation (M&S)

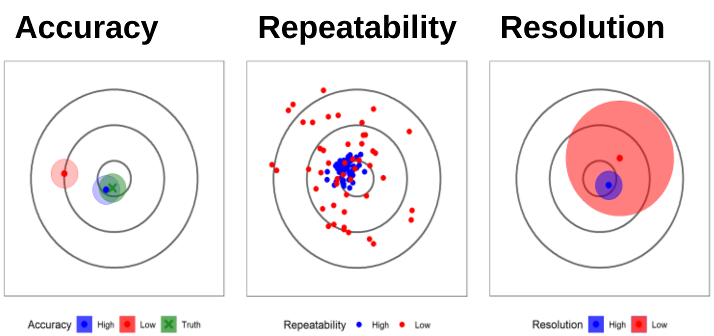
- Validation is the process that determines whether a model has sufficient fidelity for a specific intended use

- Validation of M&S is critical to understand the trustworthiness of M&S results

- Validation should be performed using Scientific Test and Analysis Techniques (STAT) to ensure rigor and consistency of assessment

What Dimensions Matter in Validation

Fidelity



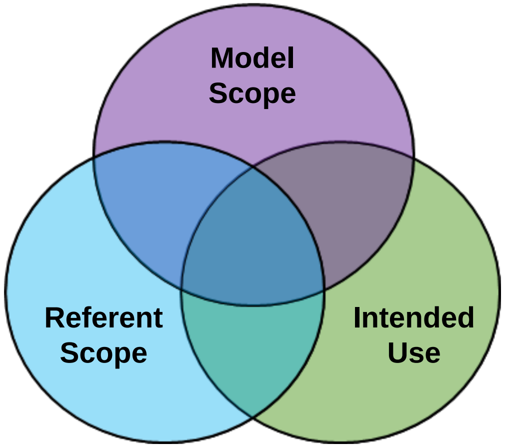
Fidelity quantifies model similarity to referent in 3 dimensions of **consistency**

Authority

Authority Level	Relevant Referent
9	Operational Real-World Data
8	Live System Test Data
7	Prototype Field Test Data
6	HWIL & SWIL Data
5	Lab-Scale System Test Data
4	Integrated Component Lab Test Data
3	Component Lab Test Data
2	First Principles/Physics Predictions
1	SME Judgement

Referent Authority Ranking quantifies trust in **baseline** of comparison

Scope



Scope quantifies degree to which model and referent represent the **same system**

Model Validation Levels (MVLs)

- MVLs provide:
  - A measure for model developers to continuously improve their models
  - A means for decision makers to better understand the risk with making decisions based on models

- In order to accomplish this, MVLs need to:
  - Be usable
  - Be comprehensive
  - Have mathematical rigor and consistency

- MVLs are automatable
  - Input can be limited to data files and intended use
  - Responsive to new data or requirements changes

- MVLs apply to predictive behavior models

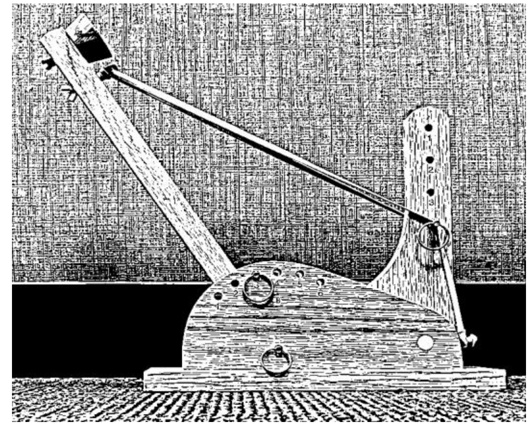
MVL Metric

- MVLs validate a model against a supporting body of knowledge
- Resulting MVLs score models on a 1 to 9 scale
- MVL starts at the trust level of the data, and is decremented according to fidelity and scope
- Models can only be as valid as the most authoritative data that it is assessed against

MVL of:	Is as trustworthy as:
9	Operational Real-World Data
8	Live System Test Data
7	Prototype Field Test Data
6	HWIL & SWIL Data
5	Lab-Scale System Test Data
4	Integrated Component Lab Test Data
3	Component Lab Test Data
2	First Principles/Physics Predictions
1	SME Judgement

Example Model: Catapult

Model



Referent

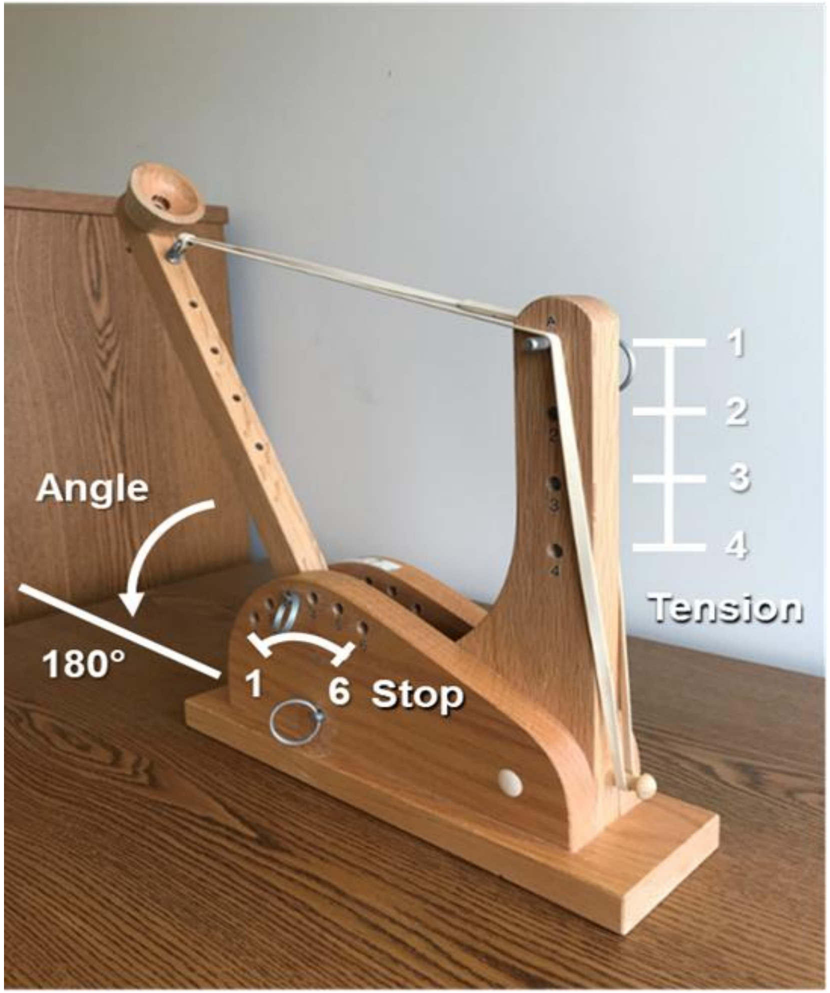
- Python Model
- Physics prediction of distance launched
  - Calibrated with component test data
- Stochastic output
  - Fixed variance independent of input conditions
- Environmental variations not modeled

Set up and Definitions

- Initial set up requires defining the mission scope
  - Technical boundaries
  - Environmental boundaries
  - Measures of Performance

Catapult Model Scope

- Model must reflect system operation for
  - Launch Angles: 160° - 180°
  - Tension Settings: 1 – 4
  - Stop Position: 1 – 4
  - Level Environments: no elevation change
- Measures of Performance
  - Distance Launched



Data Collection and Input

- Models are assessed against the full body of knowledge on system performance
- Referent Authority is used as a weighting factor
  - More trusted data carries more weight for validation
- Data from multiple sources are combined with Bayesian methods

Catapult Body of Knowledge

- Live System Test: Level 8
  - from tests conducted during DOE training courses
- SME input: Level 1
  - from experienced class instructors

Authority Level	Relevant Referent
9	Operational Real-World Data
8	Live System Test Data
7	Prototype Field Test Data
6	HWIL & SWIL Data
5	Lab-Scale System Test Data
4	Integrated Component Lab Test Data
3	Component Lab Test Data
2	First Principles/Physics Predictions
1	SME Judgement

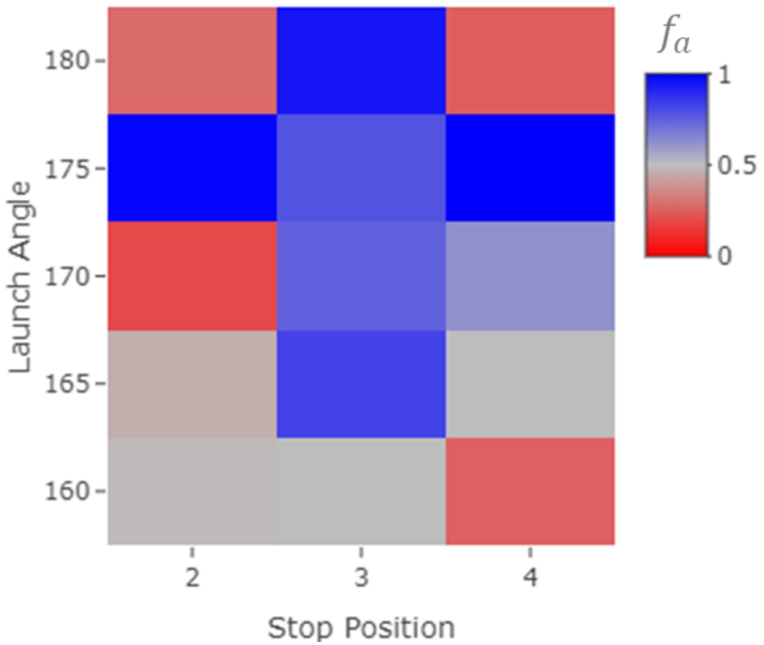
Fidelity Assessment

- Fidelity is assessed across the scope
- Assessment compares model performance to referent data
  - Model mean performance and variability are compared to referent mean and variability

$$Fidelity: \ f = f_a f_v = e^{-\frac{1}{2}\left(\frac{\bar{x}_m - \bar{x}_r}{s_r^*}\right)^2} e^{-\frac{(s_m^* - s_r^*)^2}{s_m^* s_r^*}}$$

Catapult Fidelity Assessment

- Average accuracy score 0.55
- Average variability score 0.54
- Fidelity assessment revealed poor performance in edge cases



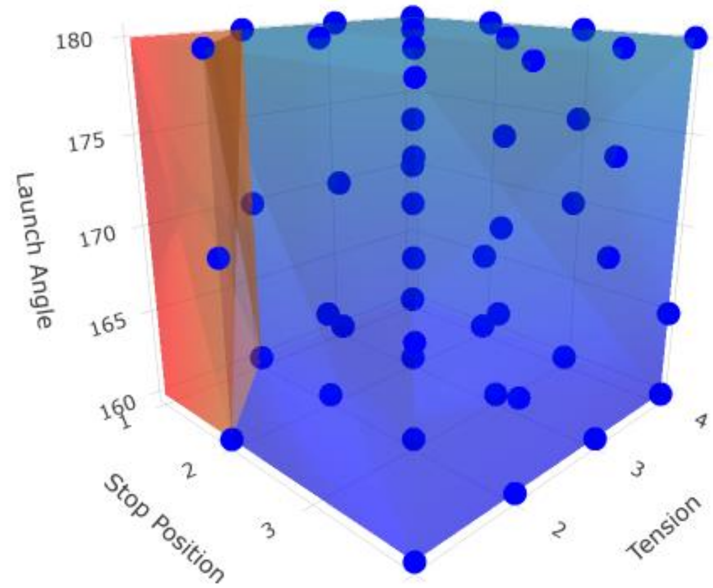
Initial demonstrations could suggest an effective model. Exhaustive fidelity assessment shows poor accuracy for edge cases

$\bar{x}_m = model\ mean$
$\bar{x}_r = referent\ mean$

$s_m^* = model\ standard\ deviation$
$s_r^* = referent\ standard\ deviation$

Scope Assessment

- Scope coverage considers data volume and density
  - Does data support the entire span of the model space?
  - Is there sufficient data throughout the model space?
- Scope is used to down-weight model authority



Catapult Scope Coverage

- Volume coverage: $C_v = 0.981$
  - Every factor fully covered
  - Some factor combinations not fully represented
- Density coverage: $C_d = 0.998$
  - Over 200 data points. Dense coverage throughout space
- Overall coverage: $C = C_v C_d$

Resulting MVL

- Process: Fidelity is calculated across the scope of the model. Fidelity is weighted according to referent authority, and scope coverage metrics to reach a final MVL score
- MVL score ranges from 1 to 9, and is understood as comparable to referent authority
  - Example: An MVL of 7 suggests that a model's outputs are as trustworthy as prototype field test data

MVL of:	Is as trustworthy as:
9	Operational Real-World Data
8	Live System Test Data
7	Prototype Field Test Data
6	HWIL & SWIL Data
5	Lab-Scale System Test Data
4	Integrated Component Lab Test Data
3	Component Lab Test Data
2	First Principles/Physics Predictions
1	SME Judgement

Catapult MVL

- Highest referent authority: Level 8
- Loss due to fidelity: 2.21
  - Accuracy loss: 1.09
  - Precision loss: 1.12
- Loss due to coverage: 0.04
- **Resulting MVL: 5.74**

*MVL component losses are based on relative impact

MVL Interpretation

- Achievable score: 8 (supported by Live System Test)
- Calculated MVL: 5.74

- The model is comparable to lab-scale system test data in terms of trustworthiness
- Significant losses from fidelity may indicate need for model improvement

MVL are decision support tools

- Possible model use decision impacts:
  - Invest in model improvement, addressing fidelity in edge cases
  - Down-scope model, validating for use in high-fidelity cases
  - Accept low-fidelity model if performance is sufficient for use case

Takeaways

- MVLs are a comprehensive assessment of a model that provide an interpretable measure of trust
- MVLs provide diagnostic tools for enabling improvement of models or awareness of model weaknesses
- MVLs are automatable, facilitating continuous model improvement and informed decisions from models

STAT COE | Delivering Insight to Inform Better Decisions

Visit, www.AFIT.edu/STAT
Email, AFIT.ENS.STATCOE@us.af.mil