Development of a statistically-based validation assessment framework to quantify model confidence, model acceptability, and validation recommendations

B. M. Wilson, A. Koskelo
Los Alamos National Laboratory

ASME V&V Symposium
May 13-17. Las Vegas, NV

Managed by Triad National Security, LLC for the U.S. Department of Energy’s NNSA
Outline of the talk

Motivation for model acceptability and validation recommendations

Outline our validation assessment framework, including
• Model accuracy and confidence
• Model acceptability
• Validation recommendations

Demonstration using direct-drive inertial confinement fusion

Conclusions and caveats
What are the questions asked of validation by customers (a LANL perspective)?

**Code and Model Developers**
- What accuracy can I expect?
- What physics need improvement?

**Program Managers**
- Is the code acceptable for our needs?
- Do the code predictions meet this set of tolerances?

**End-Users**
- What confidence do I have in the code to predict XYZ physics?
- In what regimes can I expect good accuracy?
- How good is good enough?

**Experimentalists**
- What experiments or diagnostics are needed to better constrain the validation assessment?
What questions should validation assessment answer?

**Validation:**

*The process of determining the degree to which a model is an accurate representation of the real world from the perspective of the intended uses of the model.*

**Validation should determine the degree to which**

1. A model is an accurate representation of the real world (**model confidence**).
2. A model can be used for its intended uses (**model acceptability**).
3. A model, experiment, or validation can be improved, revised, or modified to better characterize model confidence and model acceptability (**validation recommendations**).

We suggest degree infers a quantitative, statistical assessment of confidence.
Suggested components of a validation assessment

1. **Model Accuracy:**
   Determination of a model’s accuracy for a set of model prediction QoI as compared to experiment (our most reliable representation of the real world).

2. **Model Confidence:**
   Quantification of the degree of confidence in a model’s accuracy. Model confidence should consider all validation uncertainties, including model inputs, numerical discretization, and experimental variability.

3. **Model Acceptability:**
   Statistical acceptance or rejection of a model given validation evidence, model confidence, and model requirements (e.g. model-use tolerances) based on the intended uses of the model.

4. **Validation Recommendations:**
   Statistical recommendations based on validation results, such as future validation strategies, resource allocation, adequacy for regimes and applications based on customers intended uses.

How does the validation assessment framework fit into validation?

- Physical Experiments
  - Evaluate QoI, including experiment uncertainties
    - Experiment QoI PDF
  - Acceptance Criteria
- Computational Model
  - Evaluate QoI, including simulation uncertainties
    - Simulation QoI PDF

- Determine Model Confidence
- Accept
- Release Model
- Validate Next Model
- Reject
- Design Additional Experiments
- Update or Refine Computational Model
- Validation Recommendations

Model Accuracy: Determination of a model’s accuracy ($\langle e_\phi \rangle$) for a set of model prediction QoI ($\phi_S$) as compared to experiment ($\phi_D$).

$$\langle e_\phi \rangle = \langle \phi_D \rangle - \langle \phi_S \rangle$$

Model Confidence: Quantification of the degree of confidence in a model’s accuracy. Model confidence should consider all validation uncertainties, including model inputs, discretization, and experimental variability.

$$pdf(e_\phi) = pdf(\phi_D) \times pdf(\phi_S)$$

$$\langle e_\phi \rangle \pm u_{V,\phi} \approx \langle e_\phi \rangle \pm \sqrt{\sum u_i^2}$$
Model Acceptibility

Statistical acceptance or rejection of a model given validation evidence, model confidence, and model requirements based on the intended uses of the model.

$H_0$: The model accuracy is within our model-use tolerances. $P(H_0) = P(\{e_\phi\} < \Delta\phi)$

$H_a$: The model accuracy is not within our model-use tolerances. $P(H_a) = P(\{e_\phi\} > \Delta\phi)$

**Accept $H_0$:** Given the validation evidence, the model predictions are sufficient for our needs.

**Reject $H_0$:** Given the validation evidence, the model predictions are not sufficient for our needs.

Model-use tolerances $\Delta\phi$ should be defined by model developers, experimentalists, analysts, and end-users.

Validation Recommendations inform future validation, model use and development, and experimental design

**Comparison Ambiguity Hypothesis:**

\( J_0: \) The validation uncertainty magnitude is larger than the model-use tolerances.

\[ P(J_0) = P\left( |e_\phi - \langle e_\phi \rangle| > \Delta \phi \right) \approx P(u_{V,\phi} > \Delta \phi) \]

Sources: experimental uncertainty, model inputs, discretization issues, comparison uncertainties, etc.

**Comparison Systematic Error Hypothesis:**

\( K_0: \) The comparison error is not zero.

\[ P(K_0) = P(\langle e_\phi \rangle \neq 0) \]

Sources: diagnostic bias, facility design flaws, model form errors, numerically unresolved solutions, etc.
An example using direct-drive inertial confinement fusion (ICF)

On ICF in general:
“The ultimate goal of these experiments is to ignite a self-sustaining burn wave of fusion fuel, producing more energy than is delivered to the target—an event called ignition.”
Edward Moses, previous LLNL principal associate director for NIF and Photon Science

[Diagram showing Laser Energy, Ablator Blowoff, Ablation, Compression, Ignition, Burn]

Ablation: High energy lasers rapidly heat the surface of the fusion target forming a surrounding plasma envelope.

Compression: Fuel is compressed by the rocket-like blowoff of the hot surface material.

Ignition: During the final part of the laser pulse, the fuel core reaches 20 times the density of lead and ignites at 1E8 K.

Burn: Thermonuclear burn spreads rapidly through the compressed fuel, yielding many times the input energy
Model Accuracy
\[
\langle e_\phi \rangle = \langle \phi_D \rangle - \langle \phi_S \rangle
\]

Model Confidence
\[
pdf(e_\phi) = pdf(\phi_D) \cdot pdf(\phi_S)
\]
\[
\approx \bar{u}_{V,\phi} \approx \sqrt{\sum u_i^2 - \sum u_{i,j}^2}
\]
Model accuracy and confidence are insufficient to quantify model acceptability and validation recommendations

- scattered laser light
- ablation front trajectory
- ablation front velocity

• Is the model acceptable when the validation uncertainty overlaps the model-use tolerances?
• Is there a statistically significant systematic error, such as a model-form error?
• Is our validation ambiguous due to large uncertainties?

These questions must be answered using hypothesis tests for model acceptability and validation recommendations.

Model Confidence

Model Tolerances
Δφ_p = ±2.5% φ_{D,max}
Δφ_r = ±5% φ_{D,r}
Δφ_v = ±5% φ_{D,v}

Model-use tolerances Δφ should be defined by model developers, experimentalists, analysts, and end-users.
Demonstration of Model Acceptability

Given the validation evidence,
- there is over 95% probability of an acceptable model before main laser.
- there is over 95% probability the model is unacceptable for most QoI during main laser.

How can we improve our validation assessment, experiment, or model?
- Reduce uncertainty (which uncertainties)?
- Acquire more data (in what regimes)?
- Improve models (model-form error) or post processing?

Acceptability is not universal and can be regime and QoI dependent.
Demonstration of Validation Recommendations

Given the validation evidence,

- there is over 95% probability of a systematic error during primary laser drive.
  - Other experiments suggest a model form error due to missing physics, i.e. cross beam energy transfer (CBET)
- 95% probability of large uncertainties for the velocity QoI.
  - Uncertainties are dominated by experimental sources.

1. CBET models are being included in xRAGE.
2. Including exp. with varying CBET-sensitivity.
3. Velocity measurements should be re-assessed.
Common ICF calibrations help mitigate CBET deficiencies at the expense of predictability

Experiment Simulation

- scattered laser light
- ablation front trajectory
- ablation front velocity
Conclusions

- Presented a framework of hypothesis tests for quantifying essential components of a validation assessment:
  - Model Accuracy and Model Confidence (using V&V 20-2009)
  - Model Acceptability (hypothesis testing)
  - Validation Recommendations (hypothesis testing for systematic errors and ambiguity)
- Demonstrated framework on a direct-drive ICF validation example:
  
  **Model Acceptability**
  - High probability (>95%) the model is not acceptable given the current validation evidence and model-use tolerances during the main laser pulse.
  - Acceptability is not universal and can be regime and QoI dependent.

  **Validation Recommendations**
  - High probability (>95%) of a systematic error in the validation, likely a model form error (CBET).
    1. Include CBET capabilities in the xRAGE laser package (Mazinis).
    2. Supplement validation experiments with CBET sensitivity experiments
  - High probability (>68%) of ambiguity in velocity due to large experimental uncertainties.
Questions
Caveats when using hypothesis tests

The conclusions for all validation assessments (and hypothesis tests) are underwritten by the available data and results.

- Acceptability of a model in validation is underwritten by the available evidence.
  - We can never accept a hypothesis using hypothesis testing (i.e. we are unable to test ALL cases).
- Acceptability of a model is not universal; it is regime, QoI, and physics dependent.
- A poorly defined hypothesis test may give inconsistent results as more (consistent) data is acquired.
- Poorly acquired or characterized data, insufficient modelling or validation rigor can result in weak or misleading validation conclusions,
- A validation decision (or hypothesis test) can always be reversed by the collection of better, alternate, or more data.
Model Accuracy: How accurate is the model given the validation evidence?

• Model accuracy, $\delta_{\text{model, } \phi}$, of our QoI, $\phi$, is quantified by comparing to experiment (ASME VV20-2009):

$$\delta_{\text{model, } \phi} \geq e_\phi \pm u_{V, \phi}$$

comparison error

$$e_\phi = \phi_{\text{Measured}} - \phi_{\text{Simulated}}$$

validation uncertainty

$$u_{V, \phi} = f \left( u_{X, \phi} \right)$$

Validation uncertainty ($u_V$) is composed of many uncertainty sources:

• Experimental Uncertainty ($u_D$): uncertainty from facility, operations, or diagnostics

• Numerical Uncertainty ($u_N$): uncertainty from translation of the mathematical model to a numerical implementation

• Input Uncertainty ($u_I$): uncertainty of model inputs from the experiments (e.g. initial conditions, mat. prop., etc.)

• Comparison Uncertainty ($u_C$): uncertainty from comparing simulation SRQ to experimental SRQ

• Inherited Model Error: errors inherited from uncharacterized contributing sub-models (e.g. eos, opacity, etc.)
Numerical uncertainty quantification is complicated by Adaptive Mesh Refinement (AMR)

- Grid Convergence Index (GCI) methods are the “best” quantitative methods
- AMR is not well suited for GCI

- **Error from a high resolution grid (16x) is used to confirm GCI uncertainty**
  - AMR nominal and minimum cell size refined between grids
  - GCI uncertainty calculated from 3 coarsest grids (0.25x, 0.5x, 1x)
  - Error calculated from highest resolution grid (16x)

A GCI estimate of $u_{N,\phi}$ provides nearly complete coverage of estimated error $e_\phi$.

Regions of higher uncertainty are due to insufficient refinement.

Observed convergence $\sim 1.0$ ($L_1$ Norm)
The input uncertainty is estimated by latin-hypercube sampling (LHS)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>𝑢_ψ</td>
<td>±10 μm</td>
<td>±2μg</td>
<td>±1%</td>
<td>±1%</td>
<td>±3.5 μm</td>
<td>±1%</td>
<td>±1%</td>
<td>±1%</td>
<td>±1%</td>
<td>±1%</td>
</tr>
</tbody>
</table>

pdf(ψ)

Uncertainties from input parameters are propagated into the model using LHS

LHS gives the set of possible model responses for a SRQ (trajectory example shown)

Michel2013 Model

Uncertainties estimated from 68%, 95%, and 99% coverage of SRQ distribution coverage

- incident laser light
- scattered laser light
- ablation front trajectory
- ablation front velocity
Capsule geometry is the largest contributor to input uncertainty

- scattered laser light
- ablation front trajectory
- ablation front velocity

Experiment

Simulation

\[ \Delta \phi_p = \pm 2.5\% \phi_{D,max} \]
\[ \Delta \phi_r = \pm 5\% \phi_{D,r} \]
\[ \Delta \phi_v = \pm 5\% \phi_{D,v} \]
Large scattering deficiencies propagate into all QoI

- Scattered laser light
- Ablation front trajectory
- Ablation front velocity

Scattered Laser Light
- High model accuracy during triple-picket pulses.
- Scattering underpredicted during the primary drive pulse.

Ablation Front Trajectory
- Scattering deficiency propagates into ablation front QoI.
- Comparison error increases with time due to increased acceleration

Why does the nominal CH model do so much better?
xRAGE verification has begun for many analytic solutions

- **Shock propagation**
- **Inverse Bremsstrahlung laser deposition**
- **Laser Ablation**

Laser energy deposited at the critical-density surface.

## Initial results are promising; much more required for validation

<table>
<thead>
<tr>
<th>Experiment</th>
<th>$R_{\text{beam}}/R_0$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Schmitt2018</td>
<td>0.5, 0.58, 0.7</td>
</tr>
<tr>
<td>Froula 2012</td>
<td>0.5, 0.6, 0.75, 0.88, 1.0</td>
</tr>
<tr>
<td>Froula 2012</td>
<td>0.5, 0.6, 0.75, 0.88, 1.0</td>
</tr>
<tr>
<td>Michel 2013</td>
<td>0.78</td>
</tr>
</tbody>
</table>

![Graph showing $\phi_r$ vs $t$ (ps) with markers for SFC 3, SFC 4, and XRF C1.](image)

- **Experiment:**
  - Self-emission radiograph

- **xRAGE:**
  - Density
  - Electron Temperature
How do we reconcile validation differences, strengths, and weaknesses in 1d, 2d, and 3d?

**Physics fidelity**
- Increased dimensionality can converge to different physics.
- AMR becomes more important as new 2d/3d features appear:
  - Laser non-uniformities
  - Local RT/RM/jetting structures

**Mesh Independence**
- Unable to obtain mesh independence with increasing dimensionality due to increasing computational cost.
- 2d/3d features may be artifacts of mesh and AMR.

**Validation Considerations**
- High computational cost limits our ability to use Monte-Carlo sampling for UQ.