CAN SIMULATION CREDIBILITY BE IMPROVED USING SENSITIVITY ANALYSIS TO UNDERSTAND INPUT DATA EFFECTS ON MODEL OUTCOME?

J. Myers¹, M. Young², D. Goodenow¹, A. Keenan², M. Walton², L. Boley²

1. NASA Glenn Research Center
2. Wyle Science, Technology & Engineering Group
Its Really About Model Credibility! Achieving a high level of belief or trust in the model

- **NASA-STD-7009**
  - Standard for Models and Simulations (M&S)

- **Eight Factors**
  - Verification
  - Validation
  - **Input Pedigree**
  - Results Uncertainty
  - Results Robustness
  - Use History
  - M&S Management
  - People Qualifications

### Generic STD-7009 Credibility Assessment

<table>
<thead>
<tr>
<th>Credibility Assessment Factors</th>
<th>Evidence</th>
<th>Technical Review</th>
<th>Overall Score</th>
<th>Sufficiency Threshold</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Score*</td>
<td>Weight*</td>
<td>Threshold*</td>
<td>Score*</td>
</tr>
<tr>
<td>Verification</td>
<td>2</td>
<td>0.20</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Validation</td>
<td>2</td>
<td>0.25</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Input Pedigree</td>
<td>2</td>
<td>0.10</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>Results Uncertainty</td>
<td>0</td>
<td>0.10</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>Results Robustness</td>
<td>2</td>
<td>0.10</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Use History</td>
<td>1</td>
<td>0.15</td>
<td>2</td>
<td>N/A</td>
</tr>
<tr>
<td>M&amp;S Management</td>
<td>2</td>
<td>0.05</td>
<td>3</td>
<td>N/A</td>
</tr>
<tr>
<td>People Qualifications</td>
<td>4</td>
<td>0.05</td>
<td>3</td>
<td>N/A</td>
</tr>
</tbody>
</table>

* Maximum = 4; where 0=insufficient evidence and 4=highest fidelity/rigor achievable
+ Minimum = 0.05, maximum = 0.25 and sum of all weights must equal 1.0
Problem with “Input Pedigree” Ranking

• A complex model integrating many factors can have a wide array of input pedigree scores across the range of model parameters
  – Source, quality of the data, type of data

• From NASA-STD-7009: Input Pedigree (Ranking) involves the evaluation of all data that is used as input for the current M&S results. It includes not only data that is unique to the model, but also data that is produced by other simulations.

• Implication: Model and Simulation Input Pedigree Ranking is equivalent to the lowest input pedigree of the model parameters (i.e. logical AND).
Question:

- In practice, the Input Pedigree ranking of complex multi-parameter models tends to be low, resulting in lower than expected model and simulation credibility.

- Question: Is this evaluation of credibility true for all applications and simulations of the model?

- Considerations:
  - Application credibility differs from the model credibility with respect to input pedigree.
  - Application credibility can be more precisely determined using Sensitivity analysis (also known as Results Robustness) to address relative importance of input parameters.
• Saltelli: “Sensitivity Analysis is the study of how variation in the output of a model can be apportioned, qualitatively or quantitatively, to different sources of variation (input) and how the given model depends upon the information fed into it.”

• Partial Rank Correlation Coefficient (PRCC) Analysis
  – Provides the linear relationships between two variables (one input parameter and one output parameter) when all linear effects of other variables are removed after rank transformation.
  – Rank Transformation: transforms non-linear monotonic relations to linear.

• **KEEP IN MIND** the difference in a contributing parameter and a sensitive parameter
  – Many parameters contribute substantially to the mean output of the model
    • Low sensitivity may indicate a “DC-signal effect” over the range of model application and parameter variance
  – Parameter variance affecting model output (magnitude and variance) indicates a sensitive parameter
Using IMM for ISS mission 6 crew members

• IMM Provides probabilistic analysis of 99 medical condition occurrences and impact to mission outcomes
• Context: 6-month, 6 crew (M-4, F-2), ISS med kit, and evacuation is possible
• Output:
  – Crew Health Index (crew available time – time lost due to medical events)
  – Probability of consideration of evacuation
Sensitivity Analysis Results: Crew Health Index for 6 Month Mission to the International Space Station

- Visual Impairment
- Skin Laceration
- Dental Caries
- Dental Abscess
- Eye Chemical Burn
- Hearing Loss
- Nephrolithiasis
- Skin Abrasion
- Lower Extremity Stress Fracture
- Stroke Cerebrovascular
- Dental Exposed Pulp
- Eye Infection
- Traumatic Hypovolemic Shock
- Sepsis
- Barotrauma Ear Sinus Block
- Hip Proximal Femur Fracture
- Urinary Tract Infection
- Small Bowel Obstruction
- Eye Abrasion Foreignbody
- Appendicitis
- Angina Myocardial Infarction
- Back Injury
- Wrist Fracture
- Finger Dislocation
- Atrial Fibrillation
Sensitivity Analysis Results: Consideration of Evacuation for 6 Month Mission to the International Space Station

EYE CHEMICAL BURN
ACUTE PANCREATITIS
SMOKE INHALATION
ACUTE CHOLECYSTITIS BILIARY COLIC
ABDOMINAL INJURY
EYE PENETRATION FOREIGN BODY
VISUAL IMPAIRMENT
CHEST INJURY
SHOULDER DISLOCATION
ACUTE DIVERTICULITIS
APPENDICITIS
ACUTE COMPARTMENT SYNDROME
ACUTE DIVERTICULITIS
SHOULDER DISLOCATION
CHEST INJURY
VISUAL IMPAIRMENT
EYE PENETRATION FOREIGN BODY
ABDOMINAL INJURY
ACUTE CHOLECYSTITIS BILIARY COLIC
SMOKE INHALATION
ACUTE PANCREATITIS
EYE CHEMICAL BURN
Interpretation: Weighted Averaging

- Weighted averaging of the sensitivity scores uses the PRCC output ($W_{PRCC}$) to “weigh” parameter pedigree ($IP_i$).
- For this analysis assume:
  - One input parameter ranked at 1, the rest are 3.
  - Originally a strict “Std-7009” Input Pedigree Score = 1.

\[
IP_{total} = \frac{\sum W_{PRCC,i} IP_i}{\sum W_{PRCC,i}}
\]

<table>
<thead>
<tr>
<th>Eye Chemical Burn - $IP_{total}$</th>
<th>100 Conditions</th>
<th>25 Most Influential Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI</td>
<td>2.944</td>
<td>2.918</td>
</tr>
<tr>
<td>Consideration of Evacuation</td>
<td>2.772</td>
<td>2.935</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>VIIP - $IP_{total}$</th>
<th>99 Conditions</th>
<th>25 Most Influential Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>CHI</td>
<td>2.963</td>
<td>2.959</td>
</tr>
<tr>
<td>Consideration of Evacuation</td>
<td>2.691</td>
<td>2.545</td>
</tr>
</tbody>
</table>
Interpretation: Threshold Specification

• Set a threshold value below which the contribution of the parameter is considered to be small
  – Pedigree of data below this level is not considered in the assessment of simulation input pedigree

• Assume conservative threshold to be a PRCC = 1/n = 0.01 and a more lenient threshold to be PRCC = 2/n = 0.02
  – n = the number of parameters

• For this analysis assume
  – One input parameter ranked at 1, the rest are 3
  – Strict “Std-7009” Input Pedigree Score = 1
Threshold Analysis Results

<table>
<thead>
<tr>
<th>CHI</th>
<th>$W_{PRCC}$</th>
<th>IP - Con</th>
<th>IP - LCon</th>
</tr>
</thead>
<tbody>
<tr>
<td>VIIP</td>
<td>-0.41</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Eye Chem Burn</td>
<td>-0.07</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Dental Carries</td>
<td>-0.13</td>
<td>1</td>
<td>1</td>
</tr>
</tbody>
</table>

Conservative Threshold $+/-.01$

Less Conservative Threshold $+/-.02$

<table>
<thead>
<tr>
<th>Consideration of Evacuation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$W_{PRCC}$</td>
</tr>
<tr>
<td>-----------</td>
</tr>
<tr>
<td>VIIP</td>
</tr>
<tr>
<td>Eye Chem Burn</td>
</tr>
<tr>
<td>Dental Carries</td>
</tr>
</tbody>
</table>
Conclusions

• **Proof of concept analysis indicates**
  – Sensitivity analysis can be used to better characterize the input pedigree ranking of specific simulations
  – One analysis is insufficient to characterize the input pedigree ranking for all possible analyses – Case by case basis

• **Method used to evaluate input pedigree requires further investigation**
  – Current sample approaches are reasonable but not optimized
  – Quantification should be repeatable and consistent with other communications components of 7009 to maintain decision making information quality

• **Future efforts will address**
  – More appropriate definitions of cutoff thresholds, allowing a distinction in numerical noise and statistical significance
INTEGRATED MEDICAL MODEL OVERVIEW

Thank you!

Questions?
CHI – All Conditions
Consideration of Evacuation – All Conditions