Selection of Validation Metrics for Aviation Seat Models

Presented to: The Fifth Triennial International Aviation Fire and Cabin Safety Research Conference

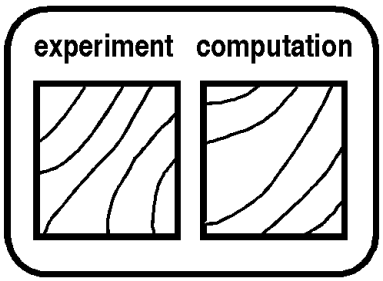
By: David Moorcroft, FAA CAMI

Date: Oct. 30th, 2007

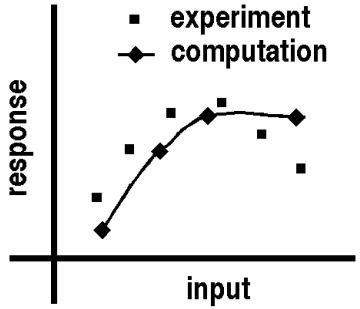
Background

- **With the increased use of numerical models for seat design and certification, there is a fundamental need to show that the model is an accurate representation of the real world.**
  - A process called validation.
- **Validation metrics calculate the error between simulation and experimental results.**
- **Specification and use of validation metrics is important because different error metrics give different scores for the same time-history pairs.**
- **Need an automated and consistent procedure.**

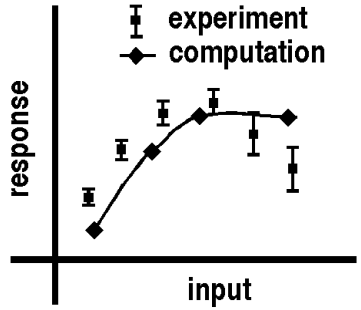
Sequence of Validation Metrics



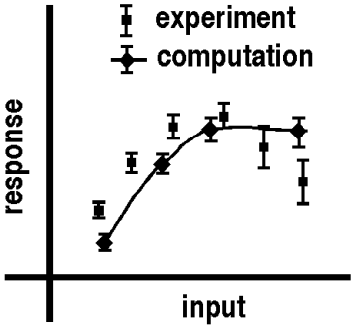
(a) Viewgraph Norm



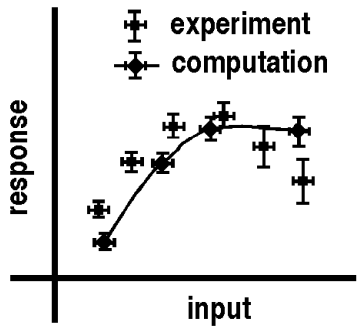
(b) Deterministic



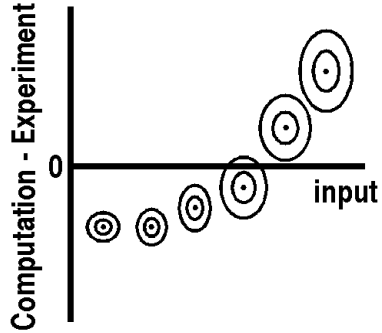
(c) Experimental Uncertainty



(d) Numerical Error



(e) Nondeterministic Computation



(f) Statistical Comparison

Trucano, et al 2002

Overview

- **Quantitative curve shape metrics:**
  - Three components: peak, phasing, and shape.
  - Components sometimes combined.
  - Need consistent values:
    - If 10% magnitude (peak) error is “good”, a 10% shape error should also be considered “good”.
- **Results should be consistent with Subject Matter Expert (SME) opinions.**
- **Increased use of metrics, but selection rationale is rarely specified.**

Goal

- **Evaluate curve shape metrics with multiple methods.**

- **Define selection criteria (rationale) for the choice of a curve shape metric that is appropriate for the aviation seating community.**

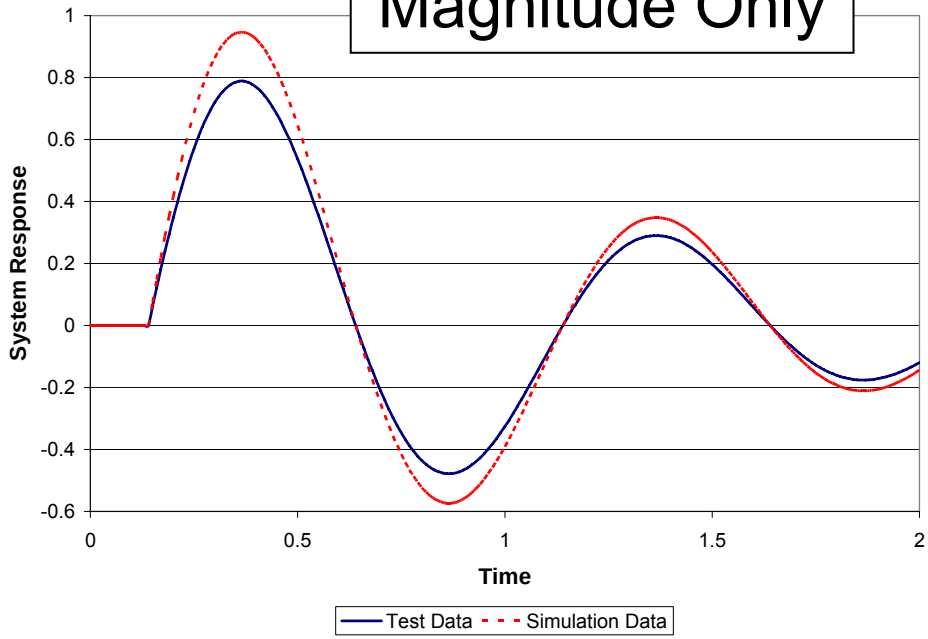
Evaluation Methods

- **Comparison to Idealized Waveforms**
  - Magnitude only
  - Phase only
  - Error at high magnitudes vs. low magnitudes
- **Comparison to Head Acceleration Time History and HIC**
- **Ability to Discriminate Curves**
- **Comparison to Subject Matter Expert Opinions**

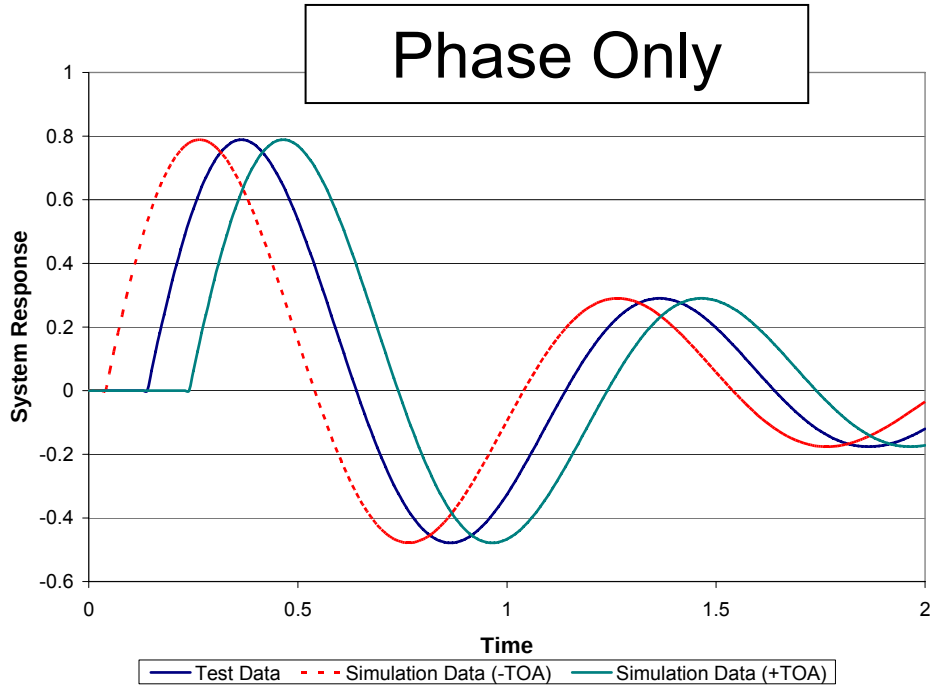
4 Curve Shape Metrics Evaluated

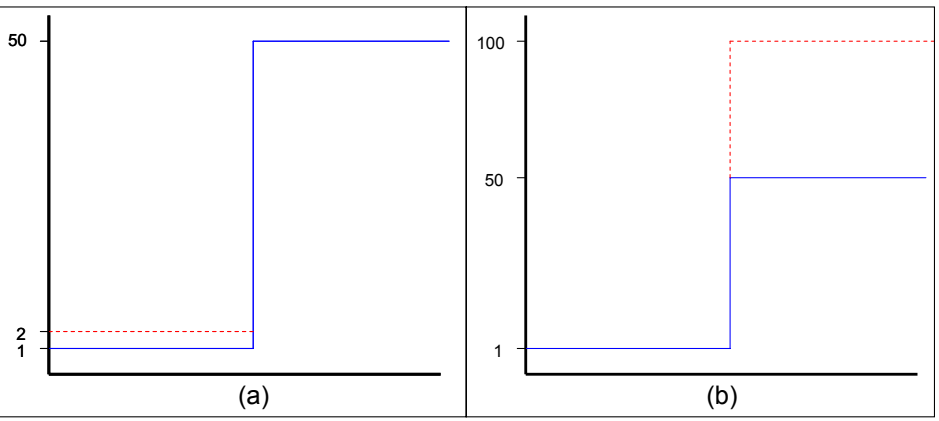
- **Sprague and Geers (S&G)**
  - General purpose curve shape metric
  - Implemented in a spreadsheet
- **Weighted Integrated Factor (WIFac)**
  - Automotive curve shape metric
  - Implemented in a spreadsheet
- **Global Evaluation Method (GEM)**
  - Automotive curve shape plus peak and timing
  - Requires ModEval (stand alone program)
- **Normalized Integral Square Error (NISE)**
  - Biomechanics curve shape plus magnitude and phase
  - Implemented in a spreadsheet

Magnitude Only

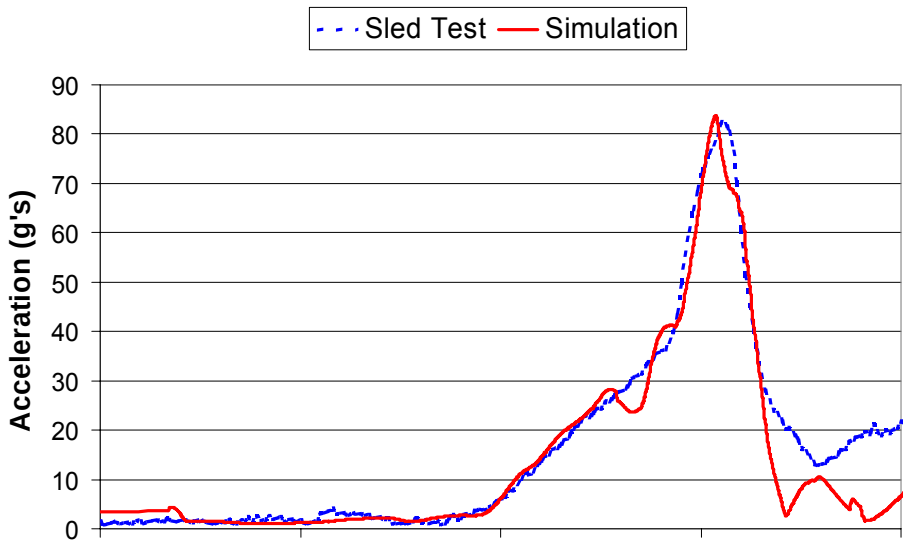


Phase Only





Weighting



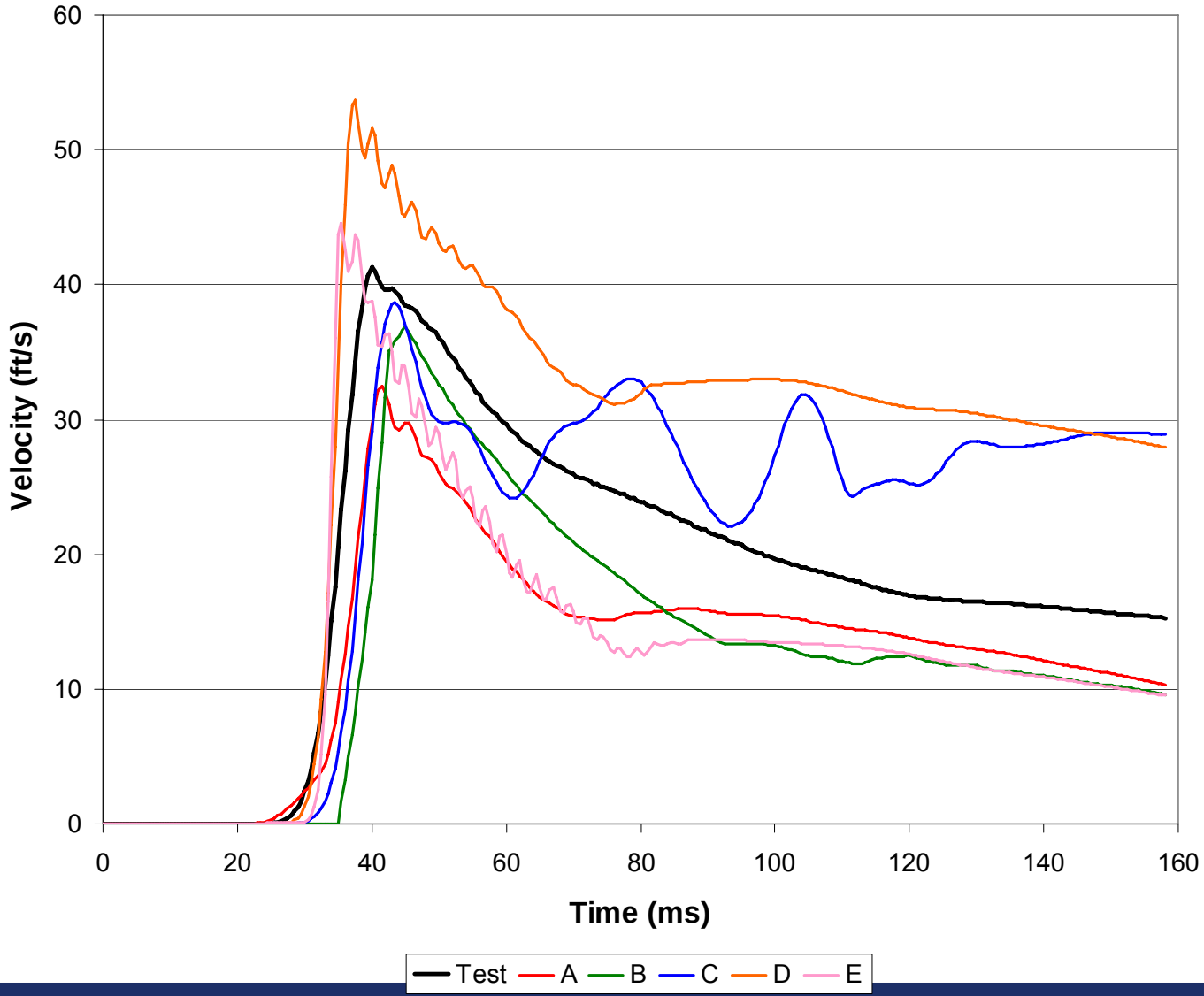
Head Acceleration

Ideal Waveforms + Head Acc.

Scenario	Ref. Error	S&G	WIFac	GEM	NISE
Mag.	20%	20.0	16.7	10.9	1.6
+ Phase	~20%	19.5	55.2	5.9	18.2
- Phase	~20%	19.5	55.0	5.4	18.2
Weighting (L)	2.0%	0.6	9.8	0.2	0.0
Weighting (H)	98%	100.0	49.8	56.4	20.0
Head Acc.	6.3%	9.9	33.1	3.6	2.9

Weighting ref. = area under curve, Acc. ref. = rel. error on HIC

Discrimination



Discrimination

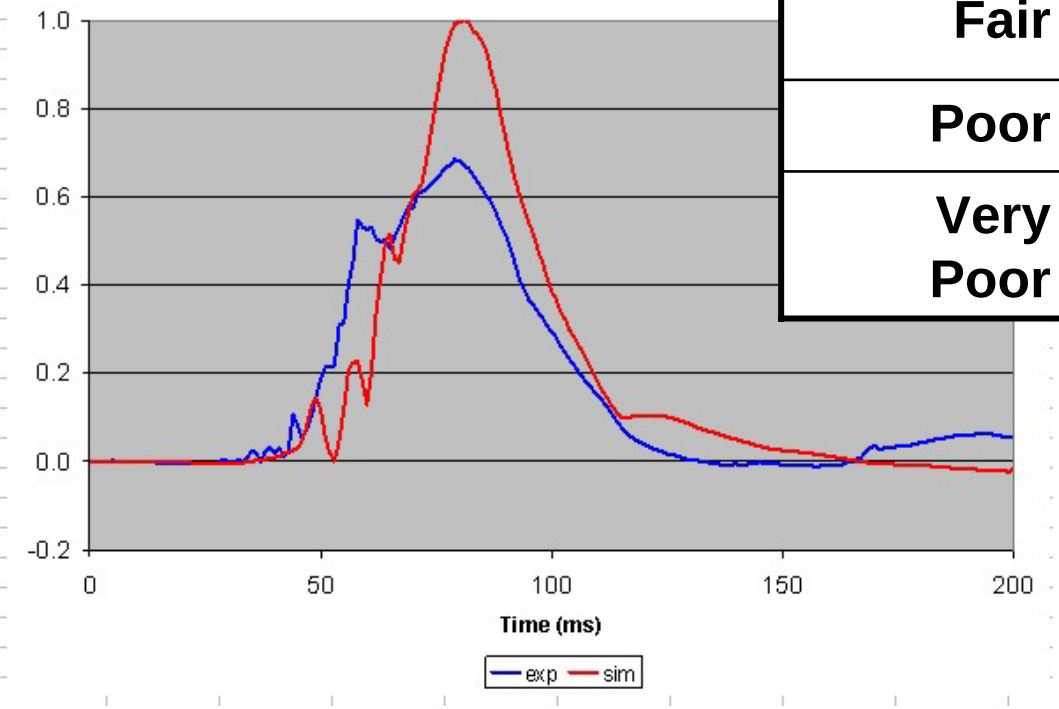
	S&G	WIFac	GEM	NISE
Model A	29	30	16	6.3
Model B	26	34	11	6.1
Model C	20	32	10	6.8
Model D	45	35	19	6.5
Model E	24	33	12	7.7
Mean	29	33	13	6.7
Coef. Var.	0.33	0.06	0.29	0.09

SME Details

- **16 experts (industry, gov't, academia) submitted evaluations of 39 test/simulation time history curves.**
- **Evaluations consisted of a score (excellent, good, fair, poor, very poor) for magnitude, phase, shape, and overall agreement.**
- **The data represent accel, vel, pos, angle, force, and moment time histories derived from both occupant & structural responses.**
- **Data normalized such that highest peak = 1.**

Example Curve (Pair 19/SME 1)

	Mag.	Phase	Shape	Overall
Excellent				
Good		X		
Fair			X	X
Poor	X			
Very Poor				



SME Data Analysis

- **Qualitative scores converted to quantitative:**
  - Excellent = 1
  - Good = 2
  - Fair = 3
  - Poor = 4
  - Very Poor = 5
- **Basic statistical calculations computed for each test/simulation pair (average, mode, st dev, etc.).**
  - Mode represents the most frequent response.

How do SMEs determine the Overall score?

- **Magnitude score = Overall score: 25/39 (pairs)**
- **Phase score = Overall score: 20/39**
- **Shape score = Overall score: 31/39**

- **Worst score from Mag/Phase/Shape: 28/39**
- **Best score from Mag/Phase/Shape: 19/39**
- **Average score from Mag/Phase/Shape: 28/39**

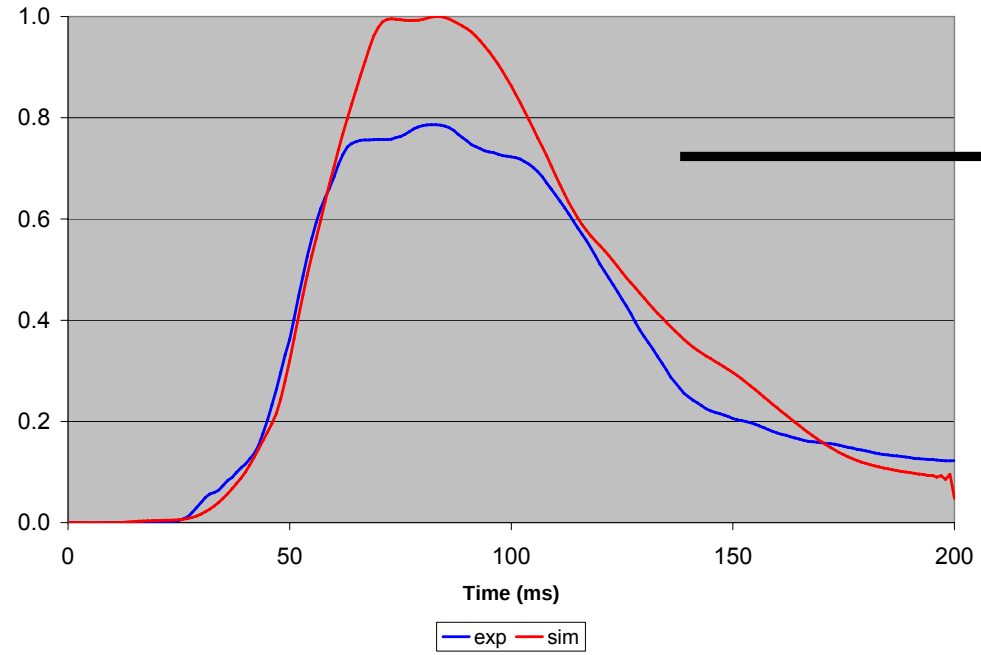
- **Magnitude = Phase = Shape = Overall: 13/39**

Observations

- **Curve shape includes both magnitude (peak) and phasing (timing).**
  - Can have good magnitude without good shape, but not good shape without good magnitude.
  - Can match time of peak with poor shape, but not good shape with poor timing.

- **Magnitude scores are not consistent within individual SME or group.**

Inconsistent Magnitude (Mag.)



Pair 2

Mag. Mode = Poor

Lower Mag. = 0.79

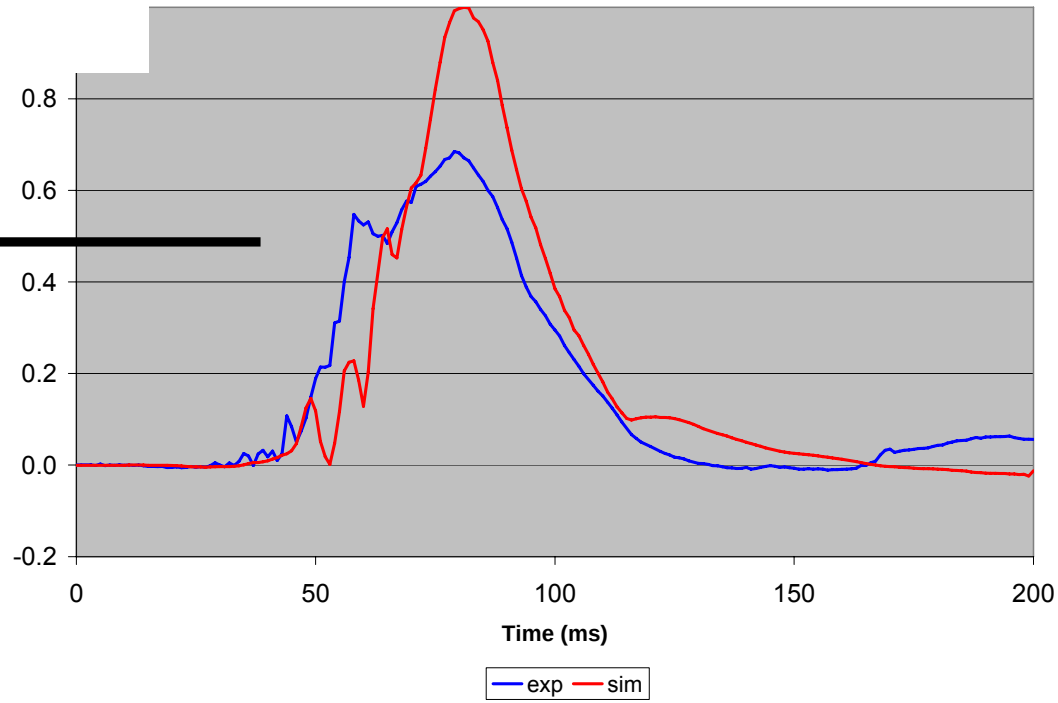
Error = 21%

Pair 19

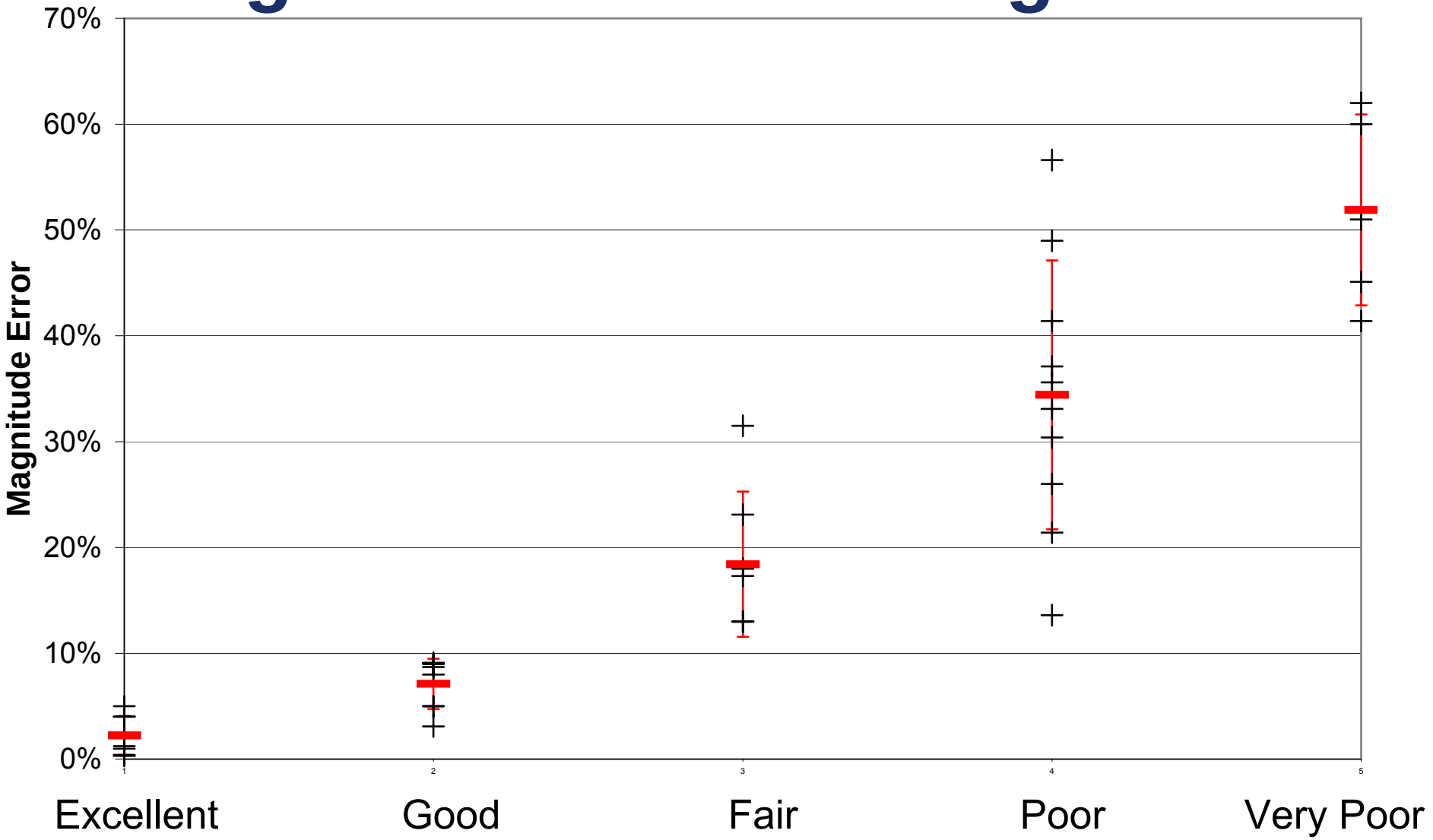
Mag. Mode = Fair

Lower Mag. = 0.69

Error = 31%



Mag. Error vs. SME Mag. Score



Mag. % Error vs. SME Mag. Score

	Avg. Diff.	St Dev	Avg - 1 St Dev	Avg + 1 St Dev	Suggested Range (%)
Excellent	2.2	1.7	0.5	3.9	0 – 4
Good	7.1	2.4	4.7	9.5	4 – 10
Fair	18.4	6.9	11.6	25.3	10 – 20
Poor	34.4	12.7	21.7	47.1	20 – 40
Very Poor	51.9	9.0	42.9	60.9	40 +

Phasing

- **Defined for SME Evaluation as the "timing of events."**
- **Time of the peak is typically used within a relative error.**
- **Definition of a reference time allows for a time independent error calculation.**
  - Simple relative error ($\Delta t / t_T$).
  - 5 ms difference at 50 ms (10%) vs. 150 ms (2.5%).
  - For ref = 100 ms ($\Delta t / t_{ref}$), error = 5% regardless of location in time history.

Comparison of Phasing Error to Mag.

	Low (% error)	High (% error)	Suggested Mag. Range
Excellent	0	5	0 - 4
Good	0	8	4 - 10
Fair	1	40	10 – 20
Poor	2	30	20 - 40
Very Poor	42	42	40 +

Metric Avg. vs. SME Shape

	S&G	WIFac	GEM	NISE	Suggested Mag Range
Excellent	4.5	14.9	2.7	0.8	0 – 4
Good	12.9	28.1	11.1	3.3	4 – 10
Fair	25.9	45.4	23.9	14.4	10 – 20
Poor	32.1	48.7	31.7	25.2	20 – 40
Very Poor	65.6	74.2	33.6	78.0	> 40

Curve Shape Results

- **S&G most closely reproduced the reference errors for idealized waveforms.**
- **S&G and GEM performed best in the discrimination evaluation.**
- **S&G, GEM, and NISE were all consistent with the SME evaluations.**
  - Curve shape error matched the error ranges suggested from the magnitude data.

Metric Evaluation Rationale

- **Idealized waveforms allow for a better understanding of the underlying features of the various metrics.**
  - Comparison to absolute error.
- **Use of head acceleration allows for comparison with relative error of HIC.**
- **Discrimination between various simulation results is beneficial.**
  - i.e., when used within an optimization routine.

Metric Evaluation Rationale (2)

- **If metrics are to be used as a stand in for expert opinion, it is important for the results to be consistent with the Subject Matter Expert opinions.**
- **When combining magnitude error, timing of peaks, and shape, it is critical that the individual error scores are consistent.**
  - i.e., 10% is “good” for all features.
  - Apples to apples comparison.

Curve Shape Recommendation

- **Simple, deterministic metric.**
  - Easy to implement in a spreadsheet.
  - Limited number of seat tests.
- **Error metric biased towards the experiment.**
  - Consistent with certification activities.
- **Appropriate results for idealized curves.**
- **Metric results consistent with SME values.**
- **Sprague & Geers metric meets these specifications and appears to be the best choice for validating numerical seat models.**