OBJECTIVE RATING METRIC FOR DYNAMIC SYSTEMS

Saeed Barbat
Yan Fu
Zhenfei Zhan
Ren-Jye Yang
Ford Motor Company
United States of America
Christian Gehre
PDB – Partnership for Dummy Technology and Biomechanics
Germany
Paper Number 13-0448

ABSTRACT

Computer Aided Engineering (CAE) has become a vital tool for product development in the automotive industry. Various computer programs and models are developed to simulate vehicle crashworthiness, dynamic, and fuel efficiency. To maximize the effectiveness and the use of these models, the validity and predictive capabilities of these models need to be assessed quantitatively.

For a successful implementation of CAE models as an integrated part of the current vehicle development process, it is necessary to develop an objective metric that has the desirable metric properties to quantify the discrepancy between physical tests and simulation results. However, one of the key difficulties for model validation of dynamic systems is that most of the responses are functional responses, such as time history curves. This calls for the development of an objective metric that can evaluate the differences of the time history as well as the key features, such as phase shift, magnitude, and slope between test and CAE curves.

In this paper, four state-of-the-art objective rating metrics are investigated. Multiple dynamic system examples for both tests and CAE models are used to show their advantages and limitations. Further enhancements are proposed to improve the robustness of these metrics. A new combined objective rating metric is developed to standardize the calculation of the correlation between two time history signals of dynamic systems. Multiple vehicle safety case studies are used to demonstrate the effectiveness and usefulness of the proposed metric for future ISO Technical Specification and Standard for the TC22/SC10/SC12/WG4 “Virtual Testing” Working Group.

INTRODUCTION

Prototype tests to evaluate safety performance of a new vehicle in order to meet current and future safety requirements are on the rise. Computer modeling and simulations are playing an increasingly important role in reducing vehicle prototype tests and shortening product development time. To achieve these goals, the validity and predictive capabilities of the computer models for various vehicle dynamic systems must be assessed objectively, quantitatively, and systematically.

Model validation is the process of comparing model outputs with experimental observations in order to assess the validity or predictive capabilities of computer models. The fundamental concepts and terminology of model validation have been established mainly by various standard committees and professional societies ([1], [2], [3], [4], [5]).

One of the critical tasks to achieve quantitative assessments of models is to develop a validation metric that has the desirable metric properties to quantify the discrepancy between functional or time history responses from both physical tests and simulation results ([6], [7], [8]). However, the primary consideration in the selection of an effective metric should be based on the application requirements. In general, the validation metric shall be a quantitative measurement to judge whether a computational model is adequate for its intended usage.

In this paper, four validation metrics for dynamic responses are investigated and they are: CORrelation and Analysis (CORA) metric [9], Enhanced Error Assessment of Response Time

Histories (EEARTH) metric [10], model reliability metric [11], and Bayesian confidence metric [12]. Several dynamic responses for both test and CAE model are used to show some limitations of these metrics. Further improvement of the CORA corridor rating and EEARTH metric are proposed to improve their robustness. Finally, a combined objective rating metric based on the improved CORA corridor metric and EEARTH is proposed to standardize the calculation of the correlation between two signals of dynamic systems. Multiple vehicle safety case studies are used to demonstrate the effectiveness and usefulness of the proposed metric for future ISO Standard.

SCOPE

The scope of ISO TC22/SC10/12/WG4 "Virtual Testing" Working Group was to provide a validated metric to calculate the level of correlation between two non-ambiguous signals (e.g. time-history signals) obtained from a physical test and a computational model of the same test. The defined metric shall be primarily aimed at vehicle safety applications.

The objective was to develop a fully documented metric instead of the development and provision of rating software.

This paper gives a general overview of the recent work. It is also an excerpt of the ISO documents ISO PDTR 16250 [13] and ISO TS 18751 [14] prepared by this expert group.

METHOD

The work on the new standard started with a literature review to determine the state-of-the-art metrics in this specific area. Black box approaches such as commercial rating software without fully documented algorithms or algorithms that are protected by intellectual property rights were excluded because of the aims of the ISO working group.

Ideal metric characteristics

There are many ideal metric characteristics that would be desirable in model assessment of dynamic systems ([6], [7], [8]). The most important ones for vehicle safety applications are:

(1) objective – produces same result regardless who conducts the assessment,

(2) generic – reflects differences in the full distribution of the simulation and experimental outcomes and key features like phase, magnitude, and slope,

(3) robust – produces consistant results with different sampling rates,

(4) symmetric – produces same result when the experiment and simulation outcomes switch,

(5) simple – easy to understand and use,

(6) contains clear physical meaning and Subject Matter Experts (SMEs)' knowledge,

(7) under uncertainty – accounts for data uncertainties in both the experiments and numerical simulations.

Pre-selected metrics

Based on the above ideal metric characteristics, four different metrics were considered for the future standard development by this expert group. An intense validation program helped to identify the most appropriate algorithms.

Algorithms that only analyze local features of a signal, e.g. peak, time of peak etc., are not considered. The approach of this working group was to develop a metric that analyzes complete signals including its local features.

CORA The objective rating tool CORA uses two independent sub-ratings, a corridor rating and a cross-correlation rating to assess the correlation of two signals [9].

The corridor rating calculates the deviation between both curves with the help of user-defined or automatically generated corridors. The cross correlation rating analyzes specific curve characteristics, such as phase shift, size, and shape of the signals. This combination of two completely independent ratings helps to compensate for each other's disadvantages.

The CORA rating tool is also trying to separate engineer's knowledge from the objective rating metric by using external parameters. It offers the possibility to fine-tune the evaluation to the specific needs of the applications by adjusting those metric parameters to reflect the SMEs' knowledge of the applications.

EEARTH The EEARTH metric is based on the Error Assessment of Response Time Histories (EARTH) [15] that provides three independent error measures: phase, magnitude, and topological. The phase error deals with the overall error in timing between two functional responses when considering all the points of the responses. Magnitude error is defined as the difference in amplitude of the two functional responses when there is no time lag between the two. Topological error deals with error associated with the shape of the functional responses, such as the number of peaks, valleys, and slope. A very unique feature of the EARTH metric is using dynamic time warping (DTW) to separate the interaction of phase, magnitude, and topological errors. DTW is an algorithm for measuring discrepancy between time histories and was first used in context with speech recognition in the 1960's [16]. The time warping technique aligns peaks and valleys as much as possible by expanding and compressing the time axis according to a given cost function [17]. Since the ranges of three errors are quite different and no single error can provide a quantitative model assessment alone, the original EARTH metric employs a linear regression method to combine the three errors into one score. A numerical optimization method is employed to identify the linear coefficients so that the resulting EARTH rating can match with the SMEs' ratings closely for a specific application. However, the resulting linear combination of the EARTH metric is mainly numerically based and application dependent, therefore, it may not be scalable to other applications.

In order to provide one intuitive rating and improve the robustness of the metric with different sampling rates while maintaining the advantages of the original EARTH metric, an enhanced EARTH metric called EEARTH is developed. The major enhancements include:

(1) developing an integrated calibration process to incorporate physical-based thresholds and SMEs' knowledge to provide phase score, magnitude score, slope score, and the combined EEARTH rating all in the standard “0” to “1” range;

(2) using a distance-only cost function for DTW instead of both distance and slope-based cost function in the original EARTH to improve the robustness of magnitude scores with different sampling rates;

(3) eliminating DTW on slope curves so that the slope error is calculated directly from the difference between the two slope curves calculated from the shifted and truncated test and CAE curves to improve the robustness of slope scores with different sampling rates.

Hence, the EARTH was enhanced by simplifying the algorithms and reducing the influence of the signal’s sampling rate on the rating score.

Model reliability metric A model reliability-based validation metric was developed for dynamic system applications [11]. The difference between CAE and test curves is taken as the validation feature. The threshold factor is defined by SMEs' experience, the lower and upper bounds of the threshold interval are defined as the product of the threshold factor and the absolute maximum amplitude of the reference signal. The model reliability metric is represented by the probability that the observed difference is within the lower and upper bounds of the threshold interval. If a pre-defined reliability target is met, the model is acceptable. Since this difference time-history curve has better normality than those of the test and CAE curves, a normal distribution of the difference can be assumed and the model reliability metric can be simply calculated.

There are only two adjusting parameters: threshold factor and reliability target, and both have clear physical meanings. The model reliability metric is one of the simplest metrics for dynamic system applications, and it is very easy to understand and interpret.

Bayesian confidence metric The Bayesian interval hypothesis testing method has been demonstrated to provide more consistent model validation results than a point hypothesis testing method [18]. Jiang and Mahadevan [19] derived a generalized explicit expression to calculate the Bayes factor based on interval-based hypothesis testing for multivariate model validation. Similar to the model reliability metric, the difference curve between the test and CAE curves is selected as the validation feature. After a prior density function is assumed, using Bayes’ theorem and assumptions given in [20] and [21], the Bayes factor for the multivariate case is equivalent to the volume ratio of the posterior density of testing data under null and alternative hypotheses.

The Bayesian measure of evidence that the computer model is valid may be quantified by the posterior probability of the null hypothesis. Using the Bayes theorem, the confidence in the model based on the validation data can be obtained. Note that expert's opinion of the model accuracy may be incorporated in the confidence quantification in term of a prior ditribution. The decision maker or model user has to decide what threshold is acceptable.

Metric evaluation

Time history signals of forces, moments, accelerations, deflections, and angles are the most common types of signals obtained in vehicle safety applications. Various pairs of signals of those physical responses were used to analyze the pre-selected metrics in detail. The metrics must differentiate between different levels of correlation. Furthermore, they should use the whole domain of the rating scale, usually between "0" and "1". The assessment of the metrics was based on SMEs' experiences.

Selection of the most appropriate metrics

The Bayesian confidence metric and the model reliability metric can provide overall scores on whole time history curves, but they cannot identify key features like phase, magnitude, and slope.

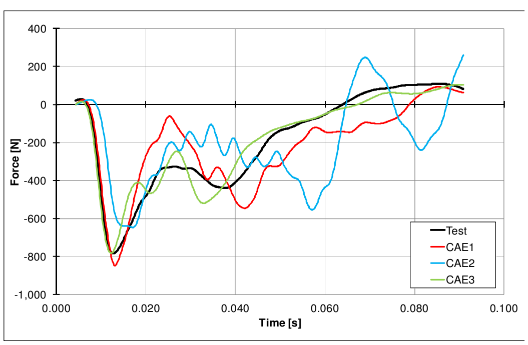

Figure 1. Force obtained in test and simulations.

The limitations of the CORA cross correlation and Bayesian confidence rating are shown in Table 1. The signals shown in Figure 1 are assessed by using the CORA and Bayesian confidence metrics. The signals are defined and evaluated in the plotted time domain.

The CORA cross correlation rating cannot differentiate between the three CAE signals because it requires signals that are defined before and beyond the interval of evaluation to calculate reasonable results.

Bayesian confidence metric grades the responses more dramatically and extremely. This is because the Bayesian hypothesis testing examines the mean of the difference distribution instead of the full difference distribution and the standard deviation of the mean of the difference is much smaller than the standard deviation of the difference. Therefore, it is more likely to give "1" score when the mean of the difference distribution is within the threshold interval, and give "0" score when the mean of the difference distribution is outside of the threshold interval.

Table 1.
Different metric ratings of force curves

	CAE1	CAE2	CAE3
CORA Total rating	0.452	0.371	0.577
CORA Corridor	0.654	0.491	0.903
CORA Cross correlation	0.250	0.250	0.250
Bayesian confidence rating	1	1	1

Since EEARTH analyzes the same characteristics of signals as the CORA cross correlation metric, but without this specific limitation, it was chosen as part of the proposed ISO metric.

Finally, two unique metrics, the CORA corridor metric and EEARTH were chosen for the proposed ISO metric.

PROPOSED ISO METRIC

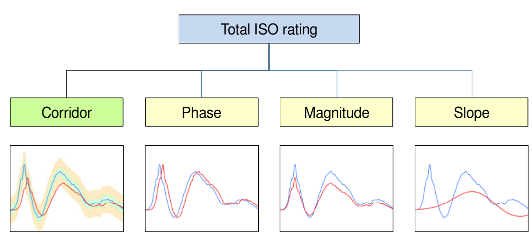

Figure 2. Structure of the proposed ISO metric.

The approach to the proposed ISO metric is to combine different types of algorithms to provide reliable and robust assessments of the correlation of two signals. The calculated score must represent a reasonable assessment for poor and for good correlations. As mentioned above, the

CORA corridor method and EEARTH are chosen. The new metric has been fully validated using responses from multiple vehicle passive safety applications.

Table 2.
Weighting factors of the sub-scores

Sub-metric	Weighting factor
Corridor	0.4
Phase	0.2
Magnitude	0.2
Slope	0.2

Figure 2 shows the structure of the proposed ISO metric. While the corridor method calculates the deviation between curves with the help of automatically generated corridors of constant width, the EEARTH method analyzes specific curve characteristics such as phase shift, magnitude, and shape. Hence, the proposed ISO metric has the advantage to compensate for the limitation of one algorithm by the other.

The total score of the proposed ISO rating metric adds up the four individually weighted sub-scores. The four weighting factors are shown in Table 2.

Corridor score

The corridor sub-metric calculates the deviation between two signals by means of corridor fitting. The two sets of corridors of constant width, the inner and the outer corridors, are defined along the test curve (reference). If the evaluated CAE curve is within the inner corridor bounds, a score of "1" is given and if it is outside the outer corridors, the score is set to "0". The assessment declines from "1" to "0" between the bounds of inner and outer corridors resulting in three different rating zones as shown in Figure 3. This transition is set to be quadratic for this proposed ISO metric.

The compliance with the corridors is calculated at each specific time of the whole interval of evaluation, and the final corridor score of a signal is the average of all scores at the specific times.

The absolute half width of the corridors is calculated by using the absolute maximum amplitude of the reference signal within the interval of evaluation and relative width factors of inner and outer corridors. The philosophy of the proposed ISO corridor approach is to use a narrow inner corridor and a wide outer corridor [22]. It limits the number of "1" ratings to only good correlations and gives the opportunity to distinguish between poor and fair correlations. If the outer corridor is too narrow, too many curves of a fair or moderate correlation would get the same poor rating of "0", like signals of almost no correlation with the reference.

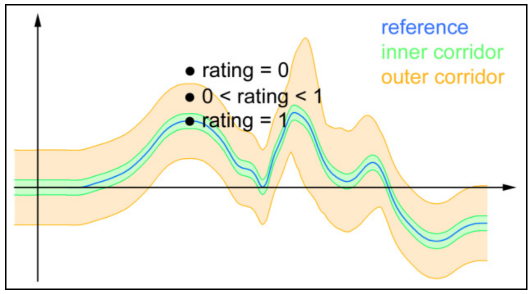

Figure 3. Rating zones of the corridor metric [9].

Phase score

The phase sub-metric is used to measure the phase lag between the two analyzed time histories. The maximum allowable percentage of time shift is pre-defined. In this step, the initial CAE curve is shifted left then right one step at a time to the original test curve, and the cross correlation between the truncated test curve and the shifted and truncated CAE curve is calculated until reaching the maximum allowable time shift limits. The best phase score is "1", which means there is no need to shift CAE curve to reach the maximum cross correlation between the initial test and CAE curves. If the time shift is equal to or greater than the maximum allowable time shift threshold, then the phase score is "0". In between, the phase score is calculated by a regression method.

Magnitude score

The magnitude sub-metric is a measure of discrepancy in the amplitude of the two time histories. It is defined as the difference in amplitude of the two time histories when there is no time lag between them. Before calculating the magnitude error, the difference between the time histories caused by error in phase is minimized by using DTW. The best magnitude score is "1", which means there is no difference in the amplitudes after phase shift and DTW. If the magnitude error is equal to or greater than the maximum allowable magnitude error threshold, then the magnitude score is "0". In between, the magnitude score is calculated by a regression method.

Slope score

The slope sub-metric is a measure of discrepancy in slope curves of the two time histories. In order to ensure that the effect of global time shift is minimized, the slope curves are calculated from the truncated test curve and the shifted and truncated CAE curve. The best slope score is "1", which means there is no difference between the two slope curves. If the slope error is equal to or greater than the maximum allowable slope error, then the slope score is "0". In between, the slope score is calculated by a regression method.

Meaning of the results

The proposed total ISO rating R ranges from "0" to "1". The higher the rating the better the correlation of the two signals. This single-rating number can be transferred to a grade that represents the goodness of the correlation by using a sliding scale (Table 3).

Table 3.
Sliding scale of the proposed total ISO rating

Grade	**Rating R**
Excellent	$R > 0.94$
Good	$0.80 < R \leq 0.94$
Fair	$0.58 < R \leq 0.80$
Poor	$R \leq 0.58$

The thresholds of R of each grade were defined based on SMEs' experiences and are only valid if none of the parameters (e.g. weighting factors, regression schemes, sampling rates, etc.) described in the proposed ISO metric ([14]) are altered.

Excellent The characteristics of the reference signal is captured almost perfectly.

Good The characteristics of the reference signal is captured reasonably well, but there are noticeable differences between both signals.

Fair The characteristics of the reference signal is basically captured, but there are significant differences between both signals.

Poor There is almost no correlation between both signals.

VALIDATION OF THE PROPOSED ISO METRIC

Similar to the evaluation of the four pre-selected metrics to be considered for an ISO standard, the validation of the proposed ISO metric was conducted with similar sets of data.

Metric parameters

The proposed ISO metric and its sub-metrics offer several parameters to adjust and validate the rating results. They were mainly used to improve the resolution of the rating domain and to improve the differentiation between signals of a similar correlation. However, all parameters are fixed in the final proposed ISO metric to gurantee comparable rating scores.

Proposed total ISO rating The weighting factors of the four sub-scores (corridor, phase, magnitude, and slope) are the only parameters to adjust the total score.

Corridor score The widths of corridors are the most important parameters to adjust this rating results. The progression of the transition between inner and outer corridors has a considerable influence on the results as well.

The type of the corridors, constant or variable width, over whole time domain may change the outcome of this sub-metric significantly.

Phase score Two parameters are used to validate the phase sub-metric: the maximum allowable percentage of time shift and the progression coefficient for the transition between "1" and "0" rating scores.

Magnitude score Similar to the phase score, two parameters are mainly influencing the results: the maximum allowable magnitude error and a progression coefficient for the transition of rating between "1" and "0".

Slope score The maximum allowable slope error and a progression coefficient are the two parameters to adjust this sub-metric.

Pre-processing of the signals

During the evaluation and validation of the proposed ISO metric, it was concluded that a few basic conditions must be kept in order to obtain correct results. This must be done by the user.

Synchronization Initially, the signals must be synchronized by physical meanings and by its timing. At each time step of the test signal, a value of the CAE signal is required.

Sampling rate The proposed ISO metric was validated with signals of 10 kHz sampling rate. The sub-metrics to evaluate magnitude and slope are especially sensitive to the signal's sampling rate.

Filtering The algorithms do not modify the original signals. It should be considered that the calculation of the correlation could be difficult when using very noisy signals.

Figure 4 shows an example of the effect of filtering. Signals A and B are derived from the same unfiltered signal and differ only by the applied filter classes. The overall correlation rating of signal B increased by 6% compared to signal A due to the application of a higher filter class.

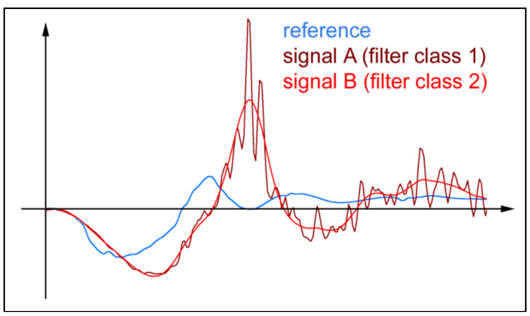

Figure 4. Differently filtered signals [9].

Interval of evaluation The assessment of the correlation should be focused on the relevant parts of the given signals. Typically, crash signals include pre-crash and post-crash phases that are usually not of interest and should be excluded from the rating. Therefore, an interval of evaluation shall be defined which describes the part of the signals that needs to be assessed. An assessment of using ratings of different sub-intervals of the same pair of signals is not allowed.

Figure 5 depicts an example of this problem. The correlation rating increases by 35% when extending the interval of evaluation from the relevant part to the whole time domain.

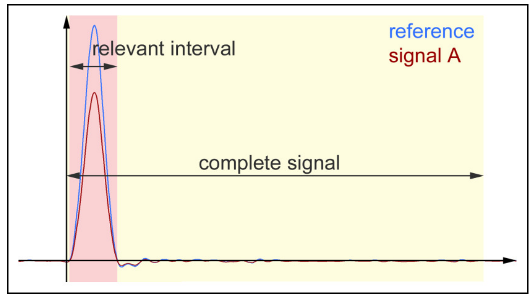

Figure 5. Different intervals of evaluation [9].

RESULTS

Four case studies that include different kinds of test and CAE signals are used to assess the potential of the proposed ISO metric. All cases are part of the mentioned ISO documents ([13], [14]) and they are defined and assessed in the time domain and fulfill all described requirements to pre-process the data.

Case 1

Figure 1 depicts a force response obtained in a test and the corresponding signals of three different CAE models. The proposed total ISO rating including the results of its sub-scores are shown in Table 4.

Table 4.
Rating of the force curves

	CAE1	CAE2	CAE3
Grade	Fair	Poor	Good
Proposed total ISO rating	0.711	0.460	0.862
Corridor score	0.654	0.492	0.904
Phase score	0.971	0.856	0.954
Magnitude score	0.738	0.372	0.929
Slope score	0.540	0.088	0.622

The rating scores reflects the different characteristics of the CAE curves. The low slope score of CAE2 correlates well with the clear shape difference of the signal compared to the test curve. The high phase score of all the CAE signals is mainly caused by the good agreement with the gradients of the signal's first peak. a The high magnitude score of CAE3 is because all peaks and valleys of the test curve are well captured. The corridor score assesses the deviation to the reference signal in the whole time domain. CAE2 resulted in the worst rating among the three CAE signals because of the clear deviation from the test. Generally, the rating differentiates between the different kinds of correlation of the three CAE signals.

Case 2

Signals of a measured torque are shown in Figure 6. The correspondig rating scores are listed in Table 5.

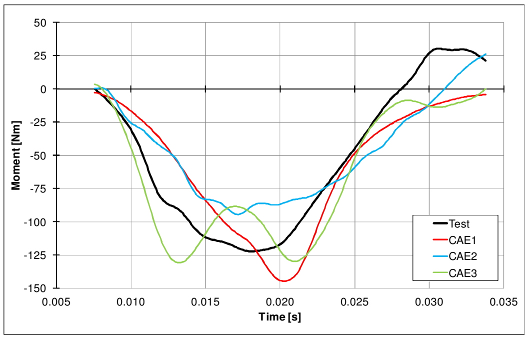

Figure 6. Moment obtained in test and simulation.

Table 5.
Rating of the moment curves

	CAE1	CAE2	CAE3
Grade	Fair	Fair	Fair
Proposed total ISO rating	0.657	0.660	0.666
Corridor score	0.539	0.538	0.556
Phase score	0.677	0.696	0.962
Magnitude score	0.840	0.798	0.735
Slope score	0.691	0.727	0.519

In spite of the different shapes of the three CAE signals, their objective rating scores are almost identical – "Fair". The twin peaks of CAE3 resulted in a low slope score. The high phase score of CAE3 is caused by the limited phase shift to reach the maximum cross correlation between test and CAE, even though the resulting maximum cross correlation number is low.

Case 3

Figure 7 shows a set of acceleration signals. The ratings are shown in Table 6.

The three CAE curves captured the gross characteristics of the test signal, but the peaks deviate. The phase and magnitude scores are the highest while the corridor scores show good correlation. The slope scores are low due to the noisy signals of the CAE curves.

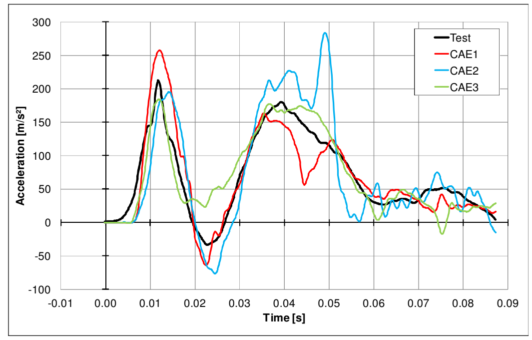

Figure 7. Acceleration obtained in test and simulation.

Table 6.
Evaluation the acceleration curves

	CAE1	CAE2	CAE3
Grade	Fair	Fair	Fair
Proposed total ISO rating	0.785	0.648	0.790
Corridor score	0.793	0.647	0.784
Phase score	0.971	0.909	0.989
Magnitude score	0.871	0.793	0.849
Slope score	0.498	0.246	0.546

This example shows that the combination of the four sub-metrics ensures reasonable ratings even if the signals are somehow difficult to handle for one of the sub-metrics.

Case 4

Figure 8 shows a set of displacement signals and Table 7 shows the corresponding rating scores.

The general characteristics of the four signals are almost identical. Therefore, the scores of phase, magnitude and slope are very high. The corridor metric does not differentiate between CAE2 and CAE3 because both signals are almost completely within the inner corridor that gives a score of "1".

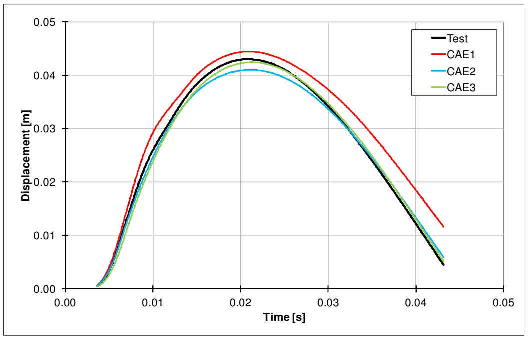

Figure 8. Displacement obtained in test and simulation.

Table 7.
Evaluation of the displacement curves

	CAE1	CAE2	CAE3
Grade	Good	Excellent	Excellent
Proposed total ISO rating	0.917	0.980	0.981
Corridor score	0.889	1.000	0.999
Phase score	0.911	0.962	0.937
Magnitude score	0.978	0.981	0.995
Slope score	0.918	0.957	0.976

CONCLUSION

This paper gives a brief overview of the capabilities of the proposed ISO metric. It is shown that the algorithms can handle different kinds of non-ambiguous signals of different qualities. The rating scheme is consistent and enables differentiated assessments of signals of different levels of correlation.

More detailed information and step-by-step procedures to implement the described metric in a software package are given in [13] and [14]. A set of ASCII curves to verify the implementation of the proposed metric is also provided with both ISO documents.

LIMITATIONS

The application of the developed metric requires some basic conditions:

(1) The metric is limited to non-ambiguous signals obtained from all kinds of tests associated with vehicle safety applications and the corresponding numerical simulations (CAE). The most commonly used signals in this field are time-history curves;

(2) The defined sliding scale to classify the proposed ISO rating score is only valid for the comparison of two signals. Any modification to the metric's parameters such as weighting factors, sampling rates, etc. requires a revision of the grade's thresholds;

(3) This proposed ISO metric is defined to calculate the level of the goodness of correlation between two signals only. If more than one pair of signals (e.g. whole set of signals from various channels of a test) is considered, the defined thresholds of the sliding scale are no longer valid.

REFERENCES

[1] Department of Energy – DOE; "Accelerated Strategic Computing Initiative (ASCI) Program Plan®"; Department of Energy; DOE/DP-99-000010592; Washington D.C.; USA; 2000.

[2] American Institute of Aeronautics and Astronautics – AIAA; "Guide for the Verification and Validation of Computational Fluid Dynamics Simulations"; American Institute of Aeronautics and Astronautics AIAA-G-077-1998; Reston; VA; USA; 1998.

[3] Department of Defense – DOD; "Verification, Validation, and Accreditation (VV&A) Recommended Practices Guide"; Department of Defense; Alexandria; VA; USA; 1996.

[4] American Society of Mechanical Engineers – ASME; "Guide for Verification and Validation in Computational Solid Mechanics"; American Society of Mechanical Engineers; ASME V&V 10; USA; 2006.

[5] American Society of Mechanical Engineers – ASME; "V&V 20-2009 Standard for Verification and Validation in Computational Fluid Dynamics and Heat Transfer"; available from ASME Codes & Standards website (June 28th 2012).

[6] Oberkampf, W. L.; "Overview of verification, validation and predictive capability"; Sandia National Laboratories; SAND2005-1824P; Albuquerque; NM; USA; 2005.

[7] Ferson, S., Oberkampf, W. L., Ginzburg, L.; "Model validation and predictive capability for the thermal challenge problem"; Computer Methods in Applied Mechanics and Engineering; 197 (29-32), pp. 2408-2430; 2008.

[8] Fu, Y., Zhan, Z., Yang, R.-J.; "A study of model validation method for dynamic systems"; SAE 101DM-0080; Detroit; MI; USA; 2010.

[9] Gehre, C., Gades, H., Wernicke, P.; "Objective Rating of Signals Using Test and Simulation Responses"; Paper Number 09-0407; 21st ESV Conference; Stuttgart; Germany; 2009.

[10] Zhan, Z., Fu, Y., Yang, R.-J.; "Enhanced Error Assessment of Response Time Histories (EEARTH) Metric and Calibration Process"; SAE 2011 World Congress; SAE 2011-01-0245; Detroit; MI; USA; 2011.

[11] Zhan, Z., Fu, Y., Yang, R.-J., Peng, Y.; "Development and Application of a Reliability-Based Multivariate Model Validation Method"; International Journal of Vehicle Design; 60 (3/4), pp. 194-205; 2012.

[12] Zhan, Z., Fu, Y., Yang, R.-J., Peng, Y.; "An Enhanced Bayesian Based Model Validation Method for Dynamic Systems"; ASME Journal of Mechanical Design; Vol.133, Issue 4, 041005; USA; 2011.

[13] International Organization for Standardization (ISO); "Road vehicles – Objective rating metrics for dynamic systems"; ISO/PDTR 16250; 2012.

[14] International Organization for Standardization (ISO); "Road vehicles – Objective rating metric for non-ambiguous signals"; ISO/TS 18571; draft; 2013.

[15] Sarin, H., Kokkolaras, M., Hulbert, G., Papalambros, P., Barbat, S., Yang, R.-J.; "Comparing Time Histories for Validation of Simulation Models: Error Measures and Metrics"; Transactions of the ASME – Journal of Dynamic Systems: Measurement and Control; Vol. 132, No. 6, 061401; 2010.

[16] Rabiner, L. R., Huang, B. H.; "Fundamentals of Speech Recognition"; Prentice Hall; 1993.

[17] Lei, H., Govindaraju, V.; "Synchronization of batch trajectory based on multi-scale dynamic Time Warping"; Proceedings of the Second International Conference on Machine Learning and Cybernetics; Xi'an; 2-5 November 2003.

[18] Jiang, X., Mahadevan, S.; "Bayesian risk-based decision method for model validation under uncertainty"; Reliability Engineering and System Safety; 92 (6), pp. 707–718; 2007.

[19] Jiang, X., Mahadevan, S.; "Bayesian wavelet method for multivariate model assessment of dynamical systems"; Journal of Sound and Vibration; 312 (4-5), pp. 694-712; 2008.

[20] Kass, R., Raftery, A.; "Bayes factors"; Journal of the American Statistical Association; 90 (430), pp. 773–795; USA; 1995.

[21] Migon, H. S., Gamerman, D.; "Statistical Inference-An Integrated Approach"; Arnold, a Member of the Holder Headline Group; London; UK; 1999.

[22] Kim, S.-A.; "Bewertung der Korrelation von Meßsignalen – Sensitivitätsanalyse der Steuerparameter der Bewertungssoftware CORA"; Master's thesis; TU Berlin; Germany; 2011.