Development of a Software for the Comparison of Curves During the Verification and Validation of Numerical Models

Mario Mongiardini*, Malcolm H. Ray*, Marco Anghileri**

*Worcester Polytechnic Institute,

Worcester, USA

**Politecnico di Milano,

Milan, Italy

(e-mail corresponding author: mario@wpi.edu)

Summary:

This paper describes the development of the Roadside Safety Verification and Validation Program (RSVVP), a software that automatically assesses the similarities and differences between two curves. This program was developed to assist engineers and analysts in performing curve comparison during the verification and validation process of a numerical model. RSVVP was designed to automatically preprocess the two input curves to make them comparable. Also, in order to ensure the most accurate comparison as possible, several options are available for the pre-processing of the input curves before the comparison metrics are computed. Data can be filtered and synchronized or any shift/drift effect can be removed. Once the signals have been pre-processed, the user can select to compute the values of one or more of the available sixteen different shape-comparison metrics. Any operation, from the input of the curves and selection of the pre-processing options till the final visualization of the results is accessible through an easy and intuitive graphical user interface. The numerical results are automatically saved by the program into a convenient spreadsheet format and the graphs are saved as bitmap images for any further investigation. Simple examples using an analytical shape are presented to illustrate the characteristics of the metrics. Also, the comparison of the acceleration time histories of a full-scale test involving a small car and the corresponding Ls-Dyna simulation is presented as an example of application of the metrics in the validation process of a numerical model.

Keywords:

Verification and Validation, Comparison Metrics, Finite Element models, Full-scale Crash Tests.

1 INTRODUCTION

Comparing the correspondence between curves from physical experiments and mathematical models is a very important and common technique used by scientists and engineers to determine if the mathematical models adequately represent physical phenomena. In the verification or validation of computational models an experimental and a numerical curve are compared in order to assess how well the numerical model predicts a physical phenomenon. Traditionally curves have been visually compared by matching peaks, oscillations, common shapes, etc. Although this kind of comparison gives an impression of how similar two curves are, it is based on a purely subjective judgment which could vary from one analyst to another. Validation and verification decisions need to be based as much as possible on quantitative criteria that are unambiguous and mathematically precise. In order to minimize the subjectivity, it is necessary to define objective comparison criteria based on computable measures. Comparison metrics, which are mathematical measures that quantify the level of agreement between simulation and experimental outcomes, can accomplish this goal.

Several comparison metrics have been developed in engineering. Metrics can be grouped into two main categories [*1*]: (i) deterministic metrics and (ii) stochastic metrics. Deterministic metrics do not specifically address the probabilistic variation of either experiments or calculation (i.e., the calculation results are assumed to be the same every time given the same input), while stochastic metrics involve computing the likely variation in both the simulation and the experiment response due to parameter variations. Deterministic metrics found in literature can be further classified into two main types: (a) domain-specific metrics and (b) shape comparison metrics. The domain-specific metrics are quantities specific to a particular application. For example, the axial crush of a railroad car in a standard crash test might be a useful metric in railroad safety but has no relevance in other applications. On the other hand, shape comparison metrics involve a comparison of two curves; one curve from a numerical simulation and another from a physical experiment. The curves may be time histories, force-deflection plots, stress-strain plots, etc. Shape comparison metrics assess the degree of similarity between any two curves in general and, therefore, do not depend on the particular application domain.

In roadside safety, comparisons between test and simulation results have mainly used domain-specific metrics (e.g. occupant severity indexes, changes in velocity, 10-msec average accelerations, maximum barrier deflection, etc.) [*2*]. The main advantage of this method is that the user could use the same domain-specific metrics that have already been evaluated for the experiments also to compare test and simulations results. Although the comparison of domain-specific metrics can give an idea of how close two tests or a test and a simulation are, shape-comparison metrics would be a more precise tool since they can be used to directly evaluate the basic response of the structures, like acceleration and velocity time histories. In roadside safety, domain-specific metrics are all derivative from the acceleration time histories so if the acceleration time history information is valid, any metric derived from the time history data will also assumed to be valid.

A computer program is described in this paper which automatically evaluates the most common shape-comparison metrics found in literature.[*3-13*] The program, called Roadside Safety Simulation Validation Program (RSVVP), was specifically developed to evaluate metrics used in the verification and/or validation of numerical models in roadside safety. The RSVVP code was written in Matlab [*14*].

In order to correctly evaluate the shape-comparison metrics, the program performs a series of preprocessing tasks before the actual metrics are calculated. The first part of this paper describes the preprocessing steps, the numerical implementation of the metrics and the post-processing operations. In the second part, the results obtained comparing some simple analytical curves are presented and discussed. Also, the comparison of the acceleration time histories of a full-scale test involving a small car and the corresponding Ls-Dyna simulation is presented.

2 PREPROCESSING

Since the two curves being compared may come from different sources, it is important to preprocess them in the same way to avoid any possible problem due to different separate preprocessing procedures. Some pre-processing operations like re-sampling and trimming of the two curves are essential since the curves must have the same length and be comparable point-to-point. Other preprocessing steps like filtering and sensor bias adjustments, though not strictly necessary, can play an important role in the final comparison result. For example, two identical curves that are simply shifted in time with respect to each other because the data was recorded with a different start time could produce a poor result just because of the initial offset value between them.

The RSVVP program performs the following pre-processing operations: (1) filtering, (2) re-sampling, (3) synchronizing and (4) trimming.

The following sections present a brief description of the pre-processing tasks performed by the RSVVP.

2.1 Filtering

Filtering the curves is usually necessary. For example, in the case of crash tests, the accelerations collected are characterized by high-frequency noise which has to be removed before calculating the comparison metrics by filtering the curves. The user can chose between the most common values for the Channel Frequency Class (CFC) or even define custom filter specifications if necessary. The filter function is a digital 4-pole Butterworth low-pass filter. The algorithm uses a double-pass filtering option: data are filtered twice, once forward and once backward using the difference equation in the time domain proposed by the SAE J211 specifications [15].

2.2 Re-Sampling

Since most shape-comparison metrics are based on point-to-point comparisons (i.e., the data at each sampling point is compared to the corresponding point in the other curve) the two curves must have the same sampling rate. After the data have been filtered, RSVVP checks the two sets of data to determine if they have been sampled at the same rate (within a fixed tolerance equal to 5E-6). If the curves do not have the same sampling rate, RSVVP proceeds to resample the curve which has the lower sampling rate (i.e., the bigger difference in time between two contiguous data points) at the higher rate of the other curve. The re-sampling is performed by means of a simple linear interpolation.

2.3 Synchronizing

Usually the time history curves to be compared do not start at the same time and, hence, the two curves are shifted by a fixed value along the abscissa direction. As the comparison metrics are generally point-to-point comparisons, the time shift between the two curves must be identified and corrected to ensure that corresponding points are matched during the metric evaluation.

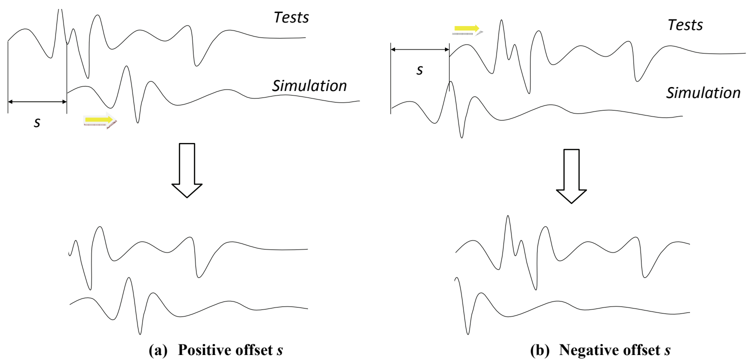


FIGURE 1: Sketch of the behavior of the shift subroutine for a (a) positive or (b) negative input s

Two different methods of synchronizing are available in RSVVP: (1) minimum area between the curves or (2) least square error method. A '*shift*' function shifts either one of the two curves by a value *s*, with a positive value of *s* meaning a forward shift for the test curve, while a negative value is equivalent to a backward shift for the simulation curve (FIGURE 1). RSVVP identifies the shift value which minimizes either the absolute area of residuals (method 1) or the sum of squared residuals (method 2). The shift value corresponding to the minimum error is the most probable matching point between the curves. In case the result is not satisfactory, the user can repeat the synchronization procedure using a different initial shift value for the minimization algorithm or using the other minimization method.

2.4 Trimming

After the two curves have been re-sampled, filtered and synchronized, the program checks that they have the same length and, in case, the longer curve is trimmed to the same size of the shorter curve. At the conclusion of these preprocessing steps, the shape-comparison metrics can be calculated.

3 METRICS

A brief description of the shape-comparison metrics used in RSVVP is presented in this section. All sixteen metrics considered in this paper are deterministic shape-comparison metrics. Details about the mathematical formulation of each metric can be found in the Appendix and in the cited literature. Conceptually, the metrics evaluated can be classified into three main categories: (i) magnitude-phase-comprehensive (MPC) metrics, (ii) single-value metrics and (iii) analysis of variance (ANOVA) metrics.

3.1 MPC Metrics

MPC metrics treat the curve magnitude and phase separately using two different metrics (i.e., M and P, respectively). The M and P metrics are then combined into a single value comprehensive metric, C. The following MPC metrics are included in RSVVP: (a) Geers (original formulation and two variants), (b) Russell and (c) Knowles and Gear [*3-7*]. The mathematical definition of each metric is shown in TABLE A1 in the Appendix. In this and the following sections, the terms m_i and c_i refer to the measured and computed quantities respectively with the "i" subscribe indicating a specific instant in time. This symbology assumes that the measured data points (i.e., m_i) are the "true" data and the computed data points (i.e., c_i) are the data points being tested in the comparison.

In all MPC metrics, the phase component (P) should be insensitive to magnitude differences but sensitive to differences in phasing or timing between the two time histories. Similarly, the magnitude component (M) should be sensitive to differences in magnitude but relatively insensitive to differences in phase. These characteristics of MPC metrics allow the analyst to identify the aspects of the curves that do not agree. For each component zero indicates that the two curves are identical. The different variations of the MPC metrics are primarily distinguished in the way the phase metric is computed, how it is scaled with respect to the magnitude metrics and how it deals with synchronizing the phase. In particular, the Sprague and Geers metric uses the same phase component as the Russell metric [*5*, *6*]. Also, the magnitude component of the Russell metric is peculiar as it is based on a base-10 logarithm and it is the only MPC metrics among those considered in this paper to be symmetric (i.e., the order of the two curves is irrelevant). The Knowles and Gear metric is the most recent variation of MPC-type metrics [*7,8*]. Unlike the previously discussed MPC metrics, it is based on a point-to-point comparison. In fact, this metric requires that the two compared curves are first synchronized in time based on the so called Time of Arrival (TOA), which represents the time at which a curve reaches a certain percentage of the peak value. In this work the percentage of the peak value used to evaluate the TOA was 5 percent, which is the typical value found in literature. Once the curves have been synchronized using the TOA, it is possible to evaluate the magnitude metric. Also, in order to avoid creating a gap between time histories characterized by a large magnitude and those characterized by a smaller one, the magnitude component M has to be normalized using the normalization factor QS.

3.2 Single-value Metrics

Single-value metrics give a single numerical value that represents the agreement between the two curves. Eight single-value metrics have been implemented in RSVVP: (a) the correlation coefficient metric, (b) the NARD correlation coefficient metric (NARD), (c) Zilliacus error metric, (d) RSS error metric, (e) Theil's inequality metric, (f) Whang's inequality metric and (g) the regression coefficient metric [9-13]. The first two metrics are based on integral comparisons while the others are point-to-point comparisons. The definition of each metric is shown in TABLE A2 in the Appendix.

3.3 ANOVA Metric

ANOVA metrics are based on the assumption that if two curves represent the same event, then any differences between the curves must be attributable only to random experimental noise. The analysis of variance (i.e., ANOVA) is a standard statistical test that can be used to assess whether the residuals between two curves can be attributed to random error [*16,17*]. When two time histories represent the same physical event, they should be identical such that the mean and the standard deviation of the residual errors are both zero. Of course, this is never the case in practical situations (e.g., experimental errors cause small variations between tested responses even in identical tests). The conventional t statistics provides an effective method for testing the assumption that the observed residual errors are close enough to zero to represent only random errors. Both Oberkampf and Ray independently proposed similar methods. In Ray's version of the ANOVA, the residual error and its standard deviation are normalized with respect to the peak value of the true curve. Using this method to compare six repeated frontal full-scale crash tests Ray proposed the following acceptance criteria [*16*]:

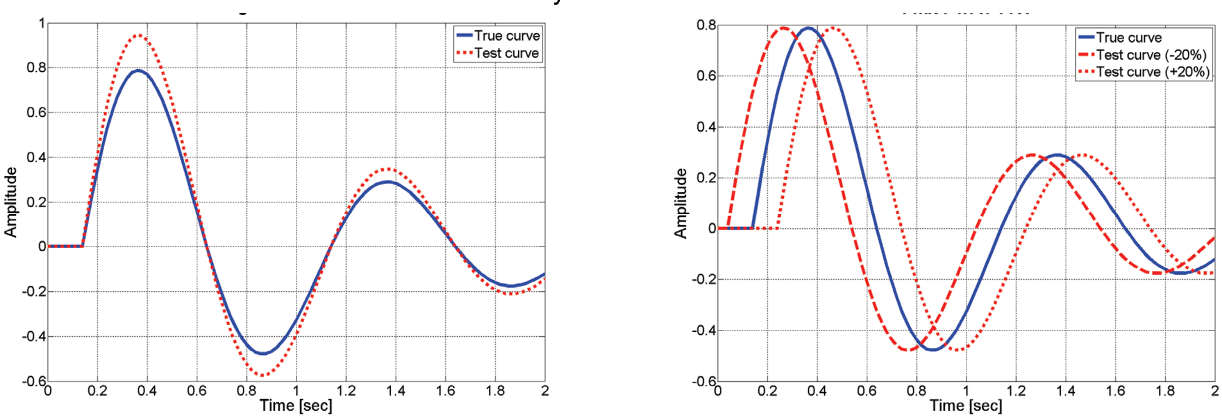
- The average residual error normalized by the peak response (i.e., $\bar{e}^{r}$) should be less than five percent.
- The standard deviation of the normalized residuals (i.e., σ^{r}) should be less than 20 percent.
- The t-test on the distribution of the normalized residuals should not reject the null hypothesis that the mean value of the residuals is null for a paired two-tail t-test at the five-percent level, $t_{0.005,\infty}$ (i.e., 90th percentile).

$$T = \frac{\sqrt{n}\,\bar{e}^{r}}{\sigma^{r}} \tag{1}$$

Where *n* is the number of samples.

4 APPLICATION TO SIMPLE ANALYTICAL CURVES

RSVVP was used to compare pairs of ideal analytical curves differing only in magnitude or phase as described in a previous work by Schwer [7]. These examples will provide some insight into the features of the different metrics calculated by RSVVP.



(a) Magnitude test **(b) Phase test**

FIGURE 2: Analytical wave forms created for a (a) the magnitude test or (b) the phase test

The baseline analytical curve used as a reference in both the magnitude and phase comparisons is referred as the "True" curve, while the curves differing respectively in phase or magnitude are referred to as the "Test" curves (FIGURE 2). Following Schwer's work, two different tests were performed: (a) a curve with the same phase but an amplitude 20 percent greater than the true curve and (b) a curve with the same magnitude but out of phase by +/- 20 percent with respect to the true curve. The analytical forms used for the magnitude-error test are shown in TABLE 1. In all cases, the sampling period was 0.02 sec, the start time was zero and the ending time was 2 sec.

TABLE 1: Analytical curves used for the magnitude and phase shift tests.

Magnitude Test	Phase Test (-20%)	Phase Test (+20%)
$\begin{cases} m(t) = e^{-(t-0.14)}\sin 2\pi(t-0.14) \\ c(t) = 1.2\cdot e^{-(t-0.14)}\sin 2\pi(t-0.14) \end{cases}$	$\begin{cases} m(t) = e^{-(t-0.14)}\sin 2\pi(t-0.14) \\ c(t) = e^{-(t-0.04)}\sin 2\pi(t-0.04) \end{cases}$	$\begin{cases} m(t) = e^{-(t-0.14)}\sin 2\pi(t-0.14) \\ c(t) = e^{-(t-0.24)}\sin 2\pi(t-0.24) \end{cases}$

4.1 MPC Metric Results

The curves used for the magnitude and phase tests are shown in FIGURE 2 and the values for the sixteen shape-comparison metrics evaluated using RSVVP are listed in TABLE 2. The M component of the MPC metrics is supposed to be insensitive to phase changes and sensitive to magnitude changes only. This is confirmed by the metric values in TABLE 2: the Geers, Geers CSA, Sprague-Geers and Knowles-Gear M components are all 20 percent and the P components are zero. Also, in all cases except for the Russell metric, the M component can be considered to be an estimate of the percent difference in the magnitude. Similarly, the phase test of this simple analytical shape confirmed that the P component of the MPC metrics is insensitive to magnitude and sensitive to phase shift. In this case, the P component of all the metrics except the Knowles-Gear metric result in scores of around 20 percent, so it can be interpreted as the percent of phase error as well. Only for the Knowles-Gear metric the P value is 62.5. This indicates that magnitude and phase scores represent different levels of error for the Knowles-Gear metrics: a 20 percent magnitude shift results in an M of 20 and a 20 percent phase shift results in a P value of 62.5. Note that the phase component for all the MPC metrics are exactly the same regardless of the direction of the phase shift, therefore providing only information about the amount of the phase shift. TABLE 2 also shows that there is very little difference between the values of each of the MPC metrics, particularly the Geers, Geers CSA and Sprague-Geers.

Lastly, the C component of the MPC metrics is simply the combination of the M and P components, obtained by taking the square root of the sum of the squares of M and P. As the Geers, Geers CSA, Sprague-Geers and Russell metrics all produce similar results there is no reason to use more than one of them. Also, as metrics that scale magnitude and phase similarly are easier to interpret, the Knowles-Geer metric is not preferred. One of the advantages of the Knowles-Gear metric is that it is formulated to account for unsynchronized signals, but if a synchronization process is used prior to making the comparison calculations, there is no need to use the time-of-arrival technique in the Knowles-Gear metric. Likewise, metrics where the score directly represents the magnitude or phase shift are easier to interpret so the Russell metric is not preferred. The Sprague-Geer MPC metric has gained some popularity in other areas of computational mechanics so it is the MPC metric recommended for roadside safety computational mechanics.

4.2 Single-Value Metrics Results

The single value metrics are listed in the middle portion of TABLE 2. The correlation coefficient, NARD correlation coefficient and regression coefficient result in a score of unity when the two curves are identical. For the magnitude test, the regression coefficient is 97.9 and both forms of the correlation coefficient are 100. Correlation suggests that two curves can be linearly transformed into each other not that they are identical curves. Two straight lines with different slopes, for example, have a 100 percent correlation. The magnitude test results show that the two correlation coefficients and the regression coefficients are not sensitive to changes in magnitude since all the three result in either perfect or nearly perfect scores. The results are similar though not as good in the phase tests. The three correlation-type single value metrics result in values between 78.9 and 81.8 indicating fairly high correlation. If the score in the phase test for these three metrics is subtracted from 100, a value near 20 is obtained indicating that these metrics are fairly direct measures of phase shift. The correlation-type metrics appear to be insensitive to magnitude shifts and directly sensitive to the amount of phase shift. It should also be pointed out that the NARD version of the correlation coefficient is identical to one minus the P component of the Geers and Geers CSA metrics and also closely related to the Sprague-Geers P component. Since the phase information detected by the correlation, NARD correlation and regression coefficients is captured equally well in the P component of the Sprague-Geers metrics, there is no reason to routinely calculate these metrics in roadside safety verification and validation activities.

The RMS is the root-mean squared error, another standard mathematical technique for comparing curves. The RMS for the magnitude test as shown in TABLE 2 is 20, the amount of the magnitude shift. The RMS score for the phase shift, however, is about 60, much greater than the 20 percent phase shift. While the RMS yields the percent shift in the magnitude test, the fact that it yields a large value in the phase test limits the diagnostic utility since for a general shape comparison it would not be clear if the difference is due to an error in magnitude or phase. The Zilliacus error metric shares a similar formulation to the RMS and results in similar values. Neither the RMS nor Zilliacus Error Factor are preferred in roadside safety verification and validation activities.

As shown in Appendix, Wang's and Theil's inequalities are very similar formulations (i.e., one using a square root of a square and the other the absolute value). In the magnitude test both yield values of 9.1 and in the phase tests values of just over 30. The two different formulations, therefore, generally will produce very similar results so there is no need to use both. Both inequalities are essentially measures of the point-to-point error between the signals as shown in their formulations in Appendix. As will be shown in the next section, the average residual error component of the ANOVA metric is essentially the same as both Wang's and Thiel's error metrics. Since these metrics are redundant with each other and the average residual error, they are not preferred for roadside safety verification and validation comparisons.

The weighted integrated factor (WiFac) value for the magnitude test was 16.7 and 48.8 for the phase test. The diagnostic value of the WiFac is not apparent to the authors so this metric is also not recommended.

TABLE 2: Comparison metrics for the analytical curves for (1) the magnitude test and (2) the phase test.

RSVVP Metric Results	Magnitude +20%	Phase -20%	Phase +20%
MPC Metrics			
Geers Magnitude	20	0.1	-0.5
Geers Phase	0	18.2	18.2
Geers Comprehensive	20	18.2	18.2
Geers CSA Magnitude	20	0.1	-0.5
Geers CSA Phase	0	18.2	18.2
Geers CSA Comprehensive	20	18.2	18.2
Sprague-Geers Magnitude	20	0.1	-0.5
Sprague-Geers Phase	0	19.5	19.5
Sprague-Geers Comprehensive	20	19.5	19.5
Russell Magnitude	13.6	0.1	-0.4
Russell Phase	0	19.5	19.5
Russell Comprehensive	12	17.3	17.3
Knowles-Gear Magnitude	20	0.0	0.0
Knowles-Geer Phase	0	62.5	62.5
Knowles-Geer Comprehensive	18.3	25.5	25.5
Single Value Metrics			
Wang's Inequality	9.1	30.7	30.6
Theil's Inequality	9.1	30.2	30.2
Zilliacus Error Metric	20	61.8	60.4
RMS Error Metric	20	60.5	60.3
WiFAC	16.7	48.8	48.8
Regression Coefficient	97.9	78.9	79.1
Correlation Coefficient	100	81.0	80.9
NARD Correlation Coefficient	100	81.7	81.8
ANOVA Metrics			
Average Residual Error	0.02	0	0
Standard Deviation of Residuals	0.09	0.26	0.26
T Score	2.08	-0.17	0.35

4.3 ANOVA Metrics Results

With the exception of Theil's and Wang's inequality factors and the Zilliacus error factor, all the metrics discussed so far are assessments of the similarity of the magnitudes or phase of the two curves being compared. The metrics proposed by Theil, Wang and Zillliacus, on the other hand are point-to-point estimate of the residual error between the two curves. Each of these methods subtracts the test from the true signal at each point in time to find the instantaneous difference between the two curves. These differences are then summed and in some fashion normalized. Both Ray and Oberkampf independently developed a more direct assessment of the residual error. Ray and Oberkampf's methods are essentially identical except Ray normalizes by the peak value of the true curve whereas Oberkampf normalized by the mean of the peaks of the test and true curves. While the other types of metrics compare the phase or magnitude of the two curves, these point-to-point error methods examine the residual error.

Ray's method has an additional advantage since it uses both the average residual error and the standard deviation of the residual error. In essence, the ANOVA method proposed by Ray examines the shape of the residual error curve resulting from a point-to-point comparison of the curves. Random experimental error by definition is normally distributed about a mean of zero and there are standard statistical tests to test the assumption that the error fits a normal distribution. The analytical shape test presented herein is not really a particularly good test of the ANOVA metric since there is no random experimental error – the differences between the curves result from the fact that the curves are in fact different though very similar analytical curves.

Nonetheless, the results of the magnitude and phase test are shown at the bottom of TABLE 2. The average residual error for both the magnitude and phase tests was near zero indicating that the average value of the error between the curves was zero. A review of the curves in FIGURE 2 shows that the curves have a symmetric oscillation above and below zero so the average distance between points on the two curves should be close to zero. The standard deviation is 0.9 in the magnitude test and 0.26 in the two phase tests.

Based on an assessment of repeated crash tests, Ray has proposed that the average residual error should be less than five percent and the standard deviation of the residual error should be less than 20 percent. By those criteria the values in TABLE 2 would indicate that the two curves could represent the same event. The third component of the ANOVA procedure is the T test which is a standard statistical test of the hypothesis that the observed error is normally distributed. For large numbers of samples, as is the case in this test, and 90 percent confidence, the critical value for the T test is 2.67. The magnitude test is close but under this critical value whereas the phase tests are well inside the acceptance range. The ANOVA test is recommended for use in roadside safety computational mechanics because it provides a direct assessment of the residual errors between the test and true curves and, thereby, provides additional useful diagnostic information about the degree of similarity or difference between the curves.

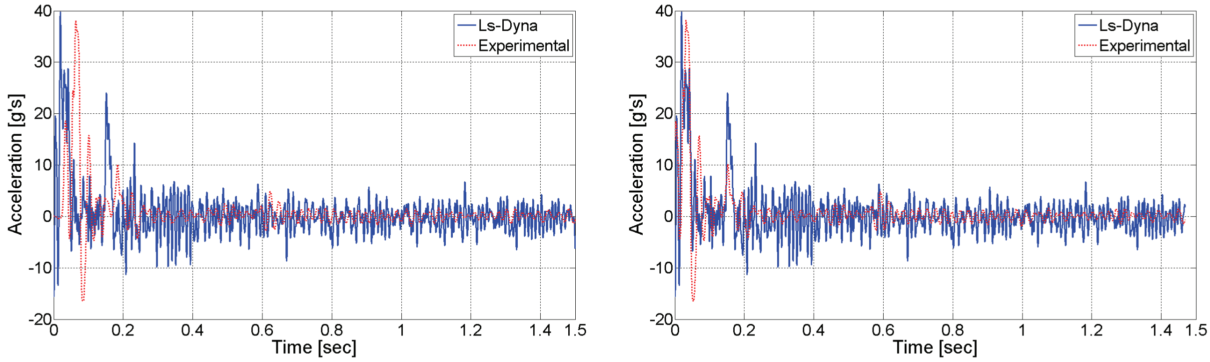
5 COMPARISON OF A FULL SCALE TEST AND AN LS-DYNA SIMULATION

The next step in this research was to use RSVVP to evaluate the comparison metrics to compare a typical roadside safety full-scale test with the corresponding numerical simulation. The lateral acceleration time history of a full-scale test with a small car impacting against a concrete rigid barrier was compared with the corresponding outcome from an Ls-Dyna simulation (FIGURE 3). As the impact was clearly limited to the first segment of the time histories, the comparison was performed on a time interval corresponding to the first 0.25 seconds of the event. In fact, RSVVP allows the user to compare the curves also on an arbitrary defined user time interval.

(a) *(b)*

FIGURE 3: Crash test with a small car, (a) Finite Element simulation and (b) and full-scale crash test.

Considering the results previously obtained in the simple case with the analytical curves, it was decided to evaluate the Sprague&Geers and the ANOVA metrics for the comparison of these time histories. In order to investigate how the synchronization of the curves before the computation of the metrics could affect the final score, the comparison metrics were evaluated using both the original non-synchronized and synchronized time histories (FIGURE 4). In both cases, the curves were first filtered using the implemented SAE CFC filtering option.



(a) non-synchronized **(b) synchronized**

FIGURE 4: Acceleration time histories: (a) non-synchronized and (b) synchronized.

From the results shown in TABLE 3 it is clear that synchronizing the curves improves the score of the comparison metrics. Indeed, this is in agreement with the general subjective judgment that the reader could probably formulate by looking at the graphs in FIGURE 4.

In either the case of the non-synchronized or synchronized curves, the Sprague-Geers metrics have been computed also using the integrals of the acceleration time histories (i.e., the velocity time histories). In both cases, the use of the velocity time histories to compute the Sprague-Geers metrics further improve the score respect to the use of the acceleration time histories. The better results obtained using the velocity time histories is not surprising as the velocity has been calculated by an integration process, which basically can be considered as a low frequency filtering. In practice, evaluating the comparison metrics with the integrals of the acceleration time histories instead of the original acceleration time histories would allow to assess the degree of match of the two curves more globally than with the original acceleration curves. In other words, using the velocity time histories the final value of the Sprague-Geers metrics seems to be less sensitive to the small localized residual errors which usually characterize the comparison of the acceleration time histories.

TABLE 3: Metrics values for comparison of the lateral acceleration time histories.

	Non-synchronized		**Synchronized**	
Sprague-Geers	Acceleration	Velocity	Acceleration	Velocity
M	*-4*	*7.1*	*-3.4*	*1*
P	*44.5*	*7.3*	*28*	*2.2*
C	*44.6*	*10.2*	*28.2*	*2.4*
ANOVA				
Mean	*0*		*0*	
Std	*0.37*		*0.24*	
t-test	*-0.46*		*-0.65*	

6 CONCLUSIONS AND DISCUSSION

This paper described the development of the RSVVP program for the evaluation of comparison metrics and provided an example of its application to both simple analytical curves and a real test/simulation curve comparison. Several pre-processing options are available for the two input curves: data can be filtered, adjusted for any bias, re-sampled to the same data acquisition frequency and synchronized to the same equivalent initial time. Pre-processing is an important step to ensure a correct comparison of two curves. For this reason, it is preferable that raw (i.e., un-preprocessed) data is used in RSVVP rather than data already processed for a crash test.

RSVVP includes sixteen separate metrics that assess the comparison between the test and true curves. The formulation of these metrics is summarized in Appendix and full details are available in the literature. A test case using a simple analytical function was performed and the results for the 16 metrics were shown in TABLE 2: Comparison metrics for the analytical curves for (1) the magnitude test and (2) the phase test.. A review of the results and formulations of these metrics show that there are really just three basic features of a shape comparison that are assessed: similarities in magnitude, similarities in phase and the shape of the residual error curve. Since many of the metrics share similar formulations, their results are often identical or very similar and there is no reason to include all the variations. The Sprague-Geers MPC metrics are recommended to assess the similarity of magnitude (i.e., the M metric) and phase (i.e., the P metric) and the ANOVA metric is recommended to assess the characteristics of the residual errors. In particular, for the Sprague-Geers metrics the use of the velocity time histories showed to give a more reliable and global assessment of the comparison.

The application of the RSVVP to compare the two time histories respectively from an experimental test and a numerical simulation showed the reliability of the program and the implemented metrics for a future use in verification and validation procedure in roadside safety. The RSVVP program will provide a convenient platform for engineers to explore the similarities and differences between both physical test and computational results in validation efforts as well as comparing the repeatability of physical experiments. The program provides all the tools needed to quickly perform the assessments between two curves.

7 ACKNOWLEDGMENTS

The authors are grateful to Dr. Leonard Schwer, Mr. David Moorcroft and other members of the ASME PTC-60 committee for providing helpful information about verification and validation metrics. Also, the authors are thankful to Dr. Chiara Silvestri for her help in finalizing and revising the manuscript and Dr. Chuck Plaxico of Battelle for his review and testing of early versions of the software. This work was made possible through the support of the National Cooperative Highway Research Program as a part of Project NCHRP 22-24.

8 REFERENCES

1 *Recommended Procedures for Verification and Validation of Computer Simulations used for Roadside Safety Applications*. Interim Report of the National Cooperative Highway Research Program (NCHRP) Project 22-24, Revised version October 2008. (Accessed March 1, 2009.) http://civil-ws2.wpi.edu/Documents/Roadsafe/NCHRP22-24/RevisedInterimReport.pdf

2 Ray, M. H., Plaxico, C.A., and Hiranmayee, K.: "Impact Performance of the G4(1W) and G4(2W) Guardrail Systems Comparison Under NCHRP Report 350Test 3-11 Conditions", In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1720*, Transportation Research Board of the National academies, Washington, D.C., 2000, pp. 7-18.

3 Geers, T.L: "An Objective Error Measure for the Comparison of Calculated and Measured Transient Response Histories", *The Shock and Vibration Bulletin*, The Shock and Vibration Information Center, Naval Research Laboratory, Washington, D.C., Bulletin 54, Part 2, June 1984, pp. 99-107.

4 *Comparative Shock Analysis (CSA) of Main Propulsion Unit (MPU)*. Validation and Shock Approval Plan, SEAWOLF Program, Contract No. N00024-90-C-2901, 9200/SER: 03/039, September 20, 1994.

5 Sprague, M. A., and Geers, T. L.: "Spectral elements and field separation for an acoustic fluid subject to cavitation", *Journal of Computational Physics*, Vol. 162, 2003, pp. 184:149.

6 Russell, D. M.: "Error Measures for Comparing Transient Data: Part I: Development of a Comprehensive Error Measure", In Proceedings of the *68th shock and vibration symposium*, 2006, pp. 175-184.

7 Schwer, L. E.: "Validation Metrics for Response Time Histories: Perspective and Case Studies", *Engineering with Computers*, Vol. 23, Issue 4, 2007, pp. 295-309.

8 Knowles, C. P., and Gear, C.W.: "Revised validation metric", unpublished manuscript, 16 June 2004 (revised July 2004).

9 Cohen, J., Cohen P., West, S.G., and Aiken, L.S.: "*Applied multiple regression/correlation analysis for the behavioral sciences*", Lawrence Erlbaum Associates, Hillsdale, NJ, 2003 (3rd ed.).

10 Basu, S., and Haghighi, A.: "*Numerical Analysis of Roadside Design (NARD) vol. III: Validation Procedure Manual*", Report No. FHWA-RD-88-213. FHWA, U.S. Department of Transportation, 1988.

11 Whang, B., Gilbert, W.E., and Zilliacus S.: "*Two Visually Meaningful Correlation Measures for Comparing Calculated and Measured Response Histories*", Report, CARDEROCKDIV-U-SSM-67-93/15. Carderock Division, Naval Surface Warfare Center, Survivability, Structures and Materials Directorate, Research and Development, Bethesda, Maryland, September, 1993.

12 Theil, H.: "*Economic Forecasts and Policy*". North-Holland Publishing Company, Amsterdam, 1975.

13 Russell, D.M.: "Error Measures for Comparing Transient Data: Part II: Error Measures Case Study", In Proceedings of the *68th shock and vibration symposium*, 2006, pp. 185-198.

14 "*MATLAB User Guide",* The MathWorks Inc, The MathWorks, Inc., 3 Apple Hill Drive, Natick, MA 01760-2098, 1994-2008.

15 SAE *J211-1 (R) Instrumentation for Impact Test—Part 1—Electronic Instrumentation*. SAE International, Warrendale PA, USA, Jul 1, 2007.

16 Ray, M. H: "Repeatability of Full-Scale Crash Tests and Criteria for Validating Finite Element Simulations", In *Transportation Research Record: Journal of the Transportation Research Board*, *No. 1528*, Transportation Research Board of the National academies, Washington, D.C., 1996, pp. 155-160.

17 Oberkampf, W.L., and Barone, M.F.: "Measures of Agreement Between Computation and Experiment: Validation Metrics", *Journal of Computational Physics*, Vol. 217, No. 1 (Special issue: Uncertainty quantification in simulation science), 2006, pp 5–36.

9 Appendix A: Analytical formulation of the comparison metrics

TABLE A1: Definition of MPC metrics

	Magnitude	**Phase**	**Comprehensive**
	Integral comparison metrics		
Geers	$M_G = \sqrt{\frac{\sum c_i^2}{\sum m_i^2}} - 1$	$P_G = 1 - \frac{\sum c_i m_i}{\sqrt{\sum c_i^2 \sum m_i^2}}$	$\sqrt{M_G^2 + P_G^2}$
Geers CSA	$M_G = \sqrt{\frac{\sum c_i^2}{\sum m_i^2}} - 1$	$P_{CSA} = 1 - \frac{\lvert\sum c_i m_i\rvert}{\sqrt{\sum c_i^2 \sum m_i^2}}$	$sign\left(\sum c_i m_i\right)\sqrt{M_{CSA}^2 + P_{CSA}^2}$
Sprague & Geers	$M_G = \sqrt{\frac{\sum c_i^2}{\sum m_i^2}} - 1$	$P_{SG} = \frac{1}{\pi} cos^{-1} \frac{\sum c_i m_i}{\sqrt{\sum c_i^2 \sum m_i^2}}$	$\sqrt{M_{SG}^2 + P_{SG}^2}$
Russell	$M_R = sign(m) \cdot Log_{10}(1 + \lvert m\rvert)$ *where* $\frac{(\sum c_i^2 - \sum m_i^2)}{\sqrt{\sum c_i^2 \sum m_i^2}}$	$P_R = \frac{1}{\pi} cos^{-1} \frac{\sum c_i m_i}{\sqrt{\sum c_i^2 \sum m_i^2}}$	$\sqrt{\frac{\pi}{4}(M_R^2 + P_R^2)}$
	Point-to-point comparison metrics		
Knowles & Gear	$M_{KG} = \sqrt{\frac{\sum\left(\frac{\lvert m_i\rvert}{m_{max}}\right)^p (\tilde{c}_i - m_i)^2}{\sum\left(\frac{\lvert m_i\rvert}{m_{max}}\right)^p (m_i)^2}}$ *where* $\tilde{c} = c(t - \tau)$ (*with* $\tau = TOA_c - TOA_m$)	$P_{KG} = \frac{\lvert TOA_c - TOA_m\rvert}{TOA_m}$	$\sqrt{\frac{10M_{KG}^2 + 2P_{KG}^2}{12}}$

TABLE A2: Definition of single-value metrics

Integral comparison metrics			
Correlation Coefficient	$\frac{n\sum c_i m_i - \sum c_i \sum m_i}{\sqrt{n\sum c_i^2 - (\sum c_i)^2}\sqrt{n\sum m_i^2 - (\sum m_i)^2}}$	***Correlation Coefficient (NARD)***	$h\frac{\sum c_i m_i}{\sqrt{\sum c_i^2}\sqrt{\sum m_i^2}}$
Weighted Integrated Factor		$\sqrt{\frac{\sum \max(m_i^2, c_i^2) \cdot \left(1 - \frac{\max(0, m_i \cdot c_i)}{\max(m_i^2, c_i^2)}\right)^2}{\sum \max(m_i^2, c_i^2)}}$	
Point-to-point comparison metrics			
Zilliacus error	$\frac{\sum \lvert c_i - m_i \rvert}{\sum \lvert m_i \rvert}$	***RMS error***	$\frac{\sqrt{\sum (c_i - m_i)^2}}{\sqrt{\sum m_i^2}}$
Theil's inequality	$\frac{\sqrt{\sum (c_i - m_i)^2}}{\sqrt{\sum c_i^2} + \sqrt{\sum m_i^2}}$	***Whang's inequality***	$\frac{\sum \lvert c_i - m_i \rvert}{\sum \lvert c_i \rvert + \sum \lvert m_i \rvert}$
Regression coefficient		$\sqrt{1 - \frac{(n-1)\sum (c_i - m_i)^2}{n\sum (m_i - \overline{m})^2}}$	