
Towards a Comprehensive Framework for Simulation-based Vehicle Systems Design Validation

**M. Kokkolaras*, G. Hulbert and P.
Papalambros**

Department of Mechanical Engineering
University of Michigan, Ann Arbor, Michigan
E-mail: {mk,hulbert,pyp}@umich.edu

*Corresponding author

Z. Mourelatos

Department of Mechanical Engineering
Oakland University, Oakland, Michigan
E-mail: mourelat@oakland.edu

R.-J. Yang

Ford Motor Company
Dearborn, Michigan
E-mail: ryang@ford.com

M. Brudnak and D. Gorsich

U.S. Army Tank Automotive Research and Development
Engineering Center, Warren, Michigan
E-mail: {mark.brudnak;david.gorsich}@us.army.mil

Abstract: We present an overview of our most recent and ongoing research efforts to develop a comprehensive framework for simulation-based vehicle design validation. Specifically we present the three major building blocks of our framework, namely i) the investigation of existing and introduction of appropriate validation metrics for comparing the dynamic responses of vehicle systems that consist of multivariate functional data, ii) the selection and robust implementation of a Bayesian interval-based hypothesis testing technique for quantifying the confidence in simulation models used for design under uncertainty and iii) the development of a sequential design optimization methodology with calibration-based validation to address the inadequacy of current validation methods for design optimization purposes. We conclude with a discussion of the techniques being developed currently that will complete the proposed framework.

Report Documentation Page			Form Approved OMB No. 0704-0188		
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.					
1. REPORT DATE 01 MAR 2011		2. REPORT TYPE N/A		3. DATES COVERED -	
4. TITLE AND SUBTITLE Towards a Comprehensive Framework for Simulation-based Vehicle Systems Design Validation(PREPRINT)			5a. CONTRACT NUMBER W56HV-04-2-0001		
			5b. GRANT NUMBER		
			5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) M. Kakkolaras; G. Hulbert; P. Papalambros; Z. Mourelatos; R.-J. Yang; M. Brudnak; D. Gorsich			5d. PROJECT NUMBER		
			5e. TASK NUMBER		
			5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Department of Mechanical Engineering University of Michigan, Ann Arbor, MI Department of Mechanical Engineering Oakland University, Oakland, MI Ford Motor Company Dearborn, MI U.S. Army Tank Automotive Research and Development Engineering Center, Warren, MI			8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) US Army RDECOM-TARDEC 6501 E 11 Mile Rd Warren, MI 48397-5000, USA			10. SPONSOR/MONITOR'S ACRONYM(S) TACOM/TARDEC/RDECOM		
			11. SPONSOR/MONITOR'S REPORT NUMBER(S) 21553		
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release, distribution unlimited					
13. SUPPLEMENTARY NOTES Submitted for publication in a Special Issue of Int'l Journal of Vehicle Design, The original document contains color images.					
14. ABSTRACT					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT SAR	18. NUMBER OF PAGES 14	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Keywords: simulation, dynamic systems, validation metrics, probabilistic principal component analysis, Bayesian interval-based hypothesis testing, confidence quantification, sequential design optimization, calibration-based validation

Reference to this paper should be made as follows: Kokkolaras et al. (2011) ‘Towards a Comprehensive Framework for Simulation-based Vehicle Systems Design Validation’, *Int. J. Vehicle Design*, Vol. x, No. x, pp.xxx–xxx.

1 Introduction

Computer modeling and simulation are the cornerstones of product design and development in the ground vehicle industry. Computer-aided engineering (CAE) tools have improved to the extent that virtual testing may lead to significant reduction in prototype building and testing of vehicle designs. In order to make this a reality, the need exists to assess confidence in the predictive capabilities of simulation models. Therefore, validation of both experimental and simulation results is critical.

Verification, validation and accreditation are very active areas of study in industry, academia, government, and professional societies (see, e.g., [22, 2, 1, 3, 17, 4]). Particular challenges arise with data associated with dynamic systems such as vehicles, as they typically appear in the form of multivariate time histories. The ability to derive helpful conclusions based on comparisons between experimental and computational multivariate functional data depends on utilizing appropriate metrics and confidence quantification methods. Our research aims at developing a comprehensive framework for validating simulation-based vehicle designs.

The first building block of a comprehensive simulation-based validation framework consist of a set of appropriate metrics that can be used in making quantitative assessments. Oberkampf and Barone list six properties that such validation metrics should satisfy [16]. One of these properties dictates that an effective metric for measuring the discrepancy between simulation model responses represented by time histories is necessary to accomplish the first step of the validation process. In [24], we reviewed existing error measures and metrics and discussed their advantages and limitations. We then proposed a combination of measures associated with three physically meaningful error characteristics (phase, magnitude and slope), utilizing cross-correlation, L_1 norms and algorithms such as dynamic time warping to quantify the discrepancy between time histories. We then showed how these measures can be used to build regression-based validation metrics in cases where subject matter expert (SME) data are available.

Despite the usefulness of the measures proposed in [24], two challenges remained unaddressed. The first one pertains to the general desire (especially of industry practitioners) of having one single value that quantifies the confidence (or not) in the agreement between experimental and computational (simulation)

data, especially when there are more than one responses of interest (multivariate data). The measures proposed in [24] can be combined to generate such a single metric (using regression) only if SME data are available. The second issue is linked to 4 of the remaining 5 properties advocated in [16], which call for useful validation metrics to account for the uncertainties related to numerical error, experimental error, experiment post-processing and the number of experiments conducted. Because of these limitations, and for the sake of brevity, we will not go into further detail of the validation metric in this paper. Instead, we will address the aforementioned limitations of the deterministic validation metric by adopting Bayesian hypothesis-testing methods for quantifying the confidence in the agreement of multivariate experimental and computational data under uncertainty.

2 Bayesian methods for confidence quantification

Validation under uncertainty, within a probabilistic context, requires quantification of the model output in terms of a statistical distribution and then effectively comparing it with the experimental data that also follows a statistical distribution. Statistical hypothesis testing is one approach to model validation under uncertainty. Hills et al. considered the uncertainty in the experimental data and used classical hypothesis testing method for model validation [10]. Zhang et al. applied Bayesian hypothesis testing for the validation of limit state-based reliability prediction models [25]. The fundamental difference between the classical and the Bayesian approach is that the former draws confidence intervals of prediction based on the statistical data analysis, while the latter assumes the model parameters themselves are random and to follow a prior distribution, usually based on the model developer's knowledge. The prior distribution will be updated once experimental data are available to obtain posterior distributions. Rebba compared the Bayesian methods with other statistical validation metrics and approaches in terms of ease of implementation, accuracy and adequacy requirements [20]. Over the past few years, many other research groups have been exploring Bayesian statistics to develop model validation methods.

Bayarri et al. referred to the discrepancy between the model prediction and the physical test results (experimental observations) as *model bias* [6]. They believed that accounting for this bias is the central issue of validation. As computer model predictions and physical test results may not be obtainable for identical inputs, they used Gaussian-based function approximations to the computer model producing a scalar output (following the work of Sacks et al. [23]) and Bayesian statistics for modeling the bias (following the work of Kennedy and O'Hagan [14]). The key advantage of this framework is the estimation of tolerance bounds for the model prediction, with the interpretation that there is a specified chance for the observations in reality to lie within the specified range. Higdon et al. demonstrated the implementation of this framework on different applications [9]. This framework was extended to handle smooth functional data by considering time as an input [5]. Complexity increases when the computer model and the physical tests produce high dimensional output. To address this issue, Higdon et al. used principal component analysis to reduce the dimensionality of the problem [8]. Following the work of Bayarri et al. and Kennedy et al., Chen et al. used the Bayesian

methodology to improve simulation-based predictive models for engineering design [7].

The research work described so far in this section provides methods to improve the predictive capability of the computer model, whereas the research group of Mahadevan uses the Bayesian methodology to quantify the agreement between the computer model prediction and physical test results. Rebba and Mahadevan use Bayesian hypothesis testing to infer how strongly the experimental data supports the null hypothesis to accept the model as opposed to the alternative hypothesis to reject the model [21]. A quantitative measure of confidence is derived using point null hypothesis testing to handle both univariate and multivariate data. Within the context of binary hypothesis testing for model validation, two hypotheses H_o and H_a need to be tested, namely, the null hypothesis ($H_o : \mu_{\text{TEST}} - \mu_{\text{CAE}} = 0$) to accept the model and an alternative hypothesis ($H_a : \mu_{\text{TEST}} - \mu_{\text{CAE}} \neq 0$) to reject the model. Jiang and Mahadevan used point null hypothesis testing for model validation [11]. It should be remembered that a rejection of the point null hypothesis would only mean that the CAE predictions and physical test results are not exactly equal. This does not automatically render the model useless, since, there is very low probability for the two numerical quantities to be equal in practice. Rebba concluded that the interval-based hypothesis test ($H_o : |\mu_{\text{TEST}} - \mu_{\text{CAE}}| \leq \epsilon$) provides more consistent model validation result than point null hypothesis tests [20]. So, Jiang and Mahadevan formulated a Bayesian interval-based hypothesis testing method for multivariate model validation [12], which was further developed and used in the automotive industry [13].

2.1 Robust implementation of Bayesian confidence quantification method

Our comprehensive framework for simulation-based validation utilizes the Bayesian methodology proposed in [11, 12, 13]. Our efforts have focused on a robust implementation of this method by making two important adjustments: we have introduced a method for calculating the variance of the multivariate data due to the application of probabilistic principal component analysis (PPCA) and a method for computing the appropriate width of the interval in the hypothesis testing procedure that eventually determines the Bayesian factor based on which confidence is calculated. Figure 1 presents a schematic of the implemented model validation framework for dynamic systems. A brief outline of the framework is provided in the remainder of this section by means of an example; for complete details, refer to [18].

Field or test data, denoted as physical test, consists of time histories of one or more data channels. Each data channel time history is scaled (normalized), so that the maximum absolute values of the different data are similar in magnitude. This procedure avoids biasing the validation framework based upon the magnitude of the data responses. The scale factors used to scale each test data channel are used to scale each of the corresponding CAE model data sets. PPCA is applied to the normalized test data to produce a rank-ordered decomposition of the test data based upon the percent of variability in the data. The user of the method can then choose the number of principal components to retain in the test data reduction, based upon the amount of information desired to be retained in the validation process. The transformation matrix obtained by PPCA is applied to the CAE

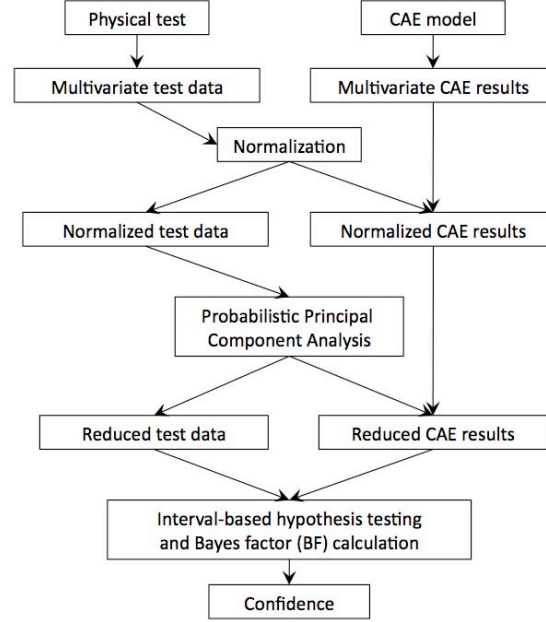


Figure 1 Bayesian confidence quantification procedure

model data to obtain a reduced set of CAE data. With both the reduced test and CAE data sets, interval-based Bayesian hypothesis testing is applied and the Bayes factor is calculated. As mentioned above, the choice of the interval width does impact the Bayes factor value. We employ a calibration procedure that equates the confidence value computed with the percent of variability captured by the included principal components from the field test data. The calibration factor that produces this equivalence is then used for the model validation assessment.

We now guide the reader through the application of the Bayesian-based confidence quantification method using an example that considers a hybrid military vehicle [19]. Figure 2 depicts a comparison of four data channels of measured field data for four responses of interest compared to normalized CAE data channels computed using a simulation model. Since we are dealing with four responses of interest, PPCA yields three principal components. During the construction of reduced data for the Bayesian-based confidence quantification, the percent of captured information as the number of used principal components increases is 62%, 86% and 99.9% for one, two and three principal components, respectively.

The confidence in the model is defined as the probability of the null hypothesis $H_o : |\boldsymbol{\mu}| \leq \boldsymbol{\epsilon}$ to be true, given the data $\mathbf{D}$ (which are the arithmetic difference between the PPCA-reduced, mean-centered test and CAE data for all responses). As is evident from the computations depicted in Figure 3, the threshold vector $\boldsymbol{\epsilon}$ has a significant impact on confidence quantification. Figure 4 illustrates the one-dimensional version of the relation between the mean of the difference between the computational and test data μ , the interval defined by ϵ , the (typically multivariate

UNCLASSIFIED

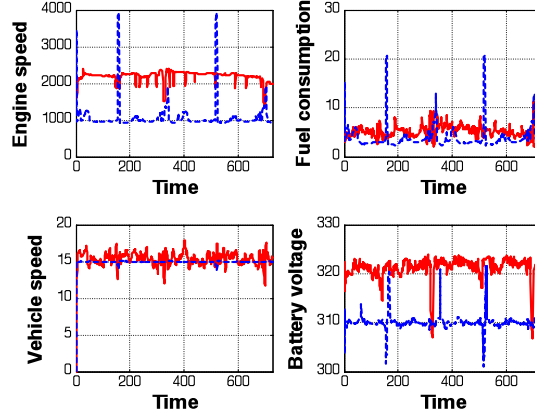


Figure 2 Comparison of field test and CAE model time histories (solid/red=test, dashed/blue=CAE)

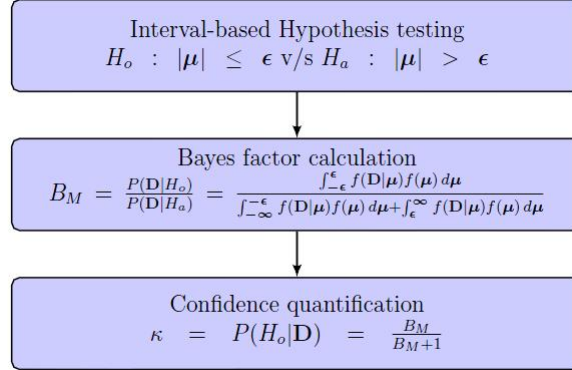


Figure 3 Interval-based hypothesis testing, Bayes factor computation and confidence quantification

normal) distribution of the available data D and the Bayes factor B_M for three different cases. A model can be deemed acceptable if $B_M \geq 1$.

Subject matter expert (SME) opinion can be used to define ϵ . In the absence of SME opinion, it becomes necessary to estimate ϵ as objectively as possible. For this purpose, we define

$$\epsilon = b\sqrt{\text{diag}(\Sigma)}, \quad (1)$$

where b is a calibration parameter and Σ is the variance of the multivariate difference between the reduced test and computational data. Figure 5 illustrates the parametric study of determining the appropriate value of the interval

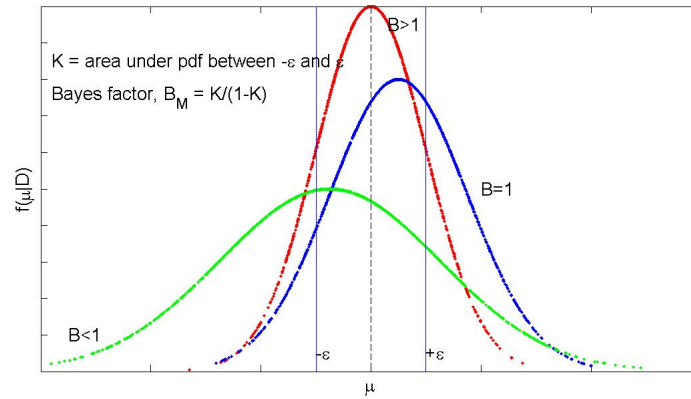


Figure 4 One-dimensional illustration of relation between interval-based hypothesis testing and Bayes factor for three different cases

calibration parameter following the procedure outlined above that aligns β with the amount of information used, i.e., principal components. As expected,

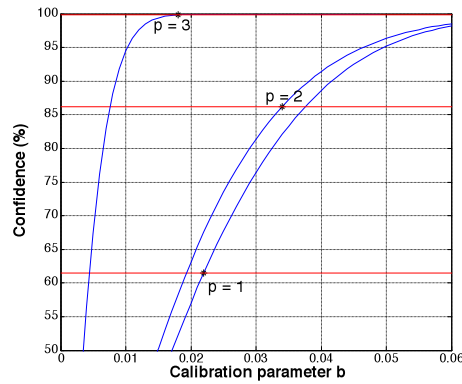


Figure 5 Choosing calibration parameter value based upon setting percent of information equal to confidence value for number of principal components

the calibration parameter value varies depending on the number of principal components employed. Using these calibration parameter values.

Figure 6 provides the confidence values computed for data from two available proving ground courses when using one, two and three principal components. The largest confidence value for the CAE models is obtained for 86% of information, that is, for two principal components. When three principal components were considered, the noise associated with the third component resulted in a reduction in model confidence. It is important to note that in all cases the confidence value is

UNCLASSIFIED

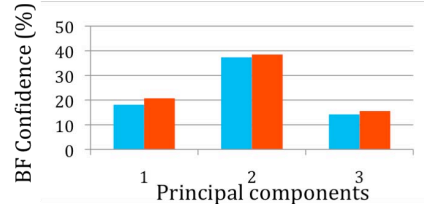


Figure 6 Confidence quantification for course 1 (left/blue) and course 2 (right/red) as a function of amount of information used (i.e., number of principal components)

below 50%, which suggests that the models should be rejected. In other words, the models need more refinement and adjustment to provide acceptable comparisons with the test data.

3 Validation and design

Computational models are not developed solely for the sake of single-point analysis. The investment in the development of simulation models aims at a return consisting of the ability to evaluate different designs under different scenarios rapidly, i.e., to conduct design optimization and robustness studies. Model validation for design optimization purposes is a daunting task. Simulation models can have a large amount of inputs. Simulation-based design optimization requires that the model is valid in a big range of both design variables (for design space exploration) and design parameters (as the variability in operating conditions, which they represent, must be addressed for robustness). A simulation-based design optimization process that uses a globally validated model, i.e., a model that has been validated for the entire range of its inputs, will always yield designs that will perform as predicted. This requires a vast amount of resources to conduct multiple tests at each point of a sufficiently large set in the extremely high-dimensional space spanned by the model inputs. Due to limited resources and time constraints, current design practice uses simulation models for optimization studies in large design spaces even though the models have been validated only at a relatively small number of points, i.e., a relatively small subset of the design space. Within this paradigm, simulation models are validated a-priori and globally, and then used, without any change or update, throughout the design optimization process.

This approach of a-priori global model validation for design optimization purposes can compromise local accuracy and do not utilize available testing resources effectively and efficiently. Numerical design optimization algorithms (both gradient-based and derivative-free) create a sequence of design iterates; validation is critical only for these iterates, and not for the entire design space. Thus, we are developing a methodology for sequential, calibration-based validation of the simulation model at, and in the vicinity of, the design candidates as they are generated during the optimization process [15]. The next section summarizes

UNCLASSIFIED

our approach by means of a contrived toy example that illustrates interesting features. The goal of such validation is confidence in the resulting design (as well as the intermediate design iterates) rather than in the global performance of the underlying simulation model. The proposed approach ensures local model accuracy as the design optimization process progresses since model calibration is conducted whenever necessary. We hypothesize that our approach utilizes available testing resources more effectively as it determines the minimum amount of tests required for validation in the targeted local domains.

3.1 Sequential design optimization with calibration-based validation

Consider a design optimization problem for which we have an initial design $\mathbf{d}$ and test data for this design point. We use the Bayesian methodology described in the previous section to quantify the confidence in the simulation model at this design point given the available test data. If this confidence C is greater or equal to a (user-specified and problem-dependent) lowest acceptable value $C_{\min}$, then we accept this design as locally valid and use it to solve the design optimization problem within a local trust region. In doing so, we assume that the CAE model is valid (i.e., $C \geq C_{\min}$) in the entire local domain. Otherwise, we update the CAE model by solving the calibration optimization problem

$$\max_{\mathbf{p}} C(\mathbf{y}_{\text{CAE}}(\mathbf{d}, \mathbf{p}), \mathbf{y}_{\text{TEST}}(\mathbf{d})), \quad (2)$$

where a set of model calibration parameters $\mathbf{p}$ is determined in order to maximize C at the center of the local trust region domain; as described in the previous sections, the confidence C depends on computational data $\mathbf{y}_{\text{CAE}}$ (which are a function of the design point $\mathbf{d}$ and the model calibration parameters $\mathbf{p}$) and the test data $\mathbf{y}_{\text{TEST}}$ (which are a function only of the design point $\mathbf{d}$). If $\max_{\mathbf{p}} C(\mathbf{y}_{\text{CAE}}(\mathbf{d}, \mathbf{p}), \mathbf{y}_{\text{TEST}}(\mathbf{d})) < C_{\min}$, where $\mathbf{p}^*$ is the optimal solution of Problem (2), we have the following options: i) terminate the design optimization procedure recognizing that we have achieved a design that is not deemed valid due to model limitations, ii) accept the model limitation and proceed acknowledging that our confidence in the simulation model (and thus the design) is below our acceptable limits, or iii) consider revisiting the model or replacing it with a higher-fidelity one.

After solving the design optimization problem, there are two cases: If the new optimal design $\mathbf{d}^*$ lies within the local trust region, then the overall design optimization problem is assumed converged to a valid design. If it lies on the boundary of the local trust region. If the optimizer found after an optimization within a local domain lies on the boundary of the local trust region, we i) set $\mathbf{d} \leftarrow \mathbf{d}^*$, ii) update the CAE model by solving Problem (2), iii) define a new local trust region centered at $\mathbf{d}$, and iv) conduct a new design optimization within the new local trust region. This sequential optimization process using local trust regions is repeated until the optimal is in the interior of the last local trust region. Note that currently, the user needs to specify the size (and shape) of the local trust region domain; we are working on a methodology that will define the size of the local trust region domain based on a minimum number of tests to be conducted at its center. Figure 7 illustrates the first iteration of the above described approach for a two-dimensional case.

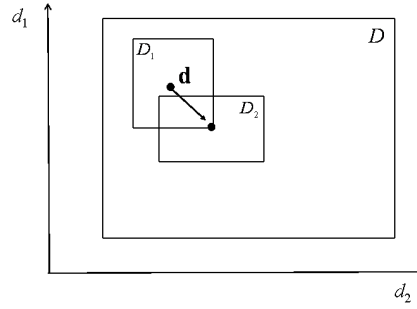


Figure 7 Schematic of sequential optimization with calibration-based validation for two-dimensional design space

To illustrate the features of the sequential design optimization approach with calibration-based validation, we contrived a toy example considering the cantilever beam depicted in Figure 8. A finite element model was developed to predict the

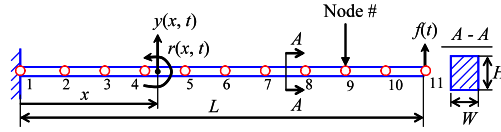


Figure 8 Cantilever beam of rectangular cross-section with tip point dynamic load $f(t)$

lateral and rotational displacements along the length of the beam. To introduce uncertainty, as well as artificial disagreement between the test and computational data, we modified the cross-section of the beam near the fixed end boundary condition, and used a torsional spring, whose stiffness was treated as a model calibration parameter, to model the fixed end condition in the finite element model (see Figure 9).

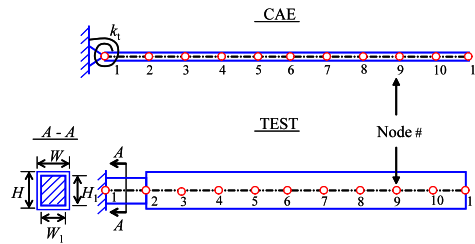


Figure 9 Cantilever beam of rectangular cross-section for test and CAE; the test uses a reduced cross-section close to the fixed end and the CAE model assumes a pinned left end with a rotational spring stiffness

UNCLASSIFIED

The design optimization is formulated to minimize the mass of the beam with respect to the cross-section dimensions subject to displacement constraints. The discrepancy between tip displacement predictions and test data as a function of time is plotted in Figure 10. It is obvious that the confidence in the simulation

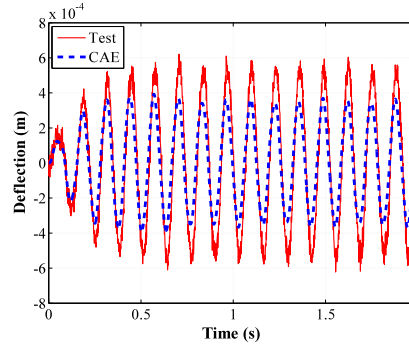


Figure 10 Comparison of tip displacement between test and CAE at initial design before calibration

model cannot be high. Indeed, the confidence in the CAE model (and thus the design) at the initial design point was found to be only 40%. After calibrating the model by solving Problem (2) confidence increased to 98%. Figure 11 depicts the confidence in the model as a function of the design variables before and after calibration at the initial design. We can see that if we were to use the model in other areas of the design spaces, model confidence can decrease dramatically. This demonstrates our previous discussion of the small value of a-priori, global model validation.

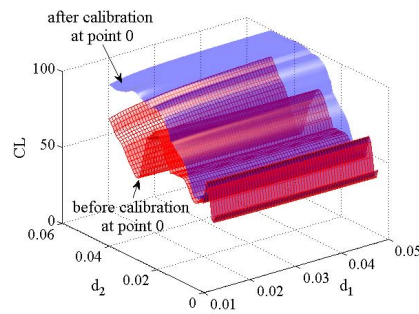


Figure 11 Confidence as a function of the design space

Figure 12 summarizes the sequential design optimization process with calibration-based validation. As already mentioned, calibration was necessary at the initial design. After that, 3 sequential optimizations were conducted within local trust regions; the optimal solution of each of these lied on the boundary of

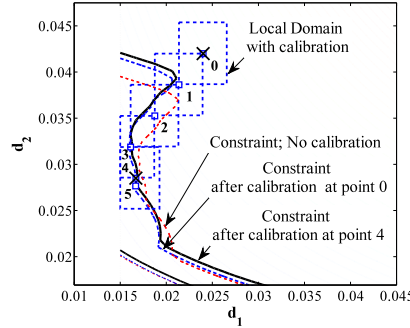


Figure 12 Intermediate and final results of the sequential design optimization process with calibration-based validation

the local trust region; the confidence value did not fall below the 80% acceptable threshold for none of these optimizations, so calibration was not necessary. At the 4th optimization, however, confidence decreased to 78%, so Problem (2) was solved, the model was re-calibrated and confidence increased to 85%. The subsequent design optimization converged to a design within the local trust region, so the overall design optimization process was terminated successfully.

4 Summary

We have presented an overview of our most recent efforts towards the development of a comprehensive validation framework for design of vehicle systems. We begun our research work with evaluating existing measures for comparing functional data (i.e., time histories) since such systems are dynamic with multiple time-dependent responses and since comparison between computational and test data is the basis of validation. After demonstrating several limitations of traditional measures, we introduced a deterministic, three-component validation metric that can be used in conjunction with subject matter expert data to build a scalar regression-based validation models.

Recognizing that the issue of uncertainty in both tests and computational models must be inevitably addressed, we proceeded to investigate the applicability of Bayesian methods for design validation. After identifying the most promising methodology, consisting of processing the available multivariate data through probabilistic principal component analysis for efficient uncertainty treatment, dimensionality reduction and significant feature (or information) extraction and then quantifying confidence based on Bayesian interval-based hypothesis testing, we developed a robust implementation by introducing a coherent variance calculation method and a technique for determining the appropriate interval width based on information content.

Lastly, we employed the implemented confidence quantification method in a novel sequential design optimization with simultaneous calibration-based validation approach. Our motivation was that typical a-priori global model validation for design optimization is only adequate when unlimited resources are

expended, which is obviously not the case due to limited resources and/or time restrictions. We have demonstrated that the use of such models can lead to both invalid and suboptimal results, and have shown that sequential optimization with calibration-based (when necessary) validation can use resources and associated information effectively.

Our current efforts focus on completing the comprehensive framework for simulation-based design validation by developing the missing techniques for determining the minimum amount of tests necessary at each stage of the sequential method and the size and shape of the associated local trust regions. Optional future work could be the development of a methodology that utilizes the data and model calibration parameter values generated during the sequential process to build models that can be used for extrapolation in domains where data cannot be obtained.

Acknowledgements

We are grateful to Mr. Sarin, Ms. Pai, Mr. Pozolo and Drs. Li, Reed, Jiang, Fu and Barbat for numerous helpful contributions and suggestions to this research.

This work has been supported partially by Ford Motor Company and by the Automotive Research Center (ARC), a U.S. Army Center of Excellence in Modeling and Simulation of Ground Vehicles led by the University of Michigan. Such support does not constitute an endorsement by the sponsors of the opinions expressed in this article.

References

- [1] Verification, validation, and Accreditation of Army Models and Simulations, 1999. Department of the Army, Pamphlet 5-11, Washington DC.
- [2] American Institute of Aeronautics and Astronautics. *Guide for the Verification and Validation of Computational Fluid Dynamics Simulations*, 1998.
- [3] American Society of Mechanical Engineers. *Council on Codes and Standards, Board of Performance Test Codes: Committee on Verification and Validation in Computational Solid Mechanics*, 2003.
- [4] American Society of Mechanical Engineers. *Guide for the Verification and Validation of Computational Solid Mechanics*, 2006.
- [5] M.J. Bayarri, J.O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R.J. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh. Computer model validation with functional output. *The Annals of Statistics*, 35(5):1874–1906, 2007.
- [6] M.J. Bayarri, J.O. Berger, R. Paulo, J. Sacks, J.A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu. A framework for validation of computer models. *Technometrics*, 49:138–154, 2007.
- [7] W. Chen, Y. Xiong, K.-L. Tsui, and S. Wang. A design-driven validation approach using bayesian prediction models. *Transactions of the ASME - Journal of Mechanical Design*, 130(2), 2008.
- [8] D. Higdon, J. Gattiker, B. Williams, and Rightley M. Computer model calibration using high dimensional output. Technical report, Los Alamos National Laboratory, 2005.

- [9] D. Higdon, M.C. Kennedy, J.C. Cavendish, J.A. Cafeo, and Ryne R.D. Combining field data and computer simulations for calibration and prediction. *SIAM Journal of Scientific Computation*, 26(2):448–466, 2004.
- [10] R.G. Hills and I. Leslie. Statistical validation of engineering and scientific models: Validation experiments to application. Technical report, Sandia National Laboratories, 2003.
- [11] X. Jiang and S. Mahadevan. Bayesian validation assessment of multivariate computational models. *Journal of Applied Statistics*, 35(1):49–65, 2008.
- [12] X. Jiang and S. Mahadevan. Bayesian wavelet method for multivariate model assessment of dynamic systems. *Journal of Sound and Vibration*, 312(4-5):694–712, 2008.
- [13] X. Jiang, R.-J. Yang, S. Barbat, P. Weerappuli, Y. Fu, and G. Li. Multivariate model validation method and process for vehicle dynamic systems. Technical report, Ford Research and Advanced Engineering, 2008.
- [14] M.C. Kennedy and A. O’Hagan. Bayesian calibration of computer models. *Journal of Royal Statistical Society, Series B*, 63:425–464, 2000.
- [15] J. Li, Z. Mourelatos, M. Kokkolaras, P. Papalambros, and D. Gorsich. Validating designs through sequential simulation-based optimization. In *Proceedings of the ASME International Design Engineering Technical Conference*, 2010. Paper no. DETC2010-28431.
- [16] W.L. Oberkampf and M.F. Barone. Measures of agreement between computation and experiment: validation metrics. *Journal of Computational Physics*, 217(1):5–36, 2006.
- [17] W.L. Oberkampf, T.G. Trucano, and C. Hirsch. Verification, validation, and predictive capability in computational engineering and physics. *Appl Mech Rev*, 57(5), 2004.
- [18] Y. Pai. Investigation of Bayesian Model Validation Framework for Dynamic Systems. Master’s thesis, University of Michigan, Ann Arbor, MI, 2009.
- [19] Y. Pai, M. Kokkolaras, G. Hulbert, P. Papalambros, M. Pozolo, Y. Fu, R.-J. Yang, and S. Barbat. Assessment of a bayesian model and test validation method. In *Proceedings of the 2009 Ground Vehicle Systems Engineering and Technology Symposium (GVSETS)*, 2009.
- [20] R. Rebba. *Model validation and design under uncertainty*. PhD thesis, Vanderbilt University, Nashville, Tennessee, 2005.
- [21] R. Rebba and S. Mahadevan. Model predictive capability assessment under uncertainty. *AIAA Journal*, 44:2376–2384, 2006.
- [22] P.J. Roache. *Verification and validation in computational science and engineering*. Hermosa Publishers, Albuquerque, NM, 1998.
- [23] J. Sacks, W.J. Welch, T.J. Mitchell, and H.P. Wynn. Design and analysis of computer experiments. *Statistical Science*, 4(4):409–435, 1989.
- [24] H. Sarin, M. Kokkolaras, G. Hulbert, P. Papalambros, R.J. Yang, and S. Barbat. Comparing time histories for validation of simulation models: Error measures and metrics. *Transactions of the ASME - Journal of Dynamic Systems, Measurement and Control*, 132(11), 2010.
- [25] R. Zhang and S. Mahadevan. Bayesian methodology for reliability model acceptance. *Reliability Engineering and System Safety*, 80(1):95–103, 2003.

UNCLASSIFIED