Statistical Model Validation for Reliable Design of Engineering Products

by

Hao Pan

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Automotive Systems Engineering)
in the University of Michigan-Dearborn
2016

Doctoral Committee:

Assistant Professor Zhimin Xi, Chair
Professor Yubao Chen
Associate Professor Yung-wen Liu
Professor Chris Mi, San Diego State University

Acknowledgements

First of all, I would like to thank my advisor, Dr. Zhimin Xi, for his dedicated guidance, support and enlightenment during my pursuit of the PhD degree in the University of Michigan-Dearborn.

I would also like to thank Dr. Yubao Chen, Dr. Yung-wen Liu and Dr. Chris Mi for serving as my committee members and their time, effort and valuable inputs for my PhD study.

I am grateful to my industrial sponsor Dr. Ren-Jye Yang from Ford Motor Company for funding this research and providing support in terms of data-sets and technical discussions. I appreciate Dr. Hongyi Xu from Ford Motor Company for exchanging views on the same problems and offering inspiration.

It is my pleasure to work with members of the Xi group, and I am thankful to Ms. Xiangxue Zhao, Ms. Ying Chen, Ms. Jiayin Qi and many others for their contribution to the improvement of my PhD dissertation.

Finally, I would like to express my gratitude to the support of my family, to my parents for their continued encouragement and care, and to my wife Jia Feng for her ever-lasting love and understanding.

Table of Contents

List of Figures

List of Tables

Abstract

Computer models have been used to simulate engineering product and system performances in applications such as vehicle crashworthiness, structural safety, thermal responses, etc. If these predictions were accurate in the product and system design space, these models can help reduce product development cycle, cut down the cost of physical tests, and identify the optimal design. However, models are built on assumptions and simplifications. Therefore, model prediction could be problematic without referring to the corresponding test data. More importantly, design errors could be created because of the model error. Model validation is to determine the degree to which the model is an accurate representation of the real world from the perspective of the intended uses of the model and is a critical process to ensure the improved design efficiency and accuracy while minimizing the overall design cost.

The objective of this dissertation is to study a systematic and practical model validation framework for the design of engineering products. To achieve this goal, five research thrusts are developed. First of all, a copula-based model bias characterization approach is developed to capture the relationship between model inputs, outputs, and the model bias. The contribution is to overcome the limitations of regression-based model bias modeling approaches including: i) the curse of dimensionality; ii) assumption of regression forms; and iii) low accuracy to the model outputs with unexplained portion of model bias defined by model parameters. Secondly, an adaptive copula-based model bias characterization approach is developed to further enhance the accuracy of the copula-based approach with the aid of clustering analysis. Thirdly, a novel validation metric for dynamic responses under uncertainty is developed so that model accuracy with dynamic responses can be quantitatively assessed considering limited test data. Fourthly, a stochastic model bias calibration and approximation approach is proposed with the aid of the developed dynamic

validation metric for reliability analysis. Finally, reliability-based design optimization is integrated with the proposed model uncertainty characterization approach for reliable design of various engineering products. Various numerical examples and practical engineering problems are employed to demonstrate the proposed model validation framework for designing reliable engineering products.

Chapter 1: Introduction

1.1 Background and Motivation

Computer models have been extensively used to simulate engineering problems such as solid mechanics, structural dynamics, hydrodynamics, heat conduction, fluid flow, acoustics, etc. These models predict various performances of interest such as stress, strain, velocity, or failure measures such as crack initiation, crack growth, fatigue life, net section rupture, critical corrosion damage, etc. Computer models are playing more and more important role to aid virtual prototyping, reduce product development cycle, and cut down the cost of performing physical tests. Not limited to engineering disciplines, applications of computer models are also popular in chemistry, biology, economics, psychology, and social science.

Although the list of applications for computer models seems endless, validity of these models needs to be examined carefully before enjoying the benefits and making design decisions based on the model prediction. A model is more abstract than the system it represents. Abstraction and assumptions are made to build the model by eliminating unnecessary details and focusing on important factors in the system. This abstraction process introduces inaccuracy and makes the model solution tractable and efficient to obtain. While on the other hand, such abstraction process requires assessment of the goodness of the model.

Model verification and validation needs to be formally defined. Various researchers from computational fluid dynamics (CFD), ground water flow, Institute of Electrical and Electronics Engineers (IEEE), software engineering and American Institute of Aeronautics and Astronautics (AIAA) provided their own definitions. Here definitions of verification and validation are adopted from [1] because of the popularity in the literature (commonly cited in [1, 2, 3, 4, 5]):

- **Verification** is the process of determining that a computer model implementation accurately represents the developer's conceptual description of the model and the solution to the model.

- **Validation** is the process of determining the degree to which a computer model is an accurate representation of the real world from the perspective of the intended uses of the model.

Verification answers the question “Have we built the model right?”. It is similar to debugging, in the sense that it intends to ensure the model does what it is designed to do. Verification focuses on comparing the elements of a simulation model of the system with the description of what the requirements and capabilities of the model were to be (i.e., see if the model matches specifications and assumptions deemed acceptable for the given purpose of application). There are many techniques that can be utilized to verify a model including, but not limited to, having the model checked by an expert, making logic flow diagrams, examining the model output under a variety of settings of the input parameters, etc.

Validation answers the question “Have we built the right model?”. It is the task of demonstrating that the model is a reasonable representation of the actual system: that it reproduces system behavior with enough fidelity to satisfy analysis objectives. Validation also focuses on determining whether the differences between the model and the system are acceptable given the intended use of the model. There are many approaches that can be used to validate a computer model. The approaches range from subjective reviews to objective statistical tests.

Verification and validation (V&V) have gained interest in various research fields and various model validation approaches have been developed. However, are these approaches sufficient to assess the goodness of the model? As a result of applying model validation approaches, how will they impact the reliability-based design in miscellaneous engineering

fields? We shall seek the answer to these questions and construct our own solution in this research if no satisfactory answers are found.

1.2 Research Thrusts and Contributions

The overall objective of this research is to study a systematic and practical uncertainty management framework for incorporating model uncertainty into reliability-based design by enhancing/developing model validation methodology. To achieve this objective, model uncertainty should be accurately approximated in the whole design space, where the design space is defined as the allowable design domain of all design variables for an engineering design practice. As a result of breaking down the research objective, the following research thrusts are formed:

- **Research Thrust 1**: Study general and robust algorithms for model uncertainty approximation at intended uses of the model: this research addresses challenges for model uncertainty approximation in the design space and further quantifying uncertainty for the model uncertainty approximation.

- **Research Thrust 2**: Study time-independent and time-dependent reliability analysis with consideration of the model uncertainty: this research task addresses challenges raised from model validation metric for dynamic systems and reliability analysis considering model uncertainty.

- **Research Thrust 3**: Study general and effective algorithms for reliability-based design by incorporating the model uncertainty: this research task addresses challenges for reliability-based design with irreducible and reducible uncertainties, statistically correlated uncertainties, and random field uncertainties.

The contributions of this research with respect to each research thrust is briefly described as follows:

- **Contribution to Research Thrust 1**: A Copula-based model bias characterization approach is developed to capture the relationship between model inputs, outputs

and the model bias, as well as to provide the prediction in new design space. Model bias is first characterized in the design space, then the model prediction is corrected by adding the characterized model bias. Through case studies it is proved that the non-linearity in the model is not affecting the potency of the proposed approach, and it is capable of handling problems with high number of design variables. The Copula-based model bias characterization approach is further enhanced by constructing an adaptive Copula-based model bias characterization approach coupling cluster analysis. Cluster analysis is performed on the raw data to group similar data points, followed by copula modeling for each group (cluster). Model prediction is then produced using information from each cluster. The final prediction is the weighted sum of every prediction from the respective cluster. The proposed approach is effective in improving the accuracy of model prediction compared to its predecessor in the sense that it is able to produce more accurate model prediction as well as narrower confidence bounds.

- **Contribution to Research Thrust 2**: A novel validation metric for dynamic responses under uncertainty is proposed. The classical U-pooling approach is extended for dynamic responses by discretizing and treating the dynamic responses as a high dimensional joint distribution. PCA is utilized to effectively represent the dynamic responses by a few random variables so that the U-pooling value can be computed more efficiently. The shape deviation is introduced in the validation metric so that the metric can still distinguish the model accuracy when the U-pooling value alone is not differentiable. Furthermore, a stochastic model bias calibration and approximation approach for dynamic system responses is proposed. It calibrates model bias using PCA so that only limited number of calibration parameters are needed and the calibration can be effectively conducted similar to the static model bias calibration. Another contribution of the proposed approach is to approximate model bias through building the response surfaces of PCA model components so that the approximate model bias keeps the same form as a PCA model which makes it possible to seamlessly integrate the dynamic model bias with the baseline model.

- **Contribution to Research Thrust 3**: Reliability-based design optimization is integrated with the proposed model uncertainty characterization approach for reliable design of engineering products. A copula-based model bias correction approach is proposed for RBDO addressing model bias calibration at available training design configurations, model bias approximation at new design configurations, reliability analysis considering model bias and design sensitivity analysis. The proposed approach is a non-causal modeling approach that conducts non-causality modeling between model bias, design variables, and the baseline model prediction. The proposed approach has the potential of being more suitable for model bias modeling compared to the regression approach because model bias is defined as the inherent model inadequacy for representing the real physical systems due to simplifications and assumptions and hence is not supposed to be fully accounted for by the defined model parameters using a causal modeling approach. Furthermore, the proposed approach employs less assumptions and is expected to be less sensitive to the dimensionality of the problem compared to the regression-based approach.

Chapter 2: Literature Survey

2.1 Model Verification and Validation History

As stated by statistician G. E. P. Box, all models are wrong, but some are useful [6]. Since computer models are, at their best, approximations of real world phenomena based on various assumptions, validation and verification (V&V) started to emerge in the literature and techniques has been developed to formally assess the goodness of the computer models. The earliest work on V&V can be traced back in [7], where the author noticed that with the advancement in computer technology in modeling complex systems, issues such as V&V were not satisfactorily solved or paid serious attention to. Sargent [8] used Fishman and Kiviat's definition [7] for V&V and did a summary on existing validation techniques. A list of validation techniques was provided, such as, subjective judgment by experts, graphical comparison, cross-validation, etc. They concluded that multiple validation techniques should be selected and the statistical test is desirable if it is feasible and economical. Dr. Balci in [9] continued their work and stated that the validation techniques available then generally fall into two categories, subjective validation techniques and statistical tests. They also provided a summary of such statistical tests. Dr. Balci was the main contributor to V&V techniques included in the 96' Department of Defense (DOD) report [2] as DOD recognized the benefits of V&V to modeling and simulation (M&S). Such benefits include, but not limited to, increasing the confidence in M&S, mitigating the risk of making the wrong decision and reducing the cost of the future verification, validation and accreditation (VV&A). It was also pointed out that it is easy to confuse verification with validation. In [10] the author emphasized various distinctions between verification and validation, and argued that verification should proceed before validation. This research will be focused on validation. Various validation methodologies will be reviewed next.

2.2 Deterministic Model Validation

Deterministic model validation aims at producing a fixed model validation outcome. This is mostly due to the fact that this type of model validation deals primarily with systems with deterministic inputs and outputs, and no statistics-based techniques are involved. The end product of deterministic model validation is often in the form of a validation metric that measures the difference between computer model and the true system response. The most intuitive and primitive approach in this area is to compare computational and experimental system response quantity (SRQ) graphically. One decides whether or not to accept the model by inspecting the difference between the two SRQ curve or response surfaces. No quantitative measure of the difference between the two quantities compared is involved. In [11] a study was done for dynamic stiffening behavior of flexible structure. A mathematical model was developed to quantify the deformation of a flexible beam. The authors validated this model by superimposing computer-simulated deformation curve onto the experimental curve image taken by a high speed camera. In [12] in order to validate the computer model for predicting slip factor of centrifugal impellers, the authors plotted the predicted values versus experimental measurements. If they agree with each other, then the collection of all the points plotted should form a line with a slope of one (base line). Similar usage of graphical comparison can be found in [13, 14] However, in [15] the authors argued that graphical comparison may be subjected to reader misinterpretation because of unknown underlying data structure. In [16] the authors argued that graphical comparison can be biased and subjective. More refined techniques are thus needed to produce plausible validation results.

More quantitative validation metrics are developed to assess the agreement between computational and experiment outcomes in terms of a comprehensive quantitative measure. For non-dynamic systems there are simple mathematical measures to quantify model adequacy such as L^2 norm. For dynamic systems more sophisticated validation metrics are needed to consider various features in the system such as magnitude, shape or phase.

A common feature to compare is magnitude. In [15] the authors discussed several magnitude-only error metrics such as mean absolute error (MAE) and the root mean square

error (RMSE). It was noted in [15] that it is impractical to set a single threshold value for these measures because the validity of the model depends on both the type of the model and its intended uses. They recommended to use a combination of various validation techniques, which echoes with the suggestions in [8].

In [17] the authors used L^1 and L^2 norms to quantify the magnitudes difference for the case where both the experimental measurements and the computer simulation results are deterministic.

Another feature to compare is shape. A continuous wavelet transform was combined with several measures in [18, 19] to evaluate the differences in the shapes of the time series. The surfaces in the time-scale domain obtained by wavelet transforms represent the first and second derivatives of the time series and, hence, can be used to quantify shape attributes, such as the slopes and slope changes of the time series.

Authors in [17] suggested the use of more sophisticated techniques that comprehensively take into account features, such as magnitude, shape, and phase. Such techniques will be discussed next.

The Sprague and Geers metric [20, 21] (denoted as SG metric) was an integration of the magnitude and phase error quantifications. Authors in [22] illustrated that SG metric is not symmetric and do not consider shape differences. And SG failed to identify discrepancy between two time histories which differ in shape and magnitude. Also in [23] the author compared SG metric with Russell's metric and pointed out that SG metric is biased.

Russell's metric [21, 23] (denoted as R metric) is similar to SG metric but different in the magnitude error factor. Russell noticed that the phase error factor is bounded between zero and one but the magnitude error factor is unbounded. In order to develop a metric that is not dominated by either error factor, Russell adjusted the magnitude error factor so that it is of the same scale as the phase error factor. In [22] the authors used the same numerical

example as used for SG metric to demonstrate that Russell's metric also failed to recognize the discrepancy between the two time series.

The Knowles and Geer's metric [21] (denoted as KG metric) used a different magnitude error factor. And in place of the phase error factor the KG metric used a time-of-arrival (TOA) metric. The time-of-arrival for a wave form is determined by the time at which the wave form attains some percentage of the maximum value, a range of 5-10% is recommended for wave forms with relative fast rise times. The rise time is a useful quantity in shock and vibration analysis.

In [21] the author compared the SG and KG metrics and concluded that although KG metric provides values comparable to SG metric, it is not able to indicate the sign of magnitude error (under or over prediction). And the use of TOA metric instead of the phase metric seems a limitation because of the requirement of characterizing the rise time. The author also expressed that SG and KG metric are not symmetric metrics. This is not a concern when compare a time series to a reference. When comparing two time series with no reference, then the metric is evaluated two ways, i.e. each time series is treated as basis, and the resulting two metrics averaged.

Compared to SG, KG and R metrics, in addition to quantifying magnitude and phase error, the EARTH metric [22] also evaluates topology error, which is discrepancy in the shapes of two time series. Discrepancy in phase is removed by shifting the time history with the number of steps before analyzing the magnitude error. There exist local timing errors as well as discrepancy in shape. In order to compensate for these, dynamic time warping (DTW) was used. The topological error is a measure of discrepancy in shape of the two time histories. The shape of a time history is defined by the slope at each point.

In order to remove the effect of phase error, time-shifting is performed before derivatives of the time shifted histories are calculated. DTW technique is applied to compensate for local timing errors and quantify the difference in magnitude of the derivatives. Unlike KG, SG and R metrics, there is no comprehensive form of the EARTH metric. Each error factor

is calculated and compared separately. When evaluations from subject matter experts (SME) are available, a regression is performed to generate comparable ratings.

In [22] the authors compared EARTH metric with several other metrics, namely, Wavelet decomposition coupled with Russell's metric, step function, ADVISER model evaluation criteria and Corridor Violation Plus Area (CVPA). They found that EARTH metric produces more consistent results than other metrics using a case study provided by the International standards Organization (ISO) working group on virtual testing (head impact test).

2.3 Statistical Model Validation

Model improvement is achieved by quantifying model bias. However this is not a simple task as a model typically has uncertainty in inputs and outputs. Quantification of uncertainty (e.g. statistics of uncertainty) is needed in model validation as stated in [7]. Such model validation is the so called statistical model validation.

Uncertainty can be classified into two categories, namely, aleatory and epistemic uncertainties [24]. Either one of or both of the SRQs to be compared can be thought of as random variables when purely aleatory uncertainty is present (e.g. a statistical distribution).

For cases where pure aleatory uncertainty is present in both experimental measurements and the computer model outputs, various techniques were developed to quantify the distances between the CDF's. In [25] the authors examined whether or not the deterministic scalar experimental SRQ is within the highest density region (HDR) of the PDF's of the computational SRQ. In [26] the authors developed a maximum horizontal distance between the two CDF's. The selection of rejection criteria is subjective.

The Kolmogorov-Smirnov statistic measures the vertical distance between the two CDF's. Observe, however, that if the SRQ has a very small variability (almost deterministic), the vertical distance could be very large even though the two CDF's are very close when their distance is measured horizontally.

Another measure of the distance between CDF's was developed in [27, 28], where they proposed to use the area between the two CDF's as a validation technique. They argued that it enjoys several advantages such as ease of interpretation, objectiveness and ability to express validation results in terms of physical units (degree, Celsius, meter, etc.). The computational SRQ is specified by the model so the associated cumulative distribution function (CDF) is assumed to be known. The authors suggested that this CDF be obtained by solving the mathematical model analytically or by propagating a large number of replicate samples in a Monte-Carlo simulation. The experimental SRQ, on the other hand, is usually provided as a collection of point values in a data set. Empirical cumulative distribution function (ECDF) was used to describe the distribution of the experimental measurements. The authors illustrated that this area metric is better than those based solely on the mean or/and variance of the data as it was able to detect the difference when the mean and variance of observations are matched but the distribution isn't. They showed a case where the traditional statistical test Kolmogorov-Smirnov (based on Kolmogorov-Smirnov statistic) fails to recognize the difference in the distributions while the area metric was able to.

In [29] the authors used Anderson-Darling test statistic as a measure of the discrepancy between two distributions. The Anderson-Darling test is one of the most powerful statistical tools for detecting most departures from normality [30]. It is a modification of the Kolmogorov-Smirnov (KS) test and gives more weight to the tails than does the KS test. The test uses a weighted quadratic ECDF statistic to measure the distance between the two CDF's.

When both aleatory and epistemic uncertainty are present, the term p-box is used to describe the set of all possible CDFs that the random variable may follow. Corresponding validation technique was developed in [17].

Another branch of statistical model validation approaches springs from hypothesis testing. For univariate quantities, t-test is used to assess the similarity between the means from computer model predictions and experimental measurements, and F-test to assess the

similarity between the variances [15, 31, 32]. Extension to multivariate quantities can be achieved by using Hotelling's T^2-test for comparing multivariate means [33, 34], and Wilk's Λ distribution for comparing covariance matrices [35]. In [36] the author stated that multivariate hypothesis test limits the inflation of type-I error present in multiple univariate tests. Also it was noted in [31] that in all these hypothesis tests normality is assumed for both the experimental SRQ and the computational SRQ. This assumption is often not valid and transformation to normality was suggested in [32]. It is possible that such transformation is not successful in transforming the data into normality. Alternatively, bootstrap method to estimate the distribution of data was suggested in [37]. In [32] the authors suggested to use univariate and multivariate tests collectively. The univariate tests can yield conflicting validation results but are able to identify which variable in the multivariate data is associated with deficiency. However, multivariate test is needed since the univariate tests do not take into account the correlation in the data and data that pass all the univariate tests may not pass the multivariate test. A closely related method to Hotelling's T^2-test is the r^2 method developed by [38] and is similar to Mahalanobis distance. The statistic r^2 measures the distance between the centroids of the two SRQ's and the computer model is rejected if the probability of r^2 being greater than the critical value is less than the significance level. The r^2 method is applicable for both univariate and multivariate cases and takes into account uncertainty in the model parameters and the model outputs. Authors in [39] further developed this method by formulating confidence intervals for the r^2 statistic. This method was implemented in a metal flanging process validation problem [40]. The r^2 method requires normality assumption. As a result, the r^2 statistic follows a X^2 distribution, and the rejection criteria is based on critical values of the X^2 distribution at a given confidence level. The authors extended this method to non-normal data by the use of maximum likelihood estimation (MLE) [41]. The rejection criteria can be determined by Monte Carlo simulations. Consequently, the computation cost increases. A drawback of the r^2 method is that the validation results are influenced by the level of prescribed type-I error. As noticed in [42], specifying the type-I error at different values can lead to different conclusions. Statistical hypothesis testing is a trade-off between type-I and II errors. And the authors suggested that the selection of significance level be an optimization problem and be balanced for the intended purposes.

In [43] it was demonstrated that Bayesian hypothesis testing methods are superior to classical hypothesis testing methods because both hypotheses (null and alternative) are considered simultaneously. And it was found that p-value used in classical hypothesis testing leads to misleading results [44] and does not yield model confidence. A combination of Bayesian methods and hypothesis testing was developed for validation purposes. Using Bayes factor ([31, 45, 46]), the authors set up validation criteria (whether the Bayes factor is above or below unity) based on the suggestion by [47]. Normality is no longer relied upon although it was used to make the calculation of Bayes factor easier than non-normal distributions. As stated in [32], explicit expression of the joint PDF for multivariate normal random variables exists but the construction of joint PDF of other multivariate non-normally distributed random variables can be cumbersome. However the joint PDF can be acquired via Nataf transformation and there are approximation methods stated in [47] that facilitate the calculation of Bayes factor once the joint PDF is available. In [43] the authors treated the Bayes factor as a random variable to address the uncertainty in model parameters. In cases where some Bayes factors support the computer model while others don't, they suggested that future work treat the model acceptance as a decision problem, considering additional test costs, cost/risk consequences of accepting the current model, etc. In [32] the authors transformed non-normal data to normality and showed the transformation helps to reduce the type-I error. In [48, 49] the authors enabled the inclusion of multiple sets of experimental data by assuming the experimental data in each set are independent. The overall Bayes factor is calculated by multiplying individual Bayes factors for each data set together. In [50] the authors indicated the difference between the r^2 technique developed by [38] and their Bayesian technique (though they are similar) is that not having enough evidence to reject a model is not the same as having enough evidence to accept the model. The numerical example in their paper demonstrated that the validation results yielded by the two techniques are the same for all experiments except one conflicting result. In [50] the authors derived model confidence based on Bayes factor and claimed being the first to derive explicit expression of the model confidence for Bayesian point-null hypothesis testing. [29, 51] made a comparison between point-null and interval based hypothesis testing, stating that as the sample size increases, the chance of rejecting a correct model also increases for point-null hypothesis testing (the effect of sample size

on the validation metric is significant). To have more consistent results, a Bayesian interval-based hypothesis testing method (BIH) was proposed in [50] and this was the first to have explicit expressions of model confidence for Bayesian interval-based hypothesis testing. In [52] the author coupled BIH with Probabilistic Principle Component Analysis (PPCA) [53] to remove correlation of data, reduce dimensionality and handle uncertainty.

2.4 Model bias characterization

With the help of model validation methodologies, one is able to tell whether a model is good or not in terms of the difference between the model and the real world. Such difference is also known as model bias and characterizing it would be a natural next step following the model validation. Characterizing the model bias is of great importance in this research. There are various approaches for model bias characterization in the model validation community. Such approaches will be characterized and briefly introduced next.

Response surface methodologies (RSM) [54] are used to characterize the model bias since they can play a critical role in the process of obtaining corrected model prediction. In [55] the authors reviewed response surface methodology and developed error bounds for capturing variation in the model bias. In [56] a methodology for quantifying model bias using response surface was developed. The authors utilized Maximum Entropy Production (MEP) principle to approximate the PDF of the model bias.

The accuracy of response surface approaches mainly depends on three factors including: i) nonlinearity of the model bias in the design space; ii) amount and location of the identified model bias in the design space; and iii) algorithm of the response surface method [57].

Due to its flexibility and ability to capture the nonlinearity of the underlying model through different settings of hyper-parameters, the Gaussian process approach remains one of the most popular approaches in meta-modeling to date, and is believed to be applicable for characterizing a large variety of different model biases. The well-known Kriging method is considered to be a special case of the Gaussian process approach. Gaussian process was used together with Efficient Global Optimization (EGO) and Efficient Global Reliability

Analysis (EGRA) to improve the efficiency of solving the RBDO problem in [58]. Information found at one design point may influence those at subsequent design points and GP was effective in modeling this relationship.

As another branch of methodologies dealing with uncertainties in model bias characterization, the Bayesian-based approaches assume a prior distribution for the model bias and apply Bayesian theorem to update the distribution based on available data so that a posterior distribution can be obtained. The work in this area can be traced back to [59], where a Gaussian Process was used to model the prior of the SRQ and the posterior was inferred using Bayes' theorem. The authors suggested performing normality transformation if normal distribution is not appropriate. They indicated that future work should extend the current technique to multivariate output. Based on [59], various subsequent works were done. Bayarri et al. [60] developed tolerance bounds for model predictions. The authors' view of validation is not simply to provide answer (yes/no) to the validity of a computer model, but rather, to evaluate the accuracy of computer model prediction for the intended use. Higdon [61] developed posteriors based on non-normal prior of parameters of the Gaussian process model. The authors' approach to validation is to assess the prediction capability of the computer model. The computer model is said to be useful if it reduces uncertainty in prediction compared to using experimental measurements only. However there is no formal and objective way to state how much reduction is enough. Chen et al. [62, 63] developed posteriors for both model bias and output using a more flexible beta distribution prior. Tolerance bounds were developed for validation purposes. The traditional criterion for validation is that the model is accepted if the interval of the model bias contains zero or if the interval of the true value of the system response quantity contains the computer model output. The authors considered this criterion problematic since it tends to reject the computer model at regions with many physical observations (and thus prediction intervals are narrow) but fails to reject the computer model at regions with few or no physical observations (and thus prediction intervals are wide). They suggested that evaluation of the associated confidence based on the validation be performed when deciding accept/reject model. They proposed an innovative criterion for validation but it suffers from the fact that it needs subjective inputs (opinions from experts, etc.). All the

above Bayesian posterior estimation techniques involve the use of Gaussian Process. The validity of this assumption needs to be examined.

2.5 Limitations of the State-of-the-Art

The deterministic model validation methodologies are often tied to single SRQs and cannot handle multiple SRQs. The predictive capability of the deterministic model validation methodologies is very limited as these methodologies are developed for the sole purpose of providing a quantitative measure of the discrepancy of the computer model. More importantly, the deterministic model validation methodologies do not consider uncertainty.

The statistical model validation often rely on some assumptions (e.g. normality). Some only concern single SRQ and cannot handle the multiple SRQs (mainly their correlation structure). Classical hypothesis testing technique is of point-null type hypothesis testing and validation results are affected by sample size [36]. The ability to handle both types of uncertainties (epistemic and aleatory) in SRQs is limited due to the dependence on the normality assumption except the r^2 method. Classical univariate hypothesis testing is sensitive to the selection of type-I error levels and is subject to accumulation of type-I error when applied to each component of a vector quantity. The Bayesian-based hypothesis testing methods do not rely on normality assumption any more, but often the posterior distribution does not have closed-form analytical expression and computation cost becomes high as a result.

Moving on to the model bias characterization approaches, the accuracy of the RSM depends on the complexity of the problem and may suffer from this dependency. As a result, the prediction accuracy can swing dramatically and the improvement can sometimes be negligible due to the high nonlinearity/complexity of the problem. The determination of the parameters used in GP can be tedious and complicated, not to mention that the GP uses normality assumption. The Bayesian methods also do have their limitations, for example, the Bayes factor often cannot be obtained in an analytical form, and the computation cost is typically high unless coupled with dimensionality reduction techniques.

All the limitations as discussed above call for further research effort in the model validation and model bias characterization. This research will begin by addressing such issue and the coming chapter 3 will be dedicated to developing a model bias characterization approach to overcome the known hurdles in the model bias characterization.

Chapter 3: Copula-based Model Bias Characterization

3.1 Introduction

Various model validation approaches have been developed over the years, and there are limitations associated to them. Some approaches do not consider model uncertainty. For those that do consider model uncertainty utilizing statistical techniques, assumptions (e.g. normality) are used to simplify the derivation/computation process but oftentimes these assumptions are not valid. When dealing with high dimension problems (consider multiple inputs/outputs simultaneously), the computation cost may rise due to the curse of dimensionality. The copula-based model bias characterization approach is proposed here to overcome these shortcomings including: i) the curse of dimensionality; ii) assumption of regression forms; and iii) low accuracy due to noisy model outputs. Before introducing the proposed approach, the concept of model bias needs to be described first.

A general relationship between the model prediction and the test data can be expressed as

$$\hat{Y}(\boldsymbol{P}, \boldsymbol{X}, \boldsymbol{\Phi}) + \delta = Y - \varepsilon \tag{3.1}$$

where $\hat{Y}$ and Y are the predicted and the measured system performances, respectively, δ is the model bias, ε is the measurement error, $\boldsymbol{P}$ is a vector of the deterministic model variable, $\boldsymbol{X}$ and $\boldsymbol{\Phi}$ are the vectors of the irreducible and reducible model random variables, respectively. The irreducible random variables $(\boldsymbol{X})$ are characterized using probability density functions (PDFs) with sufficient information. The reducible random variables $(\boldsymbol{\Phi})$ are derived from the lack of information for describing the uncertainty. For example, parameters, i.e., the mean and the variance of the PDF, or even distribution types are uncertain unless sufficient information is collected. The measurement error ε is mainly affected by the equipment accuracy and human errors. $Y - \varepsilon$ can be thought of as data obtained from physical tests (test data), while $\hat{Y}(\boldsymbol{P}, \boldsymbol{X}, \boldsymbol{\Phi}) + \delta$ can be treated as computer model outputs (CAE data).

3.2 Copulas

Due to the limitation of the model bias characterization approaches as discussed in chapter 2, a Copula-based approach is being proposed here. A copula is a general way in statistics to formulate a multivariate distribution, particularly for a bivariate distribution, with various statistical dependence patterns. Formally, a copula is a joint distribution function of standard uniform random variables. According to Sklar's theorem [64], there exists an *n*-dimensional copula C such that for all T in a real random space,

$$F(T_1, \dots, T_N) = C\big(F_1(T_1), \dots, F_N(T_N)\big) \tag{3.2}$$

where F is an N-dimensional distribution function with marginal functions $F_1, \dots, F_N$. To this date, most copulas only deal with bivariate data due to the fact that there is a lack of practical *n*-dimensional generalization of the coupling parameter [65, 66]. For multivariate data, a usual approach is to analyze the data pair-by-pair using two-dimensional copulas.

3.3 Copula Modeling of Model Bias

Applying the concept of Copula on model bias characterization, we propose an approach here to use Copula to model the distribution of the relationship between model bias and model inputs and/or model responses. There are two challenges our proposed approach aims to solve: i) selecting the best Copula; and ii) reducing dimensionality/mitigating computational cost in the construction of model bias. The solving process of the two challenges will be discussed next.

One of biggest challenges for Copula modeling is to select the best Copula suitable for the available data. The most commonly employed methods are based on a maximum likelihood approach [67, 68, 69], which relies on the estimation of an optimal parameter set. In other words, comparisons are made among copulas with given optimal parameters. The maximum-likelihood-based copula selection approach typically demands sufficient bivariate data to ensure the accurate copula selection. Hence, such an approach may not always be suitable, especially for cases when there is a lack of validation experiments for charactering the model bias at few design configurations.

Recently a Bayesian copula approach [65] was proposed to select the best copula. It was further shown that this approach provides more reliable identification of true copulas even with the small amount of samples because, unlike the maximum likelihood approach, the selection of the best copula is independent of the copula parameter estimation [65]. Hence, we employ the Bayesian copula approach for modeling statistical dependence between design variables and the model bias, between model prediction and the model bias, where the design variable indicates the controllable model variable. Determination of the best copula is done in two steps: i) selection of optimal marginal distributions; and ii) determination of an optimal copula. The marginal distribution is determined using the maximum likelihood approach. Next, we focus on the second step for determining the optimal copula using the Bayesian approach.

In the Bayesian copula approach, a set of hypotheses are first made as follows.

$$H_k: The\ data\ come\ from\ Copula\ C_k, \qquad k = 1, \dots, Q \tag{3.3}$$

The objective is to find the copula with the highest $Pr(H_k|D)$ from a finite set of copulas, where D represents bivariate data in a standard uniform space. Based on the Bayes' theorem, the probability that data come from the copula C_k is expressed as

$$Pr(H_k|D) = \frac{Pr(D|H_k)Pr(H_k)}{Pr(D)} = \int_{-1}^{1} \frac{Pr(D|H_k,\tau)Pr(H_k|\tau)Pr(\tau)d\tau}{Pr(D)} \tag{3.4}$$

where τ is the Kendall's tau, which is a non-parametric measure of the statistical dependence associated to copulas. Kendall's tau (τ) belongs to the set of each copula and the outcome is equally likely. All copulas are equally probable with respect to a given τ which reflects no preference over the copulas. The likelihood $Pr(D|H_k,\tau)$ depends upon τ and can be calculated from the copula PDF as

$$Pr(D|H_k,\tau) = \prod_{l=1}^{m} c_k(u_{1l}, u_{2l}|\tau) \tag{3.5}$$

where $c_k(\cdot)$ is the PDF of the k^{th} copula, m is the total number of coupling data, u_{1l} and u_{2l} are the l^{th} realizations of the statistically dependent bivariate variables. The normalization of $Pr(D)$ can be computed using the sum rule [70]. For more details of the Bayesian copula approach, please refer to the reference document [65].

After selecting the best Copula, the next step is to utilize the copula modeling between $\hat{Y}$ and δ to find the possible model bias δ for a realization of $\hat{Y}$ (e.g. $\hat{Y} = a$) at a new design configuration. Mathematically, this is a process to identify the conditional PDF of the model bias δ given $\hat{Y} = a$, that is,

$$C\left(F_{\hat{Y}}(\hat{y}), F_{\Delta}(\delta)\right)|\hat{y} = a \tag{3.6}$$

Meanwhile, we also know the design variable (e.g., $x_1 = a_1, x_2 = a_2, \ldots, x_j = a_j$) at the new design configuration. Thus, the possible realizations of the model bias δ must simultaneously satisfy all the conditional PDFs identified from a series of copula models. In other words, we need to find the intersection of these conditional PDFs for the model bias. Theoretically, it is very difficult. Therefore, a sampling approach is proposed for the model bias prediction. Four steps are conducted to have the empirical PDF of the model bias.

- **Step 1**: generate model bias samples (e.g. $\delta_{N,1}$, $\delta_{N,2}$,…, $\delta_{N,j}$) from the conditional PDFs obtained using Eq. (3.6)
- **Step 2**: identify lower (δ_lower) and upper (δ_upper) bounds from the model bias samples, where

$$\delta\_lower = max\left[min\left(\delta_{N,1}\right), min\left(\delta_{N,2}\right), \ldots, min\left(\delta_{N,j}\right)\right]$$

$$\delta\_upper = min\left[max\left(\delta_{N,1}\right), max\left(\delta_{N,2}\right), \ldots, max\left(\delta_{N,j}\right)\right]$$

- **Step 3**: keep model bias samples located between the lower and upper bounds
- **Step 4**: obtain an empirical PDF of the model bias from samples in step 3

3.4 Case Studies

3.4.1 Barnes Problem

The Barnes problem originated from Barnes' master's thesis [71]. It is a study of four optimization sub-problems: one high-fidelity (HF) model and three of its variants (low-fidelity, LF models), as listed in Table 3.1.

Table 3.1: Description of the Barnes problem

High-fidelity Model
$f_{high} = a_1 + a_2x_1 + a_3x_1^2 + a_4x_1^3 + a_5x_1^4 + a_6x_2 + a_7x_1x_2 + a_8x_1^2x_2 + a_9x_1^3x_2 + a_{10}x_1^4x_2 + a_{11}x_2^2 + a_{12}x_2^3 + a_{13}x_2^4 + \frac{a_{14}}{x_2+1} + a_{15}x_1^2x_2^2 + a_{16}x_1^3x_2^2 + a_{17}x_1^3x_2^3 + a_{18}x_1x_2^2 + a_{19}x_1x_2^3 + a_{20}e^{a_{21}x_1x_2}$ $g_{1high} = 1 - \frac{x_1x_2}{700} \leq 0$ $g_{2high} = \frac{x_1^2}{625} - \frac{x_2}{5} \leq 0$ $g_{3high} = \left(\frac{x_1}{500} - 0.11\right) - \left(\frac{x_2}{50} - 1\right)^2 \leq 0$
Low- Fidelity Model A
$f_{low} = a_1 + a_2x_1 + a_3x_1^2 + a_4x_1^3 + a_6x_2 + a_7x_1x_2 + a_8x_1^2x_2 + a_9x_1^3x_2 + a_{10}x_1^4x_2 + a_{11}x_2^2 + a_{12}x_2^3 + a_{13}x_2^4 + \frac{a_{14}}{x_2+1} + a_{15}x_1^2x_2^2 + a_{16}x_1^3x_2^2 + a_{17}x_1^3x_2^3 + a_{18}x_1x_2^2 + a_{19}x_1x_2^3 + a_{20}$ $g_{1low} = 0.5 - \frac{x_1x_2}{750} \leq 0$ $g_{2low} = \frac{x_1}{16} - \frac{x_2}{5} \leq 0$ $g_{3low} = \left(\frac{x_1}{500} - 0.11\right) - \left(\frac{x_2}{50} - 1\right)^2 \leq 0$
Low- Fidelity Model B
$f_{low} = a_1 + a_2x_1 + a_3x_1^2 + a_4x_1^3 + a_5x_1^4 + a_{20}$ $g_{1low} = 0.5 - \frac{x_1}{750} \leq 0$ $g_{2low} = \frac{x_1}{16} \leq 0$ $g_{3low} = \left(\frac{x_1}{500} - 0.11\right) \leq 0$

Low- Fidelity Model C
$f_{low} = a_1 + a_2 x_1 + a_3 x_1^2 + a_4 x_1^3 + a_{20}$ $g_{1low} = 0.5 - \frac{x_1}{750} \leq 0$ $g_{2low} = \frac{x_1}{16} \leq 0$ $g_{3low} = \left(\frac{x_1}{500} - 0.11\right) \leq 0$

High-fidelity model has a high-degree polynomial for the objective function and non-linear constraints while low-fidelity models have low-degree polynomials and more linear constraints. All models have two variables and they are both bounded between 0 and 70. For ease of understanding, none of the constraints was used and only the objective function was calculated to produce the model output. Out of the various configurations selected by using the Latin hypercube design (20 configurations), a portion of them (15 configurations) was used as training data to obtain the copula model, while the rest (5 configurations) was used to demonstrate the improvement the proposed approach can make.

The output calculated by the high-fidelity model is compared with that produced by one of the three low-fidelity models, as depicted in Fig. 3.1. The output from low-fidelity model A is the closest to that from the high-fidelity model, while there is significant difference in the forms of models B and C as compared to that of the HF model.

The copula approach is implemented and the error is improved significantly. Average prediction error and maximum prediction error, denoted as ε_μ and ε_{max} , respectively, are displayed in Table 3.2. Both the average prediction error and the maximum prediction error are much smaller for model A than those for model B and C. The copula has significantly larger influence on the bias associated to model B and model C as both the relative reductions of the average and maximum prediction error are considerable.

The improvement in model accuracy made by using the proposed copula-based approach is shown in Fig. 3.2. The copula-updated outputs are closer to those generated by the high-fidelity model for models A, B and C.

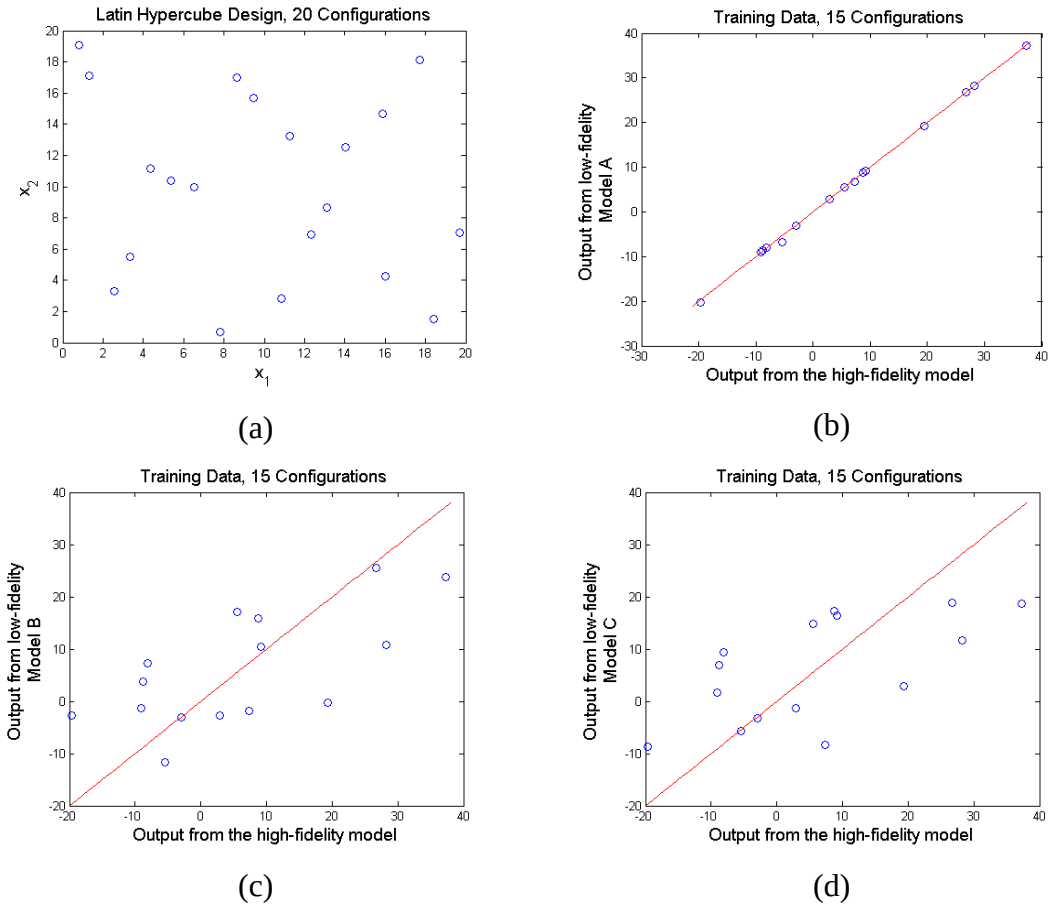


Figure 3.1: Design configurations and comparison of test data and CAE data for models A, B and C using training configurations. (a) Input configurations (training + confirmation); (b) Training data comparison for LF model A; (c) Training data comparison for LF model B; (d) Training data comparison for LF model C

Table 3.2: Improvement made by the copula-based approach for the Barnes problem

		Model A	Model B	Model C
Original Model	ε_{μ}	0.35	14.43	11.11
	ε_{max}	1.15	17.55	21.09
Corrected Model Using copula	ε_{μ}	0.23(34%)	10.17(30%)	5.86(47%)
	ε_{max}	0.74(35%)	15.63(11%)	9.00(57%)

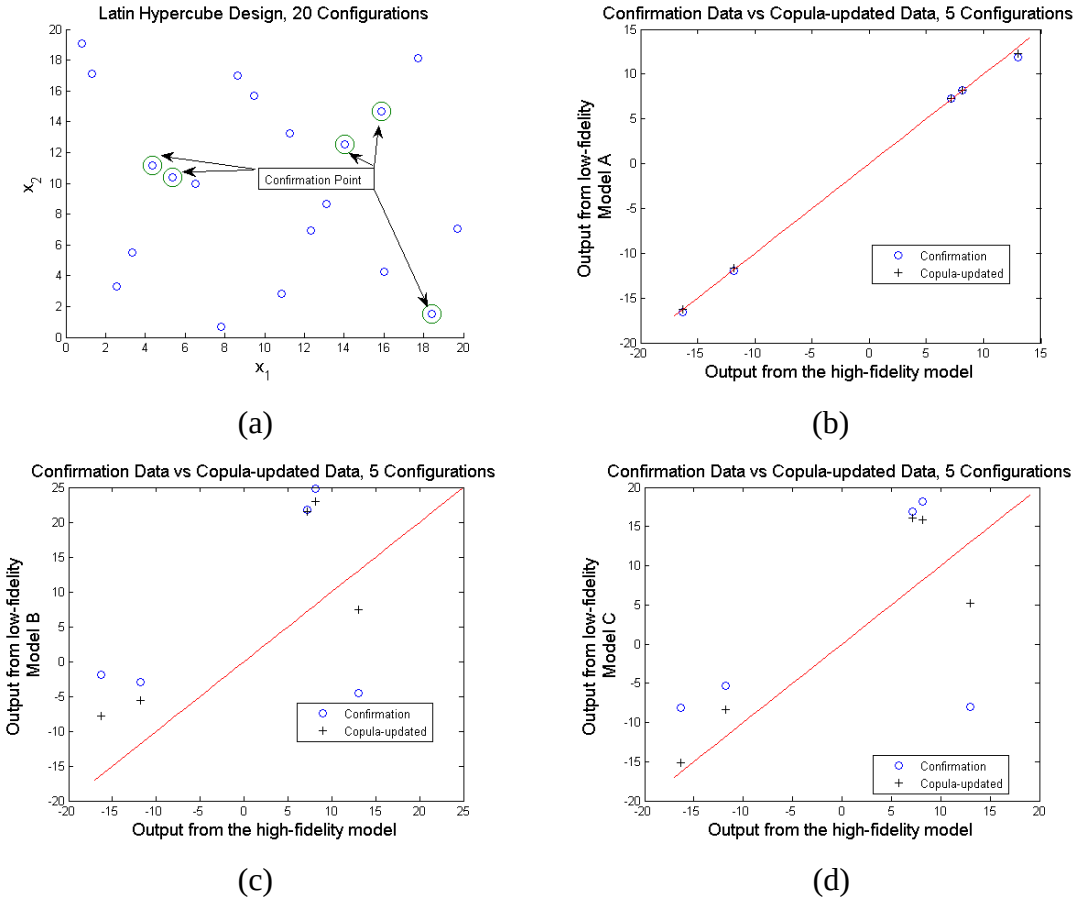


Figure 3.2: Comparison of test and cae data for models A, B and C. (a) Input configurations (confirmation configurations highlighted); (b) Confirmation data comparison for LF model A; (c) Confirmation data comparison for LF model B; (d) Confirmation data comparison for LF model C

The absolute error results presented in Tab. 3.2 are also visualized in Fig. 3.3 as bar plot.

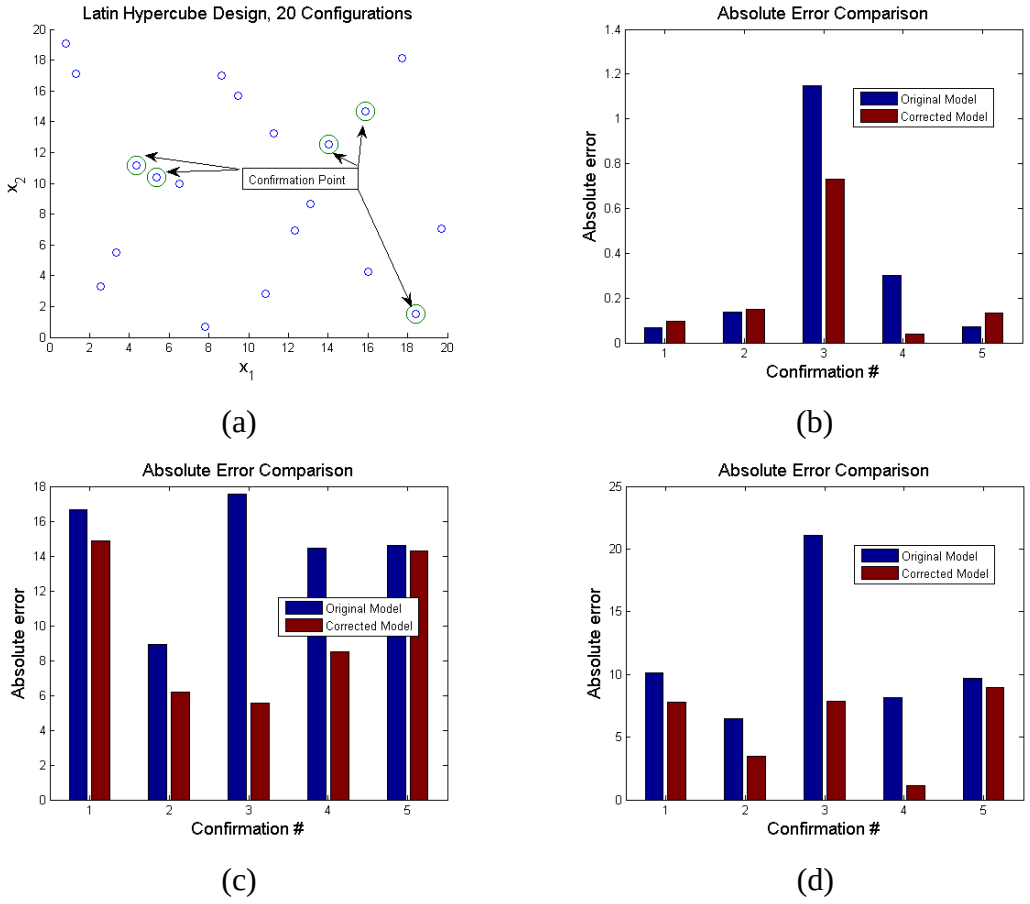


Figure 3.3: Comparison of absolute error for models A, B and C. (a) Input configurations (confirmation configurations highlighted); (b) Confirmation data comparison for LF model A; (c) Confirmation data comparison for LF model B; (d) Confirmation data comparison for LF model C

3.4.2 2001 Ford Taurus model

A 2001 Ford Taurus model is provided by the National Crash Analysis Center. Fig. 3.4 shows the physical test and model prediction for the full frontal impact. The simulation speed is 56.6 km/h against a rigid wall. For frontal impact protection, vehicle design must meet internal and regulated frontal impact requirements. In particular, vehicles must be designed to absorb enough impact energy through structural deformation and attenuate the impact force to a tolerable level in order to protect the occupants. Eight design variables are defined in Fig. 3.5 and their baseline design and design bounds are listed in Table 3.3. Chest G and Crash Distance are two key performances of interest that need to be validated.

Figure 3.4: Comparison of the full frontal impact between test and model prediction

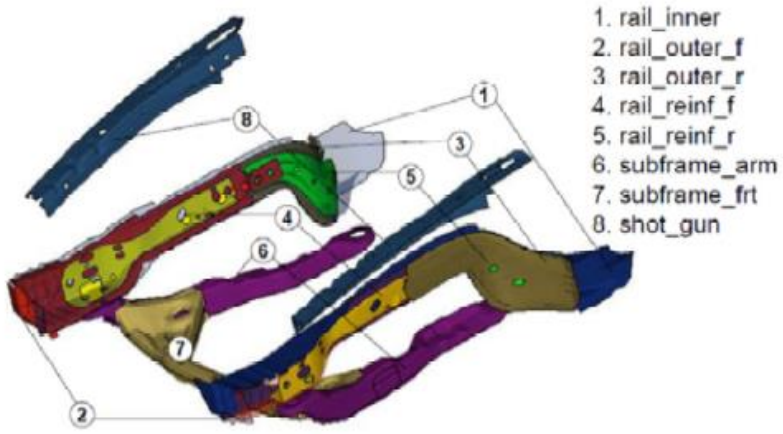
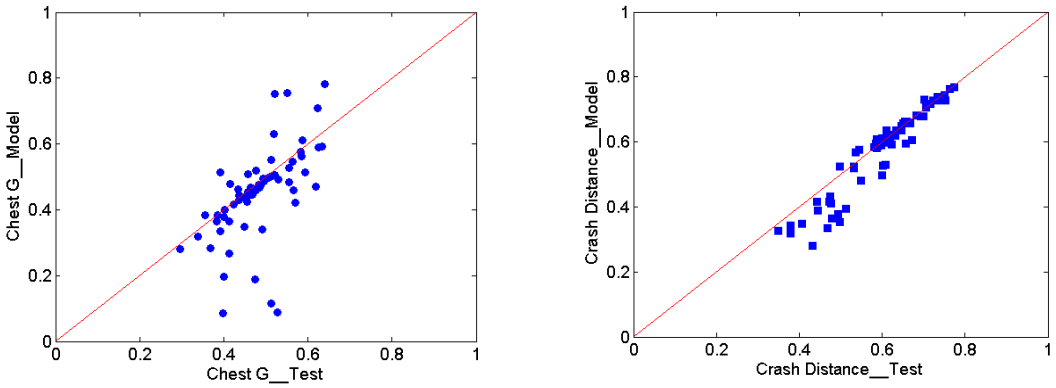


Figure 3.5: Eight design variables for main front-end structure

In this benchmark problem, 80 design configurations are generated in the design space defined in Table 3.3 where both model prediction and test values are available for the Chest G and Crash Distance. Among them, 64 design configurations are selected as training data to characterize the model bias in the design space. Fig. 3.6 shows the comparison between test and model prediction for the Chest G and Crash Distance at 64 design configurations. For Chest G, the average model prediction error is calculated as 0.0704, and the maximum prediction error could reach 0.4375. For Crash Distance, the average and maximum prediction error are calculated as 0.0328 and 0.1514, respectively. Clearly, the uncertainty from the model prediction is significant, which could mislead the designer towards a bad design for the main front-end structure.

Table 3.3: Baseline design and design bounds for main front-end structure

Design variable	Baseline	Lower bound	Upper bound
x_1	1.9	1.4	2.8
x_2	1.91	1.2	2.8
x_3	2.51	1.6	4
x_4	2.4	1.5	4
x_5	2.55	1.6	4
x_6	2.55	1.5	3.5
x_7	2.25	1.5	3.5
x_8	1.5	1.2	3



(a) (b)

Figure 3.6: Comparison between test and model prediction at 64 training design configurations. (a) Comparison for chest G; (b) Comparison for crash distance

Among 80 design configurations, we intentionally select the rest 16 design configurations to be interpolation study. In order to check the effectiveness of the proposed approach for extrapolation study, 25 extra design configurations are generated where the design variable exceeds the defined bounds in Table 3.3. Fig. 3.7 shows the initial comparison between the test and model prediction for the interpolation and extrapolation study without employing

the characterized model bias. For Chest G, the average model prediction error is calculated as 0.0551 and 0.1247 for interpolation and extrapolation, respectively. The maximum prediction error could reach 0.1393 and 0.4258, respectively. For Crash Distance, the average and maximum prediction error are 0.0496 and 0.1903 respectively for the interpolation and 0.0376 and 0.1678 for the extrapolation, respectively.

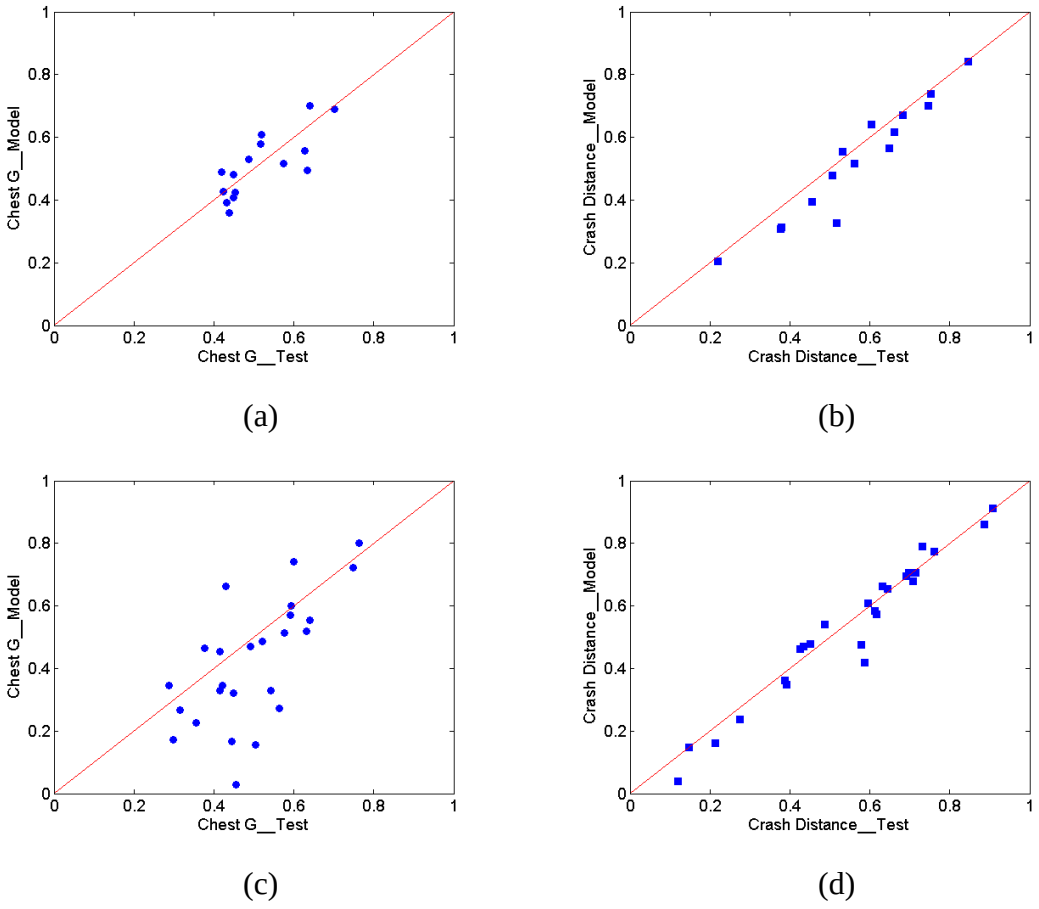


Figure 3.7: Initial comparison between test and model prediction for interpolation and extrapolation study. (a) Comparison for chest G in the interpolation study; (b) Comparison for crash distance in the interpolation study; (c) Comparison for chest G in the extrapolation study; (d) Comparison for crash distance in the extrapolation study

64 training data sets are used to calculate the model bias and the corrected model responses at 15 interpolation and 25 extrapolation design configurations. Using the copula approach, model bias is determined by the mean of the empirical PDF. To quantify the improvement

of model accuracy, Table 3.4 lists error statistics of model prediction using the copula approaches and compares their accuracy improvement with the original model, where ε_μ and ε_{max} indicate the mean and maximum of the absolute model error (|test-model prediction|). Relative change of the two error statistics in percentage produced by the copula approach is calculated and the results are shown as the numbers in the parentheses. The copula approach yields better error statistics as compared to the original model (a reduction of the error statistic, or the negative number in the parentheses indicates better accuracy). A comparison was also made between the copula approach and the traditional approach, the response surface methodology (RSM), in Table 3.4. The copula approach displays more consistent improvement of the error statistics compared to the RSM.

Table 3.4: Improvement made by the copula-based approach for the Ford Taurus model study

		Interpolation		Extrapolation	
		Chest G	Crash Distance	Chest G	Crash Distance
Original Model	ε_μ	0.055	0.05	0.125	0.038
	ε_{max}	0.139	0.19	0.426	0.168
Corrected Model using copula	ε_μ	0.053 (-3.0%)	0.034 (-31.6%)	0.088 (-29.1%)	0.038 (-1.3%)
	ε_{max}	0.120 (-13.8%)	0.131 (-30.9%)	0.228 (-46.6%)	0.130 (-22.7%)
RSM	ε_μ	0.062 (+12.6%)	0.022 (-56.9%)	0.081 (-35.5%)	0.045 (+19.6%)
	ε_{max}	0.155 (+11.2%)	0.059 (-69.1%)	0.242 (-43.3%)	0.201 (+19.5%)

The absolute error results presented in Tab. 3.4 are also visualized in Fig. 3.8 as bar plot.

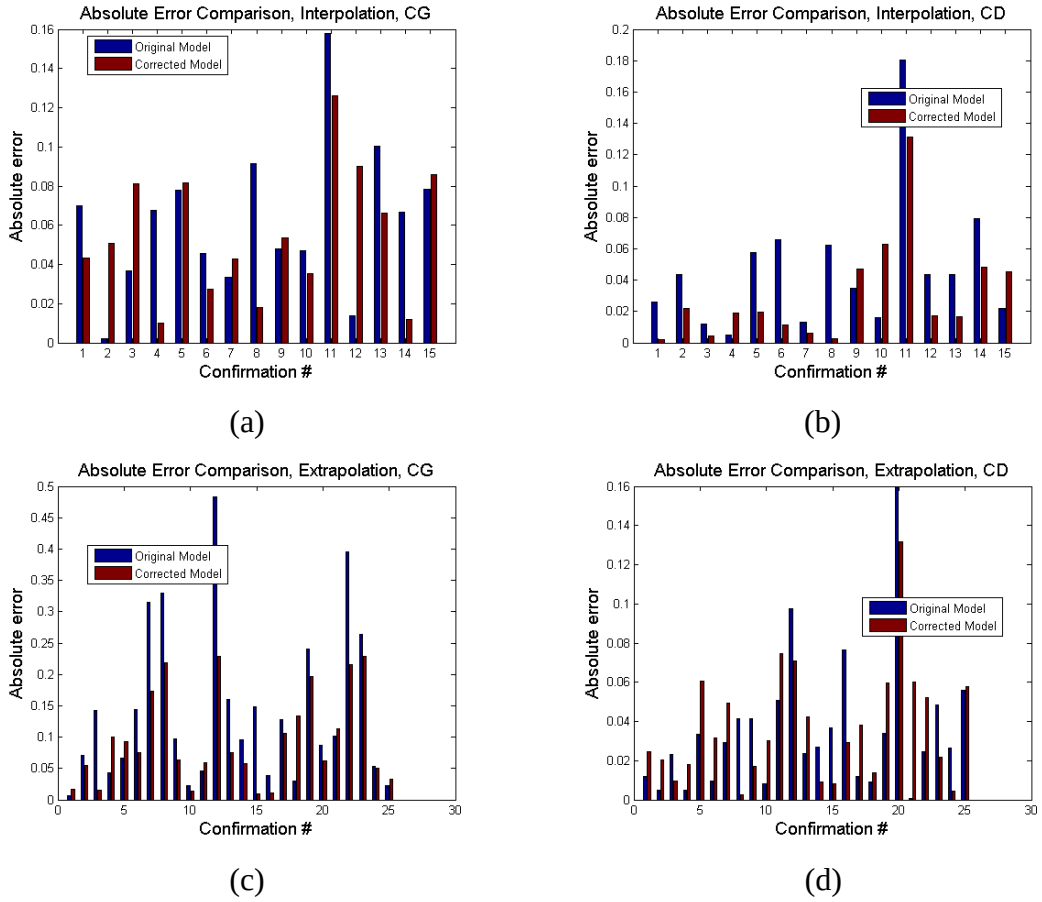


Figure 3.8: Comparison of absolute error resulted from original model and corrected model using the copula approach (a) chest G, interpolation; (b) crash distance, interpolation; (c) chest G, extrapolation; (d) crash distance, extrapolation

3.5 Summary

In this chapter, a copula-based approach was proposed to improve model accuracy by correcting the bias term. Model bias is first characterized in the design space, then the model prediction is corrected by adding the characterized model bias. The copula-based approach is capable of establishing a statistical relationship among model bias, design variables and model responses. The non-linearity in the model is not affecting the potency of the proposed approach. The two case studies used here adequately demonstrated that the proposed approach is effective in improving the accuracy of model prediction. Although the second case has a higher number of design variables, the computation time consumed

did not rise significantly. Being able to handle problems with high number of design variables is another advantage of the proposed approach.

Chapter 4: Adaptive Copula Approach for Model Bias Characterization

4.1 Introduction

The adaptive Copula modeling of model bias is proposed here as an enhancement for the Copula-based model bias characterization approach discussed in the previous chapter. Cluster analysis will be employed first to group similar data points together, followed by the copula-based approach using information from each cluster. The final prediction accumulates predictions obtained from each cluster. In this way, information contained by the available data can be better utilized and as a result, the model bias characterization can be improved and the predictive capability of the methodology can be strengthened.

4.2 Adaptive Copula Modeling of Model Bias

Important steps in the proposed approach can be summarized as those in the flow chart shown as Fig. 4.1. Test and CAE data (computer model outputs) are assumed to be available and they will be used as training data to build the clusters and copula models. Bias is then calculated by simply subtracting CAE data from the test data. Data pairs are formulated by treating each of the CAE data as the x coordinate and the corresponding bias as the y coordinate. Cluster analysis is conducted for the data pairs using k-means clustering. And a copula model will be built for each cluster. Weights associated to each cluster will be assigned to each of the confirmation data pairs (again bias vs CAE) based on the distance between the pair and the centroid of each of the clusters. It should be noted here that confirmation test data are often unavailable and as a result the confirmation bias cannot be obtained. To cope with this situation, a prediction using our previously proposed single copula-based approach [72] will be made and treated as the pseudo confirmation test data. Also using our previous approach, prediction for a specific confirmation point

using information (copula sample points) from each cluster will be made. The final prediction is the weighted sum using the weights and the predictions just mentioned.

Compared to our previously proposed approach, the proposed approach here has several enhancements, in the sense that it incorporates cluster analysis to separate data into groups based on their similarity, and uses a weighting scheme to consolidate every piece of information from the clusters, and a sampling approach for calculating expected value at the confirmation locations. All three aspects will be discussed in detail in the following sections.

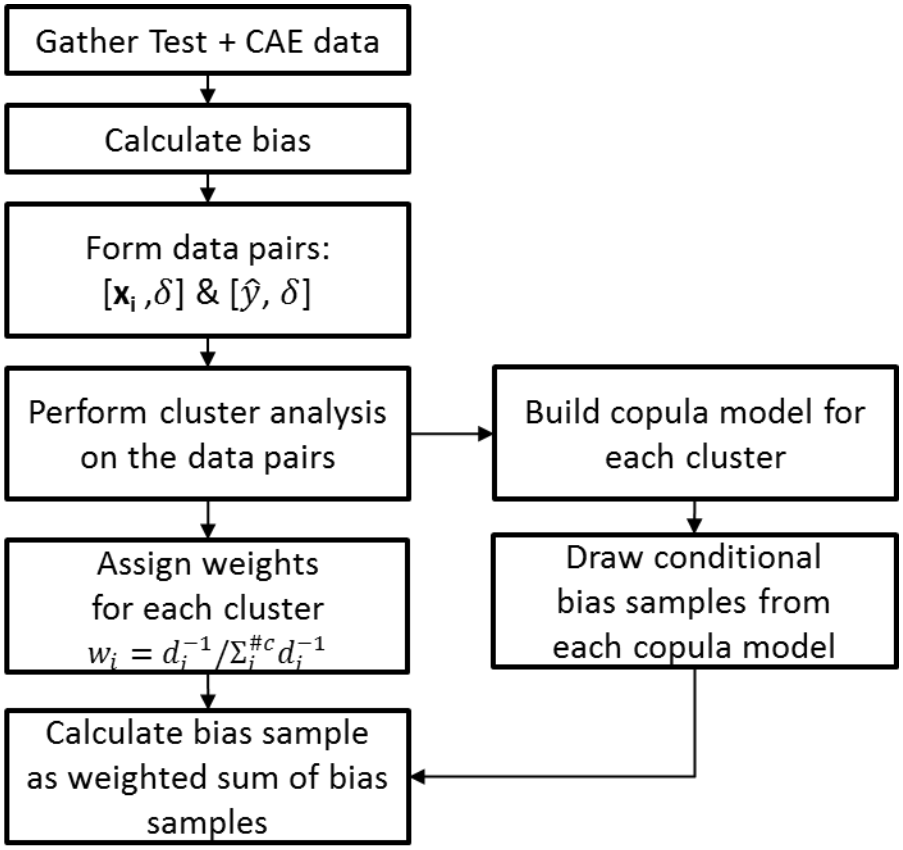


Figure 4.1: Steps of adaptive copula-based model bias characterization

4.2.1 k-means Clustering Analysis

The k-means clustering is the algorithm used here to preprocess data and group similar data points together. Doing so would enable a better fit of copula models to the data, which will be illustrated in our case studies.

The clustering procedure aims at segmenting the data in such a way that the within-cluster variation is minimized. It starts by randomly assigning objects to a number of clusters. The objects are then successively reassigned to other clusters to minimize the within-cluster variation, which is basically the (squared) distance from each observation to the center of the associated cluster. If the reallocation of an object to another cluster decreases the within-cluster variation, this object is reassigned to that cluster. With k-means, cluster affiliations can change in the course of the clustering process.

Prior to analysis, we have to decide on the number of clusters. We can tell how many segments are needed, or we may know from previous research what to look for. Based on this information, the algorithm randomly selects a center for each cluster.

Euclidean distances are computed from the cluster centers to every single object. Each object is then assigned to the cluster center with the shortest distance to it. We now have our initial partitioning of the objects into several clusters. And based on this initial partition, each cluster's geometric center (i.e., its centroid) is computed.

The distances from each object to the newly located cluster centers are computed and objects are again assigned to a certain cluster on the basis of their minimum distance to other cluster centers. Since the cluster centers' position changed with respect to the initial situation in the first step, this could lead to a different cluster solution.

The k-means procedure now repeats and re-computes the cluster centers of the newly formed clusters, and so on. In other words, steps 3 and 4 are repeated until a predetermined number of iterations are reached, or convergence is achieved (i.e., there is no change in the cluster affiliations).

Generally, k-means clustering is superior to other clustering methods such as hierarchical approaches as it is less affected by outliers and the presence of irrelevant clustering variables. Furthermore, k-means can be applied to very large datasets, as the procedure is less computationally demanding than hierarchical methods.

The following are the steps for carrying out a cluster analysis using partitioning approach:

- **Step 1**: Select k points as the initial centroids
- **Step 2**: Form k clusters by assigning each point to the closest centroid
- **Step 3**: Re-compute the centroid of each cluster
- **Step 4**: Go back to step 2 and repeat steps 2 and 3 until the centroids converge or a predetermined number of iterations has been reached

4.2.2 Determine Number of Clusters

Research has suggested several procedures for determining the number of clusters in a dataset. Most notably, the variance ratio criterion (VRC) by Calinski and Harabasz [73] has proven to work well in many situations.

The variance ratio criterion (also known as Calinski-Harabasz criterion) is formulated as

$$VRC_k = \frac{SS_B}{SS_W}\frac{(N-k)}{(k-1)} \tag{4.1}$$

where SS_B is the overall between-cluster variance, SS_W is the overall within-cluster variance, k is the number of clusters, and N is the number of observations.

The overall between-cluster variance SS_B is defined as

$$SS_B = \sum_{i=1}^{k} n_i \|m_i - m\|^2 \tag{4.2}$$

where k is the number of clusters, m_i is the centroid of cluster i, m is the overall mean of the sample data, and $\|m_i - m\|^2$ is the L^2 norm (Euclidean distance) between the two vectors.

The overall within-cluster variance SS_W is defined as

$$SS_W = \sum_{i=1}^{k} \sum_{x \in c_i} \|x - m_i\|^2 \tag{4.3}$$

where k is the number of clusters, x is a data point, c_i is the ith cluster, m_i is the centroid of cluster i, and $\|x - m_i\|^2$ is the L^2 norm (Euclidean distance) between the two vectors.

Well-defined clusters have a large between-cluster variance (SS_B) and a small within-cluster variance (SS_W). The larger the VRC_k ratio, the better the data partition. To

determine the optimal number of clusters, maximize VRC_k with respect to k. The optimal number of clusters is the solution with the highest Calinski-Harabasz index value.

It should be noted here that there are only few approaches that compares one cluster and multiple clusters. Hence, a likelihood-based approach is developed here to deal with this scenario. One cluster is preferred if the associated likelihood is higher compared to that calculated based on two clusters, and vice versa.

4.2.3 Weighting Scheme

After the cluster analysis performed on training data and the construction of the copula models associated to every cluster, weights associated to each of the copulas need to be determined so that our final weighted-sum prediction can be made. The weight assigned to cluster i is:

$$w_i = \frac{d_i^{-1}}{\Sigma d_i^{-1}} \tag{4.4}$$

where d_i is the distance between the confirmation point and the centroid of cluster i. The weight is constructed in such a way that it is larger when the distance between the confirmation point and the centroid of cluster i is small (cluster i has more influence for the final prediction than other clusters since it is closer to the point where the prediction takes place).

The final prediction is calculated as:

$$\hat{y} = \Sigma w_i \hat{y}_i \tag{4.5}$$

where $\hat{y}_i$ is the prediction made using copulas associated to cluster i.

In Eq. (4.4) the quantities d_i is calculated using Mahalanobis distance instead of the Euclidean distance to take into account the covariance information of the data. More specifically, Mahalanobis distance indicates how many standard deviations a point is from the center. It is unit-less and scale-invariant

The Mahalanobis distance is calculated as:

$$d_i = \sqrt{(x - \mu_i)^T S_i^{-1} (x - \mu_i)} \tag{4.6}$$

where $x = (x_1, x_2, \dots, x_n)$ is the coordinate of a specific confirmation data point, $\mu_i = (\mu_{i_1}, \mu_{i_2}, \dots, \mu_{i_n})$ is the coordinate of the centroid of a certain cluster, S_i is the covariance matrix and $i = 1,2$ is the index of the two closest clusters.

4.2.4 Sampling Approach for Prediction

There are cases where a confirmation data pair is completely outside of a specific copula model prediction range. In this case the weight associated to that cluster will be set to zero. This case also introduces the need to develop a new sampling scheme to overcome this situation where the number of Copula samples is low or even zero. This can potentially lead to inaccurate prediction.

Here a new sampling scheme is constructed. First the data are transformed to the copula domain. Then the best type of copula and its parameters will be determined using our previously proposed approach. Instead of generating a fixed number of samples for the entire copula, now a fixed number of samples are generated for each confirmation data pair. The sampling is done by first obtaining the conditional PDF and CDF using Eq. (4.7), where X and Y correspond to the confirmation CAE and bias transformed to the copula domain. Then a fixed number of samples will be generated from a uniform distribution on the interval [0,1]. The final samples will be obtained by calculating the values such that conditional CDFs at these values are equal to each of the random uniform samples. Finally these samples will be transformed back to the original data space.

$$P(Y = y | X = x) = \frac{P\big(Y = y, \quad X = x\big)}{P(X = x)} \tag{4.7}$$

4.3 Case Studies

4.3.1 Artificially Constructed Example Problem

In this example data are generated from two distributions. In case 1, the data group on the left are simulated samples from a bivariate Gaussian distribution with mean vector [-8,-7] and the covariance matrix [1 1.5; 1.5 3], while the data group on the right are simulated samples from a bivariate Gamma distribution with parameter $k = [1.2\ 1.8]$ and $\theta = [2\ 0.5]$. Data are shown in Fig. 4.2, where CAE stands for computer aided engineering and is essentially representing results from a computer model.

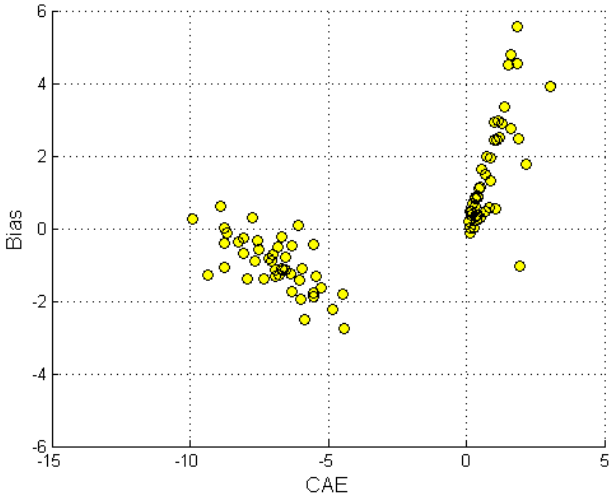


Figure 4.2: Data from case 1

In case 2, the data group on the left are simulated samples from a bivariate Gaussian distribution with mean vector [-2,-1] and the covariance matrix [1 1.5; 1.5 3], while the data group on the right are simulated samples from a bivariate Gamma distribution with parameter $k = [1.2\ 1.8]$ and $\theta = [2\ 0.5]$. Data are shown in Fig. 4.3. The only difference between case 1 and case 2 is the mean vector of the bivariate Gaussian distribution. The distance between two clusters is larger in case 1 than that in case 2.

Model prediction accuracy made by the proposed adaptive copula-based model bias characterization approach is compared with that made by its predecessor and the result is shown in Fig. 4.4. As we see, training data points are clustered into different groups and are represented by big dots using two different colors (red and yellow). The copula sample

points are shown as small black dots. Fig. 4.4 (b) corresponds to the result made by the proposed approach while Fig. 4.4 (a) corresponds to those made by its predecessor (so called single copula approach). For a specific confirmation point, prediction as well as the 95% confidence bound are shown in blue. Immediately from the graph we can see that the confidence bound produced by the proposed approach is significantly narrower than that produced by its predecessor. Besides the width of the confidence bound, the prediction accuracy is improved and is shown in Table 4.1. Using the proposed approach, the average error is about 50% less using the proposed approach, and for a specific confirmation point, the bias is characterized so that it is closer to the true bias value than the previously proposed approach.

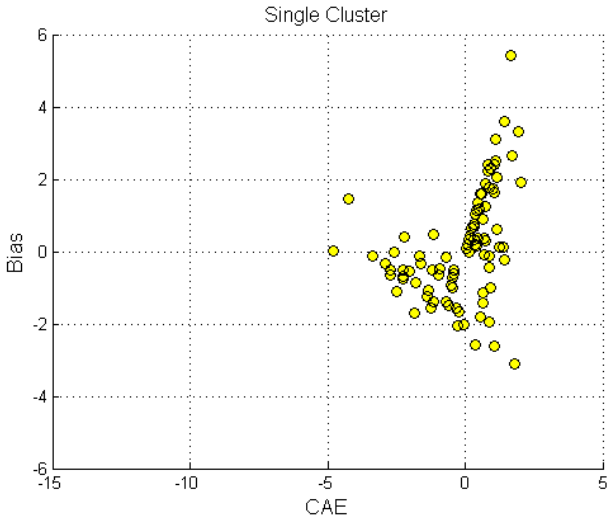
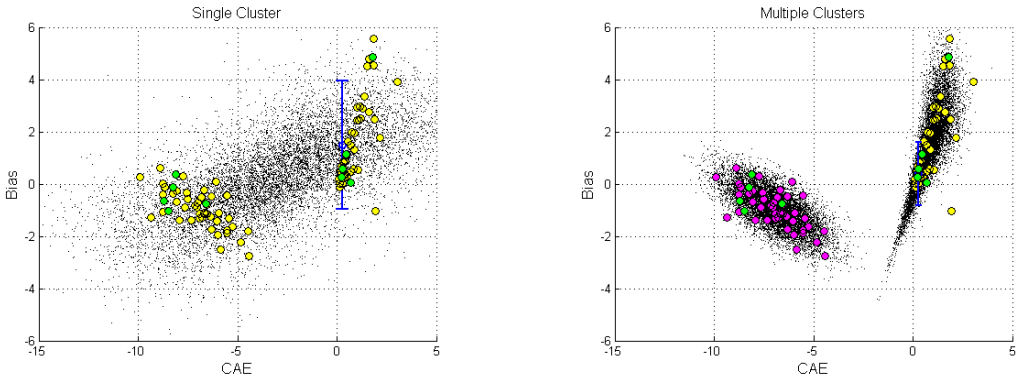


Figure 4.3: Data from case 2



(a) (b)

Figure 4.4: Results from case 1, (a) single cluster; (b) adaptive copula

Table 4.1: Average error and bias of a specific confirmation point from case 1

Average Error		Bias		
Single Copula	Adaptive Copula	True	Single Copula	Adaptive Copula
1.14	0.64	0.60	1.50	0.40

The absolute error results presented in Tab. 4.1 are also visualized in Fig. 4.5 as bar plot.

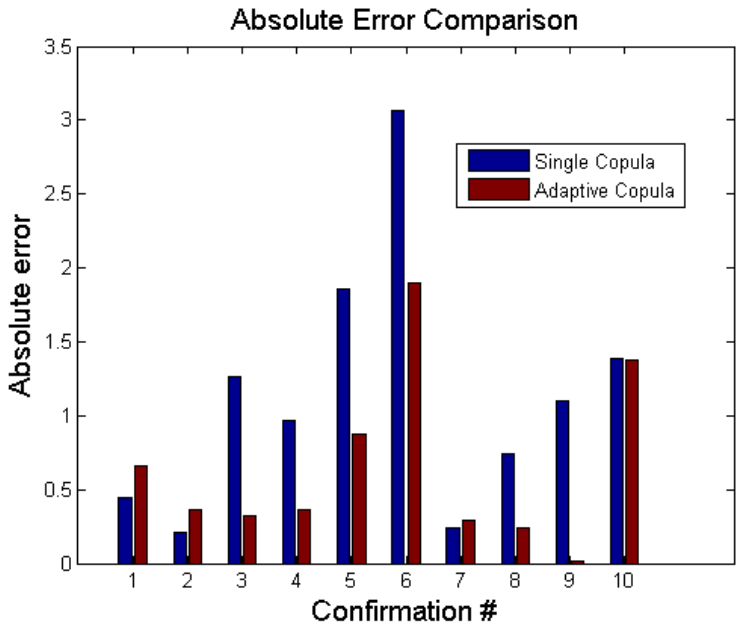
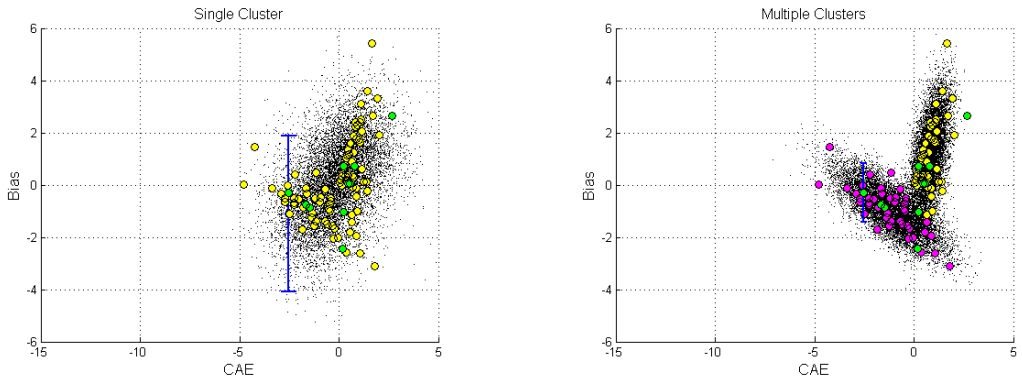


Figure 4.5: Absolute error comparison: single copula vs adaptive copula for case 1

Similar results can be observed for case 2.



(a) (b)

Figure 4.6: Results from case 2, (a) single cluster; (b) adaptive copula

Table 4.2: Average error and bias of a specific confirmation point from case 2

Average Error		**Bias**		
Single Copula	Adaptive Copula	True	Single Copula	Adaptive Copula
0.80	0.62	-0.30	-1.10	-0.29

The absolute error results presented in Tab. 4.2 are also visualized in Fig. 4.7 as bar plot.

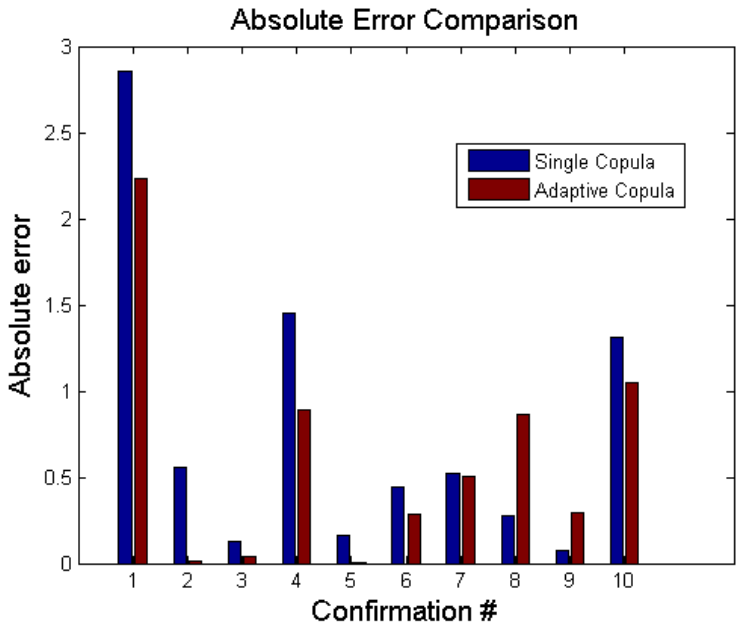


Figure 4.7: Absolute error comparison: single copula vs adaptive copula for case 2

It should be noted here that in both case 1 and case 2, there are points which are outside the range of copula sample points from a specific cluster. In this scenario, the associated weight is set to be zero.

4.3.2 Barnes Problem

The Barnes problem originated from Barnes' master's thesis [71]. It is a study of four optimization sub-problems: one high-fidelity (HF) model and its three variants (low-fidelity (LF) models A, B and C). Only LF model A is used here. The HF model and the LF model A are described in Tables 4.3 and 4.4, respectively.

Table 4.3: Description of the Barnes problem, high-fidelity model

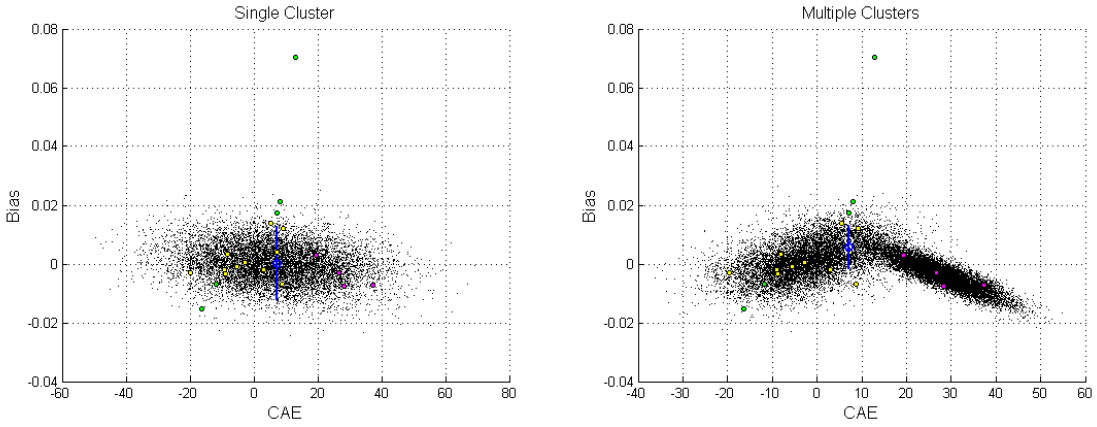
High-fidelity Model
$f_{high} = a_1 + a_2x_1 + a_3x_1^2 + a_4x_1^3 + a_5x_1^4 + a_6x_2 + a_7x_1x_2 + a_8x_1^2x_2 + a_9x_1^3x_2 + a_{10}x_1^4x_2 + a_{11}x_2^2 + a_{12}x_2^3 + a_{13}x_2^4 + \frac{a_{14}}{x_2+1} + a_{15}x_1^2x_2^2 + a_{16}x_1^3x_2^2 + a_{17}x_1^3x_2^3 + a_{18}x_1x_2^2 + a_{19}x_1x_2^3 + a_{20}e^{a_{21}x_1x_2}$
$g_{1high} = 1 - \frac{x_1x_2}{700} \leq 0$
$g_{2high} = \frac{x_1^2}{625} - \frac{x_2}{5} \leq 0$
$g_{3high} = \left(\frac{x_1}{500} - 0.11\right) - \left(\frac{x_2}{50} - 1\right)^2 \leq 0$

High-fidelity model has a high-degree polynomial for the objective function and non-linear constraints while low-fidelity models have low-degree polynomials and more linear constraints. All models have two variables and they are both bounded between 0 and 70. For ease of understanding, none of the constraints was used and only the objective function was calculated to produce the model output. Out of the various configurations selected by using the Latin hypercube design (20 configurations), a portion of them (15 configurations) was used as training data to obtain the copula model, while the rest (5 configurations) was used to demonstrate the improvement the proposed approach can make.

Table 4.4: Description of the Barnes problem, low-fidelity model A

Low- Fidelity Model A
$f_{low} = a_1 + a_2x_1 + a_3x_1^2 + a_4x_1^3 + a_6x_2 + a_7x_1x_2 + a_8x_1^2x_2 + a_9x_1^3x_2 + a_{10}x_1^4x_2 + a_{11}x_2^2 + a_{12}x_2^3 + a_{13}x_2^4 + \frac{a_{14}}{x_2+1} + a_{15}x_1^2x_2^2 + a_{16}x_1^3x_2^2 + a_{17}x_1^3x_2^3 + a_{18}x_1x_2^2 + a_{19}x_1x_2^3 + a_{20}$
$g_{1low} = 0.5 - \frac{x_1x_2}{750} \leq 0$
$g_{2low} = \frac{x_1}{16} - \frac{x_2}{5} \leq 0$
$g_{3low} = \left(\frac{x_1}{500} - 0.11\right) - \left(\frac{x_2}{50} - 1\right)^2 \leq 0$

The output calculated by the high-fidelity model is compared with that produced by the low-fidelity model A, as depicted in Fig. 4.8.



(a) (b)

Figure 4.8: Results from low-fidelity model A, (a) single cluster; (b) adaptive copula

Table 4.5: Average error and bias of a confirmation point from low-fidelity model A

Average Error		Bias		
Single Copula	Adaptive Copula	True	Single Copula	Adaptive Copula
0.065	0.0042	0.008	0.0001	0.0054

The absolute error results presented in Tab. 4.5 are also visualized in Fig. 4.9 as bar plot.

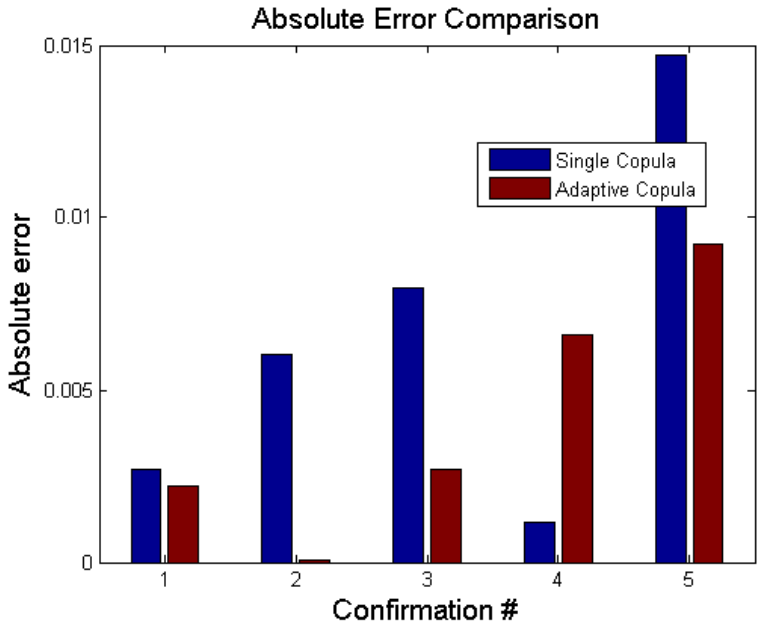


Figure 4.9: Absolute error comparison: single copula vs adaptive copula for Barnes problem high-fidelity model vs. low-fidelity model A

A hypothesis testing was used to see if the average absolute error produced by the adaptive copula approach is different than that from the single copula approach. Since there are so few sample points (5 confirmation points), bootstrapping technique was employed to generate sufficiently large number of bootstrap samples (1000). In this way, a large number average absolute error (mean value) can be produced (one mean value for each bootstrap sample, for a total of 1000 mean values in this case). A one-sample *t*-test was established with the null hypothesis H_0 set up as $H_0: \mu_{SC} - \mu_{AC} = 0$. The null hypothesis was rejected at 5% significance level indicating there is significant difference between the average absolute error produced by the adaptive copula approach and that by the single copula approach. Such statistical evidence suggests that the proposed methodology, the adaptive copula approach was able to improve the error statistic significantly compared to the previously-proposed single copula approach.

4.4 Summary

In this chapter, an adaptive copula-based approach was proposed to improve model accuracy by correcting the bias term. Cluster analysis is performed on the raw data to group similar data points, followed by copula modeling for each cluster. Model prediction is then produced using information from each cluster. The final prediction is the weighted sum of every prediction. The two case studies used in this chapter adequately demonstrated that the proposed approach is effective in improving the accuracy of model prediction compared to its predecessor which uses single copula. The proposed approach is able to produce more accurate model prediction as well as narrowed confidence bounds.

Chapter 5: Model Validation Metric for Dynamic Responses under Uncertainty

5.1 Introduction

Majority of available validation metric is designed for static system response such as the U-pooling [27] and Bayes factor [74]. Though metrics for dynamic responses are also available [75], they are specifically designed for vehicle impact application and uncertainties are not well considered in the metric. In this section we propose the statistic validation metric for dynamic responses.

5.2 Statistic Validation Metric for Dynamic Responses

The proposed validation metric is related to the U-pooling metric, hence this metric is briefly reviewed here. U-pooling metric was proposed by Ferson et al. [27] as a validation metric and has been adopted by many researchers in the study of model validation. The basic idea is to compare the cumulative distribution function (CDF) difference (i.e., the U-pooling value) between model prediction and test data in the standard Uniform space (or U-space) as shown in Fig. 5.1. The smaller the area difference, the higher of the expected accuracy of the model prediction. For static responses, each test datum y_i corresponds to one u_i value which is calculated from the CDF value of the model prediction at the same design configuration (i.e., $u_i = F_{\hat{Y}}(y_i)$ where $F(\cdot)$ is the CDF of $\hat{Y}$). To adopt the U-pooling value for dynamic responses, $F_{\hat{Y}}(y_i)$ should be calculated where y_i is a dynamic response from test and $\hat{Y}$ is an arbitrary random process from model prediction.

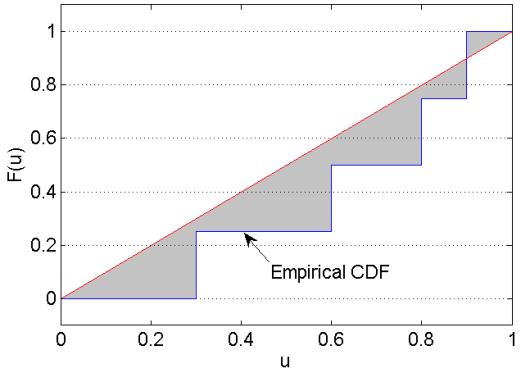


Figure 5.1: Illustration of the U-pooling value

The proposed validation metric makes use of the U-pooling approach by addressing two technical components: i) extension of the U-pooling for dynamic responses; and ii) inclusion of the shape difference in the metric. The first component is addressed by discretizing the random process into n time steps such that the calculation of each u_i equivalents to identifying the n-dimensional joint CDF value for each random process realization of y_i. The second component is addressed by quantifying the shape deviation of the model prediction from the test. Technical details are elaborated as follows.

Under the assumption of Gaussian random process, the value of u_i can be estimated numerically using a number of algorithms since there is no closed form for n-dimensional normal joint CDF [76]. Considering an arbitrary random process, such calculation could be practically difficult. Generally, the sampling approach is able to approximate the u_i by counting the number of random process realizations less than or equal to the corresponding realization of y_i from the test as shown in Eq. (5.1).

$$u_i = F_{\hat{Y}}(y_i) \cong \sum_{k=1}^{N} \frac{M_k}{N}, \text{where } M_k = \begin{cases} 1 & \hat{y}_{k,j} \le y_{i,j}\ j = 1, \dots, n \\ 0 & elsewhere \end{cases} \quad (5.1)$$

where N is the total number of random process realizations from the model. It is worth nothing that above calculation is general but computationally expensive because each u_i estimation requires a magnitude of $N \times n$ comparison operations. With several test observations of y_i, $i = 1, \dots, Q$, corresponding u_i can be calculated using Eq. (5.1) resulting in a U-pooling value as shown in Fig. 5.1.

Unlike static responses, shape difference of dynamic responses is another important attribute between model prediction and test, which is not able to be accounted for in Eq. (5.1). In other words, the same u_i value could represent different quality of the response matching as shown in Fig. 5.2a. Therefore, it is desirable to include the shape difference in the validation metric for dynamic system responses. Under the assumption that model prediction is valid and truly represents the real physical systems, test observations equivalent to corresponding random process realizations from the model prediction. Therefore, the expected shape deviation between test observations and corresponding realizations from the model should be as small as possible. Otherwise, the assumption is not valid and the magnitude of such shape deviation indicating the level of disagreement between model prediction and test. Two remaining challenges need to be addressed to include the shape difference in the validation metric: i) quantification of the shape deviation; and ii) identification of the corresponding realizations from the model.

Quantification of the shape deviation is proposed to be conducted in a normalized region as shown in Fig. 5.2b so that its magnitude is at the same level as the U-pooling value. Assign $S_i(y, \hat{y})$ as the shape deviation between y_i and the corresponding random process realization from the model, its value is the shaded area as shown in Fig. 5.2b. The corresponding realizations from the model based on test observations can be identified using Eq. (5.2).

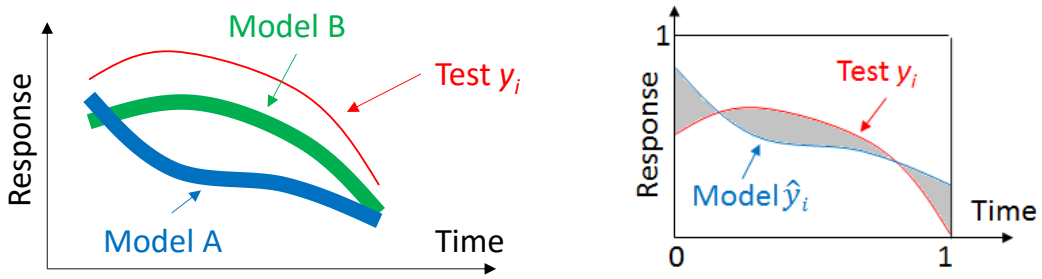
$$\hat{y}_i = F_{\hat{Y}}^{-1}\left[\frac{i}{Q+1}\right] \tag{5.2}$$

where Q is the number of test observations; i is the order of the test observations; and $F_{\hat{Y}}^{-1}(\cdot)$ is the inverse CDF of $\hat{Y}$. For example, one test observation corresponds to $F_{\hat{Y}}^{-1}(0.5)$ indicating the median of the random process or the mean for Gaussian random process. Two test observations correspond to $F_{\hat{Y}}^{-1}(0.33)$ and $F_{\hat{Y}}^{-1}(0.67)$. Identification of the corresponding realizations meeting the CDF values can be conducted from the marginal distributions.

The proposed validation metric for dynamic responses is formulated in Eq. (5.3).

$$\Psi = U(u_1, \ldots, u_Q) + \frac{1}{Q}\sum_{i=1}^{Q} S_i(y, \hat{y}) \tag{5.3}$$

where the first part is the U-pooling value and the second part is the average of shape deviation. The smaller value of Ψ indicates the better model accuracy based on available test observations. It is worth noting that calculation of the proposed metric is computationally expensive for each u_i.



(a) Shape difference with the same u_i value (b) Shape deviation in normalized region

Figure 5.2: Illustration of the shape difference for dynamic responses

Majority of the computational effort of the proposed validation metric is conducted on the basis of n-dimensional joint PDF/CDF by discretizing the dynamic responses into n time steps. Due to the statistical dependence among marginal distributions, the sampling approach may be the only feasible way for calculating the u_i value. The principle component analysis (PCA) (or proper orthogonal decomposition) is a typical method to significantly reduce the dimensionality of many correlated random variables into only a few uncorrelated random variables (or principle components) with prescribed accuracy requirement [77]. Therefore, preprocessing the dynamic system responses using the PCA can significantly improve the computational efficiency of the proposed validation metric because the computation can be conducted in a statistically uncorrelated random space.

The dynamic response $\widehat{\boldsymbol{Y}}$ can be decomposed into mean $\boldsymbol{\mu}$ and variation $\boldsymbol{\nu}$. The $\boldsymbol{k^{th}}$ random process realization can be generally expressed as

$$\hat{Y}_k = \mu + \nu_k \tag{5.4}$$

Using the PCA approach, a random process can be represented by Eq. (5.5) given sufficient number of random process realizations.

$$\hat{Y}(t) = \mu(t) + \sum_{k=1}^{n} \alpha_k \phi_k(t) \tag{5.5}$$

where t is the time; $\phi_k(t)$ is the k^{th} eigenvector (or principle component) of the covariance matrix of *n* random variables from the discretized dynamic responses; α_k is the coefficient of the k^{th} eigenvector and its value can be obtained through the inner product operation between the variation part $v(t)$ and the corresponding eigenvector as

$$\alpha_k = v(t) \cdot \phi_k(t) \tag{5.6}$$

Theoretically, all eigenvectors are required to exactly represent the random process. However, only a few important eigenvectors may be vital for the representation of the random process especially when original random variables are highly correlated. In Eq. (5.6), a dataset of the coefficient α_k of the k^{th} eigenvector can be obtained from all random process realizations. Hence, v_k can be defined as a new random variable projected to the k^{th} eigenvector (or principle component) that statistically models the coefficient data set α_k. Therefore, the random process is represented as a function of deterministic eigenvector ϕ_k and the corresponding random variables V_k as explained in Eq. (5.7).

$$\hat{Y}(t) \cong \mu(t) + \sum_{k=1}^{m} V_k \phi_k(t) \tag{5.7}$$

where m is the number of the vital eigenvectors. Marginal distributions of V_k can be easily determined using the maximum likelihood estimation (MLE) given sufficient number of random samples (e.g. $N = 1000$). It is worth noting that Eq. (5.7) essentially reduces n-dimensional random space into m dimension, where the new random variables V_k's are statistically uncorrelated and m is typically a much smaller number than n.

Joint CDF value can be easily computed if random variables are statistically independent as defined in Eq. (5.8).

$$F_v(v^*) = P(v_1 \le v_1^*, \dots, v_m \le v_m^*) = \prod_k P(v_k \le v_k^*) \tag{5.8}$$

where v_k^* is the projected value of the test observation y_i onto the k^{th} principle component. Eq. (5.8) may introduce errors if V_k's are statistically dependent because statistical independence is not ensured even though V_k's are statistically uncorrelated. Nevertheless, Eq. (5.8) significantly improves computational efficiency compared to Eq. (5.1). In

addition, the statistical dependence of V_k is rarely observed in the tested engineering problems. It is worth noting that the joint CDF value (i.e., u_i) calculated in Eq. (5.1) may not equals to the value in Eq. (5.8) because of the orthogonally rotated coordinate systems using the PCA. However, the U-pooling value still reflects the degree of agreement between the model prediction and test observations, which will be shown in the case study.

5.3 Case Studies

A driver-side occupant restraint system, a MADYMO model, is used to demonstrate the proposed validation metric before and after model validation. The model simulates a full frontal rigid barrier crash scenario at the speed of 35 mph with a 50^{th} percentile belted Hybrid III dummy in a vehicle. This represents the USA New Car Assessment Program (NCAP) test mode. Each MADYMO simulation takes approximately 7 minutes to complete on a HP C8000 workstation. There are eleven occupant response curves (summarized in Table 5.1) that are monitored and compared with the test data to evaluate the accuracy of the MADYMO model. Sixteen model input random parameters were selected on the basis of three reasons including: i) a significant range of variation shown in component tests (e.g., stiffness); ii) absence of representative component test data (e.g., friction); and iii) non-controllable parameters (e.g., impact load magnitude and location). Their lower and upper bounds were chosen based on the component tests and engineering experience, and their distributions were assumed to be uniform.

Table 5.1: Eleven Occupant Responses

Dynamic Responses	Description
Response 1	Belt Load at Anchor
Response 2	Belt Load at Retractor
Response 3	Belt Load at Shoulder
Response 4	Chest Deflection
Response 5	Chest Acceleration in X-Direction
Response 6	Femur Load Left in Z-Direction
Response 7	Femur Load Right in Z-Direction
Response 8	Head Acceleration in X-Direction
Response 9	Upper Neck Load in Z-Direction
Response 10	Upper Neck Moment
Response 11	Pelvis Acceleration in X-Direction

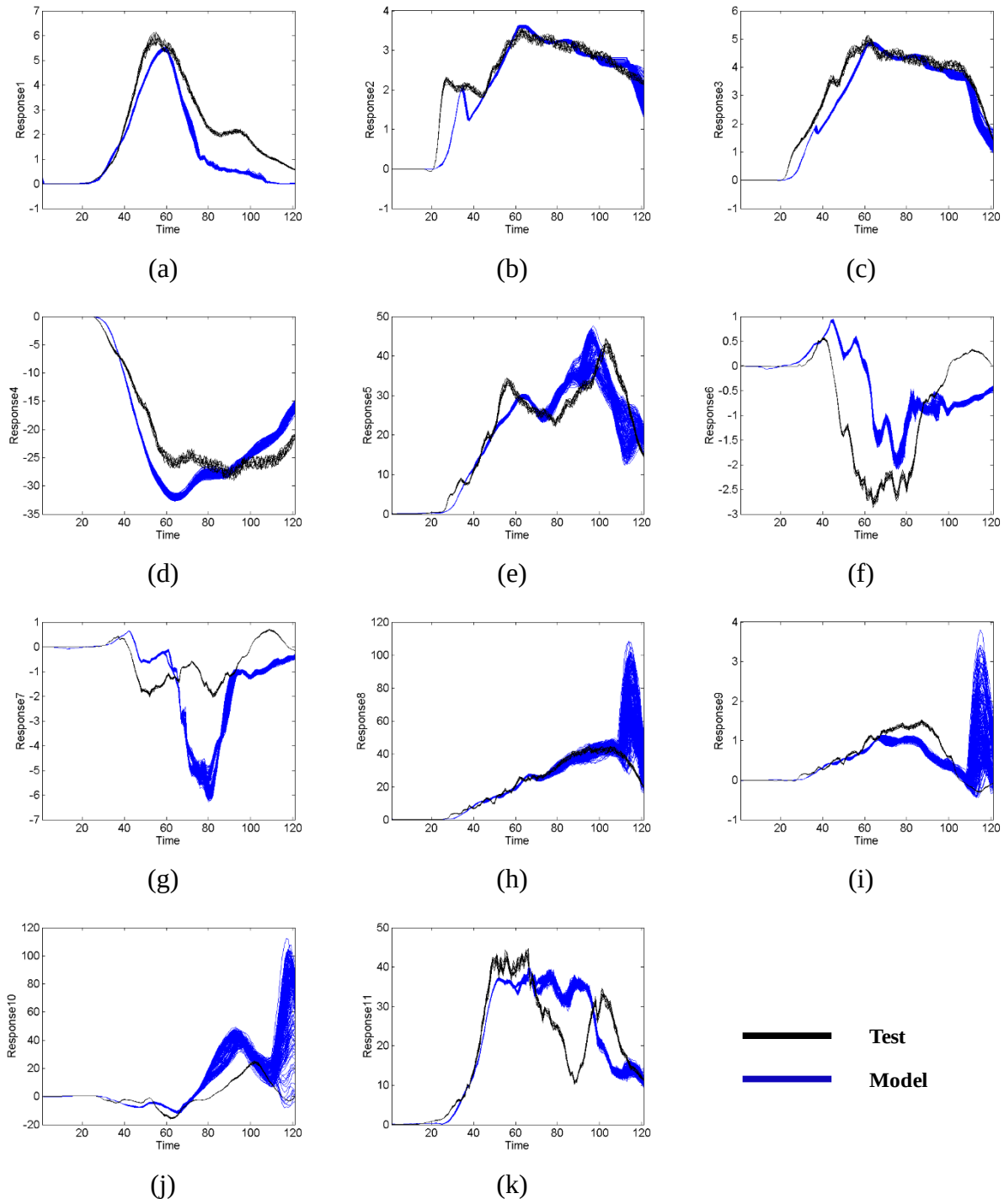


Figure 5.3: Model A vs. test

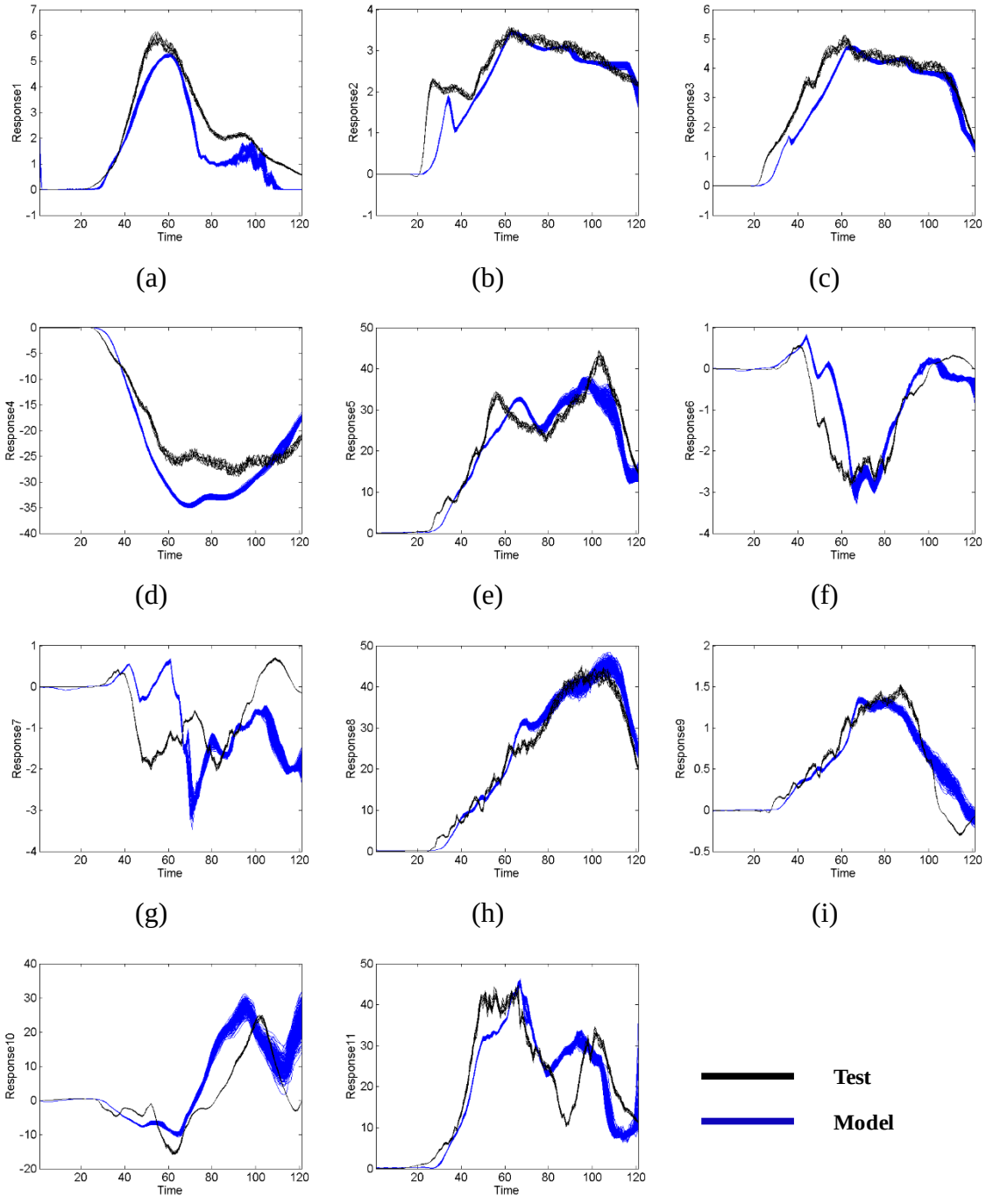


Figure 5.4: Model B vs. test

Figs. 5.3 and 5.4 show the time history responses of two model predictions (i.e., model A and B) against the same set of experimental data. The results from the model prediction consist of 200 computer simulations using the MCS, and the experimental data consist of

10 repeated tests. It is difficult to observe directly which model is more accurate. Hence, validation metric is necessary for quantitative comparison.

Table 5.2: Validation metric for model A and B

Validation Metric	Model A				Model B			
	U	U^*	S	Ψ	U	U^*	S	Ψ
Resp. #1	0.459	0.459	0.074	0.533	0.459	0.459	0.062	0.521
Resp. #2	0.459	0.459	0.048	0.507	0.459	0.459	0.059	0.518
Resp. #3	0.459	0.459	0.050	0.509	0.459	0.459	0.053	0.512
Resp. #4	0.459	0.459	0.071	0.530	0.459	0.459	0.090	0.549
Resp. #5	0.459	0.459	0.066	0.525	0.459	0.459	0.060	0.519
Resp. #6	0.459	0.459	0.144	0.603	0.459	0.459	0.076	0.535
Resp. #7	0.459	0.459	0.193	0.652	0.459	0.459	0.146	0.605
Resp. #8	0.459	0.459	0.075	0.534	0.459	0.459	0.039	0.498
Resp. #9	0.459	0.459	0.102	0.561	0.459	0.459	0.052	0.511
Resp. #10	0.459	0.459	0.175	0.635	0.459	0.459	0.091	0.550
Resp. #11	0.459	0.459	0.092	0.551	0.459	0.459	0.078	0.538
Average	**0.459**	**0.459**	**0.099**	**0.558**	**0.459**	**0.459**	**0.073**	**0.532**

Table 5.2 lists results computed from the proposed validation metric for eleven dynamic responses, where U indicates the U-pooling value, S means the average shape deviation, Ψ is the summation of them, and U^* represents the U-pooling value calculated using the PCA approach. Several valuable observations are summarized as follows. First of all, U equals to U^* for all responses demonstrating that the PCA approach for calculating the U-pooling value is equivalent to the sampling approach. It is noted in this example that 99.9% of accuracy requirement was implemented in the PCA, otherwise, some difference may be observed due to neglecting some portion of the total data variation. Secondly, the U-pooling value is relatively large (i.e., 0.459) and the same for all responses, which is mainly due to the fact that test data locate outside the boundary of the model prediction at many time steps for both models. In fact, even if only at one time step with such feature, the large U-pooling value is expected because each u_i value is calculated from the perspective of a joint CDF. Such a strict attribute may not be appreciated in this example, however, it would be very important in other applications where model prediction at each time step is critical (e.g., the strength decay of a bridge structure every year). Thirdly, model B has smaller shape deviation than model A for majority responses except for responses 2, 3 and 4.

Quantitatively, the average shape deviation of model B is reduced about 26% compared to model A. Fourthly, the top 3 responses with large shape deviation for model B is the response 7, 10, and 4.

5.4 Summary

A new validation metric for dynamic responses under uncertainty was proposed in this chapter. The classical U-pooling approach can be extended for dynamic responses by discretizing and treating the responses as a high dimensional joint distribution. The PCA is very effective for representing the dynamic responses by a few random variables so that the U-pooling value can be computed more efficiently. The shape deviation was included in the validation metric so that the metric can still distinguish the model accuracy when the U-pooling value only is not differentiable. The proposed shape deviation metric works effectively with multiple test data and distinguishes not only the mean shape difference but also the difference at corresponding percentage levels. One vehicle impact model was employed and the effectiveness of the proposed validation metric was demonstrated. Further work is to apply the metric for other engineering applications where dynamic responses are prominent.

Chapter 6: Model Bias Modeling of Dynamic System Responses

6.1 Introduction

Given limited test data, model bias needs to be effectively identified so that reliability analysis would be more accurate compared to the baseline model. In particular, it is desirable that model bias can be accurately approximated at any new design configurations where test data are not available. This chapter will first elaborate on the dynamic model bias calibration with the aid of the validation metric, then propose the dynamic model bias approximation at any new design configurations.

6.2 Dynamic Model Bias Calibration

Majority of the bias correction approach is based on the Bayesian calibration model proposed by Kennedy and O'Hagan [59] as shown in Eq. (6.1).

$$\hat{Y}(\boldsymbol{X}) + \delta = Y - \varepsilon \tag{6.1}$$

where $\hat{Y}$ is system performance prediction of the baseline simulation/analytical model, δ is model bias, Y is test data, ε is test error, and $\boldsymbol{X}$ is a vector of model parameters. In particular, model parameter $\boldsymbol{X}$ could take the form of deterministic values, irreducible random parameters (i.e., aleatory uncertainty), or reducible random parameters (i.e., epistemic uncertainty), depending on the nature and available data of the problem. The Bayesian calibration model addresses the challenge of calibrating model parameter $\boldsymbol{X}$, model bias δ, and test error ε with limited test data Y at given design configurations. Essentially, prior distributions of these calibration terms are updated to posterior distributions provided the evidence of test data Y using a Bayesian updating mechanism.

The calibration model as shown in Eq. (6.1) has been used to calibrate the non-dynamic model bias with well-defined calibration parameters. For example, either the distribution

parameters (e.g., the mean and standard deviation) or their hyper-parameters can be treated as calibration parameters with assumed distribution types of the model bias. This calibration model can be also used for dynamic model bias calibration if the calibration parameters can be properly defined.

Considering arbitrary dynamic model bias under uncertainty, the principle component analysis (PCA) (or proper orthogonal decomposition) is a typical method to characterize the dynamic randomness using only a few uncorrelated random variables (or principle components) with prescribed accuracy requirement [77]. An arbitrary dynamic model bias can be formulated as shown in Eq. (6.2).

$$\delta(t) \cong \mu(t) + \sum_{k=1}^{m} V_k \phi_k(t) \tag{6.2}$$

where t is the time; $\mu(t)$ is the mean of the dynamic model bias; $\phi_k(t)$ is the k^{th} important feature; V_k's are the uncorrelated random variables with zero means; and m is the number of the vital features. Here V_k will be considered as calibration parameters and $\mu(t)$ and $\phi_k(t)$ will be determined on the basis of the baseline model prediction and corresponding test data. In particular, sufficient model bias realizations can be obtained using the sampling approach based on Eq. (6.1). With discretization of the dynamic model bias into n time steps, an $m \times n$ matrix representing the bias can be constructed as

$$\delta(t) \cong \boldsymbol{\delta} = \begin{bmatrix} \delta_{11} & \cdots & \delta_{1n} \\ \vdots & \ddots & \vdots \\ \delta_{m1} & \dots & \delta_{mn} \end{bmatrix} \tag{6.3}$$

where δ_{ij} indicates the model bias at the j^{th} time step for the i^{th} sampled realization. The mean of the model bias is estimated as

$$\mu(t) \cong \boldsymbol{\mu} = [\bar{\delta}_{\cdot 1}, \quad \dots, \quad \bar{\delta}_{\cdot n}] \tag{6.4}$$

where $\bar{\delta}_{\cdot j}$ stands for the average of the j^{th} time step over the sampled realizations. Hence the variation part of the model bias is expressed as

$$\boldsymbol{\nu} = \begin{bmatrix} \delta_{11} - \bar{\delta}_{\cdot 1} & \cdots & \delta_{1n} - \bar{\delta}_{\cdot n} \\ \vdots & \ddots & \vdots \\ \delta_{m1} - \bar{\delta}_{\cdot 1} & \dots & \delta_{mn} - \bar{\delta}_{\cdot n} \end{bmatrix} \tag{6.5}$$

The feature vector $\boldsymbol{\phi}$ can then be obtained by solving an eigen-problem as

$$\boldsymbol{\Sigma\phi} = \lambda\boldsymbol{\phi} \tag{6.6}$$

where λ is the eigenvalue of the covariance matrix $\mathbf{\Sigma}(\mathrm{n} \times \mathrm{n})$ that is defined as

$$\mathbf{\Sigma} = \boldsymbol{\nu}^{\mathbf{T}}\boldsymbol{\nu} \tag{6.7}$$

The number of the vital features m is dependent on the required accuracy level for preserving the variation part of the matrix $\boldsymbol{\nu}$, which is determined by the ratio of the first m largest eigenvalues to the summation of all eigenvalues.

Dynamic model bias calibration is thus formulated as

$$\begin{aligned} \min \quad & \Psi\left(\hat{Y} + \delta,\ Y - \varepsilon\right) \\ S.T. \quad & L_k^{\sigma} \le \sigma(V_k) \le U_k^{\sigma} \\ & L_k^{skew} \le skew(V_k) \le U_k^{skew} \\ & L_k^{kurt} \le kurt(V_k) \le U_k^{kurt} \end{aligned} \tag{6.8}$$

where $\Psi(\cdot)$ is the validation metric function; $\sigma(V_k)$, $skew(V_k)$, and $kurt(V_k)$ are the standard deviation, skewness, and kurtosis of random variable V_k, respectively; L_k^x and U_k^x are the corresponding lower and upper bounds. It is worth noting that if Gaussian random field was assumed for dynamic model bias, only $\sigma(V_k)$ is needed in the above calibration model, otherwise, more advanced PDF approximation methods (e.g., Pearson System, Johnson System, maximum entropy principle, etc.) [78, 79, 80, 81] can be used to approximate the distribution of V_k with inclusion of the $skew(V_k)$ and $kurt(V_k)$.

6.3 Dynamic Model Bias Approximation in the Design Space

With successful calibration of dynamic model bias represented in the form of Eq. (6.2) at available design configurations, the model bias at any new design configuration can be approximated by using the response surface methods. In particular, $\mu(t)$ and $\phi_k(t)$ are deterministic for a given design configuration, and V_k follows an arbitrary distribution whose statistical moments can be extracted for building the response surface models. Four steps are summarized to approximate the dynamic model bias in the design space with the aid of the response surface methods.

- **Step 1**: identify optimal dynamic model bias using Eq. (6.8) under several design configurations;
- **Step 2**: construct response surfaces for $\mu(t)$, $\phi_k(t)$ and the central moments of V_k;
- **Step 3**: approximate $\mu(t)$, $\phi_k(t)$ and the central moments of V_k of the model bias at new design configurations on the basis of the response surfaces;
- **Step 4**: approximate the distributions of V_k at new design configurations using one of the advanced PDF approximation methods.

Response surface of the model bias plays a critical role for correcting the baseline model prediction with a more credible reliability analysis by incorporating the uncertainty of the model. Its accuracy mainly depends on three factors including: i) nonlinearity of the model bias in the design space; ii) number of identified model bias in the design space; and iii) algorithm of the response surface method. The first factor is determined by the quality of the baseline model, while the second factor is decided by available resources for conducting validation experiments at different design configurations. Intuitively, more validation experiments increase accuracy of the response surface. The third factor belongs to the study of response surface methodology [82].

6.4 Case Studies

Two case studies including a thermal problem and a beam problem are employed to demonstrate the proposed validation metric, dynamic model bias calibration and approximation, and reliability analysis.

6.4.1 Thermal Problem

A thermal problem was developed by Sandia National Laboratory for the study of model validation [83]. The subject is a safety-critical structure, which is schematically shown in Fig. 6.1. The structure is subject to a constant heat flux rate q with the thickness L. Temperature of the structure at location x is expressed in Eq. (6.9).

$$T(x,t) = T_0 + \frac{qL}{k}\left[\frac{\left(\frac{k}{\rho}\right)t}{L^2} + \frac{1}{3} - \frac{x}{L} + \frac{1}{2}\left(\frac{x}{L}\right)^2 - \frac{2}{\pi^2}\sum_{n=1}^{6}\frac{1}{n^2}exp\left(-n^2\pi^2\frac{\left(\frac{k}{\rho}\right)t}{L^2}\right)cos\left(\frac{n\pi x}{L}\right)\right] \qquad (6.9)$$

where T_0 is the initial temperature; k is the thermal conductivity; and ρ is the volumetric heat capacity. Specifically, parameters k and ρ have unit-to-unit variability due to the manufacturing tolerance and they were assumed to follow Normal distributions with the mean of 0.0820 and standard deviation of 0.0012 for k and the mean of 400000 and standard deviation of 30113 for ρ. For safety reason, probability that surface temperature exceeds a failure temperature, i.e. $T = 900℃$, should be less than 1% after exposure to a heat flux $q = 3500\, W/m^2$ at time $t = 1000\ (s)$.

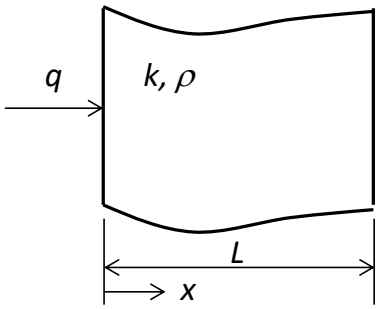


Figure 6.1: Schematic view of a safety-critical structure

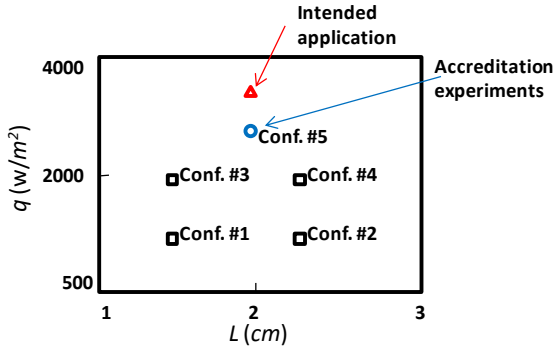


Figure 6.2: Validation experiments at five design configurations

Experiments were conducted at four design configurations as shown in Fig. 6.2, where four test data sets were available for each design configuration. Two more accreditation experiments were conducted at the 5th design configuration to verify model accuracy after model validation. Measurement error ε is ignorable in this problem.

6.4.1.1 Validation metric for the baseline thermal model

MCS was used for the surface temperature prediction with respect to the time by considering the randomness of material properties. Fig. 6.3 shows the baseline model prediction compared to four set of test data at each configuration. Qualitatively speaking, the baseline model is not accurate and its accuracy varies for different configurations. Table 6.1 shows the quantitative results of the validation metric at four configurations. From the U-pooling perspective, the baseline model shows the best accuracy for the 3rd configuration followed by the 1st configuration, and the U-pooling value cannot differentiate the accuracy for the 2nd and 4th configurations. According to the average shape deviation $\bar{S}$, the 3rd configuration again presents the best accuracy followed by the 1st, 2nd, and 4th configurations. Overall, the baseline model presents the best and worst accuracy for the 3rd and 4th configurations, respectively.

Table 6.1: Validation metric for the baseline model prediction

	U	$\bar{S}$	Ψ
Conf. #1	0.2999	0.1116	0.4116
Conf. #2	0.3956	0.1295	0.5251
Conf. #3	0.2362	0.0779	0.3142
Conf. #4	0.3957	0.1422	0.5380

6.4.1.2 Model bias calibration and approximation for the baseline thermal model

The proposed calibration model in Eq. (6.8) was employed to calibrate the dynamic model bias at the four training configurations. With predetermined mean $\mu(t)$ and features $\phi_k(t)$ of the dynamic model bias, statistical moments of V_k were calibrated to minimize the validation metric Ψ so that the agreement between the corrected model prediction and test data was maximized. The calibration history of Ψ is listed in Table 6.2 at the four training configurations. Obviously, accuracy of the corrected model prediction was significantly improved as compared to the validation metric in Table 6.1. Fig. 6.4 further shows the comparison between the corrected model prediction and the test data.

Table 6.2: Optimization history of the validation metric for the thermal problem

Iteration	Conf. #1	Conf. #2	Conf. #3	Conf. #4
0	0.099	0.088	0.098	0.103
1	0.085	0.086	0.092	0.102
2	0.070	0.039	0.050	0.037
3	0.069	0.034	0.043	0.034
4	0.069	0.033	0.042	0.034
Optimal	**0.069**	**0.033**	**0.042**	**0.034**

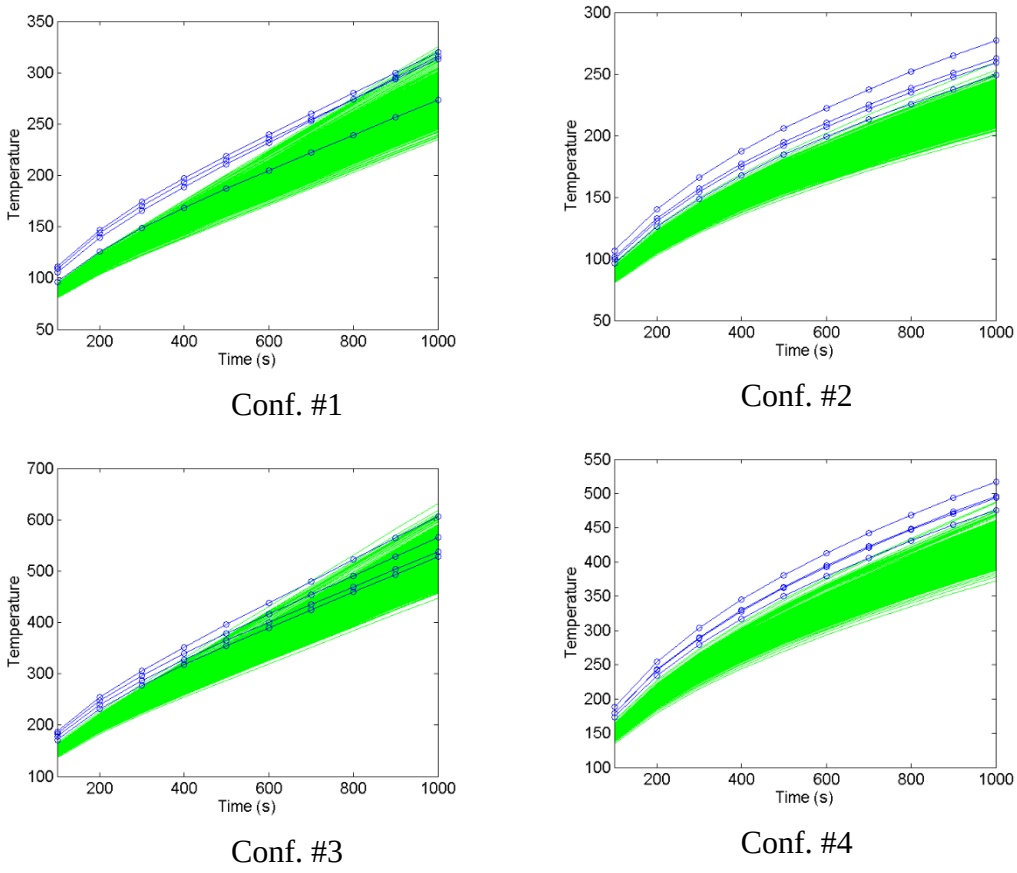
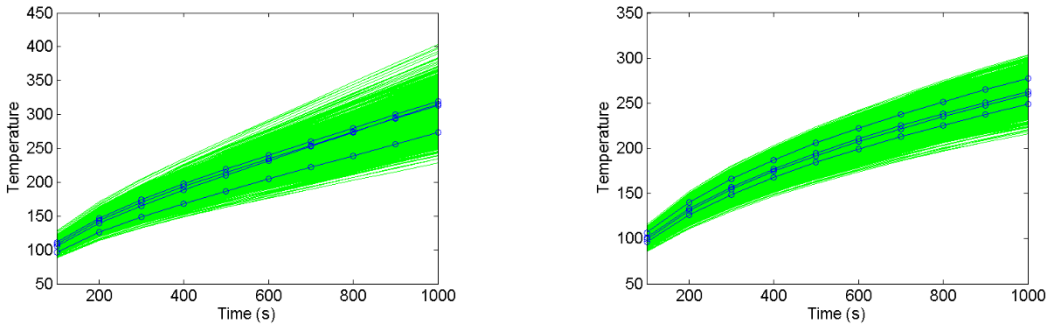
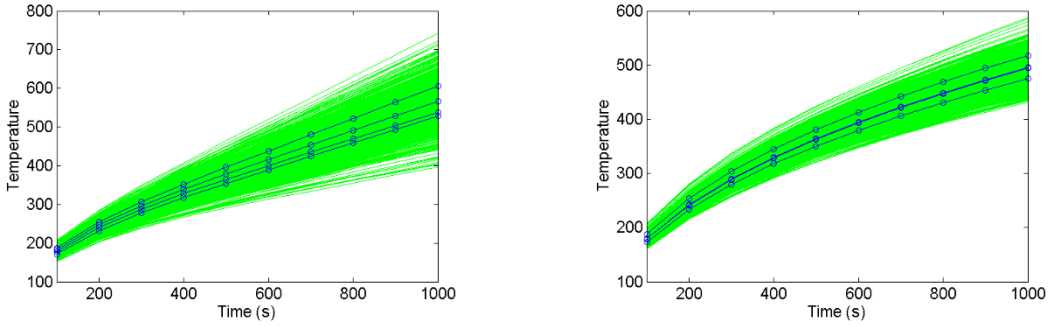


Figure 6.3: Comparison between baseline model prediction and experiment



Conf. #1 Conf. #2



Conf. #3 Conf. #4

Figure 6.4: Comparison between corrected baseline model prediction and experiment

With successful calibration of the dynamic model bias at four training configurations, response surfaces of $\mu(t)$, $\phi_k(t)$ and statistical moments of V_k were constructed as functions of the heat flux rate q and structure thickness L using the moving least square method. The approximate model bias at the 5th configuration is shown in Fig. 6.5 where one thousand random samples of the bias were generated according to Eq. (6.2) and the centered solid line is the mean of the bias. To demonstrate the effectiveness of the proposed bias approximation, the bias was employed to correct the baseline model prediction at the 5th configuration. Two sets of accreditation experiments were used to verify the accuracy improvement of the baseline model as shown in Fig. 6.6. Quantitatively, the value of Ψ was reduced 47% indicating significant accuracy improvement of the baseline model after adding the approximate model bias.

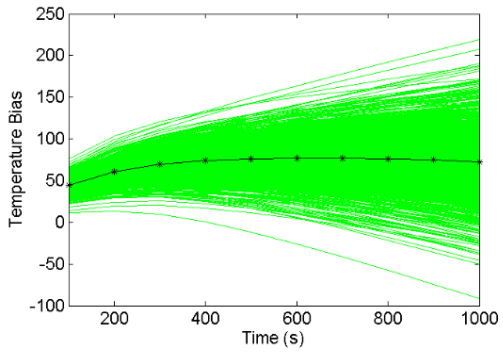
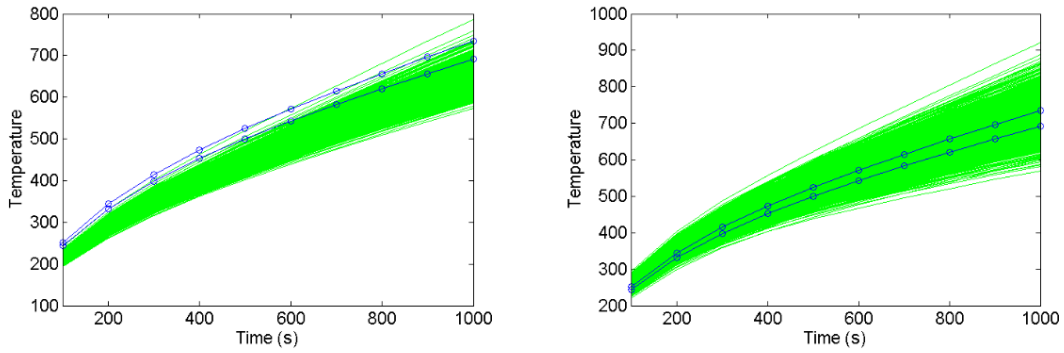


Figure 6.5: Approximate model bias of Conf. #5



Baseline model Corrected baseline model

Figure 6.6: Comparison between model prediction and experiment at Conf. #5

6.4.2 A corroded beam problem

A corroded beam problem [84, 85] was modified to further demonstrate the proposed approach. As shown in Fig. 6.7, the cross section of the beam is rectangular with initial width of a_0 and height of b_0. Both the width and height decrease at a rate of r_1 and r_2 respectively due to the corrosion. The beam is subject to its own weight and a random load F at the center of the beam. The failure occurs when the external stress exceeds the strength of the beam which decays over the time. The true limit sate function is formulated in Eq. (6.10) with assumed true decay rate r_1 and r_2.

$$g = \left[\frac{FL}{4} + \frac{\rho a_0 b_0 L^2}{8}\right] - \left[\frac{(a_0 - 2.2r_0 t)(b_0 - 3r_0 t)\sigma_u}{4}\right] \qquad (6.10)$$

where σ_u is the ultimate strength; ρ is the density; L is the length of the beam; t is the time of year; $r_1 = 1.1r_0$ and $r_2 = 1.5r_0$. It is assumed in this example that the true decay rate cannot be measured in real operation conditions and the baseline model employs a nominal decay rate r_0 on the basis of historical data. Thus, the baseline model of the limit state function which is employed for reliability prediction is formulated as

$$g = \left[\frac{FL}{4} + \frac{\rho a_0 b_0 L^2}{8}\right] - \left[\frac{(a_0 - 2r_0 t)(b_0 - 2r_0 t)\sigma_u}{4}\right] \tag{6.11}$$

In short, Eq. (6.10) is the true model that is pretended to be not known and is only used for reference and Eq. (6.11) will be used for reliability analysis for a time period of 30 years. Model parameters and their properties are listed in Table 6.3.

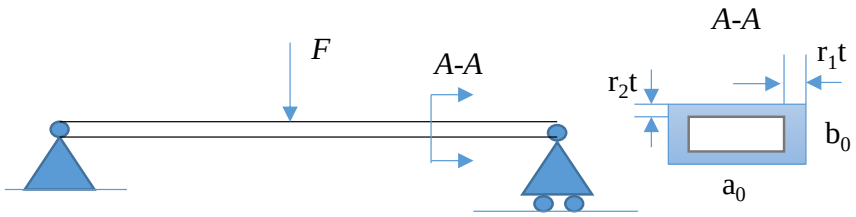


Figure 6.7: Schematic view of a corroded beam problem

Table 6.3: Model parameters and properties of the corroded beam problem

Variable	Mean	Standard deviation	Distribution
a_0	0.2 m	0.01 m	Normal
b_0	0.04 m	0.004 m	Normal
σ_u	2.4e8 Pa	2.4e7 Pa	Lognormal
F	200,000 N	40,000 N	Gaussian
L	5 m	0	Deterministic
ρ	78.5 kN/m^3	0	Deterministic
r_0	5e-5 m/year	0	Deterministic

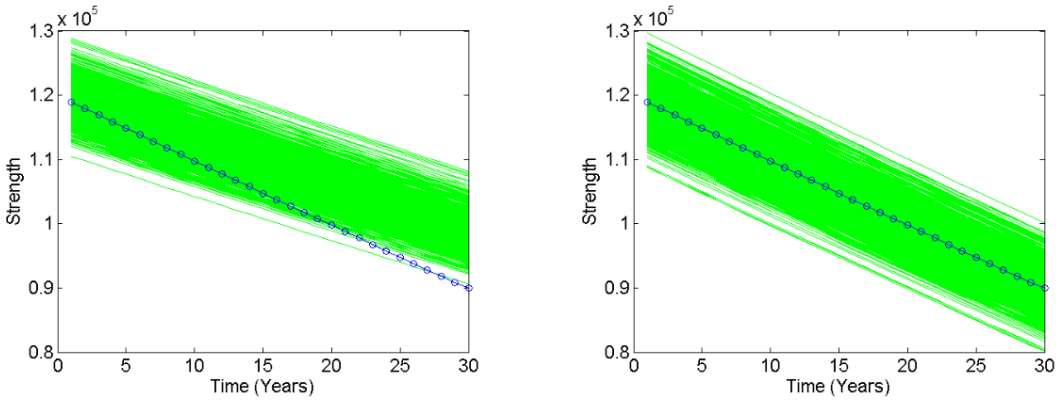
For model bias calibration, one set of virtual test data of the beam strength was obtained at four training design configurations where the means of a_0 and b_0 were downsized. In particular, the size of the beam at four training configurations is listed in Table 6.4. The virtual test data of the beam strength were obtained from Eq. (6.12) where only one random

realization was extracted using MCS considering the randomness of a_0, b_0 and σ_u at each configuration. Fig. 6.8a shows such an example of the comparison between the baseline model prediction and one virtual test data for the 1st configuration. With successful calibration of the dynamic model bias, the corrected baseline model prediction shows much higher degree of agreement as shown in Fig. 6.8b. Similar work was conducted for other training configurations and finally response surfaces were built for the dynamic model bias as functions of a_0 and b_0.

$$strength = (a_0 - 2.2r_0t)(b_0 - 3r_0t)\sigma_u/4 \tag{6.12}$$

Table 6.4: Four training configurations

Configuration	Variable	Mean	Standard deviation	Distribution
Conf. #1	a_0	0.1 m	0.01 m	Normal
	b_0	0.02 m	0.004 m	Normal
Conf. #2	a_0	0.15 m	0.01 m	Normal
	b_0	0.02 m	0.004 m	Normal
Conf. #3	a_0	0.1 m	0.01 m	Normal
	b_0	0.03 m	0.004 m	Normal
Conf. #4	a_0	0.15 m	0.01 m	Normal
	b_0	0.03 m	0.004 m	Normal



a. Before bias calibration b. After bias calibration

Figure 6.8: Comparison of model prediction and virtual test for the 1st configuration

6.5 Summary

This chapter proposed a stochastic model bias correction approach for dynamic system responses. One contribution here is to propose a calibration model for dynamic model bias using the PCA so that only limited number of calibration parameters are needed and the calibration can be effectively conducted similar to the static model bias calibration. The other contribution is to propose the dynamic model bias approximation through building the response surfaces of PCA model components (i.e., $\mu(t)$, $\phi_k(t)$, and V_k) so that the approximate model bias keeps the same form as a PCA model which makes it possible to seamlessly integrate the dynamic model bias with the baseline model. The effectiveness of the proposed approach was well demonstrated by two case studies with dynamic system responses.

Chapter 7: Reliability-based Design Optimization Incorporating Model Uncertainty

7.1 Introduction

Reliability based design optimization (RBDO) has been widely employed for engineering product design to minimize the design cost while satisfying the reliability constraints. Typically, RBDO is conducted based on a simulation model (e.g., a finite element model, an analytical model, a computer fluid dynamics model, etc.) and the model is assumed to be accurate in representing the real physical system. However, this assumption may not be valid in many realistic engineering design applications, which could result in significant reliability prediction errors. Hence, the RBDO that is solely based on the simulation model without referring to the test data is less useful for practical engineering product design. Model calibration, hence, has been widely employed to improve the model prediction accuracy by tuning (or calibrating) some model parameters [86, 87, 88, 89, 90]. However, the improved model accuracy, strictly speaking, is only for the calibrated design configurations and the accuracy may or may not be maintained for the unexplored new design configurations. In other words, prediction capability of the model is uncertain because model accuracy is not quantitatively assessable at new design configurations. The designer may fall into the trap of continuously calibrating the model, but the new design can only be certified with the corresponding test data instead of the model, which significantly reduces the role of the simulation model in engineering product design process. The deficiency of this statistical model calibration is mainly attributed to three reasons: i) model prediction error is not quantitatively assessable at specific new design configurations; ii) a calibration model typically neglects model bias which is the inherent model inadequacy for representing the real physical system due to model assumptions and simplifications; and iii) tuned or calibrated model parameters may deviate from their true values and thus cause worse model accuracy at new design configurations.

Traditionally, the research on reducing the model error was proposed to revise the model conceptually for credibility improvement of the model. From the model development perspective, the key advantage of revising the model conceptually is that accuracy of the model could be significantly improved. However, this approach is practically difficult and may not be feasible in reality due to three reasons: i) identification of the root cause for model inaccuracy is complicated particularly for large scale engineering systems; ii) fundamental modification of the model is time consuming, costly, and yet may not be practical; and iii) there is no perfect model which can represent the real physical system without any model bias. Therefore, the objective of this chapter is to propose an effective model bias correction approach to quantify and reduce model errors in RBDO. In particular, that is to propose a copula-based bias correction approach as compared to the regression-based approach for RBDO with three distinct differences. First of all, the proposed approach performs bias modeling using statistical dependency relationship instead of building a regression functional model. Essentially, the late approach is a causal modeling approach with a pre-assumption that model bias can be fully explained by the defined model parameters. Whereas, the proposed approach is a non-causal modeling approach which acknowledge the fact that model bias could also be affected by unknown factors other than the defined model parameters. Secondly, the proposed approach is a data-driven approach without artificial assumptions about the relationship such as the covariance function or basis function assumptions in the regression-based approach. Thirdly, unlike the regression-based approach suffering from the curse of dimensionality, the proposed approach is expected to be less sensitive to the dimensionality of the problem.

7.2 Reliability-based Design with Model Uncertainty

RBDO is composed of two sub-problems, reliability analysis and design optimization. Reliability analysis evaluates probability constraints at a given design configuration. Design optimization seeks for an optimal design subject to the probability constraints. Despite extensive efforts made in the RBDO methods, model uncertainty (i.e., uncertainty of the model bias) is typically ignored such that probability constraints may not be accurately evaluated in each design iteration even if the reliability analysis method is very

accurate. This section thus proposes a general approach for RBDO that can deal with model uncertainty using any reliability analysis method. The RBDO can be formulated as

$$\begin{aligned} &minimize\ y(\boldsymbol{d}) \\ &subject\ to \\ &P\left(\hat{Y}_i\left(\boldsymbol{X},\ \boldsymbol{P}\right) + \delta_i(\boldsymbol{X}) \le T_i\right) = F_{G_i}(0) \ge \Phi(\beta_i),\ i = 1, \dots, NP \\ &d^L \le d \le d^U,\ x \in \mathbb{R}^N,\ P \in\ \mathbb{R}^L \end{aligned} \tag{7.1}$$

where $y(\boldsymbol{d})$ is the objective function; $\boldsymbol{d} = \mu(\boldsymbol{X})$ is the mean vector of the controllable design parameter vector $\boldsymbol{X}$; $\boldsymbol{P}$ is the uncontrollable random parameter vector and does not change at different design configurations; $\hat{Y}_i(\cdot)$ is the model prediction of the i^{th} constraint; $\delta_i(\cdot)$ is the model bias of the i^{th} constraint; T_i is the performance limit of the i^{th} constraint; β_i is the ith reliability target; $\Phi(\cdot)$ is the cumulative distribution function (CDF) of a standard normal distribution; NP, N, and L are the number of probabilistic constraints, design parameters, and random parameters, respectively. The probability constraint, $F_{G_i}(0)$, is expressed as

$$F_{G_i}(0) = \int_{\hat{Y}_i(\boldsymbol{X},\boldsymbol{P})+\delta_i(\boldsymbol{X})-T_i \le 0} \int f_{X,P}(\boldsymbol{x},\boldsymbol{p}) d\boldsymbol{x} d\boldsymbol{p} \tag{7.2}$$

where $f_{X,P}(\boldsymbol{x},\boldsymbol{p})$ is the joint probability density function (PDF) of $\boldsymbol{X}$ and $\boldsymbol{P}$.

Compared to the traditional RBDO, the key difference is to calculate the model bias δ_i at each design iteration and incorporate it into the reliability analysis. Three technical components should be addressed including: i) approximation of model bias at defined design space; ii) reliability analysis considering model bias; and iii) design sensitivity analysis considering model bias. Technical details of each technical component are elaborated in the following subsections.

7.2.1 Model bias approximation in the design space

A two stage process is requested for model bias approximation in the design space. First of all, model bias needs to be identified at some training design configurations. Then, a bias meta-model can be constructed using different bias modeling approaches so that model bias can be approximated in the defined design space.

7.2.1.1 Model bias calibration at training design configurations

The majority of the bias calibration approaches are based on the Bayesian calibration model proposed by Kennedy and O'Hagan [59] as shown in Eq. (7.3).

$$\hat{Y}(\boldsymbol{X}, \boldsymbol{P}) + \delta(\boldsymbol{X}) = Y - \varepsilon \tag{7.3}$$

where $\hat{Y}(\cdot)$ is system performance prediction from the baseline simulation model, which is equivalent to the $\hat{Y}_i(\cdot)$ in Eq. (7.1) without specifying the i^{th} constraint; $\delta(\cdot)$ is the model bias that needs to be calibrated at available design configurations; Y is available test data at corresponding design configurations; ε is test and measurement error. It is worth noting that model bias δ is modeled as a function of only $\boldsymbol{X}$ not $\boldsymbol{P}$ because the late is not changeable at different design configurations. As a consequence, model bias is not deterministic for a given set of deterministic design variable $\boldsymbol{X}$ due to the unexplained randomness of $\boldsymbol{P}$ in the model. Unlike the classical bias calibration approach using the regression-based approach, Eq. (7.3) is employed here to calibrate the model bias individually at each design configuration instead of for all available design configurations. In the regression-based approach, the functional structure of $\delta(\cdot)$ (e.g., a linear model or a nonlinear model of $\boldsymbol{X}$) needs to be assumed, then model coefficients are calibrated. This regression-based approach essentially assumes a new model (i.e., the bias model) that needs to be validated and has inherent limitations.

To perform the bias modeling without making additional assumptions, a copula-based approach is proposed, which requires to calibrate the model bias at each specific design configuration as

$$\min U\left(\left(\hat{Y} + \delta + \varepsilon, Y\right)\middle|\delta\right) \tag{7.4}$$

where $U(\cdot)$ is the U-pooling metric to quantify the degree of agreement between simulation and test data considering uncertainties. U-pooling metric was proposed by Ferson et al. [27] as a validation metric and has been adopted by many researchers in the study of model validation. The basic idea is to compare the cumulative distribution function (CDF) difference (i.e., the U-pooling value) between model prediction and test data in the standard Uniform space (or U-space). The smaller the area difference, the higher of the expected accuracy of the model prediction. For a specific static system response, each test datum Y_i

corresponds to one u_i value which is calculated from the CDF value of the corrected model prediction (i.e., $\hat{Y} + \delta + \varepsilon$) at the same design configuration (i.e., $u_i = F_{\hat{Y}+\delta+\varepsilon}(Y_i)$, where $F(\cdot)$ is the CDF of $\hat{Y} + \delta + \varepsilon$). With available test data Y, provided estimation of test and measurement error ε, and uncertainty quantification (UQ) of $\hat{Y}$ with appropriate uncertainty modeling of $\boldsymbol{X}$ and $\boldsymbol{P}$, the only unknown quantity is the model bias δ which needs to be calibrated. It is recommended not to calibrate the model bias coupled with other unknown parameters (e.g., from $\boldsymbol{X}$, $\boldsymbol{P}$, and ε if any) because the solution is not unique and wrong bias calibration could jeopardize the bias modeling in the design space [56, 91]. At each specific design configuration, uncertainty of the model bias δ (i.e., model uncertainty) could be modeled as an arbitrary distribution by the Pearson system [92], an assumed normal distribution, or a constant value depending on the amount of test data and desired generality. If there is only one test datum, however, it may be desirable to model the bias as a constant value because there is no sufficient information to uniquely determine two distribution parameters (e.g., mean and standard deviation) of the model bias using Eq. (7.4). Here, model bias δ is modeled as a normal distribution for simplicity if there are more than two test data at each design configuration.

7.2.1.2 Model bias modeling using a copula-based approach

The main idea is to build general statistical relationships between model bias δ, baseline model prediction $\hat{Y}$, and design parameters $\boldsymbol{X}$ in the design space using calibrated model bias across various design configurations. Since model bias is modeled as a Normal distribution in general, the distribution parameter (i.e., mean and standard deviation) is extracted for building the statistical relationship. For notation simplification, δ is still used representing either mean or standard deviation of the model bias in this subsection.

A copula is a general way in statistics to formulate a multivariate distribution with various types of statistical dependence in a standard uniform space. To date, most copulas only deal with bivariate data due to the fact that there is a lack of practical n-dimensional generalization of the coupling parameter [65, 66]. One way to deal with multivariate data is to analyze the data pair-by-pair using two-dimensional copulas. According to a set of bivariate copula models between the expected model bias δ (i.e., mean of model bias), the

baseline model prediction $\hat{Y}$, and the model design parameters $\boldsymbol{X}$, it is feasible to predict the possible model bias at any new design configuration. For example, copula modeling between $\hat{Y}$ and δ allows us to identify the possible model bias δ for a realization of $\hat{Y}$ at a new design configuration. Mathematically, this is a process to identify the conditional PDF of the model bias δ given $\hat{y} = a$ that is,

$$f_{\Delta|\hat{Y}}(\delta|\hat{y} = a) = \frac{c\left(F_{\Delta}(\delta),\, F_{\hat{Y}}(\hat{y})\right)}{f_{\hat{Y}}(\hat{y} = a)} \tag{7.5}$$

where $F(\cdot)$ is the CDF of M and δ, $c(\cdot)$ is the copula PDF defined in the standard uniform space. Meanwhile, we also know the design variable (e.g., $x_1 = a_1, x_2 = a_2, \dots, x_j = a_j$) at the new design configuration. Thus, the possible realizations of the model bias δ could also be predicted from other conditional PDFs from a series of copula models between δ and design variables $\boldsymbol{X}$. These predictions should not be exclusive but inclusive because each conditional PDF prediction is obtained only based on the bivariate copula. Hence, the final PDF prediction of the model bias can be approximated as

$$f(\delta) \cong \Sigma_{j=1}^{R} w_j \times \frac{c\left(F_{\Delta}(\delta),\, F_{X_j}(x_j)\right)}{f_{X_j}(x_j = a_j)} + w_{R+1} \times \frac{c\left(F_{\Delta}(\delta), F_{\hat{Y}}(\hat{y})\right)}{f_{\hat{Y}}(\hat{y} = a)} \tag{7.6}$$

where w_i $(i = 1, \dots, R+1)$ is the weight of the i^{th} conditional PDF, and R is the number of design variables. The weight w_j is calculated as

$$w_j = |\rho_j| / \sum_{j=1}^{R+1} |\rho_j| \tag{7.7}$$

where ρ_j is the correlation coefficient of the j^{th} copula. Hence, higher weight is assigned to the copula with stronger statistical dependence. The weight would be zero if the copula has zero correlation coefficient because it is reasonable not to use the copula to predict the model bias from the uncorrelated variable such as x_j. It is noted that the PDF of δ can be an arbitrary distribution with a closed form solution depending on the combination of copulas in Eq. (7.6). To determine the optimal copula in Eq. (7.6), the maximum likelihood estimation (MLE) approach [67] or the Bayesian copula approach [65] can be employed based on the calibrated model bias at available training design configurations. For the MLE approach, an optimal parameter set is calculated for each candidate copula and the one with the largest likelihood value would be the optimal copula. The Bayesian copula approach is

more preferable with the lack of samples (i.e., bias samples at training design configurations) because it generally provides more reliable identification of the true copula [65]. Four representative copulas (i.e., Clayton, Gaussian, Frank, and Gumbel) are employed in this study.

It is worth noting that model bias prediction, either mean or standard deviation, in Eq. (7.6) follows a certain distribution given a set of deterministic design variables $\boldsymbol{X}$, indicating the epistemic uncertainty of model bias due to the lack of test data, effects from the uncontrollable random parameters $\boldsymbol{P}$, and unknown inherent inadequacy of the model for representing the real physical system. For one prediction of the model bias corresponding to a set of deterministic design variable $\boldsymbol{X}$, the expected model bias can be calculated in Eq. (7.8).

$$E(\delta) = \int \delta f(\delta) d\delta \tag{7.8}$$

7.2.2 Reliability analysis considering model bias

With approximate model bias at new design configurations, reliability analysis should be performed for the corrected model prediction (i.e., $\hat{Y} + \delta$) instead of the baseline model prediction (i.e., $\hat{Y}$). Essentially, another source of uncertainty (i.e., δ) should be included in reliability analysis. Since the model bias δ is modeled by a distribution with uncertain distribution parameters, any available reliability analysis strategies and methods that can handle both aleatory and epistemic uncertainties can be used to perform reliability analysis for the corrected model prediction. Here, two approaches of computing the reliability are proposed: i) the expected reliability; and ii) the reliability distribution. Since there are abundant literature dealing with both aleatory and epistemic uncertainty in reliability analysis, the objective of this section is to illustrate the difference of the two approaches considering the model bias.

Calculation of the expected reliability considers the expected uncertainty of δ (i.e., expected mean and standard deviation of δ calculated in Eq. (7.8)) at a specific design configuration. If MCS were used, sufficient random samples of $\boldsymbol{X}$ and δ would be

generated to calculate the corrected model prediction of $\hat{Y} + \delta$, then the expected reliability could be easily calculated from the ratio of safe trials over the total trials. Calculation of the reliability distribution treats realizations of model bias uncertainty individually. If MCS were used, sufficient random samples of $\boldsymbol{X}$ would be firstly generated to calculate the baseline model prediction $\hat{Y}$. Next, sufficient random samples of δ would be generated from a set of model bias distribution parameters to calculate one reliability value. Then, above step should be repeated for many sets of model bias distribution parameters to obtain many reliability values. Finally, a reliability distribution can be obtained considering the epistemic uncertainty of the model bias.

In summary, the 1st approach obtains the expected reliability considering the expected uncertainty from δ, whereas, the 2nd approach computes the reliability distribution which may be more useful for safety critical structure design where confidence bounds of the reliability prediction can be provided. Other than the MCS, many advanced reliability analysis methods, such as the MPP-based approaches, the EDR method, the PCE methods, etc., can be employed to significantly improve the computational efficiency of the reliability analysis.

7.2.3 Design sensitivity analysis considering model bias

This section presents the reliability sensitivity analysis with respect to the mean change of the design variable $\boldsymbol{X}$. The sensitivities of reliabilities with respect to the mean of the i^{th} design variable are computed using the finite difference method (FDM) as shown in Eq. (7.9).

$$\frac{\partial R_k}{\partial \mu_{X_i}} \cong \frac{R_k\left(\mu_{X_i} + \Delta\mu_{X_i}\right) - R_k\left(\mu_{X_i}\right)}{\Delta\mu_{X_i}}, for\ k = 1, \dots, NP \tag{7.9}$$

In particular, computation of the reliability change due to the mean shift of $\boldsymbol{X}$ is critical and its solution is determined by two factors: i) PDF change of M; and ii) PDF change of δ. A practical concern in sensitivity analysis using the FDM is the computational efficiency and it is desirable not to run extra simulations in order to maintain the efficiency of RBDO. It is clear that no extra simulations are required for the second part because the calculation of δ is based on a constructed statistical model in Eqs. (7.6) and (7.8). Therefore, the only

concern is the PDF change of $\hat{Y}$ due to the mean change of the design variable $\boldsymbol{X}$. Many advanced probability analysis methods require only a few simulations at a set of samples of the input random variables for quantifying the PDF of $\hat{Y}$. The remaining problem is how to freely evaluate the shifted $\hat{Y}$ values at the shifted set of samples. For probability analysis methods with assisted response surfaces (e.g., the EDR and PCE methods), such evaluations do not require extra simulations because the shifted $\hat{Y}$ values can be freely estimated from the response surfaces. Other methods (e.g., direct MCS and expansion methods) may need extra simulations to conduct the sensitivity analysis. Nevertheless, this issue belongs to traditional RBDO sensitivity calculation. It is worth noting that no extra simulations are required because of the consideration of the model bias δ in RBDO. Here, a perturbed mean of the design variable X is identified with a common perturbation size of 0.1%.

7.3 Case Studies

7.3.1 A modified vehicle side impact response

The objective of this case study is to compare the model bias approximation using the proposed copula-based approach and a traditional regression-based approach such as the moving least square method. One of the vehicle side impact responses, i.e., the lower rib deflection, [93] was modified for the comparison. All required model parameters are shown in Table 7.1 with defined lower and upper design bounds. To focus the problem on bias accuracy comparison, uncertainties of model parameters are not considered in this case study. A true model and a low fidelity baseline model of the lower rib deflection are defined in Table 7.2. In particular, the term ξ in the true model indicates the unexplained portion other than the defined model parameters that contributes to the true response, which models the inherent inadequacy of the defined model parameters to represent the ground truth due to a series of assumptions and model simplifications. The response of G_3 defined in Table 7.2 is equivalent to $Y - \varepsilon$ in Eq. (7.3). If the test and measurement error ε is ignored, G_3 means the test value of the lower rib deflection at a specific design. In this example, the quantity of ξ is modeled as a Normal distribution with zero mean and a certain standard deviation (STD) σ_ξ. In other words, a small σ_ξ (e.g., a value close to 0) means that defined model parameters and the functional form can truly represent the real physical systems

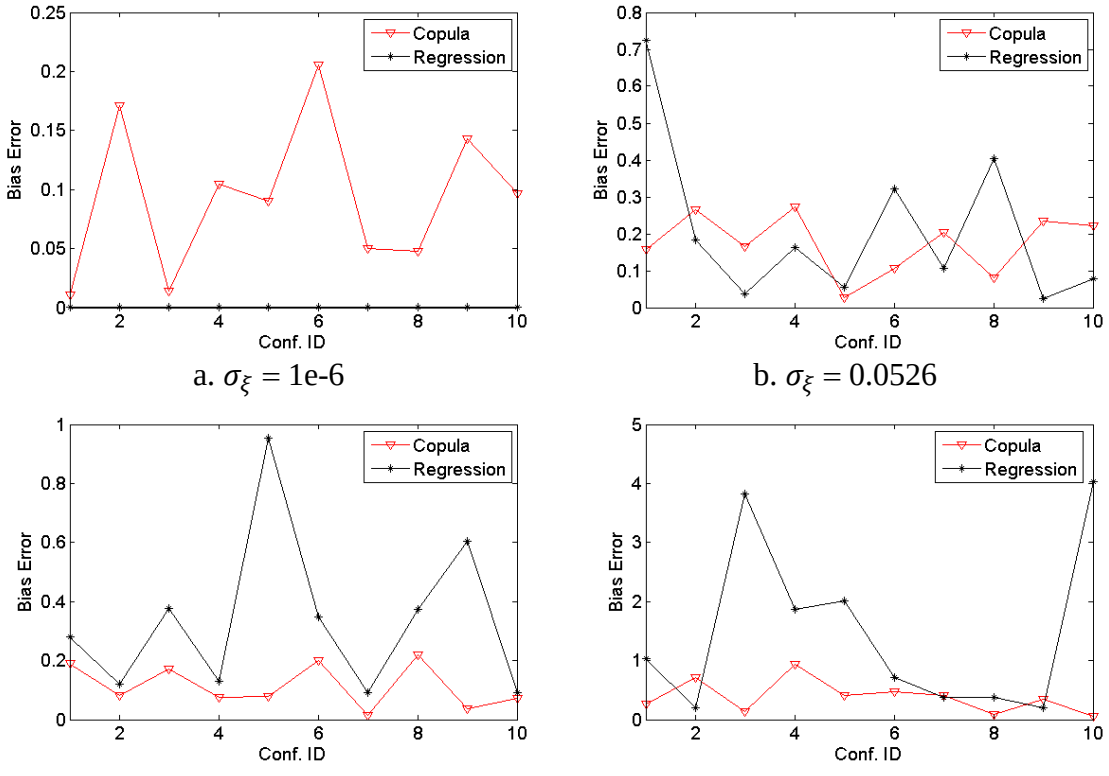
without any error, and a large σ_ξ (e.g., 1) indicates that there is a certain level of unexplained portion by the model parameters. For a given design configuration $\boldsymbol{X}$, the bias error δ_3 (i.e., model bias) is calculated as $G_3 - M_3$.

Table 7.1: Parameter lower and upper bounds for the lower rib deflection model

Model Parameters	d^L	d^U
X_1	0.500	1.500
X_2	0.500	1.500
X_3	0.500	1.500
X_8	0.192	0.345
X_{10}	-25	25

To compare the model bias approximation for an assigned σ_ξ using two different approaches, four steps were conducted and summarized as follows. First of all, Latin hypercube sampling (LHS) was employed to generate 45 design configurations. The baseline model responses of M_3 were explicitly calculated, and the true responses G_3 were correspondingly computed with addition of a random sample ξ for each design configuration. Hence, model bias δ_3 at 45 design configurations were easily calculated as $G_3 - M_3$. Secondly, model bias at 35 design configurations were randomly selected as training data sets to build the bias approximation model in the design space using two different approaches. Thirdly, model bias at the remaining 10 configurations were employed as confirmation to calculate the model bias absolute prediction errors by two different approaches. Figure 7.1 shows the model bias errors when the STD of ξ is increased. When ξ is insignificant (i.e., Fig. 7.1a), the true model bias is well represented by a simple regression model. Hence, model bias error from the regression approach is almost zero indicating that the regression approach approximate the model bias without any error at 10 confirmation configurations. However, as long as ξ is significant and particularly when its STD increases, the copula-based approach shows much stable and better model bias approximation accuracy compared with the regression approach. The reason is because the observed model bias at 35 training configurations cannot be fully explained by defined model parameters $\boldsymbol{X}$, which is very common considering many realistic engineering validation problems. Finally, the mean and maximum bias prediction

error from 10 configurations were computed and results are shown in Fig. 7.2 when the STD of ξ increases from 0 to 1. It is clearly observed that the regression approach is much more sensitive to the unexplained portion of ξ in the bias modeling, resulting in a much higher error increasing rate for both average and maximum bias prediction error. In this particular problem, the copula-based approach shows consistent advantages in terms of the accuracy than the regression approach when the STD of ξ is higher than 0.05. It is worth noting that even the STD of ξ increases to 1, the contribution of ξ on the true model response G_3 is still trivial.



a. σ_ξ = 1e-6 b. σ_ξ = 0.0526

c. σ_ξ = 0.1579 d. σ_ξ = 0.4211

Figure 7.1: Model bias prediction errors at 10 confirmation configurations by copula and regression approaches with respect to different STD of ξ

Table 7.2: Formulations of a true model and a baseline model for the lower rib deflection

Constraint		Formulations
Rib Deflection (Lower)	Baseline model	$M_3 = 40 - 9x_2 - 11x_1x_8 + 0.10x_3x_{10}$
	True model	$G_3 = 46.36 - 9.9x_2 - 12.9x_1x_8 + 0.11x_3x_{10} + \xi$

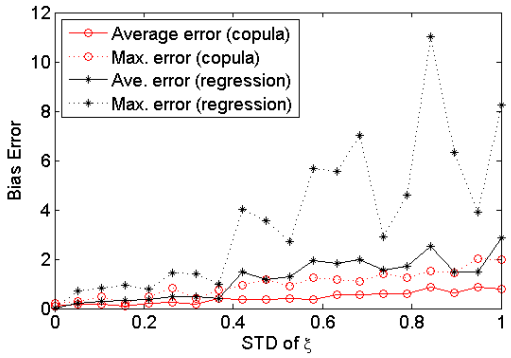


Figure 7.2: Comparison of average and maximum model bias prediction error of copula and regression approaches when the STD of ξ increases from 0 to 1

7.3.2 A 2001 Ford Taurus frontal impact model

A 2001 Ford Taurus model provided by the National Crash Analysis Center was employed for RBDO. Figure 7.3a shows the physical test and model prediction for the full frontal impact. The simulation speed is 56.6 km/h against a rigid wall. For frontal impact protection, vehicle design must meet internal and regulated frontal impact requirements. In particular, vehicles must be designed to absorb enough impact energy through structural deformation and attenuate the impact force to a tolerable level in order to protect the occupants. Eight design variables of the main front-end structure are defined in Fig. 7.3b and their baseline design and design bounds are listed in Table 7.3. Chest G (CG) and Crush Distance (CD) are two key performances of interest that need to meet the reliability constraints in RBDO.

a. Vehicle full frontal impact from test and FEA

b. Eight design variables of the main front-end structure

Figure 7.3: A 2001 Ford Taurus model with eight design variables for the main front-end structure

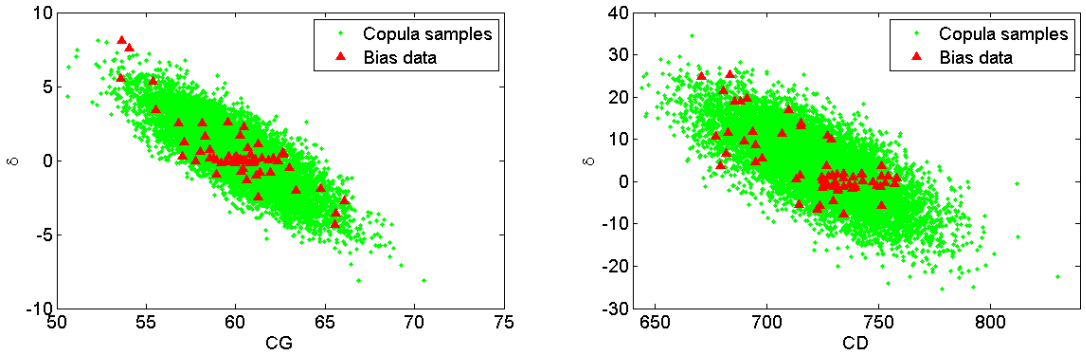
Table 7.3: Baseline design and design bounds for the main front-end structure

Design variable	Baseline	Lower bound	Upper bound
x_1	1.90	1.4	2.8
x_2	1.91	1.2	2.8
x_3	2.51	1.6	4.0
x_4	2.40	1.5	4.0
x_5	2.55	1.6	4.0
x_6	2.55	1.5	3.5
x_7	2.25	1.5	3.5
x_8	1.50	1.2	3.0

To demonstrate the proposed work without conducting real prototype testing as shown in Fig. 7.3a, meta-models (i.e., regression models) of the CG and CD were constructed from a high fidelity model (i.e., a FEA model). To characterize the model bias of the meta-model in the design space, a total of 64 design configurations were generated using Latin hypercube sampling (LHS) in the design space where CG and CD predictions are available from both regression and FEA models. Figure 7.4 shows the copula modeling between $\hat{Y}$ and δ for both CG and CD based on the 64 training design configurations. It is observed that the meta-model predictions of both CG and CD are negatively correlated with the model bias. In other words, the meta-model tends to underestimate and overestimate the true performance (i.e., FEA results) when the prediction is relatively small and large, respectively. Considering other copula models between design variables $\boldsymbol{X}$ and δ, model bias at any new design configuration can be estimated using Eqs. (7.6) and (7.8).

According to the meta-model, an RBDO problem is formulated in Eq. (7.10) to minimize the mean of the structure weight whiling satisfying 99% reliability constraints for both performances of interest, where all design variables are assumed to follow Normal distributions and standard deviations are set to be 5% of the baseline mean value. RBDO history is shown in Table 7.4 where weight of the structure is significantly reduced and the reliability constraint of CD is active. In particular, the value of c_i in Table 7.4 is defined as the subtraction of the performance value at the 99% level by the performance target value. Hence, the constraint is inactive, active, or does not meet the reliability target if the value is less than, equal to, or larger than zero, respectively. However, the reliability estimation may not be accurate without considering the model bias of CG and CD. Ten virtual test

data of CG and CD were obtained at the optimal design identified in Table 7.4 and they are compared with the UQ results from the meta-model as shown in Fig. 7.5a and Fig. 7.5c. Even though CG is inactive, CG prediction from the meta-model has large discrepancy with ten virtual test data. The consequence of such discrepancy is not significant in this setting because the CG performance target (i.e., 65) is relatively high. On the other hand, the consequence of ignoring the model bias of CD is more severe because two out of ten virtual test data exceeding the performance target (i.e., 750) indicating an extremely small probability that 99% reliability constraint is truly met. To demonstrate the copula-based approach, model bias of these two performances were approximated using Eqs. (7.6) and (7.8), and they were added to the baseline model prediction. The comparison results are shown in Fig. 7.5b and Fig. 7.5d and it is clearly observed that the agreement between the corrected model prediction and the virtual test data has been significantly improved. In particular, the expected reliability of the CD is about 92% based on the corrected model at the optimal design.



a. CG prediction vs. model bias b. CD prediction vs. model bias

Figure 7.4: Meta-model prediction vs. model bias at 64 training design configurations

$$\begin{aligned} &\text{Minimize:} \quad \mu_{weight} \\ &\text{Subject to:} \\ &\qquad P\left(CG \le CG_{target}\right) \ge 99\%,\ CG_{target} = 65 \\ &\qquad P\left(CD \le CD_{target}\right) \ge 99\%,\ CD_{target} = 750 \\ &\qquad LB_{x_i} \le \mu_{x_i} \le UB_{x_i} \end{aligned} \tag{7.10}$$

where

$$x_i \sim N(\mu_{x_i}, \left(0.05\mu_{x_i}^{baseline}\right)^2)$$

$$weight = 6.01x_1 + 3.17x_2 + 2.08x_3 + 1.24x_4 + 1.46x_5 + 4.37x_6 + 3.55x_7 + 2.31x_8$$

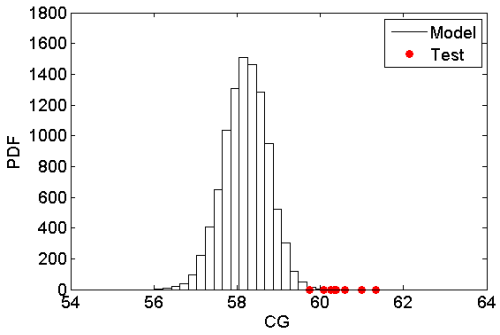
$$CG = 84.70 - 7.77x_6 + 0.76x_5x_8 + 0.98x_7 - 13.13x_1 - 1.00x_2x_5 + 4.09x_1x_6 - 0.32x_4x_8 + 0.29x_1x_5$$

$$CD = 922.51 - 2.56x_6x_7 + 0.66x_4 - 88.27x_1 + 13.93x_1^2 - 1.27x_3x_6 + 0.47x_4x_5 - 8.20x_2x_6 - 4.69x_4x_8$$

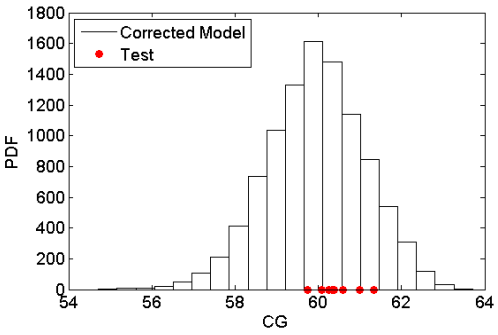
Table 7.4: Design history of RBDO without considering model bias

Iteration	Weight	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	c_1	c_2
1	51.97	1.90	1.91	2.51	2.40	2.55	2.55	2.25	1.50	-3.93	-7.82
2	44.94	2.35	1.87	1.60	2.01	1.60	1.98	1.50	1.20	-6.11	0.88
3	44.96	2.32	1.98	1.60	1.63	1.60	2.06	1.50	1.20	-5.93	0.04
4	44.81	2.20	2.19	1.60	1.50	1.60	2.07	1.50	1.20	-5.73	0.52
5	44.88	2.17	2.13	1.60	1.50	1.60	2.17	1.50	1.20	-5.46	0.08
6	44.88	2.21	2.06	1.60	1.50	1.60	2.16	1.50	1.20	-5.52	0.04
7	44.86	2.28	2.07	1.60	1.50	1.60	2.06	1.50	1.20	-5.84	0.25
8	44.88	2.23	2.06	1.60	1.50	1.60	2.14	1.50	1.20	-5.60	0.06
9	44.89	2.24	2.04	1.60	1.50	1.60	2.14	1.50	1.20	-5.61	0.01
10	44.89	2.24	2.04	1.60	1.50	1.60	2.14	1.50	1.20	-5.62	0.00
Optimal	**44.89**	**2.24**	**2.04**	**1.60**	**1.50**	**1.60**	**2.14**	**1.50**	**1.20**	**-5.62**	**0.00**

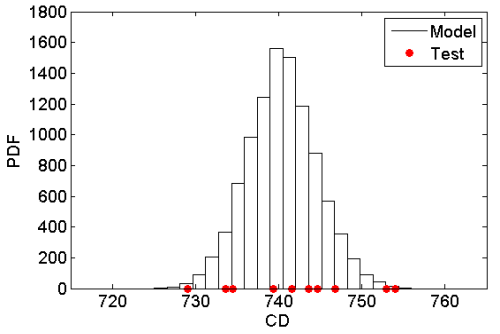
With consideration of the model bias, RBDO was conducted again and its design history is shown in Table 7.5. The optimal solution is slightly different where the weight is slightly increased because of the change of x_1, x_2, and x_6, and the constraint of CD is still active. Ten virtual test data were obtained from this optimal design and the comparison with the model prediction is shown in Fig. 7.6. It is observed that the agreement for both CG and CD responses is much better than the scenarios without considering the model bias. Furthermore, ten virtual test data of CD are all less than 750 and there is no statistical evidence that the reliability constraint of CD is not met.



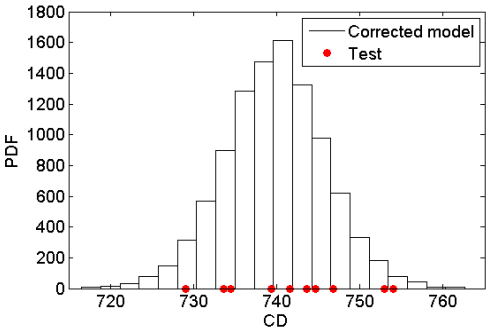
a. CG model prediction vs. virtual test without considering model bias



b. CG model prediction vs. virtual test considering model bias

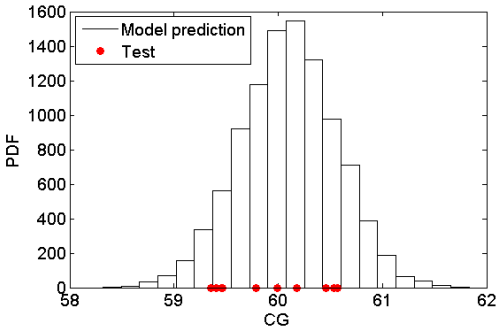


c. CD model prediction vs. virtual test without considering model bias

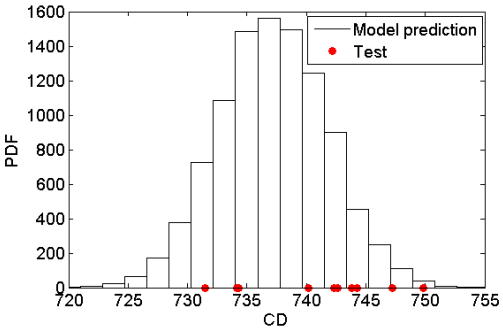


d. CD model prediction vs. virtual test considering model bias

Figure 7.5: Model predictions vs. virtual tests at an optimal design from the meta-model



a. CG model prediction vs. virtual test



b. CD model prediction vs. virtual test

Figure 7.6: Model predictions vs. virtual tests at optimal design considering model bias

This RBDO example was also conducted by another research team with the exactly same design setting, but a regression model bias modeling approach using the Gaussian process regression (GPR) model. The optimal solution was reported in a reference paper [94] and was included in Table 7.6 for comparison. It is observed that different model bias modeling technique will affect the optimal results. The reason should be that the approximate model bias at each design iteration (i.e., bias distribution) is different from different bias modeling

approaches even though the same data set were utilized for the bias modeling. For this particular problem, the weight from the copula-based approach is less than the GPR approach without statistical evidence of violating the reliability constraints. Another insightful observation is that ignoring model bias in RBDO has a high risk to produce unreliable optimal design, even though the goal of RBDO is to pursue a reliable design. On the other hand, considering model bias in RBDO could produce a conservative and reliable design, especially when reliability distribution is considered with a high confidence interval for the desired reliability.

Table 7.5: Design history of RBDO considering model bias

Iteration	weight	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	c_1	c_2
1	51.97	1.90	1.91	2.51	2.40	2.55	2.55	2.25	1.50	-3.18	-3.70
2	45.77	2.24	2.24	1.60	1.99	1.60	2.07	1.50	1.20	-4.97	0.26
3	45.54	2.11	2.28	1.60	1.64	1.60	2.26	1.50	1.20	-4.44	-0.05
4	45.45	2.11	2.27	1.60	1.50	1.60	2.29	1.50	1.20	-4.36	-0.21
5	45.41	2.11	2.28	1.60	1.50	1.60	2.26	1.50	1.20	-4.40	0.05
6	45.41	2.13	2.27	1.60	1.50	1.60	2.24	1.50	1.20	-4.45	-0.06
7	45.40	2.13	2.27	1.60	1.50	1.60	2.24	1.50	1.20	-4.44	0.02
8	45.40	2.13	2.26	1.60	1.50	1.60	2.25	1.50	1.20	-4.44	0.01
10	45.40	2.15	2.23	1.60	1.50	1.60	2.25	1.50	1.20	-4.46	0.01
11	45.40	2.16	2.21	1.60	1.50	1.60	2.24	1.50	1.20	-4.43	0.00
Optimal	**45.40**	**2.16**	**2.21**	**1.60**	**1.50**	**1.60**	**2.24**	**1.50**	**1.20**	**-4.46**	**0.00**

Table 7.6: RBDO optimal solutions with three different approaches

Design variable	**Optimum with the meta-model**	**Optimum considering model bias (GPR)**	**Optimum considering model bias (copula)**
x_1	2.24	1.96	2.16
x_2	2.04	2.26	2.21
x_3	1.60	1.96	1.60
x_4	1.50	2.93	1.50
x_5	1.60	1.67	1.60
x_6	2.14	2.24	2.24
x_7	1.50	1.60	1.50
x_8	1.20	1.20	1.20
Weight	**44.89**	**47.30**	**45.40**

7.4 Summary

In this chapter a copula-based model bias correction approach was proposed for RBDO. Four technical components were addressed including: i) model bias calibration at available training design configurations; ii) model bias approximation at new design configurations; iii) reliability analysis considering model bias; and iv) design sensitivity analysis. The proposed approach is a non-causal modeling approach that conducts non-causality modeling between model bias, design variables, and the baseline model prediction. The proposed approach is more suitable for model bias modeling compared to the regression approach because model bias is defined as the inherent model inadequacy for representing the real physical systems due to simplifications and assumptions and hence is not supposed to be fully accounted for by the defined model parameters using a causal modeling approach. Furthermore, the proposed approach employs less assumptions and is expected to be less sensitive to the dimensionality of the problem compared to the regression-based approach. Two vehicle case studies showed that the proposed approach approximates model bias more accurate and stable than the regression-based approach when there is an unexplained portion in the model bias, and hence could result in a better RBDO solution. It should also be noted that the proposed approach has the potential to be applied to any design problem (i.e. robust design [95, 96] and multidisciplinary design [97, 98]).

Chapter 8: Conclusion

In this research a systematic and practical model validation framework for reliability-based design optimization is developed. The contributions of the research are listed as follows.

- Contribution 1: A Copula-based model bias characterization approach is developed to capture the relationship between model inputs, outputs and the model bias, as well as to provide the prediction in new design space. Model validation have been developed over the years and they do have their limitations. Some approaches do not consider model uncertainty. For those that do consider model uncertainty utilizing statistical techniques, assumptions (e.g. normality) are used to simplify the derivation or the computation process but oftentimes these assumptions are not valid. When dealing with high dimension problems (consider multiple inputs/outputs simultaneously), the computation cost may rise due to the curse of dimensionality. The copula-based model bias characterization approach is proposed here to overcome these shortcomings. The copula-based approach is capable of establishing a statistical relationship among model bias, design variables and model responses. The non-linearity in the model is not affecting the potency of the proposed approach. Two case studies adequately demonstrated that the proposed approach is effective in improving the accuracy of model prediction. The proposed approach is also proved to be able to handle problems with high number of design variables.

- Contribution 2: An adaptive copula-based model bias characterization approach is developed to further enhance the accuracy of the copula-based approach with the aid of clustering analysis. Cluster analysis is employed first to group similar data points together, followed by the copula-based approach using information from each cluster. The final prediction accumulates predictions obtained from each cluster. Two case studies adequately demonstrated that the proposed approach is effective in improving

the accuracy of model prediction compared to its predecessor which uses single copula. The proposed approach is also able to produce narrower confidence bounds thus reducing uncertainty associated to the prediction. The adaptive copula-based model bias characterization approach turns out to be a better way of utilizing information contained by the available data.

- Contribution 3: A novel validation metric for dynamic responses under uncertainty is developed to assess model accuracy with dynamic responses considering limited test data. It also enables the generalization of model bias characterization from static responses to dynamic responses. Majority of available validation metric is designed for static system response. Though metrics for dynamic responses are also available, they are specifically designed for vehicle impact application and uncertainties are not well considered in the metric. The proposed statistic validation metric for dynamic responses addresses the two challenges. The classical U-pooling approach is extended for dynamic responses by discretizing and treating the responses as a high dimensional joint distribution. The PCA is applied to represent the dynamic responses by a few random variables so that the U-pooling value can be computed more efficiently. The shape deviation was included in the validation metric so that the metric can still distinguish the model accuracy when the U-pooling value only is not differentiable. The proposed shape deviation metric works effectively with multiple test data and distinguishes not only the mean shape difference but also the difference at corresponding percentage levels. One vehicle impact model was employed and the effectiveness of the proposed validation metric was demonstrated.

- Contribution 4: A stochastic model bias calibration and approximation approach is proposed with the aid of the developed dynamic validation metric for reliability analysis. Given limited test data, model bias needs to be effectively identified so that reliability analysis would be more accurate compared to the baseline model. In particular, it is desirable that model bias can be accurately approximated at any new design configurations where test data are not available. A calibration model for dynamic model bias using the PCA is established so that only limited number of

calibration parameters are needed and the calibration can be effectively conducted similar to the static model bias calibration. The dynamic model bias is then approximated through building the response surfaces of PCA model components so that the approximate model bias keeps the same form as a PCA model which makes it possible to seamlessly integrate the dynamic model bias with the baseline model. The effectiveness of the proposed approach was well demonstrated by two case studies with dynamic system responses.

- Contribution 5: Reliability-based design optimization is integrated with the proposed model uncertainty characterization approach for reliable design of various engineering products. RBDO is typically conducted based on a simulation model which is assumed to be accurate in representing the real physical system. Such assumption may not be valid in many realistic engineering design applications, which could result in significant reliability prediction errors. Model calibration has been widely employed to improve the model prediction accuracy by tuning some model parameters. However, the improved model accuracy is only for the calibrated design configurations and the accuracy may or may not be maintained for the unexplored new design configurations. An effective model bias correction approach to quantify and reduce model errors in RBDO is proposed here to address these challenges. The proposed approach is a non-causal modeling approach that conducts non-causality modeling between model bias, design variables, and the baseline model prediction. The proposed approach is more suitable for model bias modeling compared to the regression approach because model bias is defined as the inherent model inadequacy for representing the real physical systems due to simplifications and assumptions and hence is not supposed to be fully accounted for by the defined model parameters using a causal modeling approach. Furthermore, the proposed approach employs less assumptions and is expected to be less sensitive to the dimensionality of the problem compared to the regression-based approach. Two vehicle case studies showed that the proposed approach approximates model bias more accurate and more stable than the regression-based approach when there is an unexplained portion in the model bias, and hence could result in a better RBDO solution.

The successful accomplishment of the current research will have significant impact on other fields of engineering science where computer simulation models are employed. With high fidelity computer simulation models obtained from the proposed research, three significant industrial impacts are expected including: i) reduce expensive physical tests in engineering design; ii) reduce time, cost, and development risk associated with full-scale testing of engineering products; and iii) further enhance the capabilities and robustness of reliability-based design in various industrial applications.

Areas of future work are described here to overcome the shortcomings of the proposed approaches and/or extend the applicability of them.

- The copula-based model bias characterization approach could be extended to systems with dependent inputs. Currently, the copula-based approach deal with systems with multiple inputs by treating one input at a time assuming each input is independent. The correlation between the inputs may potentially has impact on the effectiveness of the proposed approach and should be studied. PCA could be applied here to transform the inputs, computer model outputs and bias into a space spanned by the PCs which are uncorrelated to each other.

- The copula-based model bias characterization approach could also be extended to deal with multiple dependent responses. Again the copula-based approach deal with multiple system responses individually assuming there is no correlation between them. Transformation techniques such as PCA could be employed here to de-correlate the responses.

- The classification of the quality of the model could be developed based on the value of the validation metric. The proposed validation metric is able to differentiate if a model has significant bias or not. But it lacks the ability to tell how good a model is. Subject matter experts' opinion and empirical results could be used here to help interpret the value produced by the validation metric. This could help the decision making process as priority can be created based on the interpreted values of the validation metric and

improvement effort can be directed to the corresponding aspects of the computer model with severe model bias.

- More case studies need to be conducted to study the integration of model bias modeling and various design problems. In this research only the RBDO was utilized to demonstrate the integration of the model bias characterization due to the limitation of resources and time. Compatibility issues need to be carefully studied when integrating other design problems and the model bias characterization approach.

References

[1] American Institute of Aeronautics and Astronautics, "Guide for the verification and validation of computational fluid dynamics simulations," 1998.

[2] Department of Defense, "Verification, validation, and accretidation (VV&A) Recommended Practice Guide," 1996.

[3] B. Thacker, S. Doebling, F. Hemez, M. Andersen, J. Pepin and E. Rodriguez, "Concepts of model verification and validation," Los Alamos National Laboratory, 2004.

[4] American Society of Mechanical Engineers, "Guide for verification and validation in computational solid mechanics," 2006.

[5] W. Oberkampf, T. Trucano and C. Hirsch, "Verification, validation, and predictive capability in computational engineering and physics," *Applied Mechanics Reviews*, vol. 57, pp. 345-384, 2004.

[6] G. Box and N. Draper, Empirical Model-building and Response Surfaces, WILEY, 1987.

[7] G. Fishman and P. Kiviat, "The statistics of discrete-event simulation," *Simulation*, vol. 185, 1967.

[8] R. Sargent, "Validation of simulation models," in *Winter Simulation Conference*, 1979.

[9] O. Balci, "Guidelines for successful simulation studies," in *Winter Simulation Conference*, 1990.

[10] P. Roache, Verification and Validation in Computational Science and Engineering, Hermosa Publishers, 1998.

[11] Y. Sugawara, K. Shinohara and N. Kobayashi, "Quantitative validation of dynamic stiffening represented by absolute nodal coordinate formulation," in *ASME 2009*

International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, 2009.

[12] X. Qiu, D. Japikse, J. Zhao and M. Anderson, "Analysis and validation of a unified slip factor model for impellers at design and off-design conditions," *Journal of Turbomachinery*, vol. 133, 2011.

[13] I. Arias, J. Knap, V. Chalivendra, S. Hong, M. Ortiz and A. Rosakis, "Numerical modelling and experimental validation of dynamic fracture events along weak planes," *Computer Methods in Applied Mechanics and Engineering*, vol. 196, pp. 3933-3840, 2007.

[14] Y. Liu, W. Chen, P. Arendt and H. Huang, "Experimental testing and validation of a magnetorheological (MR) damper model," *Journal of Vibration and Acoustics*, vol. 131, 2009.

[15] D. Mayer and D. Butler, "Statistical validation," *Ecological Modelling*, vol. 68, 1993.

[16] W. Oberkampf and T. Trucano, "Verification and validation benchmarks," *Nuclear Engineering and Design*, vol. 238, 2008.

[17] W. Oberkampf and C. Roy, Verification and Validation in Scientific Computing, Cambridge University Press, 2010.

[18] K. Danai, J. McCusker, T. Currier and D. Kazmer, "Validation of dynamic models by image distances in the time-scale domain," in *ASME 2009 Dynamic Systems and Control Conference*, 2009.

[19] K. Danai, D. Kazmer and J. McCusker, "Validation of dynamic models in the time-scale domain," *Journal of Dynamic Systems, Measurement, and Control*, 2010.

[20] M. Sprague and T. Geers, "Spectral elements and field separation for an acoustic fluid subject to cavitation," *Journal of Computational Physics*, vol. 184, pp. 149-162, 2003.

[21] L. Schwer, "Validation metrics for response histories: perspectives and case studies," *Engineering with Computers*, vol. 23, pp. 295-309, 2007.

[22] H. Sarin, M. Kokkolaras, G. Hulbert, P. Papalambros, S. Barbat and R. Yang, "A comprehensive metric for comparing time histories in validation of simulation models with emphasis on vehicle safety applications," in *ASME 2008 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, 2008.

[23] D. Russell, "Error measures for comparing transient data," in *68th Shock and Vibration Symposium*, 1997.

[24] Y. Liu, W. Chen, P. Arendt and H. Huang, "Toward a better understanding of model validation metrics," *Journal of Mechanical Design,* vol. 133, 2011.

[25] S. Mahadevan and J. McFarland, "Error and variability characterization in structural dynamics modeling," *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2621-2631, 2008.

[26] I. Babuska, F. Nobile and R. Tempone, "A systematic approach to model validation based on Bayesian updates and prediction related rejection criteria," *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2517-2539, 2008.

[27] S. Ferson, W. Oberkampf and L. Ginzberg, "Model validation and predictive capability for the thermal challenge problem," *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2408-2430, 2008.

[28] S. Ferson, W. Oberkampf and L. Ginzberg, "Validation of imprecise probability models," *International Journal of Reliability, Quality and Safety Engineering,* vol. 3, pp. 3-22, 2009.

[29] R. Rebba and S. Mahadevan, "Computational methods for model reliability assessment," *Reliability Engineering & System Safety,* vol. 93, pp. 1197-1207, 2008.

[30] M. Stephens, "ECDF statistics for goodness of fit and some comparisons," *Journal of the American Statistical Association,* vol. 69, pp. 730-737, 1974.

[31] R. Rebba and S. Mahadevan, "Model predictive capability assessment under uncertainty.," *AIAA Journal,* vol. 44, pp. 2376-2384, 2006.

[32] S. Mahadevan and R. Rebba, "Validation of models with multivariate output," *Reliability Engineering and System Safety,* vol. 91, pp. 861-871, 2006.

[33] O. Balci and R. G. Sargent, "A methodology for cost-risk analysis in the statistical validation of simulation models," *Communications of the ACM,* vol. 24, pp. 190-197, 1981.

[34] S. Mahadevan and J. McFarland, *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2467-2479, 2007.

[35] K. V. Mardia, J. T. Kent and R. F. Gunst, Multivariate Analysis, London: Academic Press, 1979.

[36] R. Rebba, "Model validation and design under uncertainty, PhD Thesis," Vanderbilt University, 2005.

[37] S. Mahadevan and J. McFarland, *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2467-2479, 2007.

[38] R. G. Hills and T. G. Trucano, "Statistical validation of engineering and scientific models: background, Technical Report," Sandia National Laboratories, 1999.

[39] J. B. Weathers, R. Luck and J. W. Weathers, "An exercise in model validation: Comparing univariate statistics and Monte Carlo-based multivariate statistics," *Reliability Engineering and System Safety,* vol. 94, pp. 1695-1702, 2009.

[40] W. Chen, L. Baghdasaryan and J. Cao, "Model validation via uncertainty propagation and data transformations," *AIAA Journal,* vol. 42, 2004.

[41] R. G. Hills and T. G. Trucano, Sandia National Laboratories, 2002.

[42] T. Buranathiti, J. Cao, W. Chen, L. Baghdasaryan and Z. Xia, "Approaches for model validation: Methodology and illustration on a sheet metal flanging process," *ASME Journal of Manufacturing Science and Engineering,* vol. 128, 2006.

[43] R. Zhang and S. Mahadevan, "Bayesian methodology for reliability model acceptance," *Reliability Engineering & System Safety,* vol. 80, pp. 95-103, 2003.

[44] R. Rebba, S. Huang, Y. Liu and S. Mahadevan, "Statistical validation of simulation models," *International Journal of Materials and Product Technology,* vol. 25, pp. 164-181, 2006.

[45] R. Rebba, S. Huang and S. Mahadevan, "Validation and error estimation of computational models," *Reliability Engineering and System Safety,* vol. 91, pp. 1390-1397, 2006.

[46] S. Mahadevan and R. Rebba, "Validation of reliability computational models using Bayes networks," *Reliability Engineering and System Safety,* vol. 87, 2005.

[47] R. E. Kass and A. Raftery, "Bayesian factors," *Journal of the American Statistical Association,* vol. 90, pp. 773-795, 1995.

[48] S. Sankararaman, Y. Ling and S. Mahadevan, *Engineering Fracture Mechanics,* vol. 78, pp. 1487-1504, 2011.

[49] S. Sankararaman and S. Mahadevan, "Model validation under epistemic uncertainty," *Reliability Engineering and System Safety,* vol. 96, pp. 1232-1241, 2011.

[50] X. Jiang and S. Mahadevan, "Bayesian validation assessment of multivariate computational models," *Journal of Applied Statistics,* vol. 35, pp. 49-65, 2008.

[51] X. Jiang and S. Mahadevan, "Bayesian inference method for model validation and confidence extrapolation.," *Journal of Applied Statistics,* vol. 36, pp. 659-677, 2009.

[52] Y. Pai, "Investigation of Bayesian model validation framework for dynamic systems, Master's thesis," University of Michigan, 2009.

[53] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis," *Journal of the Royal Statistical Society Series B,* vol. 61, pp. 611-622, 1999.

[54] Z. Xi, Y. Fu and R. Yang, "An ensemble approach for model bias prediction," *SAE International Journal of Materials and Manufacturing,* 2013.

[55] M. Papila and R. T. Haftka, "Pointwise Bias Error Bounds for Response Surface Approximations," *AIAA journal,* vol. 43, pp. 1797-1807, 2005.

[56] Z. Xi, Y. Fu and R.-J. Yang, "Model bias characterization in the design space under uncertainty," *International Journal of Performability Engineering,* vol. 9, pp. 433-444, 2013.

[57] Z. Jiang, W. Chen, Y. Fu and R.-J. Yang, "Reliability-based design optimization with model bias and data uncertainty," *SAE International Journal of Materials and Manufacturing,* 2013.

[58] A. Bichon, S. Mahadevan and M. Eldred, "Reliability-based design optimization using efficient global reliability analysis," in *The 50th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics and Materials Conference,* 2009.

[59] M. Kennedy and A. O'Hagan, "Bayesian calibration of computer models," *Journal of the Royal Statististical Society Series B,* vol. 63, 2001.

[60] M. Bayarri, J. Berger, R. Paulo, J. Sacks, J. Cafeo, J. Cavendish, C. Lin and J. Tu, "A framework for validation of computer models," *Technometrics,* vol. 49, pp. 138-154, 2007.

[61] D. Higdon, C. Nakhleh, J. Gattiker and B. Williams, "A Bayesian calibration approach to the thermal problem," *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2431-2441, 2008.

[62] W. Chen, Y. Xiong, K. Tsui and S. Wang, "Design-driven validation approach using Bayesian prediction models," *ASME Journal of Mechanical Design,* vol. 130, 2008.

[63] W. Chen and S. Wang, "Bayesian validation of computer models," *Computer Methods in Applied Mechanics and Engineering,* vol. 198, 2009.

[64] A. Sklar, "Fonctions de Répartition à N Dimensions et Leurs," *Publications de l'Institut de Statistique de l'Université de,* vol. 8, pp. 229-231, 1959.

[65] D. Huard, G. Evin and A. Favre, "Bayesian Copula selection," *Computational Statistics & Data Analysis,* vol. 51, pp. 809-822, 2006.

[66] B. Roser, An Introduction to Copulas, New York: Springer, 1999.

[67] J. Fermanian, "Goodness-of-fit tests for copulas," *Journal of Multivariate Analysis,* vol. 95, pp. 119-152, 2005.

[68] X. Chen and Y. Fu, "Pseudo-likelihood ratio tests for semiparametric multivariate copula model selection," *La Revue Canadienne de Statistique,* vol. 33, pp. 389-414, 2005.

[69] V. Panchenko, "Goodness-of-fit test for copulas," *Physica A: Statistical Mechanics and its Applications,* vol. 355, pp. 176-182, 2005.

[70] E. Jaynes and G. Bretthorst, Probability Theory: The Logic of Science, Cambridge University Press, 2003.

[71] G. Barnes, "A comparative study of nonlinear optimization codes: thesis for M.S. in engineering," University of Texas at Austin, 1967.

[72] Z. Xi, H. Pan, Y. Fu and R. Yang, "A copula-based approach for model bias characterization," *SAE International Journal of Passenger Cars - Mechanical Systems,* vol. 7, pp. 781-786, 2014.

[73] T. Calinski and J. Harabasz, "A dendrite method for cluster analysis," *Communication of Statistical Theory Methods,* vol. 3, pp. 1-27, 1974.

[74] Y. Ling and S. Mahadevan, "Quantitative model validation techniques: new insights," *Reliability Engineering and System Safety,* vol. 111, pp. 217-231, 2013.

[75] H. Sarin, M. Kokkolaras, G. Hulbert, P. Papalambros, S. Barbat and R. Yang, "Comparing time histories for validation of simulation models: error measures and metrics," *Journal of Dynamic Systems, Measurement and Control,* vol. 132, 2010.

[76] A. Genz, Computation of Multivariate Normal and t Probabilities, Springer, 2009.

[77] Z. Xi, B. D. Youn and C. Hu, "Random field characterization considering statistical dependence for probability analysis and design," *Journal of Mechanical Design,* vol. 132, 2010.

[78] K. Pearson, "Mathematical contributions to the theory of evolution, X: supplement to a memoir on skew variation," *Philosophical Transactions of the Royal Society A,* vol. 197, pp. 443-459, 1901.

[79] H. E. Daniels, "Saddlepoint approximations in statistics," *The Annals of Mathematical Statistics,* vol. 25, pp. 631-650, 1954.

[80] E. T. Jaynes, "Information theory and statistical mechanics," *Physical Review,* pp. 620-630, 1957.

[81] N. L. Johnson, S. Kotz and N. Balakrishnan, Continuous univariate distributions, New York: Wiley, 1994.

[82] R. H. Myers and D. C. Montgomery, Response surface methodology: Process and product in optimization using designed experiments, New York: Wiley & Sons, 1995.

[83] K. J. Dowding, M. Pilch and R. G. Hills, "Formulation of the thermal problem," *Computer Methods in Applied Mechanics and Engineering,* vol. 197, pp. 2385-2389, 2008.

[84] C. Andrieu-Renaud, B. Sudret and M. Lemaire, "The PHI2 method: A way to compute time-variant reliability," *Reliability Engineering & System Safety,* vol. 48, pp. 75-86, 2004.

[85] Z. Hu and X. Du, "Time-dependent reliability analysis with joint upcrossing rates," *Structural and Multidisciplinary Optimization,* vol. 48, pp. 893-907, 2013.

[86] D. Higdon, C. Nakhleh, J. Gattiker and B. Williams, "A Bayesian calibration approach to the thermal problem," *Computer Methods and Applied Mechanical Engineering,* vol. 197, pp. 2431-2466, 2008.

[87] V. Pakrashi, B. Basu and A. O'Connor, "Structural damage detection and calibration using a wavelet-kurtosis technique," *Engineering Structures,* vol. 29, pp. 2097-2108, 2007.

[88] J. McFarland, S. Mahadevan, V. Romero and L. Swileir, "Calibration and uncertainty analysis for computer simulations with multivariate output," *AIAA Journal,* vol. 46, pp. 1253-1265, 2008.

[89] M. Manfren, N. Aste and R. Moshksar, "Calibration and uncertainty analysis for computer models - A meta-model based approach for integrated building energy simulation," *Applied Energy,* vol. 103, pp. 627-641, 2013.

[90] B. D. Youn, B. C. Jung, Z. Xi, S. B. Kim and W. R. Lee, "A hierarchical framework for statistical model calibration in engineering product development," *Computer Methods in Applied Mechanics and Engineering,* vol. 200, pp. 1421-1431, 2011.

[91] Z. Xi and R. Yang, "Reliability Analysis with Model Uncertainty Coupling with Parameter and Experiment Uncertainties: A Case Study of 2014 V&V Challenge

Problem," *ASME Journal of Verification, Validation and Uncertainty Quantification,* 2015.

[92] Z. Xi, C. Hu and B. D. Youn, "A comparative study of probability estimation methods for reliability analysis," *Structural and Multidisciplinary Optimization,* vol. 45, pp. 33-52, 2012.

[93] B. D. Youn, K. K. Choi, R.-J. Yang and L. Gu, "Reliability-based design optimization for crashworthiness of vehicle side impact," *Structural and Multidisciplinary Optimization,* vol. 26, pp. 272-283, 2004.

[94] Z. Jiang, W. Chen, Y. Fu and R.-J. Yang, "Reliability-based design optimization with model bias and data uncertainty," *SAE International Journal of Materials and Manufacturing,* vol. 6, 2013.

[95] G. Taguchi, On Robust Technology Development: Bringing Quality Engineering Upstream, New York: ASME Press, 1993.

[96] I. E. Madu, "Robust Regression Metamodel for a Maintenance Float Policy," *International Journal of Quality and Reliability Management,* pp. 433-456, 1999.

[97] E. J. Cramer, J. E. Dennis Jr., P. D. Frank, R. M. Lewis and G. R. Shubin, "Problem Formulation for Multidisciplinary Optimization," *SIAM Journal of Optimization,* vol. 4, pp. 754-776, 1994.

[98] K. Deb, "Current trends in evolutionary multi-objective optimization," *International Journal of Simulation and Multidisciplinary Design Optimization,* vol. 1, pp. 1-8, 2007.