Reliability Engineering and System Safety 169 (2018) 579–592

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

journal homepage: www.elsevier.com/locate/ress

A new method for model validation with multivariate output

Luyi Li*, Zhenzhou Lu

School of Aeronautics, Northwestern Polytechnical University, PO Box 120, Xi'an 710072, PR China

ARTICLE INFO

Keywords:
Model validation
Multivariate output
Principal component analysis
Area metric

ABSTRACT

Traditional methods for model validation assessment mainly focus on validating a single response. However, for many applications joint predictions of the multiple responses are needed. It is thereby not sufficient to validate the individual responses separately, which ignores correlation among multiple responses. Validation assessment for multiple responses involves comparison with multiple experimental measurements, which makes it much more complicated than that for single response. With considering both the uncertainty and correlation of multiple responses, this paper presents a new method for validation assessment of models with multivariate output. The new method is based on principal component analysis and the concept of *area metric*. The method is innovative in that it can eliminate the redundant part of multiple responses while reserving their main variability information in the assessment process. This avoids directly comparing the joint distributions of computational and experimental responses. It not only can be used for validating multiple responses at a single validation site, but also is capable of dealing with the case where observations of multiple responses are collected at multiple validation sites. The new method is examined and compared with the existing *u-pooling* and *t-pooling* methods through numerical and engineering examples to illustrate its validity and potential benefits.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

With the rapid development of computers' abilities, an increasing number of computational models have been established for simulating the system behavior and supporting decision instead of the expensive physical experiments in many research fields, such as risk analysis [1–3], engineering design and performance estimation. The increased dependence on using computational simulation models in engineering presents a critical issue of confidence in modeling and simulation accuracy. Verification and validation (V&V) are the primary means to assess accuracy and reliability of computational simulations in engineering. *Verification* is the process of determining that a 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 model is an accurate representation of the real world from the perspective of the intended uses of the model [4]. That is, verification addresses the accuracy of the numerical solution produced by the computer code as compared to the exact solution of the conceptual model. In the verification, how the conceptual model relates to the "real world" is not an issue. Validation addresses the accuracy of the conceptual model as compared to the "real world", i.e., experimental measurements [4–7]. As Roache [8] stated, "verification deals with mathematics, while validation deals with physics". In general, code verification and numerical error estimation which are the two types of model verification activities should be completed before model validation activities are conducted or at least before actual comparison of computational results are made with experimental results [4–7]. The reason is clear. We should have convincing evidence that the computational results obtained from the code reflect the physics assumed in the models implemented in the code and that these results are not distorted or polluted due to coding errors or large numerical solution errors. As pointed out in literatures [7,9], if a researcher/analyst does not provide adequate evidence about code verification and numerical error estimation in a validation activity, the conclusions presented are of dubious merit. If conclusions from a defective simulation are used in high consequence system decision-making, disastrous results may occur. After decades-long development, while verification is well established, rigorous model validation assessment is not. In the present work, we also restrict our attention to model validation assessment.

Various model validation methods have presently been developed for assessing model validity. These methods are generally classified into four categories [10]: classical hypothesis testing [11], Bayesian factor [12,13], frequentist's metric [7,14] and area metric [9]. Classical hypothesis testing mainly focuses on determining which of the two alternative propositions (where the null hypothesis usually is that the model is correct and the alterative one is that the model is not correct.) is correct. It cannot assesse the quantitative accuracy of a model. The Bayesian factor approach is primarily interested in evaluating the probability (i.e.,

* Corresponding author.
E-mail addresses: luyili@nwpu.edu.cn, luyili@mail.nwpu.edu.cn (L. Li).

https://doi.org/10.1016/j.ress.2017.10.005
Received 6 July 2016; Received in revised form 6 October 2017; Accepted 7 October 2017
Available online 10 October 2017
0951-8320/© 2017 Elsevier Ltd. All rights reserved.

the belief) that the model is correct by incorporating an analyst's prior belief of model validity. Although it has some advantages over the classical hypothesis testing method, the Bayesian approach still cannot provide quantitative accuracy of the model, but gives a "yes" or "no" answer [9]. Different from the classical hypothesis testing and Bayesian factor, the frequentist's metric can help to quantitatively assess the adequacy of the given model for an application of interest. However, the main limitation of this method is that it is based on comparing means or other summary statistics (e.g. the maximum values) of the experimental data and model predictions. Thus, it considers only the central tendencies or other specific behaviors of data and predictions rather than their entire distributions. With this consideration, Ferson et al. [9] proposed an area metric method which assesses the accuracy of the given model by comparing the difference between the cumulative distribution function (CDF) of the model predictions and the empirical CDF of experimental data. This method can provide an objective assessment of the model accuracy by considering the entire distributions of the model predictions and experimental data. It generalizes the frequentist's metric method to validation assessment that focuses only on the mean behavior or other summary statistics of predictions and observations. More importantly, by employing the *u-pooling* technique to incorporate the experimental data collected at multiple validation sites into a single metric, the method in [9] can be used to assess the overall performance of a model against all the experimental data in the validation domain. Here, a validation site is a specific setting of the model input variables at which the accuracy of a model is validated against the measured quantities.

In spite of numerous advantages, Ferson's area metric method is based on comparison of the marginal distributions of the model predictions and experimental data. It, thus, is only suitable for validation assessment of models with single response or with multiple uncorrelated responses [15]. In practice, however, correlated multiple responses are often of interest. For instance, multiple response quantities, like stress, strain and displacement, etc., often are predicted simultaneously from the same experiment at a single location. These different quantities are correlated, since they are based on the same inputs. On the other hand, the interested model responses from the same experiment may change with location (in space or time coordinate). In this case, there is a strong correlation between any pair of response quantities from the same experiment. Model validation assessment for multiple correlated response quantities is much more complicated than that for single response, as it needs to consider both uncertainty and correlation of the multiple responses in the validation assessment process.

Validation assessment with "multivariate data" has drawn attention of the scientists in meteorological and climate community for decades [16–19], where verification and validation are combined in one verification process. Many methods, such as the minimum spanning tree (MST) histograms, multivariate rank (MVR) histograms and bounding boxes (BBs) [16,17], have been developed in this community to assess the multidimensional ensemble reliability of their forecasts within the context of high-dimensional multivariate data. The bounding boxes in BBs approach are defined by the minimum and maximum values of an ensemble forecast. They, thus, are unduly affected by outliers and may fail in characterizing the bulk ensemble properties appropriately [16]. Both the MST histograms and MVR histograms are based on the reliability criterion that the ensemble members should be statistically indistinguishable from the observations. This is similar to that underlying the u-pooling technique. From that, insights can be gained into the average dispersion characteristics of the multidimensional ensemble forecasts.

In the context of engineering community, classical hypothesis testing and Bayesian factor have been extended to validation assessment of models with multivariate output [20,21]. These methods, however, inevitably inherit their disadvantages in univariate case. That is, they cannot provide quantitative accuracy of the model, but give a "yes" or "no" answer. One intuitive and straightforward method which can account for both uncertainty and correlation of the multiple responses is directly comparing the joint CDF of the model predictions and the multivariate empirical CDF of experimental data. Nevertheless, in many engineering cases, experimental observations are usually very sparse due to the expense of full-scale physical experiments. This means that the correlation structure of experimental data is poorly known. Thus, the multivariate empirical CDF used for capturing correlation information in the data will be a very poor representation of what actually exists in the real physical system. Even if the experimental data are sufficiently collected, the method would still suffer severely from the "curse of dimensionality" when computing the high-dimensional integration in the metric.

To provide a quantitative assessment for the accuracy of the computational models with correlated multivariate output with considering both the uncertainty and correlation, Li et al. [15] proposed two validation metrics (A validation metric is a formal measure of the mismatch between predictions and experimental data.) by extending the idea of *area metric* and *u-pooling* method. They are probability integral transformation (*PIT*) *area metric* for a single validation site and *t-pooling metric* for multiple validation sites. Both the metrics are based on the multivariate PIT theorem. Although this method can take into account both the uncertainty and correlation of the multivariate output in the process of model validation assessment, it still requires estimating the joint CDF of model responses to transform the multivariate experimental observations. This is often unavailable for high-dimensional response space.

In this paper, a new method is proposed for the validation assessment of models with multiple correlated responses based on the principal component analysis (PCA) and the idea of area metric. By the PCA, the correlated multiple output responses are transformed into a set of orthogonal principal components (PCs) with the first few PCs containing nearly all the variability of the multiple outputs. Then the area metric can be applied to each PC to obtain the corresponding validation metric value. The total model validation metric is obtained by aggregating these validation metric values of PCs with their corresponding PCA weights. With the proposed method, validation assessment of models with correlated multiple high-dimensional responses can be decomposed into a series of model validation assessments in independent one-dimensional space. This allows considering both the uncertainty and correlation of the multiple responses without estimating their joint CDF. Thus, it is more feasible for the practical applications.

The rest of the paper is organized as follows. Section 2 briefly reviews the existing area metric and u-pooling technique. The theory of the proposed method is detailed in Section 3. In Sections 4 and 5, an illustrative mathematical example and a simple risk analysis model are respectively used to show the advantages of the proposed method by comparing with the existing ones. The proposed method is also applied to an engineering model where the experimental data are assumed to be sparse in Section 6. Section 7 discusses some practical issues in the application of the proposed method and the future works to be done. Finally, the conclusions come at the end of the paper.

2. Review of the area metric and u-pooling based metric

For the relevant system response quantity (SRQ) y, let $F^m(y)$ denote the CDF of y predicted by the computational model at a single validation site, and $S_n^e(y)$ represent the empirical CDF of the corresponding experimental data of y. In $F^m(y)$ and $S_n^e(y)$, the superscripts m and e represent "model" and "experiment", respectively and the subscript n is the sample size of the data set (Note that since the experimental data are often sparse, empirical CDF is usually used to represent their uncertainty which is an exact representation of the data regardless of the amount of data, and does not require any assumptions [9]. In the case where predictions are also sparse, empirical CDF is also used for describing the uncertainty of predictions). As shown in Fig. 1, area metric uses the area between the prediction distribution $F^m(y)$ and the data distribution $S_n^e(y)$ as the measure of the mismatch between the model and the data.

Download English Version:

https://daneshyari.com/en/article/5019299

Download Persian Version:

https://daneshyari.com/article/5019299

Daneshyari.com