Reliability Engineering and System Safety 131 (2014) 40–52

Contents lists available at ScienceDirect

Reliability Engineering and System Safety

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

Variable-fidelity model selection for stochastic simulation

Joshua Mullins, Sankaran Mahadevan *

Department of Civil and Environmental Engineering, Vanderbilt University, TN 37235, United States

ARTICLE INFO

Article history:
Received 1 June 2013
Received in revised form
11 June 2014
Accepted 23 June 2014
Available online 2 July 2014

Keywords:
Stochastic simulation
Resource allocation
Model selection
Model discrepancy
Computational effort
Time dependent simulation

ABSTRACT

This paper presents a model selection methodology for maximizing the accuracy in the predicted distribution of a stochastic output of interest subject to an available computational budget. Model choices of different resolutions/fidelities such as coarse vs. fine mesh and linear vs. nonlinear material model are considered. The proposed approach makes use of efficient simulation techniques and mathematical surrogate models to develop a model selection framework. The model decision is made by considering the expected (or estimated) discrepancy between model prediction and the best available information about the quantity of interest, as well as the simulation effort required for the particular model choice. The form of the best available information may be the result of a maximum fidelity simulation, a physical experiment, or expert opinion. Several different situations corresponding to the type and amount of data are considered for a Monte Carlo simulation over the input space. The proposed methods are illustrated for a crack growth simulation problem in which model choices must be made for each cycle or cycle block even within one input sample.

1. Introduction

Stochastic simulation plays a critical role in the prediction of system performance and estimation of reliability in complex engineering systems. In this context, the purpose of the simulation is to propagate all available information forward to a system-level output quantity of interest (QoI) while properly accounting for all the uncertainties that are present at each level of the hierarchy. Standard uncertainty propagation techniques result in a probability distribution of the QoI which incorporates aleatory uncertainty in the model inputs. An additional important need is to incorporate epistemic uncertainty arising from data and model uncertainties into the distribution of the QoI. Data uncertainty arises from sparse, imprecise, missing, subjective, or qualitative data, and also from measurement and data processing errors. Model uncertainty may arise due to model form assumptions, model parameters, and solution approximations. Data uncertainty may also affect the estimation of model parameters. If the various types of uncertainty are represented in a probabilistic format, then the presence of epistemic uncertainty leads to a stochastic model prediction even for a given input vector. The amount of computation time required to incorporate all the uncertainties may be very large when using conventional methods such as Monte Carlo simulation (MCS) with high fidelity system models. In some problems the required time can become so large that the problem is completely intractable.

The situation explored in this paper is one in which a stochastic simulation must be performed for the purpose of system-level uncertainty quantification (UQ) and reliability analysis. A comprehensive analysis should accurately predict the full distribution of an output QoI by including all known uncertainties. In such a case, we assume that a high fidelity model already exists, but it is too expensive to evaluate at every sample point. Once cheaper models are developed, the high fidelity model is still available, but we must decide when to use it efficiently in order to obtain results of desired accuracy within a reasonable amount of time. With this goal in mind, this paper develops a multi-fidelity model selection methodology that combines the use of both efficient simulation and surrogate modeling. The proposed framework uses surrogate models to inform the model selection decision at each random sample of the MCS (or each spatial location or time step, depending on the problem) and then executes a single selected model combination at this input. In this way, we can account for the possibility that different models may be adequate in different domains (including cheaper vs. expensive models, and even models with competing physical hypotheses). The proposed methodology accommodates different types of information about the QoI (such as actual observations, expert opinion etc.).

A simple mathematical example is first implemented to demonstrate a situation in which no prior information is available about the appropriate ranking of fidelities among candidate models. In such a case, we must have additional information about the QoI in order to define the relative accuracies in terms of a discrepancy. Otherwise, we cannot make an informed decision

* Corresponding author.

http://dx.doi.org/10.1016/j.ress.2014.06.011
0951-8320/

based on computational effort alone. Next, a richer engineering example is used to demonstrate the proposed methods for a more complicated simulation where inputs vary both spatially and temporally. Additionally, this second example establishes the model selection approach for a case where the ranking of fidelities among the candidate models is known a priori.

2. Background

There are two basic classes of approaches available to manage computationally intractable UQ problems; either the number of samples required can be reduced in an intelligent way, or the model being evaluated can be simplified so that less time is needed for each sample. Methods of efficient stochastic simulation with respect to the number of samples have been explored in studies on reliability analysis and design optimization. One inexpensive way of propagating input variability and/or parameter uncertainty through a system model is a first-order Taylor series expansion, which requires only $n+1$ function evaluations for n uncertain variables. This method is referred to as a first-order second moment (FOSM) approach in the reliability analysis literature [1]. Other reliability analysis approaches take advantage of the idea that sometimes only a particular point on the distribution of the QoI is of interest (e.g., probability that stress or deformation exceeds a particular value). This type of analysis typically uses Newton-like optimization methods to search in an equivalent uncorrelated standard normal space for the most probable point (MPP) on a limit state related to the QoI [2,3]. The failure probability is then approximated via the first-order reliability method (FORM) or the second-order reliability method (SORM) [1].

Alternatively, within the context of MCS for reliability analysis, methods such as importance sampling modify the sampling distribution to ensure that more samples fall within a region of interest, thereby reducing the total number of samples needed for the analysis. For example, Harbitz's importance sampling approach [4] creates a sampling distribution centered at the MPP; adaptive methods are also available to update the importance sampling distribution after ever few samples [5,6]. Because each of the aforementioned approaches searches only in a region of interest, they can be restrictive if the goal of the analysis is to determine the entire distribution of the QoI. To calculate the entire distribution, these methods may be applied at several regions of interest and interpolated (note: interpolation introduces additional error and uncertainty), or we must revert to a full MCS.

If a full MCS is to be performed, it may be infeasible to evaluate a high-fidelity physics model (e.g. nonlinear finite element analysis with a very fine mesh) for every Monte Carlo sample, so the class of approaches aimed at reducing computation time per sample is utilized instead. Cheaper models (in terms of CPU time per evaluation) which may be in the form of mathematical surrogate models (also referred to as response surfaces or metamodels), reduced order models, or reduced physics models have been pursued in this regard. Common surrogate models include simple regression models, Gaussian process (GP) or Kriging models [7,8], polynomial chaos expansion models [9], support vector machines [10], and neural networks [11]. However, additional error is introduced to the system prediction by these surrogates, so it is preferable to make selective use of the high-fidelity model at some sample points as allowed by the computational budget. In this case, there is a clear difference in fidelity between the full model and the surrogate, and the decision only involves handling the tradeoff between accuracy and computational expense.

On the other hand, there are also instances in which multiple competing physics-based models are available for the same prediction, but it is not obvious which of them represents reality more accurately for the application of interest. For example, one physical phenomenon may be more dominant in one region of the input space than another. This situation has been addressed by quantifying the discrepancy between the model prediction and some performance benchmark [12]. Since it is not clear which model is providing the better estimate of reality, this benchmark must come from an additional piece of information, most commonly a physical observation, known exact solution, or expert opinion. After a benchmark is selected, the decision is again a tradeoff of accuracy vs. computational expense. To develop a methodology for model selection, it must first be clear whether the ranking of fidelities among candidate models is consistent over the entire domain or whether it may change as a function of the inputs. Once this distinction is made, the goal is to select among available models in an intelligent and efficient manner.

Given these various scenarios, the general model selection problem can be posed as a decision based on one or more of the following criteria: (1) parsimony vs. accuracy in regression, (2) discrepancy compared to a benchmark, and (3) computational expense. The problem of selecting among multiple regression models has frequently been addressed by considering the first of these criteria. In several existing metrics based on the information theory, accuracy is indicated by the sum of squares of residuals or the maximum likelihood with respect to training data, and parsimony is indicated by the number of terms in the model. Both of these components are included within Mallows' C_p statistic [13], the Akaike information criterion [14] based on information entropy [15], the Bayesian information criterion [16], and the minimum description length [17]. Each of these is addressing the tradeoff between bias and variance in available models, since additional complexity will reduce the residuals (i.e. variance) but also risks "overfitting", which may increase bias. Typically, the outcome of this problem is the choice of a single model from a set, or possibly a new model which averages a set of available models.

When the models are not statistical regression models, but rather physics-based models, these metrics, based on the accuracy vs. parsimony criterion, can be difficult and inappropriate to employ for a couple of reasons. First, the forms of these models may be complex and in some cases impossible to write in an analytical form, so it will be difficult to define the parsimony of the model. Second, different physical hypotheses may attribute different physical mechanisms as causes for the observed behavior, which makes the associated models difficult to compare with respect to parsimony, and they cannot be combined in a natural way. Therefore, it is more appropriate to look only at model discrepancy and computational expense when addressing this selection scenario. The tradeoff between accuracy (w.r.t. a benchmark) and computational effort in physics-based models has been addressed in the system design literature. It is possible to develop a more accurate model by introducing additional phenomenological features (i.e. improve the model form) and/or by improving the quality of the numerical approximation to the solution (e.g. discretization refinement). Available methods [18–20] assign utilities to the candidate models based on expected performance and explore the tradeoff between utility and the associated costs (both model building cost and execution cost). The use of multiple models with varying degrees of fidelity is also studied in the design optimization literature; this is referred to as model management [21]. Lower fidelity models to evaluate the objective and constraints include surrogate models or reduced-order models [22,23].

This paper develops a model management framework for *uncertainty quantification*, based on model discrepancy and computational effort, in the presence of both aleatory and epistemic uncertainty. Model discrepancy is probabilistically quantified for different model choices and traded off against computational effort to develop an optimization-based model selection criterion