Reliability Engineering and System Safety 111 (2013) 273–286

Contents lists available at SciVerse ScienceDirect

Reliability Engineering and System Safety

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

A sequential approach for stochastic computer model calibration and prediction

Jun Yuan, Szu Hui Ng*

Department of Industrial and Systems Engineering, National University of Singapore, 117576 Singapore, Singapore

ARTICLE INFO

Article history:
Received 31 May 2012
Received in revised form
9 November 2012
Accepted 10 November 2012
Available online 22 November 2012

Keywords:
Computer model calibration
Stochastic computer simulation
Guassian process model
Bayesian analysis
Parameter uncertainty
Uncertainty quantification

ABSTRACT

Computer models are widely used to simulate complex and costly real processes and systems. When the computer model is used to assess and certify the real system for decision making, it is often important to calibrate the computer model so as to improve the model's predictive accuracy. A sequential approach is proposed in this paper for stochastic computer model calibration and prediction. More precisely, we propose a surrogate based Bayesian approach for stochastic computer model calibration which accounts for various uncertainties including the calibration parameter uncertainty in the follow up prediction and computer model analysis. We derive the posterior distribution of the calibration parameter and the predictive distributions for both the real process and the computer model which quantify the calibration and prediction uncertainty and provide the analytical calibration and prediction results. We also derive the predictive distribution of the discrepancy term between the real process and the computer model that can be used to validate the computer model. Furthermore, in order to efficiently use limited data resources to obtain a better calibration and prediction performance, we propose a two-stage sequential approach which can effectively allocate the limited resources. The accuracy and efficiency of the proposed approach are illustrated by the numerical examples.

1. Introduction

Computer models are generally designed to simulate real processes. In real applications, most processes are quite complex, making it impossible to develop realistic analytical models of the system, thus computer models become one of the best choices to represent these real complex systems. When the computer model is used to predict the behavior of the real system so as to assess and certify the system's reliability and safety that eventually is used for better decision making, it is important to improve the model's predictive accuracy and capability. In computer model analysis, both validation and calibration are closely related to the model's predictive performance. Validation is a process to confirm whether the computer model precisely represents the real process, while calibration is a process to adjust the unknown input parameters by comparing the computer model output with the real observed data so as to ensure that the computer model fits well to the real process. These unknown input parameters are usually unobservable or unmeasurable in the real process but need to be specified in the computer model. One possible relationship among calibration, validation and prediction is provided by [1] which shows that both calibration and validation are important to improve the model's predictive performance. However, in this discussion, we focus on providing a calibration procedure to update the model. A brief discussion is also provided for validation based on our proposed calibration procedure. More details about validation in scientific computing can be found in [1] and references therein. More specifically, we propose a calibration procedure to adjust the unknown parameters and obtain the predictive distributions of the calibrated model. The calibrated model usually has better predictive performance as it fits better to the real process with more accurate unknown calibration parameter values. These obtained predictive distributions can be further used for validation and eventual prediction purposes. Then the final calibrated and validated model can be applied for intended use. The importance of calibration has been recognized in many practical models that are used to evaluate the reliability and safety of complex physical and social systems, such as nuclear radiation release model [2], hydrologic model [3], and biological model [4]. For instance, disease transmission models are usually built to model the spread of the infectious diseases such as influenza A (HIN1) virus in order to find the best strategy to control their spread. Within the computer model, the illness attack rate is required to be set before using the model. However, its value is usually unknown in the real transmission process especially for a new outbreak, and the value may change based on the geographic area and population. This attack rate is known as the calibration parameter. Its value may significantly influence

* Corresponding author.
E-mail addresses: jyuan@nus.edu.sg (J. Yuan), isensh@nus.edu.sg (S.H. Ng).

0951-8320/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.ress.2012.11.004

the predictive performance of the computer model and the subsequent accuracy of the decision making. Therefore, it is necessary to estimate or adjust this value before the computer model can be used for further analysis. The procedure to adjust this parameter to improve the predictive accuracy is known as calibration.

Various approaches have been proposed for computer model calibration [5,6]. One type of calibration approach is to find an effective and efficient algorithm which can be directly applied with limited data resources, such as the stochastic approximation methods discussed by [7]. However, this approach may not be efficient when the computer models are extremely time consuming and computationally expensive as it usually requires a relatively large number of simulation runs before converging to the optimal calibration parameter value. Another popular and much more efficient approach is to use surrogates, also known as emulators and metamodels, where simpler and faster statistical approximations are used instead of the original complex computer models. More discussion on using and developing surrogates for computer models can be found in [8,9].

A computer model is used to approximate the real process and the surrogate model is used to approximate the computer model. There are various uncertainties arising in calibrating the computer model and using the surrogate model to predict the behavior of the real process, such as parameter uncertainty, model inadequacy, observation error, etc. When the model is used for decision making, not only the point estimator but also the uncertainty information about the estimator is required to make more informed decisions and to have a better assessment of risk. Uncertainty quantification is an important issue for system's risk assessment and in conveying the credibility and confidence in system's reliability in order to support making better decisions. Therefore, it is important to account for various uncertainties in predicting the behavior of the real process. Within these uncertainties, calibration parameter uncertainty sometimes may have significant effects on overall predictive uncertainty, which plays an important role in decision making. Therefore, it is important to consider this uncertainty and its effects on the subsequent prediction in many real applications. [2] proposed a Bayesian approach for computer model calibration using the Gaussian process (GP) as a surrogate model. Their approach takes into account all sources of uncertainty in the computer model analysis including the calibration parameter uncertainty. However, their discussion is based on deterministic computer models. Hence the inherent stochastic error introduced in the stochastic computer model is not accounted for. Different from the deterministic model, simulation outputs from a stochastic model may be different for the same input levels. In most applications, the real systems of interest are often stochastic in nature. The stochastic models are usually required to assess the probability distribution of the outcome of interest and the expected output is a typical measure of performance of such systems. With the increasing application of the stochastic computer model, it is important to consider this stochastic error in the calibration and prediction so as to improve the calibration accuracy and the predictive performance. In this paper, we extend current work using surrogates for calibration to stochastic computer model calibration and prediction, which accounts for all the uncertainties mentioned in [2] and also the inherent stochastic error. [4] also discussed the calibration of the stochastic computer model using a surrogate based Bayesian approach. In their study, the Markov chain Monte Carlo (MCMC) method was used to simulate values from the posterior distribution of the calibration parameter and to obtain the predictive values. Their approach requires many millions of samples to guarantee the convergence, so it is time consuming to implement although they attempt to reduce the computational burden by simplifying the surrogate model. Similarly in [10], they also discussed the Bayesian calibration approach using the MCMC method. In this paper, different from [4] that use MCMC to approximate the calibration parameter distribution and the predictive distribution, we derive these distributions by estimating some parameters that are empirically shown to have insignificant effects on the predictive performance. We derive the predictive distributions for both the computer model and the real process which quantify the predictive uncertainties. We also derive the predictive distribution for the inadequacy term which measures the discrepancy between the computer model and the real process. This can be used for computer model validation such as in the validation procedure proposed by [11]. The derived posterior distributions provide analytical calibration and prediction results. Meanwhile, the proposed approach can significantly reduce the computational time compared to MCMC and this result is illustrated by the numerical examples. Furthermore, to improve the calibration and prediction process and to support better decisions, we extend this work to address the problem of allocating limited resources to reduce the predictive uncertainty of both the computer model and the real process. We propose a two-stage sequential approach for stochastic computer model calibration and prediction so as to obtain a better set of real and computer data required to build accurate surrogate and sequentially learn and improve the model by effectively allocating resources to select more appropriate points to sample at.

The purpose of this article is to develop an efficient sequential calibration and prediction procedure for the stochastic computer models. This includes the collection of computer model and real process data, the development of the surrogate model, the calibration and prediction of the stochastic computer model, the allocation of the resources to improve the model for eventual use in prediction, and the validation of the calibrated computer model, so as to achieve the improved computer model's prediction for decision makers. This paper is organized as follows: In Section 2, we describe the model used for calibration and prediction. In Section 3, we explain the proposed Bayesian calibration approach and derive the predictive distributions of the computer model, the real process and the model inadequacy term. In Section 4, we propose a two-stage sequential approach for stochastic computer model calibration and prediction and provide the implementation steps. Finally in Section 5, two numerical examples are provided to illustrate the proposed approach.

2. Model formulation

2.1. Stochastic model

Based on the model proposed by [2], the relationship between the real observation and simulation output can be represented by

$$z_i = \zeta(x_i) + e_i = S(x_i,\theta) + \delta(x_i) + e_i, \quad (1)$$

where z_i is the ith real observation at input level x_i, $\zeta(x_i)$ is the true output from the real process, $S(x_i,\theta)$ represents the "true" simulation output at x_i and the optimum calibration parameter θ. Similar to previous works, we see the optimum θ as the value that best fits the computer model output to the real process. This does not necessarily equate to the true value of the calibration parameter if it exists in the real process. As the purpose is to improve the predictive performance of the computer model, we focus on finding the best parameter value that matches the computer model to the real process. $\delta(x_i)$ is the model inadequacy or discrepancy term, which is considered to be independent of $S(x_i,\theta)$, and e_i is the observation error.