Computers and Structures 89 (2011) 2377–2387

Contents lists available at ScienceDirect

Computers and Structures

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

A forecasting metric for predictive modeling

Sez Atamturktur [a,*], François Hemez [b], Brian Williams [c], Carlos Tome [d], Cetin Unal [e]

[a] Department of Civil Engineering, Clemson University, SC, USA
[b] XCP 1-Division, Los Alamos National Laboratory, NM, USA
[c] CCS 6-Division, Los Alamos National Laboratory, NM, USA
[d] MST 8-Division, Los Alamos National Laboratory, NM, USA
[e] CCS DO-Division, Los Alamos National Laboratory, NM, USA

ARTICLE INFO

Article history:
Received 27 August 2010
Accepted 27 June 2011
Available online 23 July 2011

Keywords:
Modeling and simulation
Verification and validation
Model calibration
Bayesian inference
Extrapolation
Predictive maturity

ABSTRACT

In science and engineering, simulation models calibrated against a limited number of experiments are commonly used to *forecast* at settings where experiments are unavailable, raising concerns about the unknown forecasting errors. Forecasting errors can be quantified and controlled by deploying statistical inference procedures, combined with an experimental campaign to improve the fidelity of a simulation model that is developed based on sound physics or engineering principles. This manuscript illustrates that the number of experiments required to reduce the forecasting errors to desired levels can be determined by focusing on the proposed forecasting metric.

Published by Elsevier Ltd.

1. Introduction

This manuscript is concerned with simulation models used to forecast predictions in support of high-consequence decision-making on the performance of engineering systems, i.e. in the context of certification. Instead of relying on virgin models, i.e. models that are not calibrated or bias-corrected, we envision certification to be applied through a combined experimental and numerical campaign that relies on simulation models calibrated and bias corrected against experimental measurements. We are particularly interested in the quantification and control of errors associated with the forecasting predictions of these calibrated and bias corrected simulation models.

In certification, the purpose of simulation models is to reduce the number of required experiments, and is best illustrated by considering two extreme cases.

(1) *Purely empirical certification*: the absence of a sound simulation model where certification is only obtained based on experimental measurements.
(2) *Purely model-based certification*: the availability of a 'perfect' simulation model where certification needs practically no experimental measurements.

In a purely empirical approach, forecasting is commonly achieved by constructing a function that best fits the discrete settings of the calibration experiments. Then, the best-fitted function is exercised to make forecasting predictions at untested settings. Both the experimental uncertainty and the uncertainty in the curve-fitting process can be considered by making forecasting predictions that are "best estimates" with quantified uncertainties. In purely empirical certifications, the number of experiments necessary to train the best-fitted function can rapidly become prohibitive. Purely empirical certifications are further challenged with the fact that experiments are typically time-consuming and expensive and thus, even in the best cases are only available in limited numbers. Moreover, obtaining the measurements at the desired experimental settings may be prohibitive due to policy regulations or simply infeasible due to technical limitations. This resulting inevitable experimental scarcity is the primary reason behind the increased reliance on modeling and simulation in various scientific and engineering fields.

The availability of simulation models incorporating sound physics or engineering principles can significantly reduce the number of required experiments for certification. In the most extreme example, if the model can perfectly reproduce reality, the dependency on experimentation can be eliminated altogether. However, simulation models are naturally impaired by imprecise model parameters (known unknowns) and inaccuracies in the interpretation of the underlying physics or engineering principles (unknown unknowns). Therefore, experimental evidence is routinely required

* Corresponding author. Tel.: +1 864 656 3003; fax: +1 814 863 4789.
E-mail addresses: sez@clemson.edu (S. Atamturktur), hemez@lanl.gov (F. Hemez), bwilliams@lanl.gov (B. Williams), tome@lanl.gov (C. Tome), cu@lanl.gov (C. Unal).

0045-7949/$ - see front matter Published by Elsevier Ltd.
doi:10.1016/j.compstruc.2011.06.010

to improve the simulation model fidelity. Hence, in model-based certification, experiments are used to calibrate and bias-correct the physics-based simulation model instead of training an arbitrary best-fitted function.

Model calibration is achieved through the comparison of a cohort of model predictions with a family of experimental measurements. This comparison has two main objectives: (1) to reduce the uncertainty in the imprecise input parameters (known unknowns) and (2) to estimate the errors due to inadequate or missing physics (unknown unknowns). The manner in which we distinguish between these two interrelated objectives, parameter calibration versus discrepancy bias, is explained in Section 2 by a simple example.

The number of experiments needed to successfully achieve these two objectives of model calibration is heavily dependent upon the fidelity of the modeled physics and engineering principles to reality. If the simulation model lacks a vital principle, parameter, or interaction between principles or parameters in its formulation (i.e., large unknown unknowns), the fundamental ability of this model to capture the phenomena of interest is compromised. Indeed, attempts to improve the fidelity of an overly crude simulation model through an experimental campaign and model calibration would fail to yield satisfactory results. Hence, a crucial first step in certification involves assessing the suitability *of a simulation model* for use in forecasting. In Section 3, we outline three important assertions that we contend play a foundational role in determining the suitability of simulation models for forecasting.

Because the purpose of simulation models is to predict in lieu of experiments, such models calibrated against a reduced number of experimental measurements are routinely applied to ***forecast*** at untested settings, which brings up the issue of *unknown forecasting errors*. However, these unknown forecasting errors can be reduced by model calibration which improves the fidelity of the simulation model through parameter calibration (known unknowns) and bias-correction (unknown unknowns). Naturally, as the number of experimental measurements available for model calibration is consistently increased, the forecasting errors would be consistently reduced. Once a sufficient number of experiments are obtained, and the forecasting errors are reduced to desired levels, allocating resources to experimentations would have diminishing returns; thusly allocating further resources to experiments would not be justified. In Section 4, we explain our approach to estimating the forecasting errors of a given simulation model for a given experimental campaign. Estimating forecasting errors is of particular importance in science and engineering, especially when such forecasts are used to determine the expected performance level of an engineering system under worst-case scenarios. Such applications are common in the context of certification.

In this manuscript, we illustrate both our approach (1) for estimating forecasting errors for a given set of experimental measurements (Section 5) and (2) for determining the quantity of experimental measurements required to assess the usefulness of a simulation model and to reduce forecasting errors to the desired levels (Section 6). In Sections 5 and 6, we illustrate the merit of this procedure using two distinct material models representing data-rich and data-poor situations. Compelling as this approach is, a set of premises must be satisfied for the proposed approach to be applicable. In Section 7, we discuss the underlying premises and limitations of the proposed approach.

2. Estimation of model discrepancy

The central philosophy of model calibration is to improve the accuracy of model predictions by exploiting a collection of available experimental measurements. Thus, model calibration invariably requires the comparison of large numbers of simulation runs against experimental measurements. Over the past two decades, model calibration has evolved into two strategies, which differ in the methods through which they improve model accuracy. The first type is the parameter calibration approach that captures the inaccuracy of the model parameters. The second type is the bias correction approach that captures the inadequacy of the physics model. These two fundamental concepts are combined together in the landmark study of Kennedy and O'Hagan [1]. Kennedy and O'Hagan's [1] approach can simultaneously calibrate model parameters and correct discrepancy bias.

The parameter calibration approach has two distinct paradigms used for defining the improved parameter values. In the first approach, calibration is considered as an optimization problem. The objective function, which constitutes some form of the disagreement between the experimental measurements and model predictions, is minimized over a subset of model parameters appropriately selected based on their uncertainty and sensitivity. If uncertainty in the experimental measurements and numerical predictions are included, the problem becomes calibration under uncertainty. The second approach to calibration is Bayesian inference, which explicitly acknowledges the uncertainty in the model parameters. In Bayesian inference, calibration is achieved by reducing the uncertainty in the models parameters and in turn reducing the uncertainty in the model output. Therefore, Bayesian inference is considered to be more refined in the way it handles uncertainty, compared to optimization-based procedures and thus is preferred for the present study.

The present manuscript adopts a Bayesian implementation of Kennedy and O'Hagan's method [1]. This implementation, derived from Higdon et al. [2], is built into a computer code called Gaussian Process Model – Simulation Analysis (GPM–SA) at the Los Alamos National Laboratory. It is deeply rooted in the following relation:

$$\hat{y}_{obs}(x) = y_{sim}(x,\theta) + \delta(x) + \varepsilon(x) \tag{1}$$

In Equation (1), the parameter x denotes the controlled variables. One must be careful not to mix control parameters with calibration parameters, denoted by θ in Equation (1). Control parameters, which can be controlled during experiments, define the domain of applicability. Calibration parameters on the other hand are either introduced by specific choices of assumptions or models, or they represent parameters that cannot be measured or controlled experimentally.

In Equation (1), $y_{sim}(x,\theta)$ corresponds to the model predictions, $\delta(x)$ corresponds to a discrepancy bias that represents the systematic bias, and $\varepsilon(x)$ represents the random experimental error. When these three terms are added together, they yield our best estimate for the "truth", $\hat{y}_{obs}(x)$ over the various settings of x in the domain of applicability. To reiterate Equation (1), if the discrepancy bias associated with a simulation model is known, the truth, $y_{obs}(x)$ can be computed by correcting model predictions, $y_{sim}(x,\theta)$ with the discrepancy bias $\delta(x)$.

One of the primary roles of experimental measurements is to supply information about the discrepancy bias at discrete points within the domain of applicability. We now introduce another term, x^t, which denotes the control parameter settings where the experimental measurements are available.

$$\hat{y}_{obs}(x^t) \sim y_{obs}(x^t) \tag{2}$$

With the formulation of Equation (1), one seeks to obtain the probability distribution of calibration parameters, θ by comparing model predictions $y_{obs}(x^t,\theta)$, to physical observations $y_{obs}(x^t)$ while simultaneously making an independent estimate for the discrepancy bias, $\delta(x^t)$. As seen, the discrepancy bias is fundamentally different than the commonly adapted concept of "goodness-of-fit."

Below is a conceptual example explaining the difference between goodness-of-fit and discrepancy bias. Fig. 2 illustrates the