Calibration, validation, and sensitivity analysis: What's what

T.G. Trucano[a],*, L.P. Swiler[a], T. Igusa[b], W.L. Oberkampf[c], M. Pilch[c]

[a]*Optimization and Uncertainty Estimation Department, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-0819, USA*
[b]*Johns Hopkins University, Baltimore, MD 21218, USA*
[c]*Validation and Uncertainty Estimation Processes, Sandia National Laboratories, P.O. Box 5800, Albuquerque, NM 87185-0819, USA*

Available online 19 January 2006

Abstract

One very simple interpretation of calibration is to adjust a set of parameters associated with a computational science and engineering code so that the model agreement is maximized with respect to a set of experimental data. One very simple interpretation of validation is to quantify our belief in the predictive capability of a computational code through comparison with a set of experimental data. Uncertainty in both the data and the code are important and must be mathematically understood to correctly perform both calibration and validation. Sensitivity analysis, being an important methodology in uncertainty analysis, is thus important to both calibration and validation. In this paper, we intend to clarify the language just used and express some opinions on the associated issues. We will endeavor to identify some technical challenges that must be resolved for successful validation of a predictive modeling capability. One of these challenges is a formal description of a ''model discrepancy'' term. Another challenge revolves around the general adaptation of abstract learning theory as a formalism that potentially encompasses both calibration and validation in the face of model uncertainty.

1. Introduction

Our primary goal for this paper is to explore and differentiate the principles of calibration and validation for computational science and engineering (CS&E), as well as to present some related technical issues that are important and of current interest to us. Our conclusion is that calibration and validation are essentially different. To explain what we mean by calibration and validation, we restrict our attention to CS&E software systems, called *codes* here. We then define the product (output) of the execution of a code for a given choice of input to be the resulting *calculation*. Now, one definition of *calibration* is to adjust a set of code input parameters associated with one or more calculations so that the resulting agreement of the code calculations with a chosen and fixed set of experimental data is maximized (this requires a quantitative

specification of the agreement). Compare this with the following simple definition of *validation*: that is, to quantify our confidence in the predictive capability of a code for a given application through comparison of calculations with a set of experimental data.

The foundation of our discussion below elaborates the meaning of these definitions of validation and calibration, primarily through the introduction of some mathematical formalism. Our formalism allows us to reasonably precisely argue that CS&E validation and calibration require rigorous comparison with *benchmarks*, which we precisely define in Section 2. Our discussion leads us to consider other concepts as well, including *uncertainty*, *prediction*, and *verification*, and their relationship to validation and calibration. Verification is a particularly important concept in CS&E and inevitably influences calibration and validation. We will explain why this is the case, and claim as well that validation and calibration in CS&E both *depend* on results of verification. We also claim that calibration is logically dependent on the results of validation, which is one way of emphasizing that calibration cannot be viewed as an adequate substitute for validation in many CS&E applications.

---

*Corresponding author.

E-mail addresses: tgtruca@sandia.gov (T.G. Trucano), lpswile@sandia.gov (L.P. Swiler), tigusa@jhu.edu (T. Igusa), wloberk@sandia.gov (W.L. Oberkampf), mpilch@sandia.gov (M. Pilch).

Uncertainty quantification, and therefore sensitivity analysis, is a critical challenge in both validation and calibration. A lot has already been written on this topic in the computational literature and so we mainly discuss three highly speculative issues that are atypical of previously published themes. First, we discuss a formalization of the concept of code credibility that results from the use of benchmarks in verification and validation (V&V). Credibility is intended to be an important consequence of V&V; and calibration for that matter. We raise, but do not answer, the question of how credibility might be quantified. However such quantification may be achieved, it will have uncertainty associated with it. Second, we discuss a specific area of overlap between validation and calibration that is centered on how to deal with uncertainty in the physical models implemented in a CS&E code. This is the topic of *calibration under uncertainty* (CUU). Our primary conclusion is that recent calibration research that mathematically confronts the presence of this model-form uncertainty in statistical calibration procedures is important and coupled to validation issues. We speculate on the nature of this coupling, in particular that validation provides important information to calibration accounting for model-form uncertainty. We further argue that a full exploration of this issue might lead to the investigation of abstract learning theory as a quantitative tool in validation and calibration research. Finally, we speculate that uncertainty quantification has a role in verification. The use of uncertainty quantification in verification is probably not controversial, for example in statistical testing procedures, but our belief that the results of verification studies have uncertainty that requires quantification is. We explain this issue, but do not attempt to resolve it in this paper.

It is perhaps unclear why we are presenting an entire paper that mainly speaks to the issue of separation of calibration and validation. After all, is not the overarching goal of computational science to improve the associated calculations for given applications? Therefore, is not it natural to perform calibration to achieve this purpose? We believe that it is dangerous to replace validation with calibration, and that validation provides information that is necessary to understand the ultimate limitations of calibration. This is especially true in certain cases for which high-consequence CS&E prediction is required. These cases represent significant challenges for the use of CS&E codes and inevitably increase the importance of precisely distinguishing between validation and calibration in support of these uses. Our approach in this paper to calibration and validation emphasizes a kind of logical ideal. We do not emphasize practical issues, but the interested reader can find practicalities discussed in many of our references. We do emphasize that "real validation" and "real calibration" can be argued to be somewhat removed from the formalism and logical separation we stress in this paper. Murky separation of validation and calibration in real CS&E problems highlights the need to have some kind of logical foundation for clearly under-

standing the interplay of these concepts, especially for high-consequence applications of CS&E.

Section 2 presents a discussion of definitions of the various concepts mentioned above. Section 2.2 provides an illustration of the key ideas of verification, validation, and calibration using a computational fluid dynamics example (virtually the only CS&E example in the paper). Formalism of the concepts is introduced in Section 2.3, including of the concept of a benchmark and its comparison with calculations through comparison functions, and a notional formalism for credibility. In Section 3, we review common ideas of calibration (Sections 3.1 and 3.2), introduce some current research themes that generalize these ideas when considering uncertainty in the models that must be calibrated (Section 3.3), and introduce the possibility that computational learning theory might have some interest to our problems (Section 3.4). Section 4 briefly touches upon the role of sensitivity analysis in our discussion. We primarily provide some references, discuss the early appearance of sensitivity analysis in validation, and briefly comment on the presence of sensitivity analysis in credibility measures. Section 5 concludes the paper. We have tried to provide a useful set of references.

We emphasize that this paper presents some research ideas that are in early stages and somewhat speculative, but that we feel offer promising potential paths forward in calibration and validation. We introduce enough formalism to add some precision to our presentation, but this formalism does not reduce the amount of speculation in our discussion. Nor is the formalism enlisted to, in some sense, solve a particular problem in this paper. We hope that future papers will perform this role.

2. Calibration and validation

2.1. Guiding definitions

In this paper our underlying emphasis is on CS&E supported by the US Department of Energy's Advanced Simulation and Computing program (ASC, formerly called ASCI). A description of this program is given in [1]. The CS&E software developed under the ASC program is centered on the large-scale parallel numerical solution of systems of nonlinear partial differential equations (PDEs) of great complexity. The software implementations that accomplish this are called computer codes, or simply *codes* (see below).

Within this context, the current formal definitions of V&V used by the ASC program are as follows:

- *Verification* (*ASC*) is the process of confirming that a computer code correctly implements the algorithms that were intended.
- *Validation* (*ASC*) is the process of confirming that the predictions of a code adequately represent measured physical phenomena.