Chapter 11
From Data to Decision: Evaluating the Dangers and Benefits of Reliance on a Flexible Discrepancy Term in Model Calibration

Alexander Knecht and Sez Russcher

Keywords Model Calibration, Bayesian Calibration, Uncertainty Quantification, Model Discrepancy, Verification and Validation

Conventional approaches for calibrating constitutive models of engineering systems have tended to rely on deterministic optimization methods to find a single set of parameters that produce the best fit to data. However, such deterministic approaches are inherently flawed due to their inability to incorporate uncertainties inherent in model development, ranging from input parameters to model form errors to measurement noise [1]. A Bayesian approach for model calibration has gained popularity as it offers a framework that treats model parameters as random variables with appropriately defined prior probability distributions. Rather than chasing single value parameter estimates, the goal becomes to construct a complete posterior probability distribution for each of the parameters. This distribution represents an updated state of knowledge, formally synthesizing prior beliefs about the parameters with the evidence contained in the experimental data by using the log-likelihood function [2]. The resulting posterior distribution offers a more comprehensive characterization of parameters and their associated uncertainties, a significantly more useful result than deterministic methods can provide. [3]

A crucial consideration in modern model calibration, the explicit handling of model discrepancy, has been popularized by Kennedy and O'Hagan (2001), originally referred to as model inadequacy [4]. Their framework suggests that forcing a potentially flawed model to perfectly calibrate to a limited experimental data set may lead to highly biased parameter estimates, stripping them of any predictive significance, a phenomenon known as parameter confounding [5],[8]. To mitigate this, they proposed augmenting the statistical model with the addition of the discrepancy term, comprehensively representing all systemic inadequacies introduced through experimental error, computational model limitations, or prior biases. In many cases, this framework is implemented via Gaussian Processes (GP) models, allowing the surrogate to learn model error, and calibrate much like a parameter [5]. The introduction of a discrepancy term, however, may present additional challenges, primarily risking overfitting and non-identifiability [6],[8].

We have evaluated this on a simple polynomial and trigonometric function. A function is produced for the purpose of this evaluation, with 10 experimental points chosen along the curve, $f(x) = 2x - 2(x)^2 + .8sin(2 * pi * x) + .8cos(2 * pi * x)$. The simulation is then produced, with the removal of the *cosine* term to simulate model form error. Table 1 presents preliminary results used in the simulation calibration, with a baseline test (no model error, no discrepancy) depicted on the left, and increasing complexity moving from left to right (with model error but no discrepancy in the center, with model error and with a discrepancy term to the right). Evident from this table is the dramatically better curve fit produced by the model with discrepancy, despite similarity in the parameter predictions and higher standard deviations (or lower precision) among the data calibrated with discrepancy.

In Table 1, the risk of overfitting is evident; when calibrating to a limited set of experimental data, the introduction of a highly flexible discrepancy term may naturally dominate the effects of true parameters, overly forcing the curve to fit the presented data. While this produces a very accurate representation of the currently available observed points, it handicaps future predictive power by over-constraining the model to a limited set of data. The second issue is that of non-identifiability; while a flexible discrepancy term allows the model to absorb the influence of the physical parameters, it can become difficult to distinguish between the effects of the discrepancy term and the physical parameters, particularly when data is sparse [7]. This issue, as explored by Arendt, Apley, and Chen (2012), describes the effect where a flexible discrepancy term may replicate or even counteract the impacts of the physical model parameters, thereby inhibiting parameter prediction, as seen

Alexander Knecht · Sez Russcher(✉)
Clemson University
e-mail: knecht@clemson.edu; sez@clemson.edu

Kristin B. Zimmerman, PhD (eds.), *Model Validation and Uncertainty Quantification, 2026, Vol. 3*, Conference Proceedings of the Society for Experimental Mechanics Series, https://doi.org/10.1007/97887-438-1442-9_11

Table 1 Post-Calibration Results of 3 distinct calibration scenarios, the true curve matched to the simulation curve (left), a curve with error but no discrepancy (center), a curve with error and with a discrepancy term (right)

	Perfect Match, no Discrepancy		With Model Error, without Discrepancy		With Model Error, with Discrepancy	
True Parameter Value:	Mean Prediction:	Standard Deviation:	Mean Prediction:	Standard Deviation:	Mean Prediction:	Standard Deviation:
2	1.9998	0.0399	1.996	0.2008	1.990	0.9728
-2	-2.0003	0.0200	-1.998	0.1968	-1.828	0.6624
.8	.80032	0.0183	0.798	0.0794	0.799	0.2678
RMSE:	0.0002		1.2866		0.2240	
|E|:	0.0002		1.8308		0.3275	

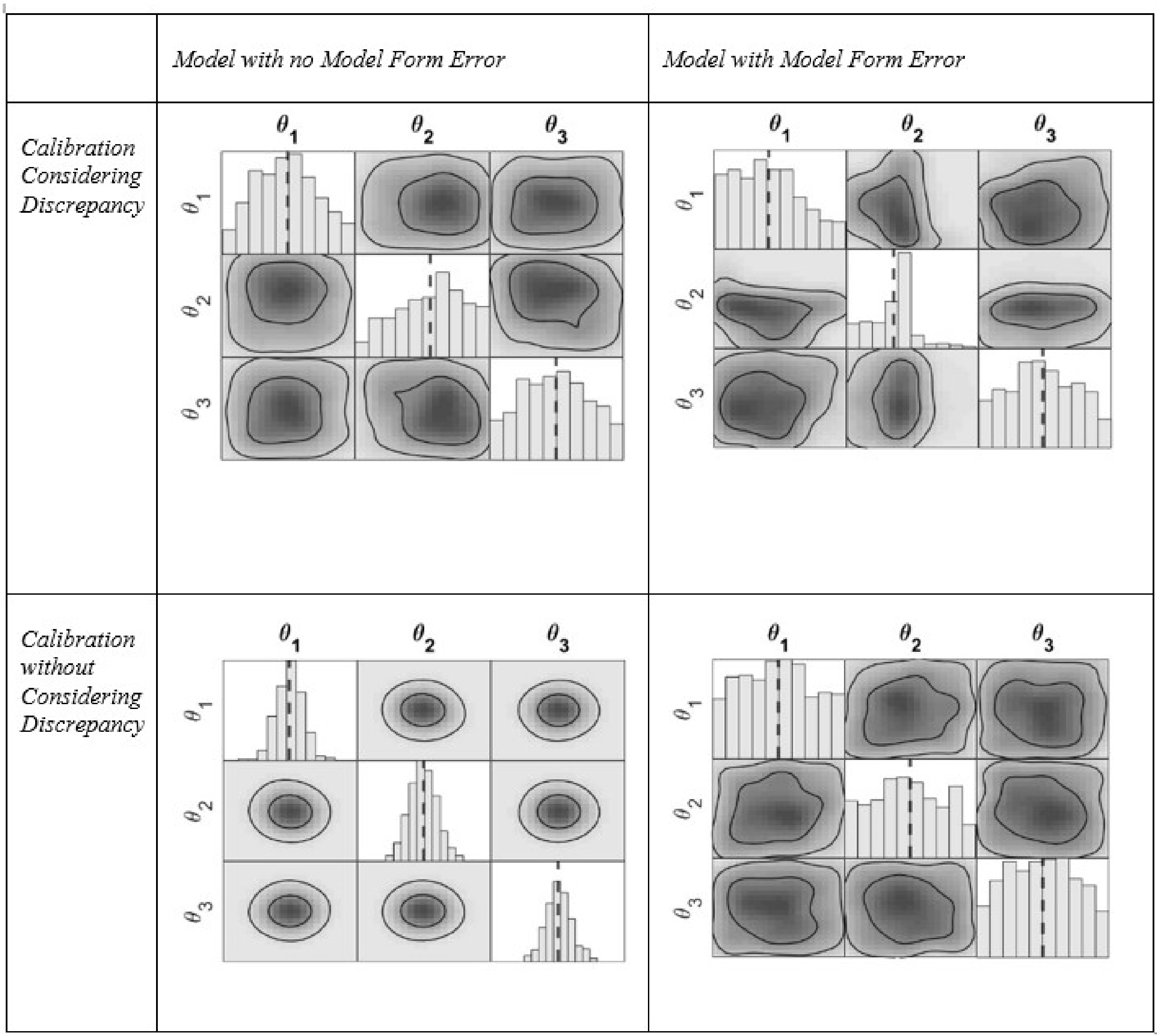

Fig. 1 Posterior prediction histograms produced in the calibration process, with and without model form error, with and without discrepancy terms

in Table 1. This can result in a posterior parameter distribution that is excessively wide or skewed, rendering it physically meaningless in calibration processes [7]. For this reason, we will present model calibration results executed both with and without the discrepancy consideration, to observe how the inclusion of a flexible, GP based discrepancy term impacts model calibration.

Using the same simulation function explored in Table 1, Figure 1 outlines the results of various calibration processes both with and without discrepancy terms and with and without model form error, to observe the effects of flexible discrepancy consideration on the distribution of posterior predictions.

Quadrants 1 and 3 represent the logical selections based on inference alone. For example, in a model with known limitations, the modeler will tend to prefer a method of accounting for those in the calibration process, therefore selecting a calibration with discrepancy when known model form error exists. Alternatively, if the model is known to be an exceptional representation of reality, the best choice will likely be a calibration without the flexible discrepancy term, choosing to accept a highly biased model for the sake of a tight posterior distribution.

References

1. Beck, J. L., & Au, S. K. (2002). Bayesian updating of structural models and their uncertainties. I: Theory. Journal of engineering mechanics, 128(4), 380-391.
2. ASME. (2006). Standard for Verification and Validation in Computational Solid Mechanics (ASME V&V 10-2006). American Society of Mechanical Engineers.
3. Sankararaman, S., & Mahadevan, S. (2015). Bayesian methodology for uncertainty quantification in material property estimation. International Journal of Materials Research, 106(1), 3-13.
4. Kennedy, M. C., & O'Hagan, A. (2001). Bayesian calibration of computer models. Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63(3), 425–464.
5. Brynjarsdóttir, J., & O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. Inverse Problems, 30(11), 114007.
6. McFarland, J., & Mahadevan, S. (2008). Error and variability in robust design. AIAA Journal, 46(10), 2586-2597.
7. Arendt, P. D., Apley, D. W., & Chen, W. (2012). Quantification of model uncertainty: Calibration, model discrepancy, and identifiability. Journal of Mechanical Design, 134(10).
8. Maupin, K. A., & Swiler, L. P. (2020). Model discrepancy calibration across experimental settings. Reliability Engineering & System Safety, 200, Article 106818