Bayesian Calibration and Uncertainty Quantification for TRACE Based on PSBT Benchmark

Chen Wang,[a] Xu Wu,[b] Katarzyna Borowiec,[a] Tomasz Kozlowski[a]
*Department of Nuclear, Plasma and Radiological Engineering, University of Illinois at Urbana-Champaign, Urbana, Illinois, USA, chenw3@illinois.edu, kb6@illinois.edu, txk@illinois.edu
†Department of Nuclear Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, USA, xuwu@mit.edu

INTRODUCTION

In the framework of BEPU (Best Estimate plus Uncertainty) methodology, the uncertainties involved in the simulations must be quantified to prove that the investigated design is acceptable. The output uncertainties are usually calculated by propagating input uncertainties through the simulation model, which requires knowledge of the model input uncertainties. However, in some best-estimate Thermal-Hydraulics (TH) codes such as TRACE, the physical model parameters used in empirical correlations may have large uncertainties, which are unknown to the code users. Therefore, obtaining uncertainty distributions of those parameters becomes crucial if we want to study the predictive uncertainty or output sensitivity.

Bayesian Calibration [1], also called inverse uncertainty quantification [2], aims at inversely quantifying the uncertainties associated with input parameters that are consistent with experimental data, and thus replacing the ad-hoc "expert judgment" or "user self-assessment" [2, 3]. This work adopts a Modular Bayesian calibration approach, which was initially proposed by Kennedy and O'Hagan [1], and then developed by Xu Wu et al. [4] for its applications in nuclear TH codes.

The Modular Bayesian approach [4] considers the presence model discrepancy during Bayesian calibration. Several TRACE physical model parameters are selected as calibration parameters in this work. Model discrepancy, also referred to as model inadequacy or model bias, accounts for the inaccuracy in computer simulation caused by underlying missing/insufficient physics, numerical approximation errors, and other errors of a computer code, even if all its parameters are fixed at their "true" values [1, 4]. Model discrepancy always exists in computer models because they are reduced representations of the reality. The consideration of model discrepancy is important because it can help avoid the "overfitting" problem in Bayesian calibration [5].

This paper uses a set of steady-state experimental data from PSBT benchmark [6], and it mainly aims at: (1) quantifying the uncertainties of TRACE physical model parameters based on experiment data; (2) quantifying the uncertainties in TRACE outputs based on inversely quantified physical model parameters uncertainties.

The rest of the paper includes a brief introduction to the PSBT benchmark and TRACE physical model parameters, an introduction to the modular Bayesian calibration approach, and results regarding the posterior distribution and predictive uncertainty.

OVERVIEW OF PSBT BENCHMARK AND TRACE PHYSICAL MODEL PARAMETERS

The void fraction measurement facility in PSBT benchmark [6] includes an electrically heated rod bundle as a partial section and full length of a PWR fuel assembly. The average void fractions data are measured at three different locations (Lower, Middle, Upper) along the heated length.

TRACE model was built according to the test assembly geometry. 74 cases in PSBT bundle test assembly 5 are selected in this study. A comparison between code simulations and experiment observations are shown in Figure 1. Each test consists of a different combination of boundary conditions, shown in Figure 2, where the x-axis is the case number. The test cases are separated into two groups: validation set and calibration set. The reason for the separation will be explained in the following part.

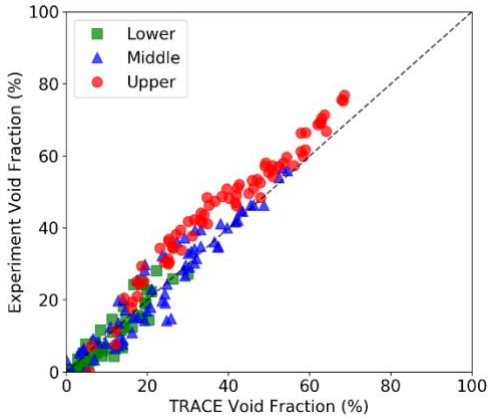

Fig. 1. Comparison of void fractions between experiment and TRACE for PSBT benchmark test assembly 5.

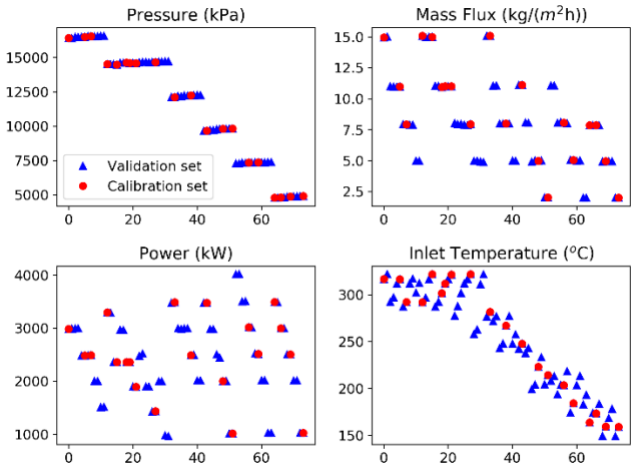

Fig. 2. Boundary conditions of test assembly 5

Uncertain inputs are selected by sensitivity analysis in previous study [2] for all the 36 physical model parameters available in TRACE. Finally, four of them which affect the outputs are used as the uncertain calibration parameters and are listed in Table 1. All these parameters are multiplicative factors with nominal value of 1.0.

Table 1. Description of Uncertain Inputs.

Parameter	Description
P1008	Single phase liquid to wall heat transfer coefficient
P1012	Subcooled boiling heat transfer coefficient
P1022	Wall drag coefficient
P1028	Interfacial drag (bubbly/slug Rod Bundle - Bestion) coefficient

THE MODULAR BAYESIAN CALIBRATION APPROACH

The key component of the modular Bayesian calibration approach is the model updating equation:

$$\mathrm{y}^{\mathrm{e}}(\mathbf{x}) = \mathrm{y}^{\mathrm{m}}(\mathbf{x}, \boldsymbol{\theta}^{*}) + \delta(\mathbf{x}) + \epsilon \qquad (1)$$

where $\mathbf{x}$ is called design variable, and $\boldsymbol{\theta}$ is the calibration variable; $\mathrm{y}^{\mathrm{e}}(\mathbf{x})$ is the experimental response, $\mathrm{y}^{\mathrm{m}}(\mathbf{x}, \boldsymbol{\theta}^{*})$ is the computer model response given $\mathbf{x}$ and the unknown true value of $\boldsymbol{\theta}$, $\delta(\mathbf{x})$ is the additive discrepancy function, and ϵ accounts for the experimental uncertainty, which is assumed to follow normal distribution $\epsilon \sim N(0, \sigma^2)$. In the following subsections, both $\mathrm{y}^{\mathrm{m}}(\mathbf{x}, \boldsymbol{\theta}^{*})$ and $\delta(\mathbf{x})$ can be formulated by a Gaussian Process (GP) model, and then the posterior distributions of calibration parameters can be explored via a proper numerical method. Only a brief introduction is included in this paper, reader may refer to [1, 4, 5, 7] for details about the modular Bayesian calibration approach.

GP Models for Computer Code and for Model Discrepancy

The GP regression model can provide an approximation of the input/output relationship of the computer model with desirable accuracy and much less computation time. It is widely used as an interpolation method, where the interpolated values are modeled by a Gaussian random process. This property of GP makes it well suited for Bayesian calibration problems.

The GP model for computer code is built based on the simulation data $\mathbf{y}^{m} = [y^{m}(\boldsymbol{x}_1, \boldsymbol{\theta}_1), \ldots, y^{m}(\boldsymbol{x}_{\mathrm{N}}, \boldsymbol{\theta}_{\mathbf{N}})]$ collected at N input sites. The design variable $\boldsymbol{x}$ represents the boundary conditions as shown in Figure 2, and it only takes the value in the calibration set $\mathbf{x}_{\mathbf{cali}}$. The reason for this treatment is that the datasets of design variable to train the GP model for computer code and for model discrepancy must be different, otherwise the computer code would play no role in Eq. 1 because GP model interpolates exactly. We selected 20 cases out of 74 as the calibration set to train the GP model for computer code. The rule for the selection is that the calibration domain should be "enveloped" by the validation domain, otherwise extrapolation might occur and make the GP model of model discrepancy inaccurate. For each case, 40 samples with respect to $\boldsymbol{\theta}$ in the range of $[0,5]^4$ by Latin Hypercube Sampling (LHS) method are drawn. The number is selected by a convergence study to make sure the GP model is accurate.

Model discrepancy is an indispensable part of Bayesian calibration when the computer code cannot predict the real process accurately. The discrepancy may be caused by underlying missing physics, numerical approximations, and other inaccuracies, which are commonly seen in best-estimate TH codes. Ignoring the presence of model discrepancy in Bayesian calibration will lead to overfitting of the calibrating parameters, and make calibrating parameters compensate for the model discrepancy.

The GP model for model discrepancy term $\delta(\mathrm{x})$ is built based on $[y^{e}(\mathbf{x_1}) - y^{m}(\mathbf{x_1}), \ldots, \mathrm{y}^{\mathrm{e}}(\mathbf{x_M}) - \mathrm{y}^{\mathrm{m}}(\mathbf{x_M})]$, where $\mathbf{x}$ is selected from the validation set $\mathbf{x}_{\mathbf{vali}}$. The GP models in this work mainly utilized the scikit-learn package in Python [8].

Posterior Distributions of Calibration Parameters

The calculation of posterior distributions is based on GP models for computer code (GP_{cc}), GP models for model discrepancy (GP_{md}), and experimental data $\boldsymbol{d}$. The posterior distribution is:

$$p(\boldsymbol{\theta}|\boldsymbol{d}, GP_{cc}, GP_{md}) \propto p(\boldsymbol{d}|\boldsymbol{\theta}, GP_{cc}, GP_{md}) p(\boldsymbol{\theta}) \qquad (2)$$

The likelihood term $p(\boldsymbol{d}|\boldsymbol{\theta}, GP_{cc}, GP_{md})$ can be expressed by:

$$\frac{\exp\left[-\frac{1}{2}[\boldsymbol{y}^{\mathrm{E}}(\mathbf{x})-\boldsymbol{y}^{\mathrm{M}}-\delta(\mathbf{x})]^{\mathrm{T}}(\boldsymbol{\Sigma}_{\mathrm{t}})^{-1}[\boldsymbol{y}^{\mathrm{E}}(\mathbf{x})-\boldsymbol{y}^{\mathrm{M}}-\delta(\mathbf{x})]\right]}{\sqrt{|\boldsymbol{\Sigma}_{\mathrm{t}}|}} \quad (3)$$

where $\boldsymbol{\Sigma}_{\mathrm{t}} = \boldsymbol{\Sigma}_{\mathrm{exp}} + \boldsymbol{\Sigma}_{\mathrm{md}} + \boldsymbol{\Sigma}_{\mathrm{cc}}$ is the summation of 3 covariance matrices: (1) Experimental error (σ = 4% void fraction); (2) GP model for computer code predicted at ($\mathbf{x}_{\mathbf{cali}}, \boldsymbol{\theta}$); and (3) GP model for model discrepancy predicted at ($\mathbf{x}_{\mathbf{cali}}$).

The posterior distribution in Eq. 2 can be calculated numerically using MCMC method. In this work, an adaptive random-walk Metropolis-Hasting algorithm is used. The posterior distribution of $\boldsymbol{\theta}$ gives us information about the statistics of those physical model parameters. It can be used to quantify the uncertainty in the computer code output $\boldsymbol{y}^{\mathrm{M}}(\boldsymbol{x}, \boldsymbol{\theta})$ due the parametric uncertainty (after calibration) in $\boldsymbol{\theta}$.

RESULTS

Following the approach outlined in the previous section, two GP models are constructed and validated. The posterior distributions for the four calibration parameters are explored by MCMC method and results are shown in Figure 3. The diagonal plots in Figure 3 are marginal distributions for each parameter fitted by kernel density estimation, and the off-diagonal plots show the pair-wise joint density distributions. From the pair-wise joint distributions we can see a non-zero correlation coefficient between parameter 'P1008' and 'P1012'. The negative correlation coefficient between these two parameters is consistent with physical phenomenon, as both will lead to higher void fraction when they get larger.

As mentioned before, ignoring the presence of model discrepancy in Bayesian calibration will typically lead to overfitting of the calibration parameters, and may make calibration parameters compensate for the model discrepancy. This phenomenon is shown in Figure 4. The left figures are posteriors without considering model discrepancy using data from calibration set, and overfitting can be observed for parameter 'P1008' and 'P1012'. One of them tends to the upper bound and the other tends to be small and narrow ranged. The figures in the right column which include model discrepancy can be seen as corrected by our available knowledge for model discrepancy.

The posterior distributions of the model responses $\mathrm{y}^{\mathrm{m}}(\mathrm{x}, \boldsymbol{\theta})$ given $p(\boldsymbol{\theta}|\boldsymbol{d})$ can be calculated by integrating $\mathrm{y}^{\mathrm{m}}(\mathrm{x}, \boldsymbol{\theta})$ with respect to $p(\boldsymbol{\theta}|\boldsymbol{d})$ on the validation set $\mathbf{x}_{\mathbf{vali}}$. The standard deviations of the posterior distributions of TRACE void fraction outputs are shown as error bars in Figure 5, and the error between the experiments and posterior means are represented by red dots. The error bars reflect the uncertainty in the outputs due to the parametric uncertainty (after calibration) in $\boldsymbol{\theta}$.

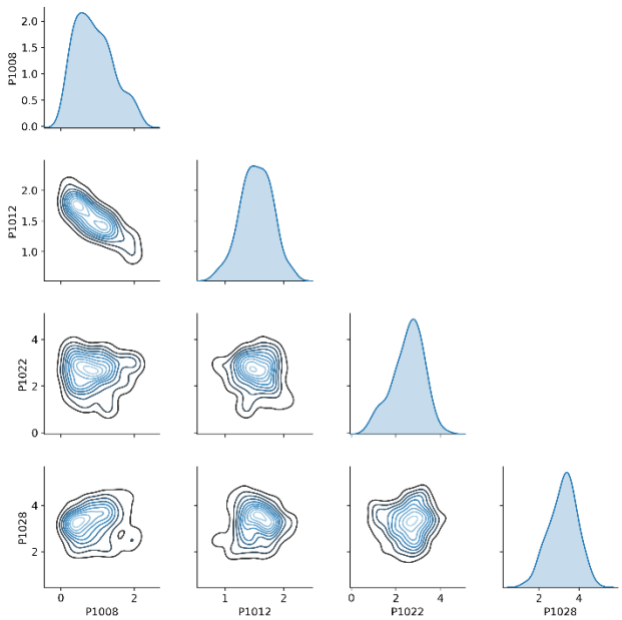

Fig. 3. Posterior pair-wise joint and marginal distributions

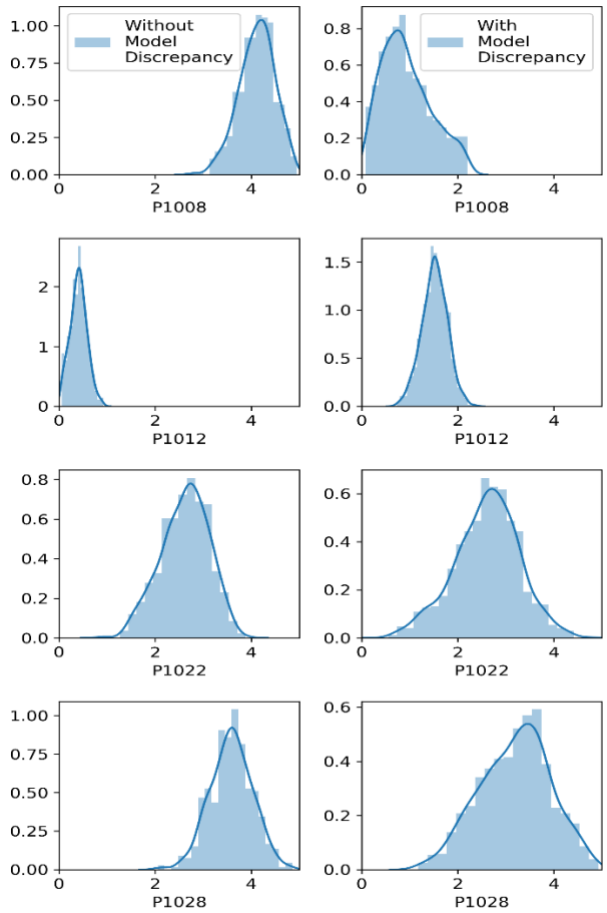

Fig. 4. Comparison of posteriors when model discrepancy is considered and not considered.

Comparing the posterior mean with the TRACE outputs using prior nominal values of $\boldsymbol{\theta}$ (denoted by blue squares in Figure 5), we can see an improvement in prediction accuracy. The posterior means are closer to experiment than original results for most of the cases, especially in the 'Upper' measuring location. The Root Mean Square Error (RMSE) for the posterior mean is 4.7 (void fraction %), and it is 5.2 (void fraction %) for the outputs using prior nominal value. The improvement is not significant here, but better results may be obtained if more experimental data are used in future.

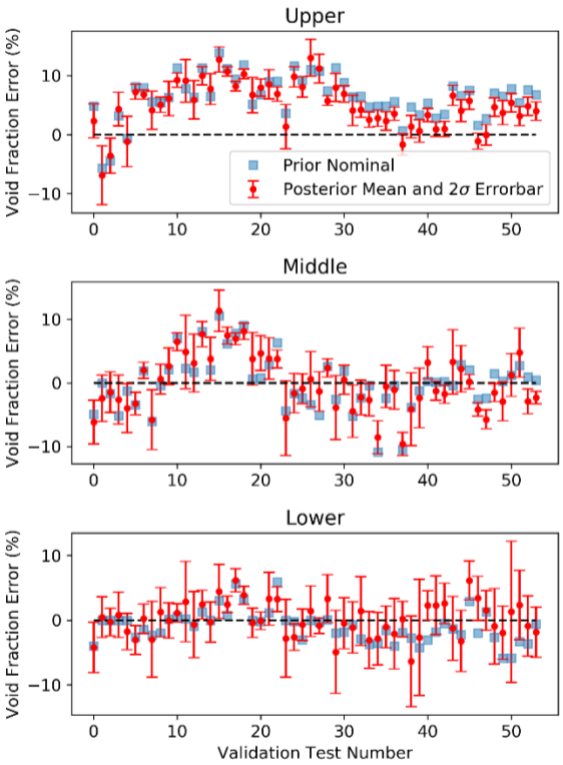

Fig. 5. Comparison of TRACE posterior mean and TRAE output with prior nominal value of $\boldsymbol{\theta}$ on the validation set.

CONCLUSIONS

This work applies the modular Bayesian calibration approach to the TRACE physical model parameters based on steady-state PSBT benchmark. The major uncertainty considered in this work is parametric uncertainty from physical model parameters that may not be measured accurately. The posterior uncertainty distributions for those parameters are calculated via the modular Bayesian calibration approach, and are then propagated through the computer code to obtain the uncertainty distribution of model responses. The comparison of the posteriors of physical model parameters obtained by different approaches shows the necessity of considering model discrepancy in Bayesian calibration. Future work will include sensitivity analysis on several model assumptions in the modular Bayesian approach and its application to high dimensional output.

REFERENCES

1. M. C. KENNEDY and A. O'HAGAN, "Bayesian calibration of computer models," Journal of the Royal Statistical Society: Series B (Statistical Methodology), 63, 3, 425– 464 (2001).
2. X. WU, T. MUI, G. HU, H. MEIDANI, and T. KOZLOWSKI, "Inverse uncertainty quantification of TRACE physical model parameters using sparse gird stochastic collocation surrogate model," Nuclear Engineering and Design, 319, 185–200 (2017).
3. C. WANG, X. WU and T. KOZLOWSKI, "Surrogate-based Inverse Uncertainty Quantification of TRACE Physical Model Parameters using Steady-State PSBT Void Fraction Data", in "Proceedings of the 17th International Topical Meeting on Nuclear Reactor Thermal-Hydraulics (NURETH17)", Xi'an, China, Sept. 3 – Sept. 9, 2017 (2017).
4. X. WU, T. KOZLOWSKI, H. MEIDANI, and K. SHIRVAN, "Inverse Uncertainty Quantification using the Modular Bayesian Approach based on Gaussian Process, Part 1: Theory," (arXiv preprint arXiv:1801.01782) (2018).
5. J. BRYNJARSDÓTTIR and A. O'HAGAN, "Learning about physical parameters: The importance of model discrepancy," Inverse Problems, 30, 11, 114007 (2014).
6. A. RUBIN, et al. "OECD/NRC Benchmark based on NUPEC PWR subchannel and bundle tests (PSBT), Volume I: Experimental Database and Final Problem Specifications." *OECD Nuclear Energy Agency* (2010).
7. P. D. ARENDT, D. W. APLEY, and W. CHEN, "Quantification of model uncertainty: Calibration, model discrepancy, and identifiability," Journal of Mechanical Design, 134, 10, 100908 (2012).
8. Scikit-learn: Machine Learning in Python, Pedregosa et al., JMLR 12, pp. 2825-2830 (2011).