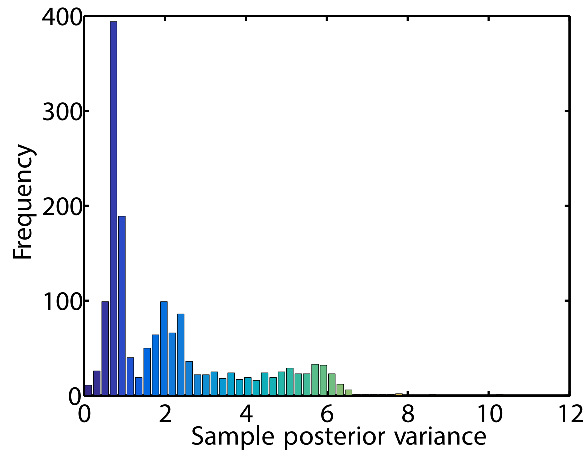

FIG. 7: Histogram of the 1600 sample posterior variances for $\{y_4, y_5\}$.

without generating $\mathbf{\Delta}^{(i)}$ and $\mathbf{E}^{(i)}$, we directly take $\mathbf{Y}^{e(i)} = \mathbf{Y}^{m(i)}$ and apply the multi-response modular Bayesian approach to calculate the negative second-order derivative of log-likelihood function [as the term inside the brackets in Eq. (12)]. The results for all 16 values of θ are plotted in Fig. 8(b). The Fisher-information-based scalar predictor $\hat{\mathcal{I}}$, taken to be the average of the 16 values, is 351.8. Because θ is scalar with a uniform prior for this example, taking the average over the 16 evenly spaced values of θ is more computationally efficient than generating random draws of θ from its prior and using Eq. (12).

To better illustrate the relationship between the preposterior and the surrogate preposterior analyses, we conduct the preposterior analysis in the same "fixed-θ" manner described in the preceding paragraph for the surrogate preposterior analysis. That is, we consider the 16 different values of θ equally spaced within $[150, 300]$ GPa, and for each specific θ value we use 100 MC replicates to calculate the sample posterior variance. A procedure identical to that describes in Section 3.1, but with θ fixed over the 100 MC replicates, was used to estimate the preposterior variance (as the average of the 100 sample posterior variances over the 16 values of θ). The results are shown in Fig. 8(a). Comparing Figs. 8(a) and 8(b), we see a clear negative correlation between the preposterior variance and the Fisher information criterion. This is expected, considering that the preposterior variance is an estimate of the actual posterior variance, which is closely related to the inverse of the Fisher information; a larger value of the former generally corresponds to a larger value of the latter.

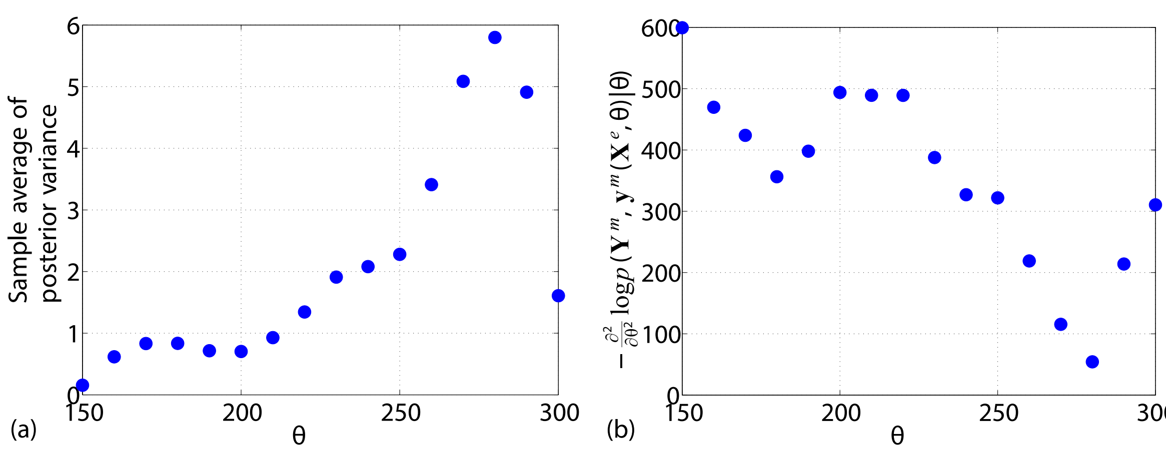

FIG. 8: 16 equally spaced values of θ, and the corresponding (a) sample average of posterior variance (unit: GPa^2) over 100 MC simulations per θ, and (b) negative second-order derivative of the log-likelihood function.

The same procedure was repeated for every other pair of responses. Table 2 shows the results of the preposterior and surrogate preposterior analyses. The ranks in columns 2 and 3 are based on the predicted identifiability; 1 corresponds to the smallest preposterior variance / largest $\hat{\mathcal{I}}$, and 15 to the largest preposterior variance / smallest $\hat{\mathcal{I}}$. The rankings provide us with predictions of which subsets are most likely to enhance identifiability (lower ranks) and which less likely (higher ranks). The "Posterior Analysis" columns in Table 2, which are taken from [11], show the actual posterior covariances that resulted from each combination of measured responses after the experiments were conducted. They serve as a basis for comparison, and ideally we would like the rankings from the posterior analyses to coincide with the predicted rankings from the preposterior and surrogate preposterior analyses. It can be seen that the preposterior analysis is in good relative agreement with the posterior. Although the values of preposterior and posterior standard deviations are off by roughly a factor of 2.5, both analyses indicate that $\{y_4, y_5\}$ together lead to the best identifiability, while $\{y_1, y_2\}$ together lead to the worst identifiability. Overall, the rankings are in very close agreement. The improvement on identifiability relative to the worst case (i.e., $\{y_1, y_2\}$ for both analyses) is also calculated and provided in the table. The relative improvements are also in very close agreement. The results provided by the surrogate preposterior analysis are also in close agreement with the posterior standard deviation results, although slightly less so than the preposterior standard deviations. The top seven pairs of responses coincide with the top seven pairs from the actual posterior analysis. Hence, the surrogate preposterior analysis would have effectively narrowed down the candidate pairs to consider in the preposterior analysis. Considering its extremely low computational cost, the surrogate preposterior analysis is a useful enhancement to the preposterior analysis for reducing the number of response pairs to consider.

The results from three analyses in Table 2 are in good accordance with the underlying physics of the system. For example, the strain y_1 and the plastic strain y_2 are perfectly correlated with each other; their values are off by a constant (equal to the value of elastic strain). Therefore, their combination adds no more information about θ and enhances identifiability little beyond using either single response. In contrast, the internal energy y_4 and the midpoint displacement y_5 follow a nonlinear relationship, and thus the degree of improvement in identifiability is substantial.

It is not surprising to observe the absolute differences between the posterior variance and preposterior variances. In the MC simulations of the preposterior analysis, the hypothetical experimental data are generated based on discrepancy functions generated from their assigned prior distribution. In contrast, the posterior variance calculation is based on the single realization that is the actual discrepancy function. Consequently, the Bayesian analysis modules inside the MC loops are hypothetical and are not expected to obtain same values of preposterior variance as the actual posterior variance. The surrogate preposterior analysis involves further approximations. However, it appears to accomplish

TABLE 2: Comparisons between posterior, preposterior, and surrogate preposterior analyses

Responses		**Posterior**			**Preposterior**			**Surrogate preposterior**	
y_i	y_j	σ_θ	Rank	Improvement	$\tilde{\sigma}_\theta$	Rank	Improvement	$\hat{\mathcal{I}}$	Rank
y_4	y_5	3.63	1	86.7%	1.4161	1	87.0%	351.8	1
y_4	y_6	3.90	2	85.8%	1.5224	3	86.0%	186.9	6
y_5	y_6	4.67	3	83.0%	1.6528	4	84.8%	218.2	3
y_3	y_6	4.86	4	82.2%	1.6618	5	84.7%	161.6	7
y_3	y_4	5.49	5	80.0%	1.4801	2	86.4%	273.0	2
y_3	y_5	6.27	6	77.1%	2.1117	6	80.6%	202.9	5
y_1	y_4	9.29	7	66.1%	3.0933	7	71.6%	209.6	4
y_1	y_3	10.13	8	63.0%	3.2583	8	70.1%	113.3	9
y_2	y_3	10.96	9	60.0%	3.6595	9	66.4%	93.34	10
y_2	y_5	15.42	10	43.4%	6.0765	11	44.2%	114.3	8
y_1	y_5	15.57	11	43.1%	5.5379	10	49.2%	90.92	11
y_2	y_4	18.35	12	33.9%	8.7816	12	19.4%	65.94	14
y_1	y_6	24.08	13	12.0%	9.1008	13	16.5%	81.19	12
y_2	y_6	26.46	14	3.3%	10.1850	14	6.5%	80.99	13
y_1	y_2	27.37	15	—	10.8931	15	—	47.09	15

its intended purpose, as the preposterior and surrogate preposterior analyses do a reasonable job of predicting the relative degree of identifiability and guide users in selecting the best set of responses to measure experimentally. Fig. 9 illustrates this by plotting the preposterior standard deviation and the Fisher information criterion $\hat{\mathcal{I}}$ versus the actual posterior standard deviation. The preposterior standard deviation is roughly in proportion to the actual posterior standard deviation, and $\hat{\mathcal{I}}$ is negatively correlated with the posterior standard deviation, which indicates that for this case study preposterior and surrogate preposterior analyses are sufficient in predicting identifiability.

We further illustrate the improvement on identifiability in Figs. 10 and 11. While neither analyzing y_4 nor analyzing y_5 alone can provide an informative posterior distribution of θ (Fig. 10), uncertainty quantification considering both y_4 and y_5 provides a much tighter posterior distribution of θ, and the mean of the posterior is close to the true value of θ [Fig. 11(a)]. In contrast, the subset of $\{y_1, y_2\}$ provides a dispersed posterior distribution of θ [Fig. 11(b)]. The level of uncertainty is well predicted by both the preposterior and surrogate preposterior analyses.

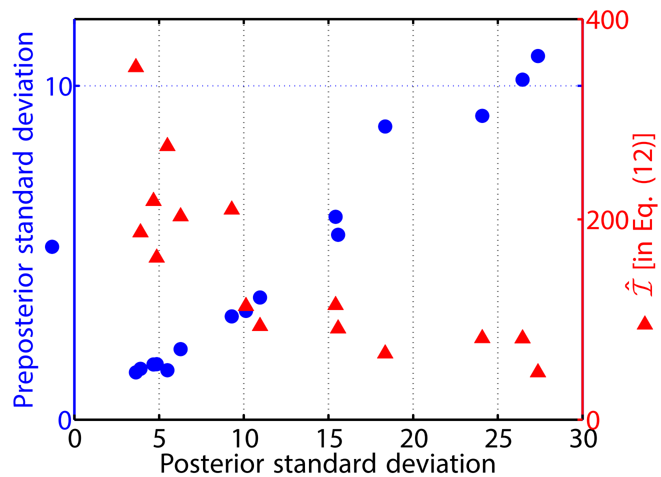

FIG. 9: Preposterior standard deviation and Fisher-information-based identifiability predictor, versus posterior standard deviation, demonstrating very high correlation.

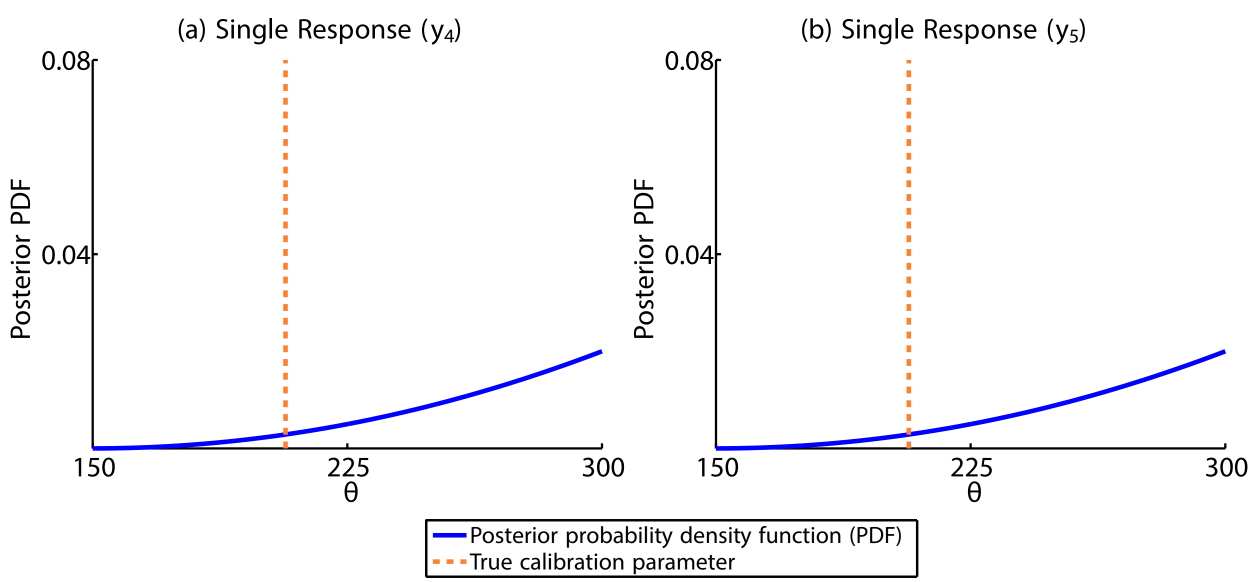

FIG. 10: Posterior distribution of the calibration parameter using a single response.

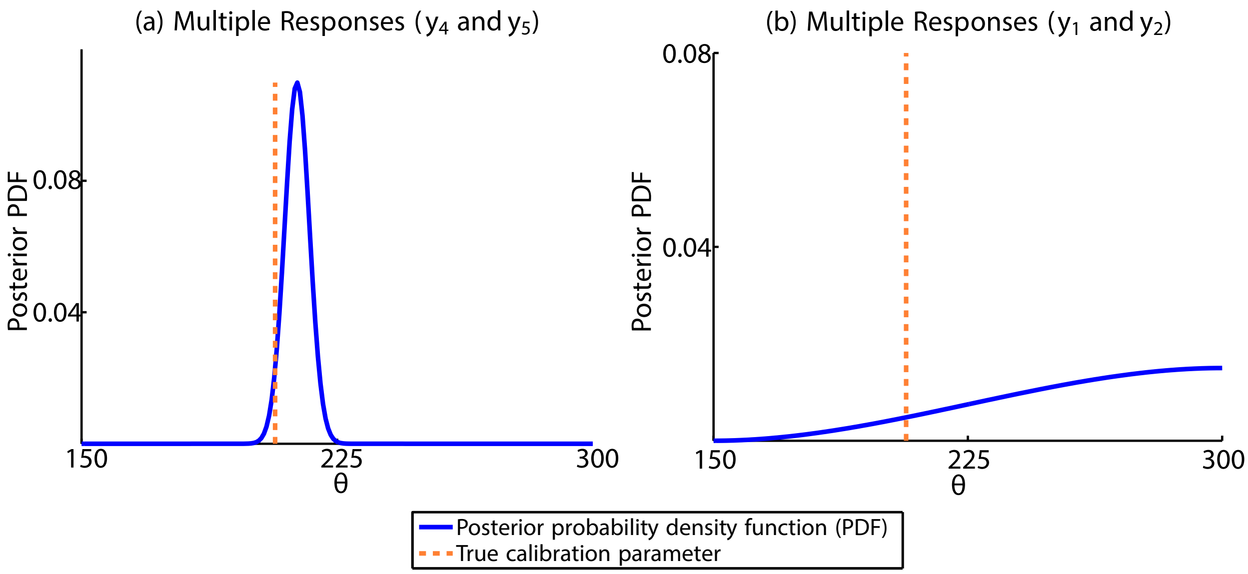

FIG. 11: Posterior distribution of the calibration parameter using multiple responses.

5. CONCLUSIONS

Identifiability is of major importance in model calibration and predictive modeling in all engineering disciplines. The degree of identifiability can be measured by the posterior covariance of the calibration parameters in a typical model uncertainty quantification framework. Earlier studies have demonstrated that identifiability can be enhanced by measuring multiple responses that share a mutual dependence on a common set of calibration parameters. However, to take advantage of this, a method is needed for predicting multi-response identifiability prior to conducting the physical experiments, to allow users to choose the most appropriate set of responses to measure experimentally. In this research, we propose a preposterior analysis that, prior to conducting the physical experiments but after conducting computer simulations, can predict the degree of identifiability that will result using different subsets of responses to measure experimentally. This is accomplished by calculating the preposterior covariance from a modular Bayesian Monte Carlo analysis of a MRGP model. To render the approach computationally feasible in engineering applications with a large number of responses, we also proposed a surrogate preposterior analysis based on the Fisher information of the calibration parameters, which is used to eliminate combinations of responses that are unlikely to provide substantial improvement in identifiability, thereby substantially reducing computational cost. The proposed methods were applied to a simply supported beam example to select which two out of six responses will best improve identifiability. Our study shows that the approach is effective in predicting which subset of responses will provide the largest improvement in identifiability. Even though there are absolute differences between the preposterior and actual posterior covariances, the relative differences and the rankings derived from them are quite consistent, indicating that the method can be used effectively to choose the best combination of responses to measure experimentally.

Future work in this direction includes examining the impact of using different priors for discrepancy function MRGP hyperparameters on the preposterior covariance. Also, the model input settings for physical experiments apparently affect identifiability; simultaneously optimizing the selected experimental responses and the design for the experimental input settings is another research direction.

ACKNOWLEDGMENTS

The grant support from the National Science Foundation (CMMI-1233403) and the Terminal Year Fellowship at Northwestern University are greatly acknowledged. Zhen Jiang would like to thank Dr. Paul D. Arendt for valuable discussions. This paper is revised based on DETC2013-12457 in the proceedings of the *ASME 2013 International*

Design Engineering Technical Conferences & Computers and Information in Engineering Conference. We are grateful for ASME to grant us the permission to publish the content of DETC2013-12457 in this paper with *International Journal for Uncertainty Quantification*.

REFERENCES

1. Kennedy, M. C. and O'Hagan, A., Bayesian calibration of computer models, *J. R. Stat. Soc. Ser. B (Statistical Methodology)*, 63(3):425–464, 2001.
2. Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D., Combining field data and computer simulations for calibration and prediction, *SIAM J. Sci. Comput.*, 26(2):448–466, 2004.
3. Reese, C. S., Wilson, A. G., Hamada, M., Martz, H. F., and Ryan, K. J., Integrated analysis of computer and physical experiments, *Technometrics*, 46(2):153–164, 2004.
4. Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J., A framework for validation of computer models, *Technometrics*, 49(2):138–154, 2007.
5. Higdon, D., Gattiker, J., Williams, B., and Rightley, M., Computer model calibration using high-dimensional output, *J. Am. Stat. Assoc.*, 103(482):570–583, 2008.
6. Chen, W., Xiong, Y., Tsui, K.-L., and Wang, S., A design-driven validation approach using bayesian prediction models, *J. Mech. Design*, 130(2):021101, 2008.
7. Qian, P. Z. and Wu, C. J., Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments, *Technometrics*, 50(2):192–204, 2008.
8. Wang, S., Chen, W., and Tsui, K.-L., Bayesian validation of computer models, *Technometrics*, 51(4):439–451, 2009.
9. Arendt, P. D., *Quantification and Mitigation of Multiple Sources of Uncertainty in Simulation Based Design*, PhD Thesis, Northwestern University, Evanston, IL, 2012.
10. Arendt, P. D., Apley, D. W., and Chen, W., Quantification of model uncertainty: Calibration, model discrepancy, and identifiability, *J. Mech. Design*, 134(10):100908, 2012.
11. Arendt, P. D., Apley, D. W., Chen, W., Lamb, D., and Gorsich, D., Improving identifiability in model calibration using multiple responses, *J. Mech. Design*, 134(10):100909, 2012.
12. Loeppky, J., Bingham, D., and Welch, W., Computer model calibration or tuning in practice, Tech. Rep., University of British Columbia, 2006.
13. Arendt, P. D., Apley, D. W., and Chen, W., A preposterior analysis to predict identifiability in experimental calibration of computer models, *IIE Transactions*, accepted, DOI: 10.1080/0740817X.2015.1064554, 2015.
14. Conti, S., Gosling, J. P., Oakley, J., and O'Hagan, A., Gaussian process emulation of dynamic computer codes, *Biometrika*, 96(3):663–676, 2009.
15. Conti, S. and O'Hagan, A., Bayesian emulation of complex multi-output and dynamic computer models, *J. Stat. Planning Inference*, 140(3):640–651, 2010.
16. Liu, F., Bayarri, M., Berger, J., Paulo, R., and Sacks, J., A Bayesian analysis of the thermal challenge problem, *Comput. Methods Appl. Mech. Eng.*, 197(29):2457–2466, 2008.
17. McFarland, J., Mahadevan, S., Romero, V., and Swiler, L., Calibration and uncertainty analysis for computer simulations with multivariate output, *AIAA J.*, 46(5):1253–1265, 2008.
18. Cressie, N., *Statistics for Spatial Data: Wiley Series in Probability and Statistics*, Wiley-Interscience, New York, 1993.
19. Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D., Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *J. Am. Stat. Assoc.*, 86(416):953–963, 1991.
20. Handcock, M. S. and Stein, M. L., A bayesian analysis of Kriging, *Technometrics*, 35(4):403–410, 1993.
21. Santner, T. J., Williams, B. J., and Notz, W., *The Design and Analysis of Computer Experiments*, Springer, Berlin, 2003.
22. Johnson, R. T., Montgomery, D. C., Jones, B., and Parker, P. A., Comparing computer experiments for fitting high-order polynomial metamodels, *J. Quality Technol.*, 42(1):86–102, 2010.

23. Barton, R. R., Metamodeling: A state of the art review, in *Proc. of the 26th Conf. Winter Simulation*, Society for Computer Simulation International, pp. 237–244, 1994.

24. Rasmussen, C. E. and Williams, C. K. I., *Gaussian Processes for Machine Learning*, The MIT Press, Cambridge, MA, 2006.

25. Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P., Design and Analysis of Computer Experiments, *Stat. Sci.*, 4(4):409–423, 1989.

26. Abramowitz, M. and Stegun, I. A., *Handbook of Mathematical Functions*, Vol. 1, Dover New York, 1972.

27. Beyer, W. H., *CRC Standard Mathematical Tables*, 29th ed., Vol. 1, CRC Press, Boca Raton, FL, 1991.

28. Robert, C. P. and Casella, G., *Monte Carlo Statistical Methods*, Springer-Verlag, New York, 1999.

29. Smith, A. F. and Gelfand, A. E., Bayesian statistics without tears: A sampling–resampling perspective, *The Am. Stat.*, 46(2):84–88, 1992.