Introduction

Carl Ehrett

August 26, 2020

1 Computer model calibration

Computer models are often imperfect reflections of the phenomena they purport to represent. Among their imperfections is the need for calibration. That is, the performance of a computer model is often dependent upon certain settings or assignments of values in the model. For example, in a model designed to estimate the height-time curve of a ball in free fall from a given height, the model output will depend on (among other things) the drag coefficient C of the ball. Given experimental data describing some observed height-time curves of the ball in question, one can use some calibration methodology to attempt to estimate C.

Though the above example relates to a case of estimating the true value of some parameter, calibration need not take this form. Calibration can also take the form of setting values which optimize the model's faithfulness to reality, even when the values being set do not correspond to any particular real features. For example, a classification model may return scores which constitute some measure of confidence of category membership. A probability calibration of the model would involve mapping these confidence scores to values in the interval $[0, 1]$ such that the resulting transformed values are plausible estimates of the probability of category membership.

Given a set of data which may be used for calibration, the available calibration methods are diverse. Often, calibration is approached in an *ad hoc* manner, by manually searching for a set of values for the parameters of interest that cause the model output to come acceptably close to the observed data. This manual approach is often improved by automating it as a grid search or stochastic search, where some measure of model performance is optimized in a brute force manner.

Even among more sophisticated approaches, calibration methods often yield a point estimate of the appropriate value(s) for the calibration parameter(s). Such approaches fail to capture resulting uncertainty regarding those estimates. The successful quantification of uncertainty in model calibration, and the propagation of that uncertainty to yield model estimates with similarly quantified uncertainty, is a major goal of this and other contemporary works in the area of computer model calibration.

Trucano et al. (2006) provide a high-level overview of the relationship of model calibration to both model validation and sensitivity analysis. Much of the field of computer model calibration takes place either within the framework of, or as a response to, the work of Kennedy and O'Hagan (2001). This work

established the dominant paradigm for Bayesian computer model calibration. In this dissertation I refer to that paradigm as the KOH framework. In KOH, one weds a set of experimental data to a computer model that stands in need of calibration, often mediated through a Gaussian process (GP) surrogate model when the original model is of high computational complexity. Systematic discrepancy between the model and the true system can also be represented by a GP, and priors are placed on the calibration parameters. GP hyperparameters are either estimated via maximum likelihood estimation (MLE) or else are also assigned prior distributions. Via integration or some form of Markov-chain Monte Carlo (MCMC, Gelfand and Smith (1990)), a posterior distribution on the calibration parameters is obtained.

Of course, not all work in computer model calibration derives from the KOH framework. For example, Craig et al. (1997) provide an influential demonstration using Bayes linear methods to calibrate a model through a combination of expert judgments (about means and variances of model parameters) and information from runs of low-fidelity versions of the model. Contemporaneously to Kennedy and O'Hagan (2001), Cox et al. (2001) offer a similar calibration framework from a frequentist perspective, including the use of GPs as meta-models for computational efficiency. The authors use MLEs of calibration parameters and GP hyperparameters. Similarly, and more in response to the KOH framework, Loeppky et al. (2006) also provide an MLE-based alternative to KOH that is designed to improve the identifiability of the calibration parameters when model discrepancy is present (i.e., when the model is a systematically biased representation of the target phenomenon). More recently, Wong et al. (2014) describe general framework for a semi-parametric frequentist approach to model calibration, using bootstrapping to provide uncertainty quantification.

Far more common in the field of computer model calibration are extensions and refinements of the KOH framework. Higdon et al. (2004) provide such an extension, adding uncertainty quantification to KOH calibration, to estimate remaining uncertainty regarding the values of the calibration parameter(s). Williams et al. (2006) further refines and exemplifies this approach. Bayarri et al. (2007b) extends KOH to effect simultaneous validation and calibration of a computer model, while Bayarri et al. (2007a) applies this approach to functional data, using a hierarchical representation of coefficients of a wavelet basis. Paulo et al. (2012) further extend the work of Bayarri et al. (2007b) to the case of multivariate model output. While most work in this area assumes that the computer model is deterministic, Pratola and Chkrebtii (2018) provide an example of non-deterministic model calibration.

Paulo et al. (2012) thus are part of a diverse array of projects aiming to shore up a weakness of the KOH framework: its difficulty in accommodating high-dimensional data and/or large data sets in a computationally efficient way. Higdon et al. (2008) focus on meeting the computational challenges of high-dimensional, large data sets by using principal components basis representations for dimensionality reduction. Whereas that work is concerned with the dimensionality of model outputs, Drignei and Mourelatos (2012) are concerned rather with the model inputs. They increase the numerical stability of model calibration by using a global sensitivity analysis to reduce the model inputs' dimensionality. Bhat et al. (2010) demonstrate the application of the KOH framework to multivariate spatial output, and Pratola et al. (2013) apply KOH to nonstationary spatial-temporal field output. Higdon et al. (2013) adapt

KOH to employ an ensemble Kalman filter rather than GPs, for improved computational efficiency. Yuan and Ng (2013) offer a sequential approach to model calibration, in which each evaluation of the high-fidelity model is selected using the previous evaluations to minimize the resulting calibration uncertainty. Such an approach reduces the total number of required model evaluations and thereby the total computational cost of the calibration procedure.

Other works have endeavored to address another known weakness of KOH: poor identifiability of the calibration parameters in the face of model discrepancy. With respect to this problem, Brynjarsdóttir and O'Hagan (2014) emphasize the importance of strong priors on the model discrepancy term when performing model calibration. And for the purpose of improving identifiability over previous approaches, Gu and Wang (2018) propose a novel stochastic process that combines elements of GPs with L_2 calibration (in which the calibration parameter is chosen so as the minimize the L_2 norm of the discrepancy term).

Some works in the area of computer model calibration are premised upon broadening the conception of calibration as relating experimental data to a computer model of the experimental phenomenon. Model calibration can be seen more generally as a method for relating two or more different sources of data with varying costs and varying levels of fidelity. In this vein Qian et al. (2006) explore a method for integrating "high" and "low" models to build a computationally efficient surrogate. Goh et al. (2013) extend KOH to cases of more than two different levels of fidelity.

2   Calibration versus model-assisted design

"Design" used here means not experimental design, but rather refers to making engineering choices to try to effect a desired outcome. Thus the sort of design considered here is related to the mathematical field of optimization. Where calibration involves setting input parameters to induce the model to approximate reality, design involves choosing input parameters to induce an engineered system to behave in some desired way. To understand the way the present work applies methods from the field of computer model calibration, it is helpful to consider the relationship between the two tasks of calibration and design.

At the highest level, calibration can be conceived as follows. One has a model $\eta : \mathbb{R}^{p+q} \to \mathbb{R}^m$, as well as a set of data $\mathbf{y}$ and corresponding $p-$dimensional inputs $\mathbf{x_y}$. We can thus consider η to be a function of two vectors, $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$ Often (but not exclusively) $\mathbf{y}$ is a set of observed experimental outcomes from a real phenomenon $f : \mathbb{R}^p \to \mathbb{R}^m$ where η is a model of f. One wishes to use $\mathbf{y}$ to select values $\boldsymbol{\theta}$ such that settings $\mathbf{t} = \boldsymbol{\theta}$ induces the output of $\eta(\mathbf{x_y}, \boldsymbol{\theta})$ to approximate $\mathbf{y}$. The selection of these values is the calibration of η.

Similarly, model-assisted design can be conceived as follows. One has a model $\eta : \mathbb{R}^{p+q} \to \mathbb{R}^m$, as well as a set of data $\mathbf{y}$ and corresponding $p-$dimensional inputs $\mathbf{x_y}$. We can thus consider η to be a function of two vectors, $\mathbf{x} \in \mathbb{R}^p$ and $\mathbf{t} \in \mathbb{R}^q$ Often (but not exclusively) $\mathbf{y}$ is a set of target outcomes one wishes to achieve in some system $f : \mathbb{R}^p \to \mathbb{R}^m$ where η is a model of f. One wishes to use $\mathbf{y}$ to select values $\boldsymbol{\theta}$ such that settings $\mathbf{t} = \boldsymbol{\theta}$ induces the output of $\eta(\mathbf{x_y}, \boldsymbol{\theta})$ to approximate $\mathbf{y}$. The selection of these values is model-assisted design using η.

Seen in this way, calibration and design are surprisingly similar undertak-

ings. This raises the possibility of applying tools from one domain to problems in the other. One of the primary goals of the present work is to adapt the KOH framework for model-assisted design, and to demonstrate its utility and flexibility in that context.

3   Model-assisted design

The present work thus must be situated in the context of model-assisted design. In model-assisted design, one optimizes a model of the system of interest with respect to some objective function. The resulting optimal model inputs are treated as estimates of the optimal design settings for the system. The present work is intended to be applicable to problems of multi-objective design.

In the field of model- and metamodel-assisted design, Sacks et al. (1989) provide a very influential discussion of strategies to accommodate computer models of high computational complexity with GP surrogates. Following in this vein, Santner et al.'s (2003) foundational work serves as a focal point for much subsequent discussion of how best to learn from computer models in conjunction with physical experiments, for purposes including but not limited to engineering design. Currin et al. (1988) and Currin et al. (1991) develop a similar approach to that of Sacks et al 1989, but with a Bayesian interpretation. Mitchell and Morris (1992) demonstrate the resulting methodology. Craig et al. (2001) discuss Bayesian methods for forecasting using a computer model under uncertainty. Simpson et al. (2008) provide an overview and retrospective of progress in the area of model- and metamodel-assisted design optimization since the work of Sacks et al. (1989), and highlight the value of employing multiple models with varying degrees of fidelity. Bartz-Beielstein and Zaefferer's (2017) is a similar, more up-to-date discussion, focusing on discrete optimization problems. Westermann and Evins (2019) offer a comprehensive and illustrative discussion of the use of metamodels for optimal design of sustainable buildings, including a focus on sensitivity analysis and uncertainty quantification.

4   Gaussian process metamodels

The present work shares with many KOH-based, BO-based, and other approaches a reliance on GP metamodels as surrogates for computationally expensive models. Long popular in geostatistics (where GP regression is referred to as *kriging*; Cressie (2015)), the use of GPs as computer model surrogates was popularized by Sacks et al. (1989). GPs are attractive in this context due to their flexibility, and the ease with which they can interpolate observations of deterministic computer model output while providing closed-form expressions for uncertainty quantification. The application of GPs in this domain is further explored by Santner et al. (2003), who include discussion of the choice of correlation function to suit the desired smoothness properties, and also include a discussion of hierarchical Gaussian random field models for cases when the user is not prepared to specify the desired smoothness. Diagnostic methods for validating a GP surrogate model are explored by Bastos and O'Hagan (2009). Looking at broader applications than metamodeling, Rasmussen et al.'s (2006) work is a seminal text for practitioners seeking to employ GPs in a machine

learning context.

Particularly relevant to the area of model calibration are explorations of the use of GPs to integrate multiple models with varying degrees of fidelity and computational complexity. Qian et al. (2008a) propose a set of Bayesian hierarchical GP models to accommodate a case when low- and high-accuracy data is available. With similar aims, Cumming and Goldstein (2009) describe methods for combining low- and high-accuracy information into a multiscale emulator, as well as providing a design strategy for using the low-fidelity model to select the points at which to evaluate the high-fidelity model. Expanding beyond the paradigm of two levels of fidelity, Goh et al. (2013) use GPs to build a model based on the results of models with several different degrees of fidelity.

Researchers have endeavored to expand the applicability of GP metamodels in numerous ways. To accommodate models that cannot be represented by stationary GPs, Gramacy and Lee (2008) implement a nonstationary GP metamodel using treed partitioning. Qian et al. (2008b) discuss the application of GP metamodels to discrete input spaces. Gratiet et al. (2016) provide a comparative analysis of GPs and polynomial chaos expansions as metamodels in the context of global sensitivity analysis. Other work has focused on a primary weakness of GPs – their poor scalability for large or high-dimensional data sets. Snelson and Ghahramani (2006) provide an influential method for generating sparse GPs to accommodate large data sets. Liu et al. (2020) review a variety of methods for producing scalable GPs, divided broadly into the categories of global approximations that are based on the full data set, and local approximations that achieve scalability by relying on only a subset of the data at each point in the domain.

5 Optimization

The mathematical underpinnings of model-assisted engineering design are found in the field of optimization. A primary strength of the present work is its success in quantifying uncertainty while undertaking calibration and design. This is an area on which much recent innovation in optimization/design has focused. Rockafellar and Wets's (1991) is an influential work describing an algorithm for using a limited set of observations (and hence under uncertainty as to appropriate stochastic model for the system) to solve a multiperiod optimization problem. Sahinidis (2004) offer a comprehensive overview of optimization under uncertainty, analyzing the strengths and weaknesses of a diverse array of approaches. Jin et al. (2003) provide a comparative analysis of a variety of metamodeling techniques with respect to their performance when optimizing under uncertainty. Peherstorfer et al. (2018) offer a survey of methods for optimization under uncertainty that take advantage of multiple models with varying levels of fidelity and computational expense.

Approaches to optimization and design vary enormously, but most methods can be classed as either gradient-based, evolutionary, or as a form of Bayesian optimization (BO). Ruder (2016) offers a brief but comprehensive overview of gradient descent optimization algorithms. Peitz and Dellnitz (2018) propose a gradient-based approach for multi-objective optimization under uncertainty in which descent directions are chosen so as to account for approximation error in the available information about the gradient of the objective function. This

leads one to identify a collection of subsets of the design space which contain the Pareto set for the objective function. Vasilopoulos et al. (2019) demonstrate a method for using gradients to locate an approximate point along the Pareto front and then "trace" the Pareto front to explore it efficiently.

Evolutionary algorithms have become popular methods for optimization in recent times, partly due to their ability to treat the objective function as a black box, without requiring gradient information. This is a feature shared with many forms of BO and with the present work. Jin et al. (2003) provide a survey of work on the use of evolutionary optimization under uncertainty. Zhou et al. (2011) provide a similar survey, focusing on multiobjective evolutionary algorithms. Deb and Gupta (2006) describe two approaches for achieving multi-objective optimization solutions that are robust to small perturbations in the input space.

The present work is a Bayesian approach to optimization problems, and thus might be considered a form of BO. However, this is misleading, as the term BO is widely used to refer more specifically to a set of techniques loosely following in the footsteps of Jones et al. (1998). In the present work, "BO" refers more specifically to this subset of Bayesian approaches to optimization. Other Bayesian approaches that do not fall within that umbrella include the work of Pelikan et al. (1999) and Pelikan (2005), who describe an application of a Bayesian framework to evolutionary optimization.

Jones et al.'s (1998) influential work uses a GP-based response surface approximation to define an acquisition function for sequential selection of evaluation points of the objective function. Vazquez and Bect (2009), Bect et al. (2012), and Chevalier et al. (2014) extend the approach of Jones, developing a stepwise uncertainty reduction methodology for sequential evaluation of the objective function to minimize resulting uncertainty regarding the optimum. This approaches, when applicable (i.e. when it is possible to evaluate the objective function adaptively), can significantly reduce the total number of objective function evaluations required for optimization.

BO's utility has been enhanced by a number of works aimed at practitioners and researchers seeking to apply BO in diverse areas. Shahriari et al. (2016) helpfully introduces the core concepts of BO, and demonstrates its usage in a variety of applications, while Frazier (2018) offers a practical tutorial on the use of BO techniques. Snoek et al. (2012) demonstrate the application of BO to optimization of the hyperparameters of machine learning models. Calandra et al. (2016) provide a demonstrative evaluation of Bayesian optimization methodologies as applied to optimization under uncertainty in the specific domain of robot locomotion parametrization. Picheny et al. (2019) demonstrate how Bayesian optimization methods can be used to identify Nash equilibria in a game theoretic context.

In addition, many efforts have been made to extend the capabilities of BO and to address its shortcomings. Lizotte (2008) reviews the weaknesses of BO – high computational complexity, poorly understood approximation of the response surface to the objective function, and failure to take advantage of the differentiability of objective functions – and demonstrates how to overcome or mitigate these weaknesses. Letham et al. (2019) develop strategies for applying BO in contexts that suffer from high levels of noise, and Snoek et al. (2015) use neural networks for basis function regression of the objective function in place of GPs, to improve the computational efficiency of BO.

References

Bartz-Beielstein, T. and Zaefferer, M. (2017). Model-based methods for continuous and discrete global optimization. *Applied Soft Computing*, 55:154–167.

Bastos, L. S. and O'Hagan, A. (2009). Diagnostics for Gaussian process emulators. *Technometrics*, 51(4):425–438.

Bayarri, M. J., Berger, J. O., and Cafeo, J. (2007a). Computer model validation with functional output. *The Annals of Statistics*, 35:1874–1906.

Bayarri, M. J., Berger, J. O., Paulo, R., Sacks, J., Cafeo, J. A., Cavendish, J., Lin, C.-H., and Tu, J. (2007b). A framework for validation of computer models. *Technometrics*, 49(2):138–154.

Bect, J., Ginsbourger, D., Li, L., Picheny, V., and Vazquez, E. (2012). Sequential design of computer experiments for the estimation of a probability of failure. *Statistics and Computing*, 22(3):773–793.

Bhat, K., Haran, M., and Goes, M. (2010). Computer model calibration with multivariate spatial output. In Chen, M.-H., Dey, D., Müller, P., Sun, D., and Ye, K., editors, *Frontiers of Statistical Decision Making and Bayesian Analysis: in Honor of James O. Berger*, pages 168–184. Springer, New York.

Brynjarsdóttir, J. and O'Hagan, A. (2014). Learning about physical parameters: The importance of model discrepancy. *Inverse Problems*, 30(11).

Calandra, R., Seyfarth, A., Peters, J., and Deisenroth, M. P. (2016). Bayesian optimization for learning gaits under uncertainty. *Annals of Mathematics and Artificial Intelligence*, 76(1):5–23.

Chevalier, C., Bect, J., Ginsbourger, D., Vazquez, E., Picheny, V., and Richet, Y. (2014). Fast Parallel Kriging-Based Stepwise Uncertainty Reduction With Application to the Identification of an Excursion Set. *Technometrics*, 56(4).

Cox, D. D., Park, J. S., and Singer, C. E. (2001). A statistical method for tuning a computer code to a data base. *Computational Statistics and Data Analysis*, 37(1):77–92.

Craig, P. S., Goldstein, M., Rougier, J. C., and Seheult, A. H. (2001). Bayesian forecasting for complex systems using computer simulators. *Journal of the American Statistical Association*, 96(454):717–729.

Craig, P. S., Goldstein, M., Seheult, A. H., and Smith, J. A. (1997). Pressure Matching for Hydrocarbon Reservoirs: A Case Study in the Use of Bayes Linear Strategies for Large Computer Experiments. pages 37–93. Springer, New York, NY.

Cressie, N. (2015). *Statistics for Spatial Data*. Wiley Series in Probability and Statistics. Wiley.

Cumming, J. A. and Goldstein, M. (2009). Small sample bayesian designs for complex high-dimensional models based on information gained using fast approximations. *Technometrics*, 51(4):377–388.

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1988). A bayesian approach to the design and analysis of computer experiments. Technical report, Oak Ridge National Lab., TN (USA).

Currin, C., Mitchell, T., Morris, M., and Ylvisaker, D. (1991). Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments. *Journal of the American Statistical Association*, 86(416):953–963.

Deb, K. and Gupta, H. (2006). Introducing robustness in multi-objective optimization. *Evolutionary Computation*, 14(4):463–494.

Drignei, D. and Mourelatos, Z. P. (2012). Parameter screening in statistical dynamic computer model calibration using global sensitivities. *Journal of Mechanical Design, Transactions of the ASME*, 134(8).

Frazier, P. I. (2018). A Tutorial on Bayesian Optimization.

Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409.

Goh, J., Bingham, D., Holloway, J. P., Grosskopf, M. J., Kuranz, C. C., and Rutter, E. (2013). Prediction and computer model calibration using outputs from multifidelity simulators. In *Technometrics*, volume 55, pages 501–512. Taylor & Francis Group.

Gramacy, R. B. and Lee, H. K. H. (2008). Bayesian treed Gaussian process models with an application to computer modeling. *Journal of the American Statistical Association*, 103(483):1119–1130.

Gratiet, L. L., Marelli, S., and Sudret, B. (2016). Metamodel-based sensitivity analysis: polynomial chaos expansions and gaussian processes. *arXiv preprint arXiv:1606.04273*.

Gu, M. and Wang, L. (2018). Scaled Gaussian stochastic process for computer model calibration and prediction. *SIAM-ASA Journal on Uncertainty Quantification*, 6(4):1555–1583.

Higdon, D., Gattiker, J., Lawrence, E., Jackson, C., Tobis, M., Pratola, M., Habib, S., Heitmann, K., and Price, S. (2013). Computer model calibration using the ensemble kalman filter. In *Technometrics*, volume 55, pages 488–500. Taylor & Francis Group.

Higdon, D., Gattiker, J., Williams, B., and Rightley, M. (2008). Computer model calibration using high-dimensional output. *Journal of the American Statistical Association*, 103(482):570–583.

Higdon, D., Kennedy, M., Cavendish, J. C., Cafeo, J. A., and Ryne, R. D. (2004). Combining field data and computer simulations for calibration and prediction. *SIAM Journal on Scientific Computing*, 26(2):448–466.

Jin, R., Du, X., and Chen, W. (2003). The use of metamodeling techniques for optimization under uncertainty. *Structural and Multidisciplinary Optimization*, 25(2):99–116.

Jones, D. R., Schonlau, M., and Welch, W. J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. Technical report.

Kennedy, M. C. and O'Hagan, A. (2001). Bayesian calibration of computer models. *Journal of the Royal Statistical Society: Series B*, 63(3):425–464.

Letham, B., Karrer, B., Ottoni, G., Bakshy, E., et al. (2019). Constrained bayesian optimization with noisy experiments. *Bayesian Analysis*, 14(2):495–519.

Liu, H., Ong, Y.-S., Shen, X., and Cai, J. (2020). When Gaussian Process Meets Big Data: A Review of Scalable GPs. *IEEE Transactions on Neural Networks and Learning Systems*, pages 1–19.

Lizotte, D. J. (2008). *Practical bayesian optimization*. University of Alberta.

Loeppky, J. L., Bingham, D., and Welch, W. J. (2006). *Computer model calibration or tuning in practice*. Department of Statistics, University of British Columbia.

Mitchell, T. J. and Morris, M. D. (1992). Bayesian design and analysis of computer experiments: two examples. *Statistica Sinica*, pages 359–379.

Paulo, R., García-Donato, G., and Palomo, J. (2012). Calibration of computer models with multivariate output. *Computational Statistics and Data Analysis*, 56:3959–3974.

Peherstorfer, B., Willcox, K., and Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization.

Peitz, S. and Dellnitz, M. (2018). Gradient-based multiobjective optimization with uncertainties. In *NEO 2016*, pages 159–182. Springer.

Pelikan, M. (2005). Hierarchical Bayesian Optimization Algorithm. pages 105–129. Springer, Berlin, Heidelberg.

Pelikan, M., Goldberg, D. E., and Cantú-Paz, E. (1999). BOA: The Bayesian Optimization Algorithm. In Banzhaf, W., Daida, J., Eiben, A., Garzon, M., Honavar, V., Jakiela, M., and Smith, R., editors, *Proceedings of the Genetic and Evolutionary Computation Conference GECCO99*, pages 525–532, Orlando, Florida. Morgan-Kaufmann.

Picheny, V., Binois, M., and Habbal, A. (2019). A Bayesian optimization approach to find Nash equilibria. *Journal of Global Optimization*, 73(1):171–192.

Pratola, M. and Chkrebtii, O. (2018). Bayesian calibration of multistate stochastic simulators. *Statistica Sinica*, 28:693–719.

Pratola, M. T., Sain, S. R., Bingham, D., Wiltberger, M., and Rigler, E. J. (2013). Fast sequential computer model calibration of large nonstationary spatial-temporal processes. *Technometrics*, 55(2):232–242.

Qian, P. Z. G., Jeff, C. F., and Stewart, W. H. M. (2008a). Bayesian hierarchical modeling for integrating low-accuracy and high-accuracy experiments. *Technometrics*, 50(2):192–204.

Qian, P. Z. G., Wu, H., and Wu, C. F. J. (2008b). Gaussian process models for computer experiments with qualitative and quantitative factors. *Technometrics*, 50(3):383–396.

Qian, Z., Seepersad, C. C., Joseph, V. R., Allen, J. K., and Wu, C. F. (2006). Building surrogate models based on detailed and approximate simulations. *Journal of Mechanical Design, Transactions of the ASME*, 128(4):668–677.

Rasmussen, C. E., Williams, C. K. I., Sutton, R. S., Barto, A. G., Spirtes, P., Glymour, C., Scheines, R., Schölkopf, B., and Smola, A. J. (2006). *Gaussian processes for machine learning*. MIT Press, Cambridge.

Rockafellar, R. T. and Wets, R. J.-B. (1991). Scenarios and Policy Aggregation in Optimization Under Uncertainty. *Mathematics of Operations Research*, 16(1):119–147.

Ruder, S. (2016). An overview of gradient descent optimization algorithms. *arXiv preprint arXiv:1609.04747*.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments. *Statistical Science*, 4(4):409–423.

Sahinidis, N. V. (2004). Optimization under uncertainty: state-of-the-art and opportunities. *Computers & Chemical Engineering*, 28(6-7):971–983.

Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The design and analysis of computer experiments*. Springer, New York.

Shahriari, B., Swersky, K., Wang, Z., Adams, R. P., and de Freitas, N. (2016). Taking the human out of the loop: A review of Bayesian optimization. *Proceedings of the IEEE*, 104(1):148–175.

Simpson, T., Toropov, V., Balabanov, V., and Viana, F. (2008). Design and analysis of computer experiments in multidisciplinary design optimization: a review of how far we have come-or not. In *12th AIAA/ISSMO multidisciplinary analysis and optimization conference*, page 5802.

Snelson, E. and Ghahramani, Z. (2006). Sparse gaussian processes using pseudo-inputs. In *Advances in neural information processing systems*, pages 1257–1264.

Snoek, J., Larochelle, H., and Adams, R. P. (2012). Practical bayesian optimization of machine learning algorithms. In Pereira, F., Burges, C. J. C., Bottou, L., and Weinberger, K. Q., editors, *Advances in Neural Information Processing Systems 25*, pages 2951–2959. Curran Associates, Inc.

Snoek, J., Rippel, O., Swersky, K., Kiros, R., Satish, N., Sundaram, N., Patwary, M., Prabhat, M., and Adams, R. (2015). Scalable bayesian optimization using deep neural networks. In *International conference on machine learning*, pages 2171–2180.

Trucano, T. G., Swiler, L. P., Igusa, T., Oberkampf, W. L., and Pilch, M. (2006). Calibration, validation, and sensitivity analysis: What's what. *Reliability Engineering and System Safety*, 91(10-11):1331–1357.

Vasilopoulos, I., Asouti, V. G., Giannakoglou, K. C., and Meyer, M. (2019). Gradient-based pareto front approximation applied to turbomachinery shape optimization. *Engineering with Computers*, pages 1–11.

Vazquez, E. and Bect, J. (2009). A sequential bayesian algorithm to estimate a probability of failure. *IFAC Proceedings Volumes*, 42(10):546–550.

Westermann, P. and Evins, R. (2019). Surrogate modelling for sustainable building design–a review. *Energy and Buildings*, 198:170–186.

Williams, B., Higdon, D., Gattiker, J., Moore, L., McKay, M., and Keller-McNulty, S. (2006). Combining experimental data and computer simulations, with an application to flyer plate experiments. *Bayesian Analysis*, 1(4):765–792.

Wong, R. K. W., Storlie, C. B., and Lee, T. C. M. (2014). A frequentist approach to computer model calibration. *Journal of the Royal Statistical Society: Series B)*, 79(2):635–648.

Yuan, J. and Ng, S. H. (2013). A sequential approach for stochastic computer model calibration and prediction. *Reliability Engineering and System Safety*, 111:273–286.

Zhou, A., Qu, B.-Y., Li, H., Zhao, S.-Z., Suganthan, P. N., and Zhang, Q. (2011). Multiobjective evolutionary algorithms: A survey of the state of the art. *Swarm and Evolutionary Computation*, 1(1):32–49.