Computer model calibration as a method for design

Carl Ehrett*
School of Mathematical and Statistical Sciences, Clemson University,
D. Andrew Brown
School of Mathematical and Statistical Sciences, Clemson University,
Evan Chodora
Department of Mechanical Engineering, Clemson University,
Mingzhe Jiang
Department of Chemical and Biomolecular Engineering, Clemson University,
Christopher Kitchens
Department of Chemical and Biomolecular Engineering, Clemson University,
and
Sez Atamturktur
Department of Architectural Engineering, Pennsylvania State University

November 26, 2018

Abstract

In traditional engineering design, material selection involves choosing a material with appropriate properties from a database of known materials, often as a matter of ad-hoc satisficing. Material design usually occurs separately, and without an eye to specific end-uses. We wed these design processes, selecting a material design by modeling its performance outcomes in a particular engineering application. We show that existing techniques for model calibration can be profitably reconceptualized as a method for optimization and applied to solve this material design problem. We demonstrate by calibrating material design parameters to desired performance targets for a wind turbine blade.

Keywords: Uncertainty quantification, Gaussian processes, optimization, Pareto optimality, wind turbines, material design

*The authors gratefully acknowledge *please remember to list all relevant funding sources in the unblinded version*

1 Introduction

Broadly, in model calibration, one may consider a model to be of the form $\eta(\mathbf{x}, \boldsymbol{\theta})$, where $(\mathbf{x}, \boldsymbol{\theta})$ comprise all inputs to the model. Input vector $\mathbf{x}$ is the collection of inputs that are known and/or under the control of the researcher. The vector of calibration inputs $\boldsymbol{\theta}$ is the collection of parameters the values of which are unknown. These must be estimated for successful simulation. Thus where f describes the true system and y an observation of that system, consider the model to be

$$y(\mathbf{x}) = f(\mathbf{x}) + \epsilon(\mathbf{x}) = \eta(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \epsilon(\mathbf{x}) \tag{1}$$

where $\delta(\cdot)$ describes the model discrepancy – ~~i.e.,~~ the bias of the model as an estimate of the real system – and $\epsilon(\cdot)$ is a mean-zero observation error, often i.i.d. Gaussian.

Much interest in the past two decades has centered on Bayesian methods for model calibration. The appeal of a Bayesian approach to model calibration is that the calibration parameters are a source of uncertainty for the model. This uncertainty should be quantified so that its effect on the model can be made explicit. One can use Bayesian methods to construct a posterior distribution of the calibration parameters which balances our prior knowledge about the calibration parameters with what can be learned from the available data, and which allows for accurate uncertainty quantification on the model outputs.

In many cases of optimization, there are no unknown input parameters for the model. Instead, there are control inputs for which it is unknown which settings achieve optimality. In the application studied here – designing a composite material for a wind turbine blade – the unknowns are the optimal ratio of the two ingredients of the composite and the optimal

thickness of the material when used in a blade. These unknown inputs can be treated as unknown parameters in a calibration problem. Rather than calibrating a computer model's unknown parameters to observations from the real system, then, we calibrate the computer model's control input settings to performance targets. The performance targets are treated as though they were observations of a real system, and the calibration minimizes the discrepancy between the computer model output and these "desired observations." The proposed methodology provides not merely a point estimate of optimal settings, but a posterior distribution which ~~includes~~ uncertainty quantification.

~~The~~ proposed methodology uses the Bayesian framework for computer model calibration provided by Kennedy and O'Hagan (2001). This area is furthered by Higdon et al. (2004), who undertake model calibration with quantification of the related uncertainty. They explicitly incorporate uncertainty regarding the computer model output, the bias of the computer model, and uncertainty due to observation error (of field data). That approach is further refined and exemplified by Williams et al. (2006). Loeppky et al. (2006) offer an MLE-based alternative to the Bayesian approach promulgated by Kennedy and O'Hagan, intending thereby to improve the identifiability of the calibration parameters in the face of model discrepancy. Bayarri et al. (2007) extend the approach of Kennedy and O'Hagan, allowing for simultaneous validation and calibration of a computer model (using the same training data). Bayarri et al. (2007) apply this methodology to functional data using a hierarchical framework for the coefficients of a wavelet representation of the functional data. Similarly, Paulo et al. (2012) apply the lessons of Bayarri et al. (2007) to computer models with multivariate output. Brynjarsdóttir and O'Hagan (2014) demonstrate the

importance of strong priors on the model discrepancy term when undertaking calibration.

Common to these approaches is the conception of calibration as a matter of aligning the computer model output to observations of the real system. In this paper, we calibrate a computer model to align with performance targets, in order to find system settings that optimize performance with respect to those targets. In Section 2 we describe the calibration framework that we adapt as a method for optimization. In Sections 3 and 4 we apply the proposed methodology to an example involving simulated data and to an application in which the goal is to find material design settings to optimize the performance of a wind turbine blade. In Section 4 we also apply the methodology to produce an estimate of the entire Pareto front of the (trivariate) system (where the Pareto front is the set of inputs such that making any of the system outputs better off would require making one or more of the outputs worse off). Section 5 concludes with discussion of the results.

2 Calibration for design

In describing the calibration framework we use in this work, when an emulator is needed we assume the use of a Gaussian process (GP) emulator. Just as a multivariate Gaussian random variable is characterized by its mean vector and covariance matrix, a Gaussian process is fully characterized by its mean function $\mu : D \to \mathbb{R}$ and covariance function $C : D \times D \to \mathbb{R}$, where D is the domain of the process. Thus for any points $\mathbf{x}, \mathbf{y}$ in the domain of the Gaussian process, $\mu(\mathbf{x})$ gives the mean of the Gaussian process at $\mathbf{x}$, and $C(\mathbf{x}, \mathbf{y})$ gives the covariance between the values of the Gaussian process at points $\mathbf{x}$ and $\mathbf{y}$. The distribution of the process at any finite number of points is multivariate Gaussian with

mean vector and covariance matrix ~~given~~ by $\mu(\cdot)$ and $C(\cdot, \cdot)$. In principle, model calibration need not rely on a GP emulator, or any other sort of emulator; one could (e.g.) complete a full Bayesian analysis via an MCMC chain that involves running the relevant computer model at each iteration of the chain. Indeed, in Section 3 we perform calibration on our example simulated data without an emulator. However, computer models are frequently too computationally expensive to allow for such profligacy. Instead, a computationally tractable emulator can be constructed using a sample of observations from the computer model. GPs are popular prior distributions on computer model output for three reasons. Firstly, because their use does not require detailed foreknowledge of the model function's parametric form. Secondly, GPs easily interpolate the computer model output, which is attractive when the computer model is deterministic. This is usually the case, although some attention in model calibration has focused specifically on stochastic computer models; ~~see~~ e.g. Pratola and Chkrebtii (2018). Thirdly, GPs facilitate uncertainty quantification through the variance of the posterior GP. This section provides brief background on Gaussian processes and their use in regression broadly and in computer model calibration specifically.

The use of GPs to produce a computationally efficient predictor $\hat{\eta}(\mathbf{x})$ of expensive computer code $\eta(\mathbf{x})$ given observations of code output at $\mathbf{X} = (\mathbf{x}_1, \ldots, \mathbf{x}_n)^T$ is promulgated by Sacks et al. (1989) and explored at length by Santner et al. (2003). Since computer code is typically deterministic, these applications differ from the focus of O'Hagan (1978) in that the updated GP is induced to interpolate the observations $\boldsymbol{\eta} = (\eta(\mathbf{x}_1), \cdots, \eta(\mathbf{x}_n))^T$. Kennedy and O'Hagan (2001) use GPs for computer model calibration. Kennedy et al. (2006) showcase this use of GP emulators for uncertainty and sensitivity analyses. Bas-

tos and O'Hagan (2009) describe both numerical and graphical diagnostic techniques for assessing when a GP emulator of a computer model is successful, as well as discussion of likely causes of poor diagnostic results. While most work in the area of GP emulation uses stationary covariance functions (in which $\mu(\cdot)$ is constant and $C(\mathbf{x}, \mathbf{x}') \equiv C(\mathbf{x} - \mathbf{x}')$ depends only on the difference between $\mathbf{x}$ and $\mathbf{x}'$, rather than on their location in the input domain) and quantitative inputs, efforts have been made to branch away from these core uses. Gramacy and Lee (2008) use treed partitioning to deal with a nonstationary computer model. Qian et al. (2008) explore methods for using GP emulators that include both quantitative and qualitative inputs.

Suppose that we have inputs $\{\mathbf{x}_i\}_{i=1}^n \subseteq \mathbb{R}^p$ scaled to the unit hypercube, and observations $y(\mathbf{x}_i) = f(\mathbf{x}_i) + \epsilon(\mathbf{x}_i)$, for $i = 1, \cdots, n$, where $f(\cdot)$ is the true system and $\epsilon(\cdot)$ is known measurement error. Then by (1) we have $y(\mathbf{x}_i) = \eta(\mathbf{x}_i, \boldsymbol{\theta}) + \delta(\mathbf{x}_i) + \epsilon(\mathbf{x}_i)$, where $\eta(\cdot, \cdot)$ is the computer model, $\boldsymbol{\theta}$ is the best setting of the vector of calibration parameters, and $\delta(\cdot)$ is the discrepancy function describing the bias of $\eta(\cdot, \cdot)$ as an estimate of $f(\cdot)$. Define the GP prior for modeling $\eta(\cdot, \cdot)$ as follows. Let the mean function $\mu(\mathbf{x}, \mathbf{t}) = c$, c a constant. Set the covariance function in terms of the marginal precision λ_η and a product power exponential correlation function:

$$C((\mathbf{x}, \mathbf{t}), (\mathbf{x}', \mathbf{t}')) = \frac{1}{\lambda_\eta} \prod_{k=1}^{p} \exp\left(-\beta_k^\eta |x_k - x_k'|^{\alpha_\eta}\right) \times \prod_{k=1}^{q} \exp\left(-\beta_{p+k}^\eta |t_k - t_k'|^{\alpha_\eta}\right) \qquad (2)$$

where each β_k describes the strength of the GP's dependence on one of the elements of the input vectors $\mathbf{x}, \mathbf{t}$, and α_η determines the smoothness of the GP.

It is common to plug in the MLEs of λ_η and *beta*$^\eta$ instead of including them in a full Bayesian analysis. In the proposed methodology, that is not merely a convenience,

but rather is essential. This is because in a full Bayesian analysis, λ_η and $\boldsymbol{\beta}^\eta$ would be calibrated to the desired observations. The resulting emulator would be trained not only only the simulator output, but also on our performance targets, which will typically be (intentionally) unrealistic. Therefore, we use values found by minimizing the negative log likelihood of the observations of the simulation with respect to λ_η and $\boldsymbol{\beta}^\eta$. We set the GP to have constant mean $c \equiv 0$, which works well when (as here) the GP is not used for extrapolation.

We similarly model the discrepancy term as a GP, also with mean zero, and with covariance function $C_\delta(\mathbf{x}, \mathbf{x}') = \lambda_\delta^{-1} \prod_{k=1}^{p} \exp\left(-\beta_k^\delta |x_k - x_k'|^{\alpha_\delta}\right)$. We use priors $\rho_k^\delta \sim \text{Beta}(1, 0.3)$, where $\rho_k^\delta = \exp(-\beta_k^\delta/4)$ for $k = 1, \ldots, p$. A Gamma prior is appropriate for λ_δ, with strength determined by the amount of prior information available. With sufficient prior information, a degenerate prior can be used. Details surrounding the choice of prior for λ_δ will be discussed below.

Where the simulation observations are $\boldsymbol{\eta} = (\eta(\mathbf{x}_1, \mathbf{t}_1), \cdots, \eta(\mathbf{x}_n, \mathbf{t}_n))^T$, the field observations are $\mathbf{y} = (y(\mathbf{x}_{n+1}), \cdots, y(\mathbf{x}_{n+m}))^T$, and $\mathcal{D} = (\boldsymbol{\eta}^T, \mathbf{y}^T)^T$, the conditional posterior distribution $\mathcal{D}|\boldsymbol{\theta}, \lambda_\eta, \boldsymbol{\rho}^\eta, \lambda_\delta, \boldsymbol{\rho}^\delta$ is normal with mean 0 and covariance $\mathbf{C}_\mathcal{D}$. $\mathbf{C}_\mathcal{D}$ is a matrix with i, j entry equal to $C((\mathbf{x}_i, \mathbf{t}_i), (\mathbf{x}_j, \mathbf{t}_j)) + I(i, j > n) \cdot (C_{obs}(\mathbf{x}_i, \mathbf{x}_j) + C_\delta(\mathbf{x}_i, \mathbf{x}_j))$ where $C_{obs}(\cdot, \cdot)$ is the (known) observation error variance.

Thus, the joint posterior density under the model is

$$\pi(\boldsymbol{\theta}, \lambda_\delta, \boldsymbol{\rho}^\delta | \mathcal{D}, \lambda_\eta, \boldsymbol{\rho}^\eta) \propto \pi(\mathcal{D}|\boldsymbol{\theta}, \lambda_\eta, \boldsymbol{\rho}^\eta, \lambda_\delta, \boldsymbol{\rho}^\delta) \times \pi(\lambda_\delta) \times \pi(\boldsymbol{\rho}^\delta) \tag{3}$$

~~Note that where~~ a discrepancy function is not included in the model, (3) simplifies to $\pi(\mathcal{D}|\boldsymbol{\theta}, \lambda_\eta, \boldsymbol{\rho}^\eta, \lambda_\delta, \boldsymbol{\rho}^\delta)$. Markov chain Monte Carlo methods are useful for evaluating (3).

Call performance targets treated as observations for the purpose of calibration "desired observations", and call the calibration procedure proposed here, which uses a Bayesian model calibration framework with desired observations, "calibration to desired observations" (CDO). Computer models are more malleable than reality, and one might worry that in calibrating a model to desired observations, the model's fidelity to reality might be mitigated. In many cases, however, one is fortunate to have (perhaps after undertaking traditional model calibration, validation and verification) a computer model such that one is confident that the model is known to be faithful to reality over a given set $\mathcal{T}$ of user specified input settings. In such a circumstance, in calibrating $t \in \mathcal{T}$ to one's desires, one does not risk calibrating the model *away* from agreement with reality, even if one's performance targets are not realistically achievable. Instead, one finds a distribution on the settings that achieve the best realistic approximation to the desired targets.

The tools of model calibration founded in the work of Kennedy and O'Hagan (2001) retain their advantages under ~~the~~ proposed methodology. Most centrally, calibration to desired observations $\mathbf{y}$ produces not merely a static optimum $t \in \mathcal{T}$, but rather a posterior distribution of $t|\mathbf{y}$ reflective of remaining uncertainty about the appropriate value of t. Such uncertainty may have its source in parameter uncertainty (uncertainty about the values of certain model inputs), code uncertainty (uncertainty about how closely the code approximates reality), and that which traditional calibration would consider observation error and model inadequacy. Of course, targets are not actually observations, so the concept of observation error does not cleanly transfer. However, a similar uncertainty would be that due to how close reality *can* come to our desired observations. The Bayesian model

calibration framework allows for the quantification of all of these uncertainties. Furthermore, by the use of informative priors on the model discrepancy and observation error, the identifiability concerns of the Kennedy-O'Hagan approach can be mitigated (Bayarri et al., 2007; Tuo and ~~Jeff~~ Wu, 2016).

Unlike in the case of field observations, when calibrating to performance targets treated as desired observations, the question arises of choosing what exactly those "observations" should be. In many cases, no objectively natural target manifests itself. Indeed, there is no barrier to the use even of impossible targets such as negative values for model outputs known to be nonnegative. Such a target observation in certain situations may be appropriate. However, in general, target observations should aim only a little beyond what is realistically achievable; only as much as is necessary to ensure the targets are at least as ambitious as any true optimum in the system. Three reasons why one should go only a little beyond that are as follows. (1) If target observations are set to be too farfetched, then the calibration can become computationally unstable due to underflow and round-off error, since any value of $\boldsymbol{\theta}$ within its support will have extremely low likelihood. (2) Increasing the distance of the desired observations from the optimal region reduces the identifiability of that region. The calibration finds the region of the parameter space with output closest to the target observations. If the entire model range is far from the target observations, then the optimal region will in relative terms be only a little closer than the rest of the model range. As a result, the identifiability of the optimal region will suffer. (3) The desired observations lose their interpretability when they delve too far into the fantastical, such as with impossibly negative values. Identifying the appropriate range of outputs for desired observations,

which exceed reality only slightly, will often require one to consult a subject matter expert.

When a target cannot be selected, calibration can be performed to each point in a grid over the region of plausible target values. That is, rather than include a desired observation of (e.g.) cost when calibrating a model, one can calibrate to performance targets under each point of a grid of "known" costs. Thus we present a comprehensive picture of optimal parameter distributions and resulting performance under a range of costs, which could inform the process of setting a budget for material construction.

It is not merely likely but often desirable that the performance targets have low probability with respect to the likelihood of the calibrated model. In this way, CDO ~~(calibration to desired observations)~~ is unlike traditional calibration. The reason for this is that if the posterior predictive distribution places substantial probability mass at regions of the parameter space that achieve the target desired observations, then the desired observations may have been insufficiently ambitious. In the wind turbine blade application considered in this work, the ideal material would (impossibly) not deform at all under load. In a different application, one might wish to design a material that deforms in a pre-specified (possible) way. In such a case, it would be appropriate to set desired observations that one indeed does hope to find as the posterior predictive mode after calibration. But in cases such as the wind turbine application, finding the desired observations to be the posterior mode would be an indication that the desired observation could potentially be outperformed, or else a warning (if the desired observation is known to be impossible, such as a material that undergoes zero deformation under load) that the model itself may be unrealistic.

Hereafter in this work, we ~~will assume that~~ the discrepancy term $\delta(\cdot)$ ~~will be modeled~~

as a mean-zero, stationary GP. In order to successfully calibrate to the optimal region of the parameter space, it is necessary either to place an informative prior on the marginal precision λ_δ of the discrepancy, or else to specify that value outright. Otherwise, identifiability issues can cause the calibration to fail. This is a longstanding concern with the Kennedy-O'Hagan framework, raised in the discussion of Kennedy and O'Hagan (2001) as well as by Bayarri et al. (2007), Tuo and ~~Jeff~~ Wu (2015), and Plumlee (2017). How informative one's prior on λ_δ will be depends upon how much one knows about the true Pareto front prior to undertaking CDO. For instance, if in a univariate case it is known with some confidence that the true optimum is nearly constant across control settings and that it occurs in the interval $[10, 11]$, then a constant desired observation of 9 could be used with an informative prior tailored to this prior knowledge of the approximate resulting discrepancy – say Gamma$(20, \text{rate} = 20)$.

When the true Pareto front cannot be estimated prior to undertaking CDO, the desired observations and the prior on the marginal precision of the discrepancy function must be set to avoid the identifiability problems of the Kennedy-O'Hagan framework. That is, where the prior on λ_δ cannot be chosen to be *accurate* (due to insufficient prior knowledge) it should be chosen to *overestimate* the precision. Otherwise, underestimation of λ_δ may lead to poor identifiability of the optimal region of the parameter space. Again consider the example with the constant optimum in $[10, 11]$, but suppose now that our prior knowledge is much more impoverished – we can confidently hold only that the optimum takes positive values under 20. Then a constant desired observation of -1 could be paired with a Exp(1) prior on λ_δ, to reflect our hope for a discrepancy of 1 while remaining open to a significantly

larger discrepancy. In such a case, by setting a prior that overestimates λ_δ, the posterior distribution of λ_δ becomes less reliable than when the prior derives from substantive prior knowledge. Nonetheless, even when λ_δ must be overestimated, the posterior distribution of $\boldsymbol{\theta}$ will still peak at the optimal region(s) of the parameter space, since overestimation of λ_δ only increases the penalty of leaving that region. And so while relying on vague knowledge of the optimum does interfere with one's ability to estimate the true discrepancy of the model from the desired observations, even in such circumstances one may still locate the posterior mode(s) of $\boldsymbol{\theta}$ and thereby the optimal settings for the model. However, if λ_δ is too highly overestimated, then MCMC may become trapped in a local mode. In short, while the proposed methodology is forgiving of overestimation of λ_δ, the identifiability of the optimal region(s) is best served by supplying as accurate and informative of a prior as possible.

When too little is known about the Pareto front of a system to select a performance target and an appropriate prior for λ_δ, a preliminary round of CDO can be used to estimate the Pareto front. For example, consider again the univariate case where we know only that the optimum is in the range $(0, 20)$. One can perform CDO with constant desired observation -1 and a prior on λ_δ that deliberately exploits the identifiability problems of the Kennedy-O'Hagan framework in order to explore large regions of the parameter space – say Exp(rate $= 0.1$). The Pareto front of the resulting predictive distribution can be used as an estimate of the true Pareto front in the vicinity of the desired observation. This preliminary estimate (unlike the estimate achieved after full CDO) does not include uncertainty quantification, since it is achieved by filtering the predictive distribution samples

Algorithm 1: Full CDO procedure including preliminary estimation of Pareto front	
1.	Set desired observations $\mathbf{d}$ to lie outside of the model range.
2.	Set a vague prior on λ_δ.
3.	Perform calibration and use the resulting posterior samples of θ to draw posterior predictive samples of the model output.
4.	Filter the resulting predictive samples to retain only the Pareto optimal samples. The remaining samples $\mathcal{P}$ serve as the estimate of the Pareto front.
5.	Select new desired observations using $\mathcal{P}$. These may be entirely new targets chosen after examining $\mathcal{P}$, or may simply be updated automatically to set each d_i to be the same (small) distance $1/\sqrt{\lambda}$ (for some λ) from $\mathcal{P}$.
6.	Set a strong (or degenerate) prior on λ_δ with mean λ.
7.	Perform calibration.

to retain only the Pareto optimal points. The preliminary estimate allows one to select a new (set of) desired observation(s) that are known to lie near the optimal region, along with an accurate and informative prior on λ_δ that reflects the estimated distance between the new desired observation(s) and the optimal region. To continue the above example, preliminary CDO may show ~~one~~ that the true optimum across the domain of the control inputs is in the interval $[10, 11]$, which puts one in a position to use the aforementioned combination of a constant desired observation of 9 and $\lambda_\delta \sim \text{Gamma}(20, \text{rate=20})$ prior, or even a degenerate prior with λ_δ set equal to 1. The full CDO process including preliminary estimation of the Pareto front is ~~shown as~~ Algorithm ~~(1)~~.

In the case of univariate output, the primary reason to perform this preliminary round of

calibration would be to improve the accuracy of the posterior discrepancy. This is because in the case of univariate output, while there may be uncertainty about the magnitude of the difference between the Pareto front and a given set of desired observations, there is no uncertainty about the direction of this difference. But in the case of multivariate output, both the magnitude and the direction of the difference may be unknown. In such a case, a poorly chosen set of desired observations may not result in calibration to the desired region of the Pareto front. Consider a case of bivariate positive output (y_1, y_2) where y_2 is continuous and strictly decreasing in y_1. Then any value of y_1 corresponds to some point in the Pareto front, and so it may be ~~antecedently~~ unclear to what performance targets the system should be calibrated. In such a case when selecting a target for optimization one often wishes to locate an "elbow" which maximizes $\frac{d^2 y_2}{dy_1^2}$, since this elbow is a point of diminishing returns when trading higher y_1 for lower y_2 values. Depending on its location, selecting (e.g.) $[0, 0]$ as a desired observation might not achieve calibration to this elbow, if some other part of the Pareto front is closer to the desired observation than the elbow is. In such a situation a rough estimate of the Pareto front (if not antecedently available) via a preliminary round of CDO using a vague prior on λ_δ can help one select a desired observation to which the closest region of the Pareto front is the region to which one wishes to calibrate. Then one may perform CDO using desired observations and a prior distribution on λ_δ that are chosen informatively, so that one will enjoy the above-described benefits both of calibrating to the desired region of the Pareto front and of improving the accuracy of one's posterior distribution of the discrepancy of the model from the desired observations. Note that when an emulator is used, a preliminary round of CDO can use

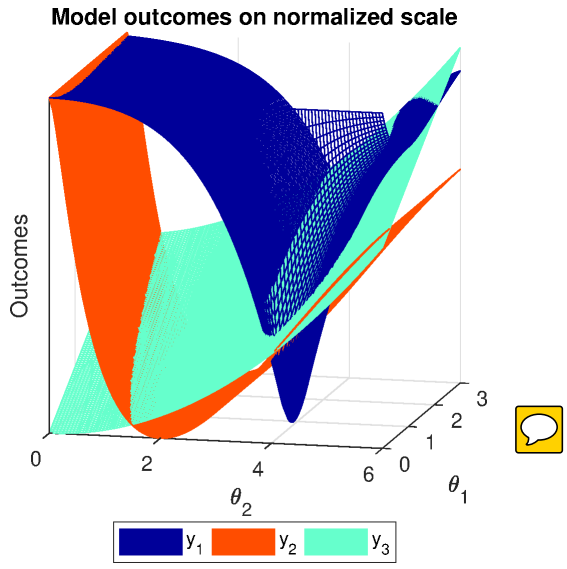

Figure 1: Example model outputs shown on a common scale.

the same set of model observations as the subsequent CDO as the training points for the emulator. So performing preliminary CDO to sharpen one's desired observation and λ_δ prior does not add to the total budget of model runs, and can thus be a computationally cheap supplement to CDO.

3 Example

To illustrate CDO, consider the following artificial problem. Let $(x, \boldsymbol{\theta})$ be the vector of inputs, with scalar control input $x \in [1.95, 2.05]$ and calibration parameters $\boldsymbol{\theta} = (\theta_1, \theta_2) \in [0, 3] \times [0, 6]$. We consider three outputs: $y_1 = \left(\theta_1 \exp\left(-\left(\theta_1 + |\theta_2 - \frac{\pi x}{2}|\right)\right) + 1\right)^{-1}$, $y_2 = \left(\theta_2^{x-1} \exp\left(-0.75\theta_2\right) + 1\right)^{-1}$, $y_3 = 15 + 2\theta_1 + \theta_2^2/4$. Figure 1 displays the y_1, y_2, and y_3 surfaces as functions of θ_1 and θ_2 at $x = 2$ on a common scale. Assuming an easily evaluated model, we have $\mathbf{y}(\mathbf{x}, \boldsymbol{\theta}) = \mathbf{f}(\mathbf{x}, \boldsymbol{\theta}) + \delta(\mathbf{x}) + \epsilon$ for desired observation $\mathbf{y}$, where $\mathbf{f}$ is the model output,

$\delta(\cdot)$ is the discrepancy function and ϵ is $N(0, 0.05)$.

We initially set the desired observations to $[0, 0, 0]$, constant as a function of x. We then estimated the Pareto front via a preliminary round of CDO in order to estimate the standardized distance of the desired observation from the Pareto front. The distance from the estimated Pareto front to the desired observation was found to be large ~~— at~~ 16 units on the standardized scale, roughly four times the diameter of the entire model range. As a result, the use of $[0, 0, 0]$ as a desired observation would lead to poor identifiability of the optimal region. This is because the desired observation is approximately the same distance from any point in the model range, relative to the distance from the desired observation to the optimal region. Therefore in order to improve identifiability of the optimal region, we updated the desired observation to lie along the same line connecting the original desired observation to the estimated Pareto front, but now closer to the latter. We chose a distance of one unit away (roughly one fourth of the diameter of the model range), approaching the estimated Pareto front as closely as possible while remaining confident that the new desired observation of $[0.71, 0.71, 17.92]$ still outperforms the true Pareto front (i.e., lies outside the model range and is such that if it were added to the model range, it would be a Pareto optimal point). We then set the discrepancy marginal precision λ_δ to 1 for subsequent CDO, corresponding to a degenerate prior informed by the estimated distance of the new desired observation from the Pareto front. Observation error $\epsilon(\cdot)$ from (1) was taken to be distributed as $N(0, 0.05)$ for all x. Figure 2 shows the resulting posterior distribution, including the marginal distributions of the calibration parameters. The sharply peaked marginals show substantial Bayesian learning compared to the prior

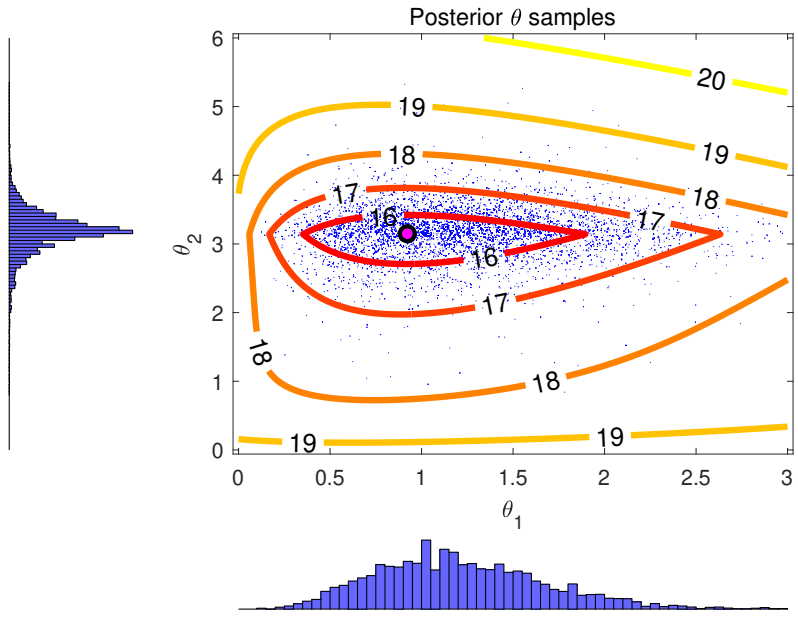

Figure 2: Posterior draws from CDO in the simulated example. The contours show, for each point in the parameter space, the Euclidean distance of the model output at that point from the desired observation (averaged across the control input range $[1.95, 2.05]$). The large dot shows the true optimum.

distributions distribution of the calibration parameters, which is uniform over the area shown in the scatterplot. The calibration successfully maps the contours of the optimal region, and peaks near the true optimum.

4 Application

In this section we describe the use of CDO for designing a material to be used in a wind turbine blade of fixed geometry. In traditional engineering design, material selection is a matter of choosing a material with appropriate properties for the project at hand from a database of known materials, often as a matter of ad-hoc satisficing. Material design

usually occurs separately, and without an eye to specific end-uses. CDO allows us to wed these design processes, selecting a material design by modeling its performance outcomes in a wind turbine blade. This calibration is mediated by a model using `ANSYS` finite element analysis software. The finite element model is treated as an accurate representation of reality.

4.1 Project background

Two primary performance targets for the design and construction of wind turbine blades are the distance (in meters) that the blade tip deflects under load from its starting position, and the angle of rotation the blade experiences under load. Within the set of materials studied here, we want each of these measures and the material cost be as close to zero as possible. The blade is to be a composite of two given materials, one serving as the *matrix* and the other the *filler*. In a composite, the matrix holds the filler together; an example would be concrete, in which a filler of loose stones is combined with a matrix of cement. For the wind turbine blade, given a fixed choice of matrix and filler, the properties of the composite depend on the volume fraction (i.e. the volume ratio of filler material to matrix material used in the composite) and the thickness of the material used to build the blade. The resulting material properties impact the performance of the blade, as well as its cost per square meter. The finite element model takes as inputs a triplet (h, v, k), where h is the operating temperature of the wind turbine (in kelvin), v is the volume fraction of the material, and k is the thickness of the material (in mm). The outputs of the model are a triplet (d, r, c), where d is tip deflection (in meters), r is rotation (in radians), and c is cost

per square meter (USD). The wind turbine should be capable of operating over the range of temperatures 230K-330K. We used CDO to find a distribution on optimal settings for v and k given outputs from the finite element simulator and desired observations.

4.2 Emulation of finite element model

The finite element simulator is too computationally expensive to be suitable for direct use in ~~(e.g.)~~ an MCMC routine. We employed a GP emulator in the manner of Williams et al. (2006). For this purpose, we drew 504 (trivariate) observations from the finite element simulator. These inputs follow a Latin hypercube sampling design (McKay et al., 1979) based on plausible ranges for the three inputs, as identified by subject matter experts. We consider the finite element observations to follow a GP with mean 0 and covariance function C as described by (2) above, with $\alpha_\eta = 2$. This choice of α_η assumes smooth, infinitely differentiable sample paths.

The hyperparameters $\lambda_\eta, \boldsymbol{\beta}^\eta$ must be estimated. To avoid our estimates being biased by calibration to the desired observations, we estimated them prior to calibration via maximum likelihood estimation. We used `fmincon()` in MATLAB to maximize the log-likelihood of the simulation observations (Equation (3) with $\mathcal{D} = \boldsymbol{\eta}$) over the joint (6-dimensional) support of $\boldsymbol{\beta}^\eta, \lambda_\eta$. The result is $\hat{\lambda}_\eta = 0.0152$, $\hat{\boldsymbol{\rho}}^\eta = (0.9358, 0.6509, 0.6736, 0.4797, 0.9673)$ where $\rho_k^\eta = \exp(-\beta_k^\eta/4)$.

4.3 Calibration of the wind turbine blade system

All model inputs were normalized to [0,1] over their supports. All model outputs were standardized so that the vector of simulation responses $\boldsymbol{\eta}$ has mean 0 and standard deviation 1. The full joint posterior density of the calibration parameters and discrepancy function hyperparameters is given in Equation (3), using the MLEs given above.

The initial desired observations were set to $[0, 0, 0]$ on the original scale, constant as a function of temperature, on a grid of temperature values. We carried out an initial round of CDO in order to update the desired observations to ones that lie close to the Pareto front. For this purpose, a total of 20,000 samples were drawn via Metropolis-Hastings-within-Gibbs MCMC, of which 4,000 samples were discarded as burn-in. During the burn-in period, the covariance of the proposal distributions for $\boldsymbol{\theta}$, λ_δ, and $\boldsymbol{\rho}^\delta$ were all periodically adjusted for optimal acceptance rates using the sample covariance of the preceding draws. As expected for the preliminary round of CDO, the posterior distribution of $\boldsymbol{\theta}$ was quite diffuse. We used the GP emulator to estimate the model output for each sample of $\boldsymbol{\theta}$ drawn. We filtered the resulting posterior predictions to retain only the estimated Pareto front. ~~Examining the estimated Pareto front, one finds a distinct "elbow"; see figure 3.~~ We selected this elbow as the target for calibration. To do so, we set the point [deflection = 0.75m, rotation = 0.09 rad, cost = \$130.34] as the desired observation (constant as a function of temperature). The elbow is the closest region of the Pareto front to this point. Based on the estimated Pareto front, the desired observation is approximately 0.2 units away on the standardized scale. Therefore, we set $\lambda_\delta = 1/0.2^2 = 25$.

In the ~~resulting CDO,~~ we employed the same MCMC approach as in the preliminary

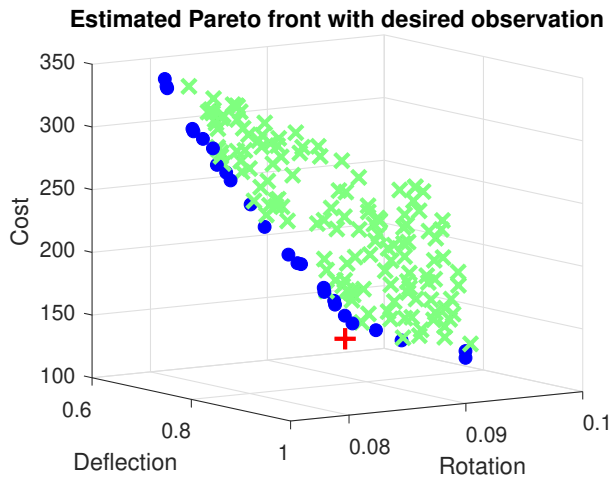

Figure 3: Each x shows a non-Pareto optimal point drawn from the predictive distribution through preliminary CDO. The dots show the estimated Pareto front. The plus sign is the desired observation selected to calibrate to the "elbow" in the Pareto front.

round, except that λ_δ was now treated as known. The marginal posterior distribution of the calibration parameters is shown in Figure 4 via ~~four levels~~ of highest density regions. The contrast of the posterior distribution with the prior, which is uniform over the area shown in the figure, indicates that significant Bayesian learning has occurred in the calibration process. The prior and posterior marginal predictive distributions of the model outputs are shown in Figure 5. The posterior marginals peak sharply near the performance target. The mean model output under the prior is [deflection = 0.76m, rotation = 0.09 rads, cost = $\$207.90/\text{m}^2$], whereas under the posterior it is $[0.76\text{m}, 0.09 \text{ rad}, \$148.68]$. Though the performance outcomes are approximately the same under the posterior distribution as under the prior, the cost per square meter has dropped dramatically. If one desires instead to prioritize gains in performance over cost, this can be accomplished by selecting desired observations that reflect those priorities.

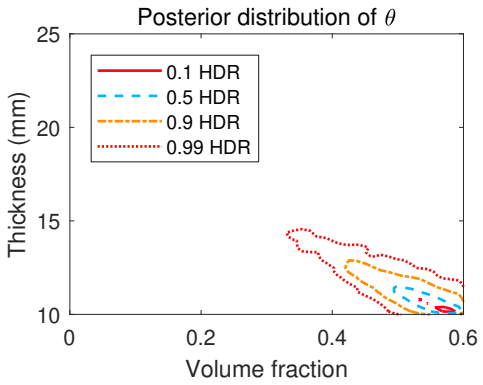

Figure 4: ~~Four levels~~ of highest density regions of the posterior distribution from calibration of the wind turbine blade system. The prior is uniform over the area shown.

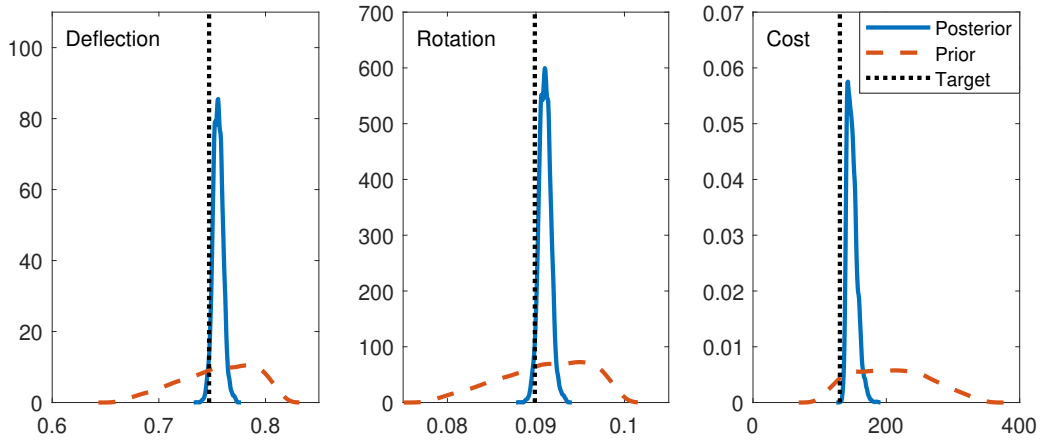

Figure 5: Prior and posterior marginal predictive distributions for each of the three model outputs. Notice that it is to be expected that the posteriors peak near the target (and not on it), since the target was intentionally chosen to lie outside the model range.

4.4 Pareto front estimation

Often in the case of a system with multivariate output, one might not antecedently have a clear target outcome. When *ceteris paribus* all outputs are to be minimized, any point in the Pareto front is optimal relative to some set of priorities. If those priorities have not been explicitly determined prior to calibration, then no particular outcome can be targeted. In determining one's priorities, it is helpful to know the Pareto front of the relevant system. For example, in a system where quality is monotonically increasing in cost, depending on one's tolerance for high cost, any point in the model range might be optimal. In low-dimensional cases, CDO may be used to achieve a holistic picture of the Pareto front by optimizing to each of a grid of performance targets. To do this, where the model output is $d-$dimensional, one may draw a grid over the range of $d-1$ of the model outputs and perform CDO to minimize the remaining output at each point of the grid. The $d-1$ outputs, at each grid point, are treated as known up to observation error (meaning that the discrepancy function $\delta(\cdot)$ is set to 0 in the dimension of these outputs). The resulting estimate is distinguished from other methods of estimating the Pareto front (including from the filtering method employed in preliminary CDO) by including uncertainty quantification.

This procedure is illustrated here using the wind turbine blade application. For simplicity, rotation has been removed as a model output, leaving a system with 2-dimensional output of deflection and cost. The range of cost is known (via preliminary CDO) to be $[\$96, \$352]$. A 20-point grid was drawn over this range of costs. For each point c in the cost grid, we used the point $(0\text{m}, \$c)$ as the performance target for calibration (constant with respect to temperature). For each such point, we then updated this initial desired

observation to improve identifiability using the rough estimate of the Pareto front from preliminary CDO using desired observation (0m, \$0). Note that only one round of preliminary CDO was needed for this purpose, rather than a separate instance at each grid point.

The result of the strategy is to provide an estimate of the response surface with included uncertainty quantification describing, for each point in the grid, the optimal achievable outcome for the output not included in the grid. Thus a decisionmaker can visualize the space of desirable possibilities with associated uncertainty metrics. They can do so without the need for ~~antecedently~~ rigorously determining their exact priorities for weighing gains in each of the outputs against one another.

The result of applying this strategy to the wind turbine blade application is shown in Figure 6. The lefthand plot shows that the posterior model output respected the "known" cost values used in the calibrations. The Pareto front for the system appears with uncertainty bands in the righthand plot. This plot visualizes a distribution on the optimal performance outcome for any cost that a decisionmaker might select as a budget for production, which would be helpful when selecting a budget. For example, the elbow around \$140 manifests itself as a potentially attractive choice, since it can be seen in the plot that prior to that point each dollar spent brings significant gains in reducing deflection. Spending above that level continues to reduce deflection, but less sharply.

5 Conclusion

We have described the use of computer model calibration under the framework established by Kennedy and O'Hagan (2001), ~~Williams et al. (2006) and Bayarri et al. (2007)~~ and

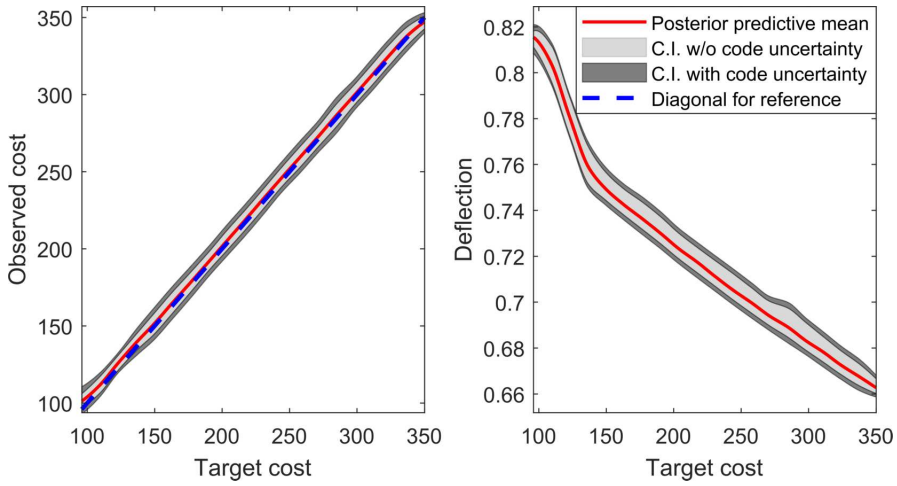

Figure 6: The lefthand plot verifies that the calibration achieved the "known" costs up to small error. The righthand plot is an estimate of the Pareto front for the system with attendant uncertainty quantification.

how it can be used to address questions of engineering design. CDO is a modification of that framework which calibrates a computer model, not to field observations, but rather to desired observations; i.e., to performance targets for the system. Unlike other methods of Bayesian optimization (e.g., Shahriari et al. 2016), CDO does not require the ability to carry out computer model observations adaptively. Instead, it can operate using a batch of observations gathered prior to (and independently of) the calibration procedure. We described the implementation of this approach in an MCMC routine along with considerations to accommodate computational instability. The use of this methodology is illustrated in the case of material design for a wind turbine blade.

We have shown thereby a variety of ways in which CDO can be used to guide decision-makers in the design process. By expropriating established tools of model calibration, CDO offers a method of optimization which is sensitive to all sources of uncertainty, and which

results in an estimate that includes uncertainty quantification.

~~As discussed earlier, the~~ methodology as described here treats the computer model as universally valid over the domain of the calibration inputs. Future work in this area will include the use of a discrepancy term capturing model bias. This would allow for simultaneous calibrations: both traditional and CDO. ~~That is, a computer model could be calibrated to real observations while also being calibrated to performance targets treated as desired observations. These two goals would not necessarily conflict, since the two calibrations take different inputs to be the calibration parameters. CDO calibrates inputs that are under operator control, and thereby would be treated as control inputs in traditional calibration.~~ Other extensions of the proposed methodology could include its application to ~~state-aware~~ calibration (Atamturktur and Brown, 2015; Stevens et al., 2018; Brown and Atamturktur, 2018), which would allow the optimal region of the calibration parameters to vary as a function of the control inputs.

SUPPLEMENTARY MATERIAL

Matlab code for CDO: This includes the example model described in Section 3, along with code to perform CDO on that system and thereby reproduce Figure 2.

References

Atamturktur, S. and D. A. Brown (2015). State-aware calibration for inferring systematic bias in computer models of complex systems. *NAFEMS World Congress Proceedings, June 21-24.*

Bastos, L. S. and A. O'Hagan (2009, nov). Diagnostics for Gaussian Process Emulators. *Technometrics 51*(4), 425–438.

Bayarri, M. J., J. O. Berger, J. Cafeo, G. Garcia-Donato, F. Liu, J. Palomo, R. J. Parthasarathy, R. Paulo, J. Sacks, and D. Walsh (2007). Computer Model Validation with Functional Output. *The Annals of Statistics 35*, 1874–1906.

Bayarri, M. J., J. O. Berger, R. Paulo, J. Sacks, J. A. Cafeo, J. Cavendish, C.-H. Lin, and J. Tu (2007, may). A Framework for Validation of Computer Models. *Technometrics 49*(2), 138–154.

Brown, D. A. and S. Atamturktur (2018, feb). Nonparametric Functional Calibration of Computer Models. *Statistica Sinica 28*, 721–742.

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

Gramacy, R. B. and H. K. H. Lee (2008, sep). Bayesian Treed Gaussian Process Models With an Application to Computer Modeling. *Journal of the American Statistical Association 103*(483), 1119–1130.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne (2004, jan). Combining Field Data and Computer Simulations for Calibration and Prediction. *SIAM Journal on Scientific Computing 26*(2), 448–466.

Kennedy, M. C., C. W. Anderson, S. Conti, and A. O'Hagan (2006, oct). Case studies

in Gaussian process modelling of computer codes. *Reliability Engineering & System Safety 91*(10-11), 1301–1309.

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

Loeppky, J. L., D. Bingham, and W. J. Welch (2006). Computer Model Calibration or Tuning in Practice. *Technometrics*.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979, may). Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code. *Technometrics 21*(2), 239–245.

O'Hagan, A. (1978). Curve Fitting and Optimal Design for Prediction. *Journal of the Royal Statistical Society. Series B 40*(1), 1–42.

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

Plumlee, M. (2017). Bayesian Calibration of Inexact Computer Models. *Journal of the American Statistical Association 112*(519), 1274–1285.

Pratola, M. and O. Chkrebtii (2018). Bayesian Calibration of Multistate Stochastic Simulators. *Statistica Sinica 28*, 693–719.

Qian, P. Z. G., H. Wu, and C. F. J. Wu (2008, aug). Gaussian Process Models for Computer Experiments With Qualitative and Quantitative Factors. *Technometrics 50*(3), 383–396.

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

Santner, T. J., B. J. Williams, and W. I. Notz (2003). *The Design and Analysis of Computer Experiments*. New York: Springer.

Shahriari, B., K. Swersky, Z. Wang, R. P. Adams, and N. de Freitas (2016, jan). Taking the Human Out of the Loop: A Review of Bayesian Optimization. *Proceedings of the IEEE 104*(1), 148–175.

Stevens, G. N., S. Atamturktur, D. A. Brown, B. J. Williams, and C. Unal (2018, apr). Statistical inference of empirical constituents in partitioned analysis from integral-effect experiments. *Engineering Computations 35*(2), 672–691.

Tuo, R. and ~~C. F. Jeff~~ Wu (2015). Efficient calibration for imperfect computer models. *Annals of Statistics 43*(6).

Tuo, R. and ~~C. F. Jeff~~ Wu (2016, jan). A Theoretical Framework for Calibration in Computer Models: Parametrization, Estimation and Convergence Properties. *SIAM/ASA Journal on Uncertainty Quantification 4*(1), 767–795.

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