Coupling material and mechanical design processes via computer model calibration

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

February 7, 2019

Abstract

In traditional engineering design, material selection consists in choosing from a database of known materials, often as a matter of ad-hoc satisficing. Material design usually occurs separately, without an eye to specific end-uses. We wed these design processes, designing a material by modeling its performance in a particular application. We show that techniques for model calibration can be reconceptualized as a method for optimization and applied to solve this material design problem. We demonstrate our proposed approach by calibrating material design settings to performance targets for a wind turbine blade and by estimating the system's Pareto front with quantificatied uncertainties.

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

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

1 Introduction

Real-world optimization problems typically involve multiple objectives. This is particularly true in the design of engineering systems, where performance outcomes are balanced against budgetary constraints. Among the complexities involved in optimizing over multiple objectives is the effect of uncertainties in various aspects of the problem. Design is guided by models known to be imperfect, systems are built using materials with uncertainty regarding their material properties, variations will occur in the construction of designed systems, and so on. These uncertainties will propogate to uncertainty in the resulting solution to a given design problem.

In traditional engineering design, one chooses a material with appropriate properties for a project from a database of known materials, which were themselves designed apart from consideration of any particular projects performance goals. By coupling models of material properties and models of the behavior of a built system, we can combine these two traditionally separate design processes under the umbrella of a unified multiple objective optimization problem.

Our 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). The approach of Higdon et al. (2004) is further refined and exemplified by Williams et al. (2006). Loeppky et al. (2006) offer a maximum-likelihood-based alternative to the Bayesian approach pro-

mulgated 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 addition to optimizing to a specific set of target outcomes, in many situations it is helpful to have a comprehensive picture of optimal outcomes while remaining agnostic with respect to specific targets. For example, in a system in which there is a trade-off between cost and performance, which region of the model range is treated as a target outcome will depend upon the budgetary constraints of the project. Decision makers in such a scenario are well-served by having a complete picture of the curve describing the optimal performance possible at any budget. More generally, in a system with multiple outputs, one wants to estimate the Pareto front of the model range, i.e., the set of points in the model range that are Pareto optimal. A point is Pareto optimal if and only if, in order to improve any one of its elements, some other

element must be made worse off. For example, if a bivariate system in a minimization problem has exactly three output points $a = (2, 0)$, $b = (2, 1)$, and $c = (1, 10)$ then a and c are each Pareto optimal, but b is not, since there is a point (namely a) that is less than or equal to b in every output and strictly less than b in at least one output. We show that the proposed methodology can be used to estimate the Pareto front of a system with uncertainty quantification.

We apply the proposed methodology both to an artificial example and to the problem of finding material design settings to optimize the performance and cost of a wind turbine blade of fixed geometry. The wind turbine blade in question is to be constructed using a composite of two materials. One design variable targeted for optimization is the *volume fraction*, or ratio of the two materials used in the composite. Another design variable is the thickness (in mm) of the composite material when used in the blade. Our material design goal is to minimize the cost per square meter of the composite material, the rotation (in radians) of the blade when under load, and the deflection (in meters) of the blade tip when under load.

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 the example involving simulated data and to the wind turbine blade design. In Section 4 we apply the proposed methodology to produce an estimate of the Pareto front of the wind turbine blade system. The estimate includes uncertainty quantification. Section 5 concludes with discussion of the results.

2 Calibration for design

2.1 Gaussian process emulators for calibration

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 determined 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 Markov chain Monte Carlo (MCMC; Gelfand and Smith, 1990) by running the relevant computer model at each iteration of the chain. Indeed, Hemez and Atamturktur (2011) compare the use of Kennedy-O'Hagan-style calibration both with and without and emulator, and 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 (Van Buren et al., 2013, 2014). 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 and hence free of measurement error. This is usually the case, although some attention in model calibration (e.g. Pratola and Chkrebtii, 2018) has focused specifically on stochastic computer models. 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. Bastos 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 and quantitative inputs, efforts have been made to branch away from these core assumptions. 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.

Whether or not an emulator is used, in calibration one may consider a computer model

to be of the form $\eta(\mathbf{x}, \boldsymbol{\theta})$, where $(\mathbf{x}, \boldsymbol{\theta})$ comprise all inputs to the model. The input vector $\mathbf{x}$ is the collection of inputs, called control 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 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 (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.

To employ an emulator, suppose that we have inputs $\{(\mathbf{x}_i, \mathbf{t}_i)\}_{i=1}^n \subseteq \mathbb{R}^p \times \mathbb{R}^q$ scaled to the Cartesian product of the $p-$ and $q-$dimensional unit hypercubes, and that we have completed computer model runs $\eta\left((\mathbf{x}_i, \mathbf{t}_i)\right)$ for $i = 1, \cdots, n$. 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:

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

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

It is common to plug in the MLEs of λ_η and $\boldsymbol{\beta}^\eta$ 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 on 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 $c = 0$, which works well when (as here) the GP is not used for extrapolation (Bayarri et al., 2007).

We similarly model the discrepancy term $\delta(\cdot)$ as a GP, also with mean zero, and with covariance function $C_\delta(\mathbf{x}, \mathbf{x}') = \lambda_\delta^{-1} \prod_{k=1}^{p} \exp\left(-\beta_k^\delta |x_k - x'_k|^{\alpha_\delta}\right)$. This is included in the model to capture systematic discrepancy between target outcomes and the feasible model range. 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. We set $\alpha_\delta = 2$, thereby assuming that the model output is infinitely differentiable.

Let $\boldsymbol{\eta} = (\eta(\mathbf{x}_1, \mathbf{t}_1), \cdots, \eta(\mathbf{x}_n, \mathbf{t}_n))^T$ be the vector of completed runs of the simulator, $\mathbf{y} = (y(\mathbf{x}_{n+1}), \cdots, y(\mathbf{x}_{n+m}))^T$ the field observations, and $\mathcal{D} = (\boldsymbol{\eta}^T, \mathbf{y}^T)^T$. Then $\mathcal{D}|\boldsymbol{\theta}, \lambda_\eta, \boldsymbol{\rho}^\eta, \lambda_\delta, \boldsymbol{\rho}^\delta$ is distributed as multivariate 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. Setting a uniform prior on the design variables $\boldsymbol{\theta}$, the joint posterior density under the model is

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

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

2.2 Calibration to desired observations

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). One might worry that calibrating a model to artificial "desired observation" will undermine the model's accuracy, since the model is essentially being fed false data. 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 $\mathbf{t} \in \mathcal{T}$ 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 most 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 our proposed methodology. Most centrally, calibration to desired observations $\mathbf{y}$ produces not merely a point optimum $\mathbf{t} \in \mathcal{T}$, but rather a posterior distribution of $\mathbf{t}|\mathbf{y}$ reflective of remaining uncertainty about the appropriate value of $\mathbf{t}$. Such uncertainty may have its source in parameter uncertainty (uncertainty about the values of model inputs other than the design variables), model form 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 Wu, 2016).

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: Firstly, 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. Secondly, 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. Thirdly, 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. For instance, since different manufacturers

have different budgets, one can calibrate to performance targets under each point of a grid of "known" costs, rather than calibrating the model to a specific desired cost. In doing so, 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 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.

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 $\delta(\cdot)$, or else to specify λ_δ 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 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$(20, 20)$.

Where the prior on λ_δ cannot be chosen to be *accurate* (due to insufficient prior knowledge of the Pareto front) 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. This is because λ_δ indirectly describes the (inverse) expected distance of the desired observations from the posterior predictive distribution. Thus, underestimating λ_δ can result in there being little penalty in leaving the region of the model range that is closest to the desired observations. When setting a prior that overestimates λ_δ, the posterior distribution of λ_δ becomes less reliable than when an accurate prior is used. 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, leading to convergence problems. 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 informative of a prior as possible.

In situations where one lacks the prior knowledge necessary to select performance targets near the Pareto front and an accurate prior for λ_δ, an alternative to simply relying on an overestimated λ_δ with a "safe" choice of performance targets is to use a preliminary round of CDO to estimate the Pareto front. For example, consider again the univariate case, supposing now that we know only that the optimal output is approximately constant somewhere in the range $(0, 20)$. One could perform CDO directly with constant desired observation -1 and an exponential prior on λ_δ with rate 2. If the true optimal output turns out to be in the lower end of its known range, this will likely work well. However, if the optimal output turns out to be high, the optimal region of the parameter space could suffer from poor identifiability. To avoid this, one can instead perform CDO still with constant desired observation -1 but with 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 exponential with rate 0.1. Though the resulting posterior distribution will have greater density in the optimal region(s) than would a uniform sampling, it will likely not center in the optimal region, instead covering a larger area of the model range. The resulting posterior predictive distribution can be filtered to retain only its Pareto front, and these remaining predictive samples 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. The preliminary estimate merely allows one to select a new set of desired observations that is 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 target and the optimal region. Performing CDO with these new targets and prior will result in a posterior distribution that concentrates on the optimal region, and the resulting posterior predictive distribution will allow one to estimate the optimal output with appropriate quantification of the uncertainty.

To continue the above example, preliminary CDO may show 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 $\lambda_\delta \sim$ Gamma(20, 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 given in Algorithm 1.

An illustration of the benefits of preliminary CDO appears in Figure 1. Suppose that, prior to calibration, we know only that the model outputs are positive. Then $(0, 0)$ is a natural choice for a performance target to use as a desired observation. However, in this case, that choice of desired observation will yield three problems. Firstly, the model range is distant from $(0, 0)$. As a result, the optimal region is, relative to the size of the model range, not much closer to the desired observation than other regions of the model range. Indeed, the farthest point in the entire model range is only 1.78 times farther from $(0, 0)$ than is the nearest point. This will lead to very poor identifiability of the optimal region. Secondly, the optimal region determined by the choice of $(0, 0)$ is somewhat arbitrary. Notice that in

Algorithm 1: Full CDO procedure including preliminary estimation of Pareto front	
1.	Set desired observations 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 $\boldsymbol{\theta}$ 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 in the usual way.

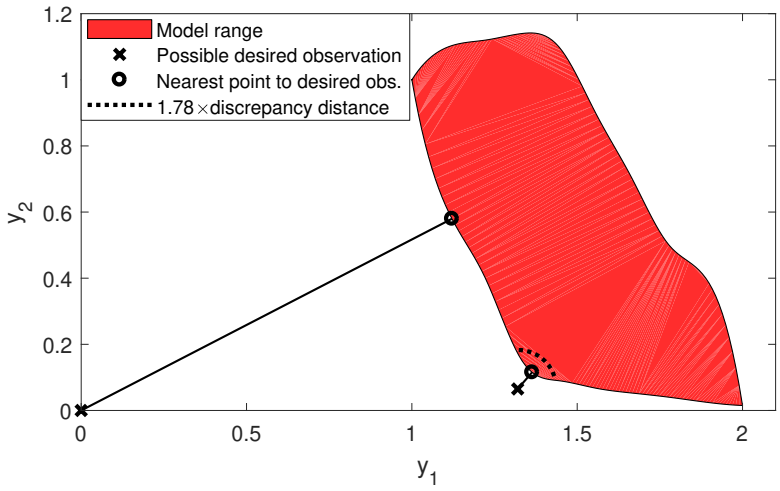

Figure 1: Two potential choices of performance targets to serve as desired observations. The distance from $(0, 0)$ to the farthest point in the model range is only 1.78 times the distance to the nearest point. By contrast, for the point (1.32,0.065), the dotted line shows the small region of the model range within 1.78 times the distance to the nearest point.

this example the entire "left" and "bottom" edges of the model range comprise the Pareto front. The point closest to $(0, 0)$ is not intrinsically superior to any other point on the Pareto front. Its uniqueness lies solely in being nearest to the origin, and that choice of desired observation was itself driven merely by our ignorance of the model range. Thirdly, since we don't know the model range, we are not in a position to set an informative and accurate prior for λ_δ. By contrast, suppose now that we have performed preliminary CDO and have available a rough estimate of the Pareto front, empowering us to choose the point (1.32,0.065) as our desired observation. This answers all of the above problems. Firstly, this point is much closer to the model range, leading to greater identifiability. Whereas the entire model range is within 1.78 times the distance of the nearest point to the origin, here the region of the model that is 1.78 times the distance to (1.32,0.065) is small. As a result,

the calibration process is able to concentrate the mass of the posterior distribution in a small optimal region. Secondly, this choice of desired observation and resulting optimum are not arbitrary. An estimate of the Pareto front allows us to see an "elbow", where allowing y_1 to increase further leads to diminishing returns in reducing y_2. This may (depending on one's goals) be an especially attractive point to which to calibrate. Or, if one's priority is to minimize y_1 at all costs, one might instead select $(.9, 1)$ as a desired observation. Regardless of one's goals, preliminary CDO empowers one to select a target that comports with those goals, as opposed to the case illustrated by selecting the origin out of ignorance of the model range. Thirdly, with a rough estimate of the Pareto front we can supply a strong (or even degenerate) prior for λ_δ. Note also that when an emulator is used, a preliminary round of CDO can use the same set of model observations as the subsequent CDO for the training points of the emulator. So performing preliminary CDO does not add to the total budget of model runs, and can thus be a computationally cheap supplement to CDO.

3 Simulated Example

As an illustration of our proposed procedure with known solution, 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 2 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

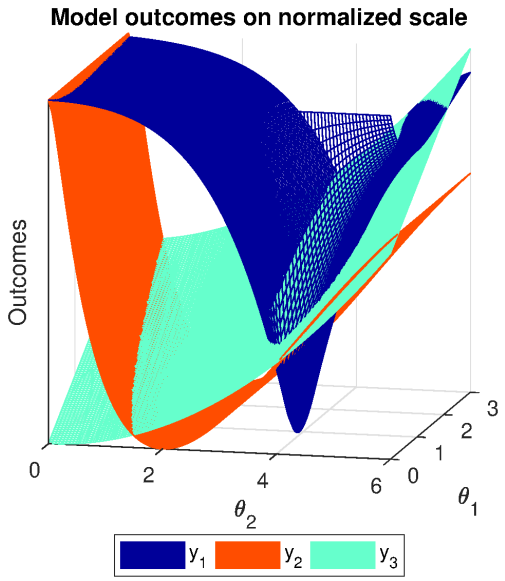

Figure 2: True outputs of the example model, shown on a common scale.

performance target $\mathbf{y}$, where $\mathbf{f}$ is the model output, $\delta(\cdot)$ is the discrepancy between the optimal region and the performance target, and $\boldsymbol{\epsilon}$ follows a $N(\mathbf{0}, 0.05I)$ distribution. Aside from easing matters computationally, $\boldsymbol{\epsilon}$ serves as "observation error" of our own performance targets, since in a typical case one can at best identify a small region within which the choice of any particular point as a performance target would be arbitrary due to our acceptable tolerance.

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 with $\lambda_\delta \sim \text{Exp}(1)$ in order to estimate the standardized distance of the desired observation from the Pareto front. We filtered the resulting posterior predictive distribution to retain only the Pareto optimal points and estimated the distance from the Pareto front to the desired observation. The distance is 16 units on the standardized scale, which is large compared to the model range, at roughly four times its diameter. 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 line connecting the original desired observation to the nearest point of 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 marginal precision of the discrepancy function 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. For comparison, we also performed CDO directly, without the use of preliminary CDO. To do so, we used our original desired observation of $(0, 0, 0)$ with a Gamma(10,10) prior deliberately overestimating λ_δ. Notice that although the Gamma prior from direct CDO and the degenerate prior $\lambda_\delta = 1$ from full CDO have the same means, this is an overestimation only in the case of direct CDO, since the two calibrations used different desired observations. While the Exp(1) prior used in preliminary CDO does pair a mean of 1 with desired observation (0,0,0), the exponential prior does not penalize low values of λ_δ, and so fails to result in overestimation of λ_δ. Figure 3 shows the resulting posterior distributions of the two calibration procedures, including the marginal

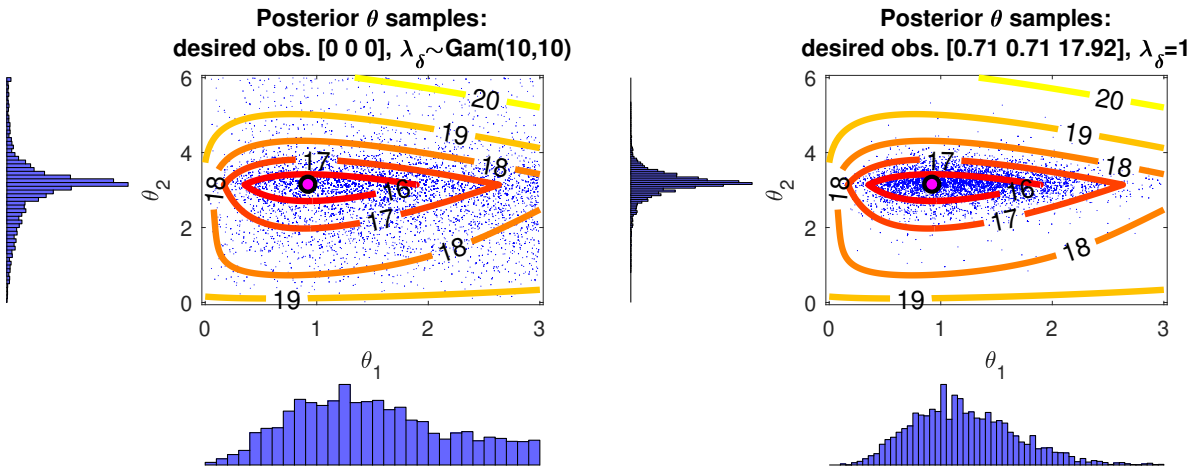

Figure 3: Posterior draws from CDO in the simulated example both without and with the use of preliminary CDO. The contours show, for each point in the parameter space, the Euclidean distance of the model output at that point from the original desired observation (0,0,0), averaged across the control input range $[1.95, 2.05]$. The large dot shows the true optimum.

distributions of the calibration parameters. The marginals in each case show substantial Bayesian learning compared to the prior distribution of the calibration parameters, which is uniform. The calibrations successfully map the contours of the optimal region, and peak near the true optimum. However, the benefits of preliminary CDO are apparent in the greater spread of the posterior distribution from direct CDO. The marginals are much more sharply peaked after using preliminary CDO, with much lighter tails. This simulation example illustrates that CDO can be used directly with little foreknowledge of a system's Pareto front, but that greater identifiability of the optimal region can be achieved using preliminary CDO.

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. The goal is to wed the typically separate tasks of material selection and material design, thereby designing a composite material to optimize performance in a turbine blade. The calibration here is mediated by a model using `ANSYS` finite element analysis software. We assume that the finite element model is 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 output of the model is the triplet (d, r, c), where d is tip deflection (in meters), r is rotation (in radians), and c is cost per square meter (USD) to construct the material. 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 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. The inputs are determined by 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), 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()` (MATLAB, 2017) 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 an evenly-spaced grid of 21 temperature values over the range [230K, 330K]. We carried out a preliminary round of CDO with a weak Exp(5) prior on λ_δ, in order to estimate the Pareto front and update the desired observations to lie close to the Pareto front and thereby improve identifiability of the optimal region. For this purpose, a total of 20,000 samples were drawn via Metropolis-Hastings-within-Gibbs MCMC (Metropolis et al., 1953; Hastings, 1970; Geman and Geman, 1984), 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. Figure 4 displays the estimated Pareto front, where we can see a distinct "elbow." 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

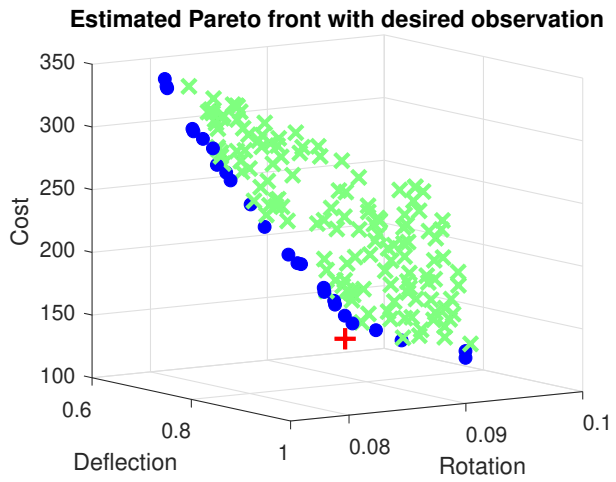

Figure 4: Each x shows a non-Pareto optimal point drawn from the predictive distribution through preliminary CDO. The dots show the estimated Pareto front. The plus sign is the desired observation selected to calibrate to the "elbow" in the Pareto front.

desired observation is approximately 0.2 units away on the standardized scale. Therefore, we set $\lambda_\delta = 1/0.2^2 = 25$.

In the subsequent CDO step, we employed the same MCMC approach as in the preliminary round, except that λ_δ was now treated as known. The marginal posterior distribution of the calibration parameters is shown in Figure 5 as contours 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 6. 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 $(0.76\text{m}, 0.09 \text{ rad}, \$148.68/\text{m}^2)$. Though the performance outcomes are approximately the

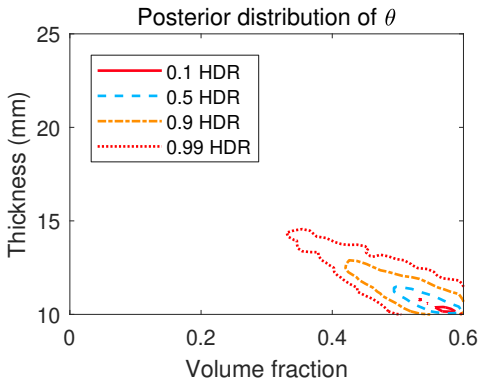

Figure 5: Contours of highest density regions of the posterior distribution from calibration of the wind turbine blade system. The prior is uniform over the area shown.

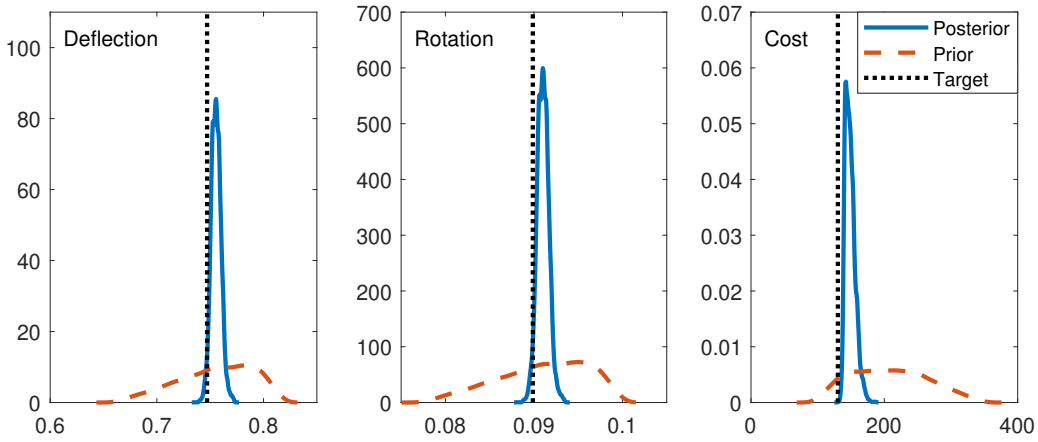

Figure 6: Prior and posterior marginal predictive distributions for each of the three model outputs. Note that it is to be expected that the posteriors peak near the target (and not on it), since the target was intentionally chosen to lie outside the model range.

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.

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.

Our proposed 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 $(0\text{m}, \$0)$. 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 decision maker can visualize the space of desirable possibilities with associated uncertainty metrics. They can do so without the need for 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 7. 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 decision maker 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.

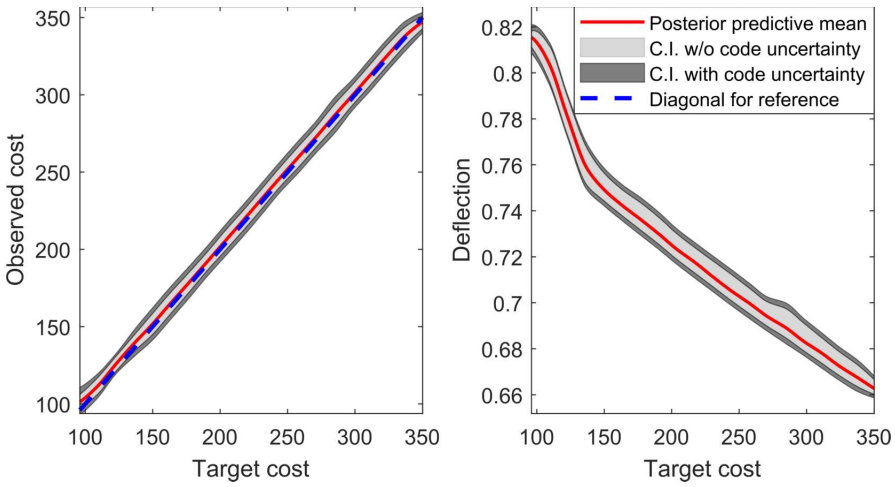

Figure 7: The lefthand plot shows the posterior predictive distribution of cost as a function of target cost, verifying that the calibration achieved the "known" costs up to small error. The righthand plot shows the posterior predictive distribution of optimal deflection as a function of target cost, and therefore is an estimate of the Pareto front for the system with attendant uncertainty quantification.

Spending above that level continues to reduce deflection, but less sharply.

The use of CDO in the wind turbine calibration case illustrates how preliminary CDO may be used, not merely to improve the identifiability of a pre-determined optimal region as in Section 3, but rather to identify a desirable region of the model range and select performance targets that calibrate to that region. In the wind turbine case, selecting $(0, 0, 0)$ as one's target determines the optimal region to be the high-cost region toward the upper-left of Figure 4, since (on the standardized scale of model outputs) that region happens to be closest to the target. If one has substantive goals that drive one to select that target, then one is well-served by optimizing to that high-cost region. But if the target $(0, 0, 0)$ is chosen arbitrarily, then the resulting optimal region is itself determined arbitrarily. The estimate of the Pareto front provided by preliminary CDO allows us to identify regions of special interest, and to select performance targets that calibrate to those regions, as illustrated in Figure 4. The use of CDO in this case also demonstrates the value of obtaining a posterior distribution on the calibration parameters, rather than just a point estimate. For example, Figure 5 shows not just that a reasonable point estimate of the optimal $\boldsymbol{\theta}$ is at (.58, 10.2mm), but also the that if the volume fraction is lowered to 0.4 it is important to simultaneously raise the thickness to 14mm. More generally, one can see in the Figure an indication of the range of $\boldsymbol{\theta}$ values that achieve results near the target, which is potentially useful when one's goal is to set tolerances (rather than a specific value) for $\boldsymbol{\theta}$. Finally, the use of CDO in the wind turbine case illustrates how the method can deliver "Pareto bands", providing not merely an estimate of the Pareto front (as in preliminary CDO) but also uncertainty bands on that estimate. Such an estimate can be of use to

decision-makers when deciding on performance and budgetary targets.

5 Conclusion

We have described the use of computer model calibration under the framework established by Kennedy and O'Hagan (2001), and 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 (a review of which was provided by Shahriari et al. 2016), CDO does not require the ability to evaluate model output 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.

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. Other extensions of the proposed methodology could include its application to so-called "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 3.

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). 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). A framework for validation of computer models. *Technometrics 49*(2), 138–154.

Brown, D. A. and S. Atamturktur (2018). 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).

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

Geman, S. and D. Geman (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence 6*(6), 721–741.

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

Hastings, W. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika 57*(1), 97–109.

Hemez, F. M. and S. Atamturktur (2011, sep). The dangers of sparse sampling for the quantification of margin and uncertainty. *Reliability Engineering & System Safety 96*(9), 1220–1231.

Higdon, D., M. Kennedy, J. C. Cavendish, J. A. Cafeo, and R. D. Ryne (2004). 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). Case studies in Gaus-

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

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

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

MATLAB (2017). *Version 9.2.0 (R2017a)*. Natick, Massachusetts: The MathWorks, Inc.

McKay, M. D., R. J. Beckman, and W. J. Conover (1979). Comparison of three methods for selecting values of input variables in the analysis of output from a computer code. *Technometrics 21*(2), 239–245.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller (1953). Equation of state calculations by fast computing machines. *The Journal of Chemical Physics 21*(6), 1087–1092.

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). 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). 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). Statistical inference of empirical constituents in partitioned analysis from integral-effect experiments. *Engineering Computations 35*(2), 672–691.

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

Tuo, R. and C. F. J. Wu (2016). A theoretical framework for calibration in computer models: Parametrization, estimation and convergence properties. *SIAM/ASA Journal on Uncertainty Quantification 4*(1), 767–795.

Van Buren, K. L., S. Atamturktur, and F. M. Hemez (2014, feb). Model selection through robustness and fidelity criteria: Modeling the dynamics of the CX-100 wind turbine blade. *Mechanical Systems and Signal Processing 43*(1-2), 246–259.

Van Buren, K. L., M. G. Mollineaux, F. M. Hemez, and S. Atamturktur (2013, jul). Simulating the dynamics of wind turbine blades: partII, model validation and uncertainty quantification. *Wind Energy 16*(5), 741–758.

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