Review of Manuscript Number TCH-19-111
Coupling material and mechanical design processes via computer model calibration

This paper lays out a method to search for the engineering design with considering the performance target. The method is developed under the computer model calibration framework proposed by Kennedy and O'Hagan (2001). The Pareto front of the system is firstly estimated. The engineering design is then searched near the Pareto front. The authors also present a simulation example and a case study. The method and the results are interesting, but there are several concerns to be addressed.

1. The methodology-wise novelty of this work does not stand out as the Bayesian approach of Kennedy and O'Hagan (2001) can be applied to the problem with minor changes. The preliminary CTO is novel but it is not mathematically formulated. It might add some methodology-wise contribution to the paper if the problem is formulated as a multi-objective optimization problem and show some statistical properties of the estimated Pareto front.

2. The method is established on the computer model calibration framework proposed by Kennedy and O'Hagan (2001). The first difference is that no observational data is required. Instead, the computer simulation model $\eta(x, \theta)$ is "calibrated" towards a deterministic function $\zeta(x)$. When $\eta(x, \theta)$ is also deterministic, $\epsilon(x)$ in Equation (1) does not incorporate any randomness. $\epsilon(x)$ is described as the tolerance of the estimated ζ and the target value on line 49 of page 8 and the author claims that all the errors within the tolerance region are considered as equivalent. In this case, why $\epsilon(x)$ is assumed to be a normally distributed random variable?

3. The second difference is that both of x and θ are observable in the reality and the target function $\zeta(\cdot)$ is assumed to be a function of x only. However, the distance to the target function δ might depend on x and θ. The authors should mention and justify the assumption that $\theta(\cdot)$ is also assumed to be a function of x only.

4. Under the formulation in Equation (1), the target function ζ could be a function of x. Then the Pareto front is also a function of x. In this case, how do the authors select the target outcome? Will θ be a function of x as well?

5. Intuitively, the proposed method is equivalent to searching for a design setting near the Pareto front of all the simulation responses with allowing for some error $\delta(x)$. If this is the case, the tolerance $\epsilon(x)$ and the discrepancy $\delta(x)$ play the similar role. Shouldn't these two terms be merged?

6. Instead of using the model calibration framework, why do not optimize the input variables of the computer simulation model directly?

7. The paper mentions the identification issue in the computer model calibration framework on page 9. It is better to move those discussion to the introduction section.

8. "True Pareto front" is mentioned in the paper. But the mathematical definition of the Pareto front is not provided. Also, it would be better that the authors show that the estimated Pareto front approximates the true Pareto front in the simulation study. I'm also curious about the relationship between the estimation bias of the Pareto front and the selection of the preliminary target outcome when the prior of δ is vague.

9. Line 39 of page 15: the motivation of using Latin hypercube sampling design needs to be justified.

10. Algorithm 1 step 5: "mean the distance" should be "the mean distance".

11. Line 49 of page 15: the sentence "we use two binary dummy variables a_1, a_2 to convert the model to univariate output with five inputs, in order to exploy a univariate GP emulator" needs more explanation.

This article was reviewed by two reviewers and myself. It considers the calibration problem but with an eye toward optimization. The authors appear to employ the statistical model of Kennedy and O'Hagan (2001). Both reviewers were critical of the simplicity of adopting this simple model given all the subsequent work. I see no problem defaulting to a simple model and using it in a creative way.

The articles main innovation seems to be in Algorithm 1. This also reveals some striking problems.

(1) I have no idea how one would "Set target outcomes y out of the feasible design space". Don't you need some idea of target outcomes first?

(2) Step 2 uses the data y to create a "posterior distribution" of theta. It is unclear how this is a posterior, as the rules of probability do not apply here. y has no distribution, it was set in Step 1. The authors appear to imply it should have the covariance listed on page 8, line 39. This is a philosophically confusing point that was not clear after several read-throughs.

(3) Steps 3-5 are vague and Step 6 has the same problem as Step 2.

Overall, I think the idea of using a calibrated model for optimization with robustness is an interesting idea. But the lack of mathematical justification or intuition in this paper makes it hard to recommend. There is simply not enough here to advocate for its publication. Despite giving some details, I am unsure the methodology in the paper could be replicated based on the vagueness present in many of the discussions. I encourage the authors to try to create more specific, mathematical arguments. For example, the "true Pareto front" is not fully described, as pointed out by a reviewer. The few details the authors provide sometimes slips into the trivial side of things. I will give a couple of examples:

(o) The discussion of the covariance functions on page 8 is routine.

(o) "Since the model output is trivariate, we use two binary dummy variables a 1, a 2 to convert the model to univariate output with five inputs, in order to employ a univariate GP emulator."