Reviewer 1:

1. Address record-to-record variability
   1. Same covariate point could have different result because of ground motion variation
   2. This leads to a stochastic GP and DoE should address this (Picheny paper is deterministic)
2. Be clear that GP is not the solution to inverse design
   1. Present the problem generally
      1. Optimization of design parameter
      2. GP as a solution
   2. The presentation should focus first on the optimal design parameter problem and then present the GP as means for efficient solution.
3. Reconsider ARD kernel
   1. If ARD chosen, explain in detail
4. Batch DoE
   1. Why not serial? Update the variance after each point without recalibrating GP?
   2. Is this guaranteed to not have overlapping points?
5. Should use more points in test-sample validation
   1. Why use test sample instead of cross validation?
      1. A sample from Yi & Taflanidis (specific for comparing different strategies)
      2. The metamodel accuracy is evaluated using a test-sample validation. This is preferred to the cross-validation setting discussed previously because it avoids dependence of the DoE validation on the specific set of simulation experiments chosen (Kleijnen and Van Beers 2022; Zhang and Taflanidis 2018). Since the objective of the validation is to compare across the different DoE strategies, the test-sample implementation is necessary for accommodating consistency.
   2. Why not use global accuracy as a metric?
   3. What are reasonings behind the convergence threshold and # of pt per batch?
6. Should solve the general problem before applying GP
   1. The difference in the design parameters identified by the different GP variants warrants more careful discussion. Why are these differences so big and which solution is closer to the actual true solution? The authors need to solve the actual problem (after properly defining it) and then compare the GP solution to that one. This is a standard practice when promoting GP-based optimization approaches.
7. Accuracy of GP is not sufficient in general

Reviewer 2:

Consider presenting design optimization as a multi-objective, perhaps with a Pareto set.

Ideas:

* Plot one of theseA comparison of a design spectrometer

  Description automatically generated
* A graph of a graph

  Description automatically generated
* Solving the optimization “by hand”
  + Define the objective function
  + Generate a 1D version of the problem
    - Run say 22 GMs at each gap ratio for about 10 gap ratios
    - Create a distribution of p(failure) at each gap ratio
    - Perform the optimization problem by hand this way
    - Identify inverse design, considering mean and mean+1std
* Create objective function that incorporates both cost and performance

One important observation is that the predictive variance in GP does not depend on y