Helton, J. C., and Davis, F. J. (2003), "Latin Hypercube Sampling and the Propagation of Uncertainty in Analyses of Complex Systems," *Reliability Engineering and System Safety*, 81, 23–69. [21]

Janusevskis, J., and Le Riche, R. (2010), "Simultaneous Kriging-Based Sampling for Optimization and Uncertainty Propagation," HAL Report: hal-00506957. Available at *http://hal.archives-ouvertes.fr/hal-00506957_v1/*. [21]

Kleijnen, J. P. C. (2007), "Risk Analysis: Frequentist and Bayesians Unite!," in *Proceedings of the 2007 INFORMS Simulation Society Research Workshop*, ed. E. Yücesan, pp. 61–65. [21]

——— (2008), *Design and Analysis of Simulation Experiments*, New York: Springer. [21,22]

Kleijnen, J. P. C., and Helton, J. C. (1999), "Statistical Analyses of Scatter Plots to Identify Important Factors in Large-Scale Simulations, 1: Review and Comparison of Techniques," *Reliability Engineering and Systems Safety*, 65, 147–185. [23]

Kleijnen, J. P. C., Mehdad, E., and Van Beers, W. C. M. (2012), "Convex and Monotonic Bootstrapped Kriging," in *Proceedings of the 2012 Winter Simulation Conference*, eds. C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher. [23]

Kleijnen, J. P. C., Ridder, A. A. N., and Rubinstein, R. Y. (in press), "Variance Reduction Techniques in Monte Carlo Methods," in *Encyclopedia of Operations Research and Management Science* (3rd ed.), eds. S. Gass and M. Fu, New York: Springer. [22]

Kleijnen, J. P. C., Van Beers, W. C. M., and Van Nieuwenhuyse, I. (2010), "Constrained Optimization in Simulation: A Novel Approach," *European Journal of Operational Research*, 202, 164–174. [22,23]

——— (2013), "Expected Improvement in Efficient Global Optimization Through Bootstrapped Kriging," *Journal of Global Optimization*, 54, 59–73. [22]

Law, A. M. (2007), *Simulation Modeling and Analysis* (4 ed.), Boston: McGraw-Hill. [21]

Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013), "Quantile-Based Optimization of Noisy Computer Experiments with Tunable Precision," *Technometrics*, 55, 2–13. [21]

Vose, D. (2000), *Risk Analysis; A Quantitative Guide* (2nd ed.), Chichester, UK: Wiley. [22]

Yin, J., Ng, S., and Ng, K. (2011), "Kriging Meta-Model With Modified Nugget Effect: An Extension to Heteroscedastic Variance Case," *Computers and Industrial Engineering*, 61, 760–777. [22]

Supplementary materials for this article are available online. Please go to http://www.tandfonline.com/r/TECH

Comment: EI Criteria for Noisy Computer Simulators

Pritam Ranjan

Deparatment of Mathematics and Statistics
Acadia University
Wolfville, NS B4P 2R6, Canada
(*pritam.ranjan@acadiau.ca*)

The article by Picheny et al. (2013) is very interesting, and I believe that the methodologies developed here are innovative and should be useful for computer experiment practitioners. I congratulate the authors for writing such a nice article. Although I like most of the ideas in this article (especially the time allocation strategy), the formulation of the improvement function may need more justification, and this is the main focus of this discussion. To recall, the key assumptions of Picheny et al. are as follows:

1. The computer simulator is noisy or nondeterministic. That is, the ith observed simulator output $\tilde{y}_i = y(x_i) + \varepsilon_i$ is noisy, where $y(x_i)$ is the unobservable true underlying deterministic simulator output for the input x_i, and $\varepsilon_i \sim N(0, \tau_i^2)$ characterizes the noise (replication error).
2. The error variance τ_i^2 (or precision $1/\tau_i^2$) of an observation can be tuned by the user, for example, $\tau_i^2 = \tau(t_i)$, where t_i is the time spent in evaluating the computer simulator at x_i.

For deterministic simulators, Jones et al. (1998) proposed an *expected improvement* (EI) criterion for choosing follow-up points in the sequential design framework with the objective of estimating the global minimum of the simulator. This EI criterion is very popular because it facilitates both the local and the global search of the input space for finding the global minimum. However, having noise in the responses changes the surrogate model and the sequential design scheme. Furthermore, the control over the response precision introduces another parameter that has to be tuned. Under these assumptions, Picheny et al. (2013) proposed the following two main contributions:

1. In the same spirit of Huang et al. (1998), a new *expected quantile improvement* (EQI) criterion—a generalization of the EI criterion in Jones et al. (1998)—is proposed for choosing follow-up trial locations from the noisy simulator.
2. A procedure for the allocation of the computational time (or equivalently the precision) for each future observation is also proposed in the new EQI formulation.

The notion of EI criteria for deterministic simulators has been extensively investigated in the computer experiment literature, however, a parallel formulation of EI for noisy simulators is a relatively unexplored area. For simplicity, I assume that the noise variance and hence the time allocation for every future observation is the same and known beforehand.

TECHNOMETRICS, FEBRUARY 2013, VOL. 55, NO. 1
DOI: 10.1080/00401706.2012.739107

The remainder of this note is organized as follows: Section 1 presents a brief (arguably somewhat philosophical) discussion of the choice of the surrogate model used in Picheny et al. (2013). The main objective of the sequential design proposed in Picheny et al. is discussed in Section 2. I discuss the formulation of the improvement function in Section 3. Simulation results for two test functions are presented in Section 4. Finally, a few concluding remarks are made in Section 5.

1. SURROGATE MODEL FOR NOISY SIMULATORS

Statisticians have always been fitting response surfaces to data with noisy outputs. The choice of surrogate model varies from simple parametric families such as polynomials to complex nonparametric techniques such as splines, kernel smoothing, and wavelets. The area of computer experiments has evolved from the fundamental assumption that the observed process is deterministic (i.e., no replication error), and thus we needed a new idea that is not restricted to the classical approach of modeling the noise. For deterministic process outputs, Sacks et al. (1989) proposed a Gaussian process (GP) model that exploits the dependency among the observations.

It is concerning that we (the computer experimenters) are mostly stuck on GP models for noisy processes, despite the plethora of statistical metamodels that can be used as surrogates. It certainly makes sense to use the valuable methodologies developed thus far in this area, however, not all of them are limited to GP models and can be used with other surrogates. For instance, Chipman, Ranjan, and Wang (in press) used an ensemble of tree models [Bayesian additive regression trees (BART)] for emulating simulator outputs in the EI-based sequential design framework for estimating the global minimum. Perhaps, more emphasis should be given to the justification of the choice of surrogate model.

2. OBJECTIVE OF THE SEQUENTIAL DESIGN

According to Section 2.1 of Picheny et al. (2013), the main objective is to minimize a deterministic simulator that is observed with noise. In Section 4, Picheny et al. develop the EQI criterion—a generalization of the EI criterion in Jones et al. (1998) and Huang et al. (2006)—for noisy simulators where the precision of future observations can be controlled. However, I suspect that the EQI criterion, as defined, appears more likely to minimize a quantity like the upper confidence limit, and not the value (or location) of the global minimum of the underlying deterministic simulator.

A few reasons for my skepticism are as follows: Picheny et al. claim that the proposed EQI approach works better if the noise variance τ^2 is large. The developed EQI criterion is more likely to favor observation repetition or clustering instead of exploration. These features are consistent with the assumption that minimizing the simulator output is not the main focus here, as it would not make much sense to optimize a very noisy function.

It is worth noting that if the objective is to determine the value and/or the location of the global minimum of the deterministic component of a noisy simulator, one might consider using a different improvement criterion for efficient sequential sampling. In Section 3, we propose two alternative improvement functions for this problem. Simulation results from my preliminary investigation on finding a good EI criterion are presented in Section 4.

3. IMPROVEMENT FUNCTION FOR GLOBAL MINIMUM

An improvement function is basically a gain (negative loss) function for a predetermined goal, and the corresponding EI is its expectation with respect to the predictive distribution. The notion of a merit-based criterion for selecting follow-up design points in a computer experiment started with Mockus, Tiesis, and Zilinskas (1978); however, sequential designs based on the improvement function received significant attention after Jones et al. (1998) and Schonlau et al. (1998) developed EI criteria for estimating the global minimum of expensive deterministic simulators via GP emulators. Note that the EI criteria is not limited to GP models [e.g., see Chipman, Ranjan, and Wang (in press)].

Though coming up with a new improvement function is easy, proposing a good one may not be straightforward. Assuming $y(\cdot)$ is the observed deterministic simulator, Jones et al. (1998) proposed the popular improvement function

$$I_1(x) = \max\left\{0,\, f_{\min}^1 - y(x)\right\},$$

for estimating both the global minimum and the minimizer, where $f_{\min}^1 = \min\{y(x_i),\, 1 \le i \le n\}$ is the minimum observed response. If the simulator is noisy, $y(x_i)$ and $f_{\min}^1$ are unobservable, and thus $I_1(x)$ cannot be used. A naive alternative is to minimize the predicted simulator output using the improvement function

$$I_2(x) = \max\left\{0,\, f_{\min}^2 - \tilde{y}(x)\right\},$$

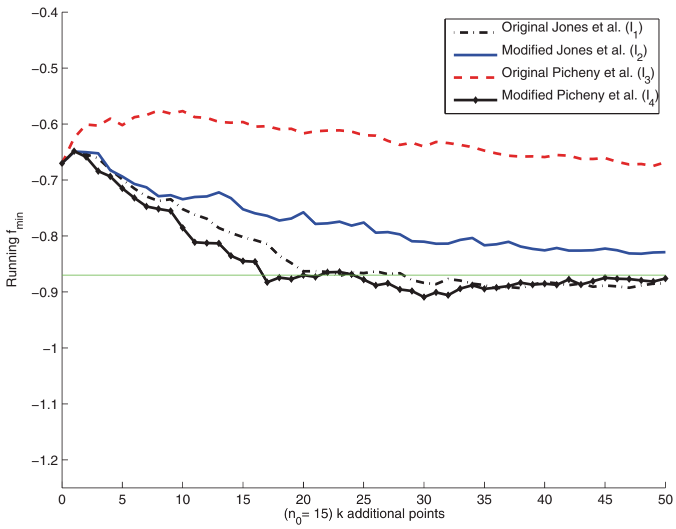
where $f_{\min}^2 = \min\{\hat{y}(x_i),\, 1 \le i \le n\}$ is the minimum predicted response at the sampled design points. The use of predicted response in $f_{\min}^2$ accounts for some noise in the process. Alternatively, the minimum predicted response over the entire input space $f_{\min}^{2*} = \min\{\hat{y}(x),\, x \in [0,1]^d\}$ could be used, which would potentially account for more process uncertainty.

Picheny et al. (2013) generalizes the EI criterion in Jones et al. (1998) by suggesting the explicit use of process uncertainty in the improvement function

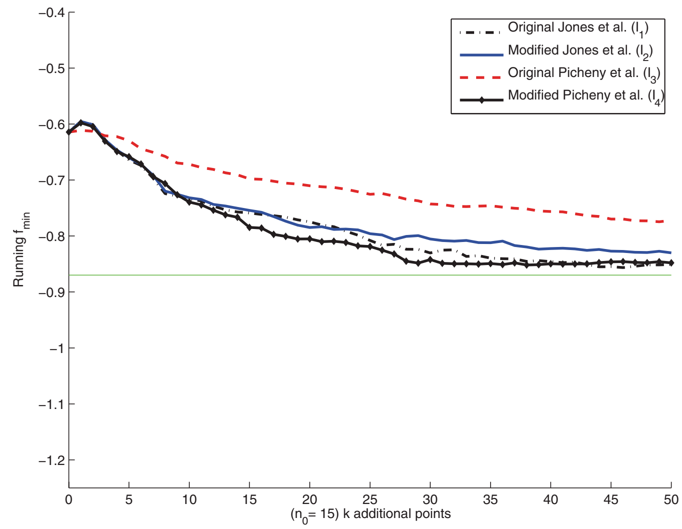
$$I_3(x) = \max\left\{0,\, \min_{1 \le i \le n} Q_n(x_i) - Q_{n+1}(x)\right\},$$

where $Q_k(x) = \hat{y}_k(x) + \Phi^{-1}(\beta)s_k(x)$ denotes the β-quantile of the predicted response based on the surrogate fit obtained using k observed data points. Note that $I_3(x)$ with $\beta = 0.5$ reduces to $I_2(x)$, and hence is a generalization. Furthermore, for deterministic simulators, $\hat{y}(x_i) = y_i$ and $s(x_i) = 0$, that is, $I_3(x)$ reduces to $I_1(x)$.

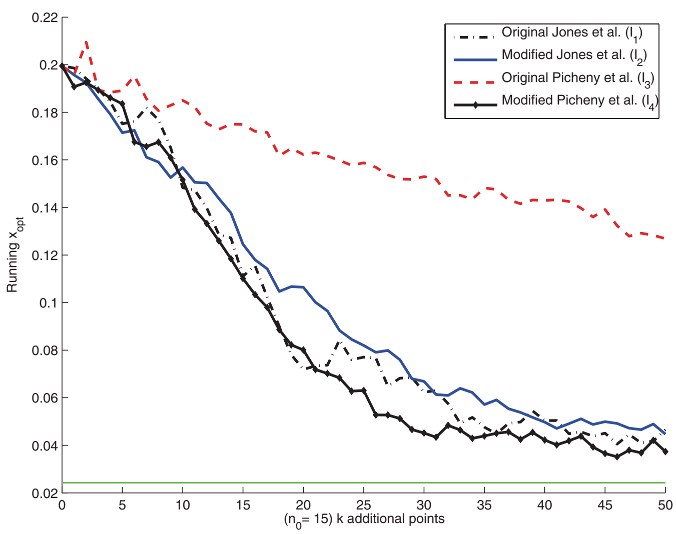
This formulation of the improvement function is certainly intuitive and very clever generalization of the improvement function in Jones et al. (1998). However, the local search component of the proposed EQI criterion, for the recommended values of $\beta \in [0.5, 1)$ (near 1), encourages sampling near minimum $\hat{y}_n(x)$ with small $s_n(x)$. Typically, the local search encourages follow-up design points near minimum $\hat{y}_n(x)$ with large $s_n(x)$. To change the weight of $s_n(x)$, we suggest reversing the quantile from upper to lower. That is, for finding the global minimum



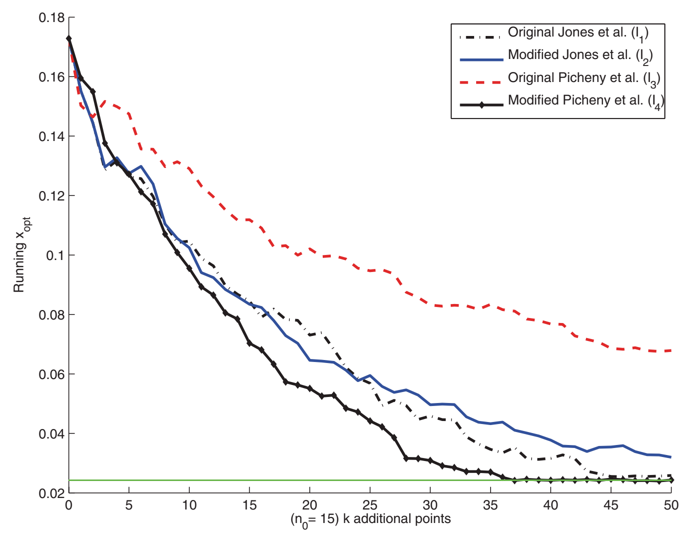
(a) Average global minimum ($\tau = 0.2$)



(b) Average global minimum ($\tau = 0.05$)



(c) Average minimizer ($\tau = 0.2$)



(d) Average minimizer ($\tau = 0.05$)

Figure 1. 1d simulator in Example 1. Average of running global minimum and minimizer with $n_0 = 15$ and $n_{\text{new}} = 50$. The horizontal line shows the target minimum [in (a) and (b)] and minimizer [in (c) and (d)]. The online version of this figure is in color.

of the underlying deterministic simulator, one can consider the following improvement function

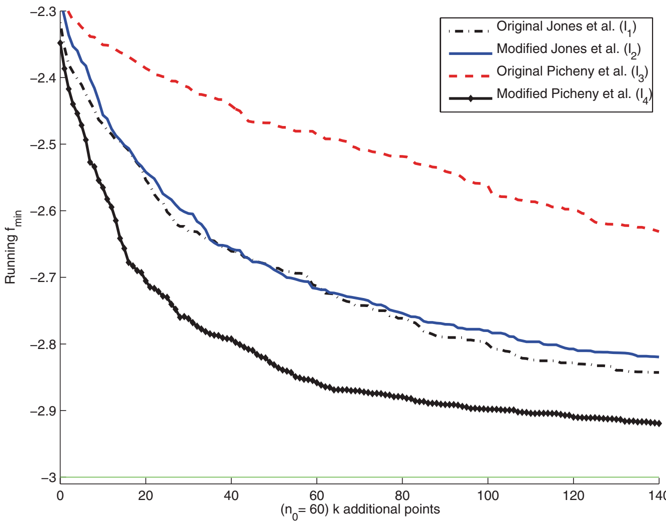
$$I_4(x) = \max\left\{0, \min_{1\le i\le n} Q'_n(x_i) - Q'_{n+1}(x)\right\},$$

where $Q'_k(x) = \hat{y}_k(x) - \Phi^{-1}(\beta)s_k(x)$ represents the $(1-\beta)$-quantile of the predicted response for $\beta \in [0.5, 1)$. Equivalently, one can consider I_3 with β-quantile for $\beta \in (0, 0.5)$ (near 0). Note that I_4 is also a generalization of I_1 and I_2.

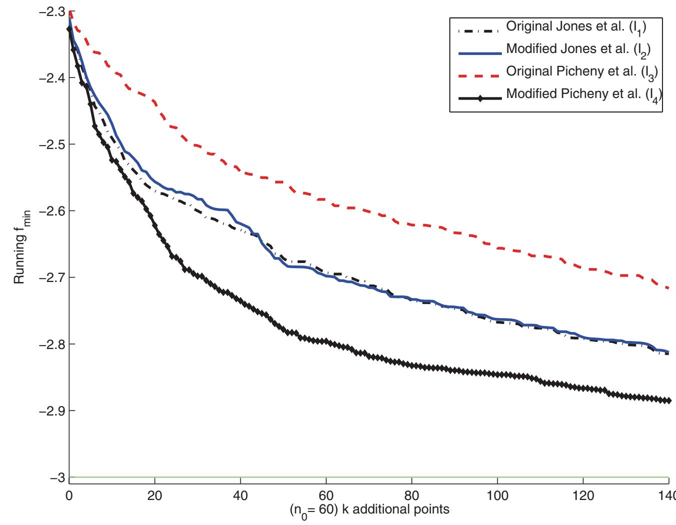
4. SIMULATION RESULTS

This section presents the performance comparison of the proposed EQI criterion in Picheny et al. with I_2- and I_4-based EI criteria in finding the global minimum of the underlying deterministic simulators. For completeness, we also use the I_1-based EI criterion with $y(x)$ replaced by $\tilde{y}(x)$. I used BART [with the same parameter settings as in Chipman, Ranjan, and Wang (in press), except "sigest $= \tau\sigma$"] instead of GP as the predictive machine. Furthermore, I used sample average of $I(x)$ values over the Markov chain Monte Carlo (MCMC) draws of response to approximate the EI. I believe that the trends in the results obtained here would not change significantly if GP was used instead of BART. The EI surface is optimized by first evaluating the EI criterion on a $500d$–random Latin hypercube design (LHD) and then finding the design point with largest EI value.

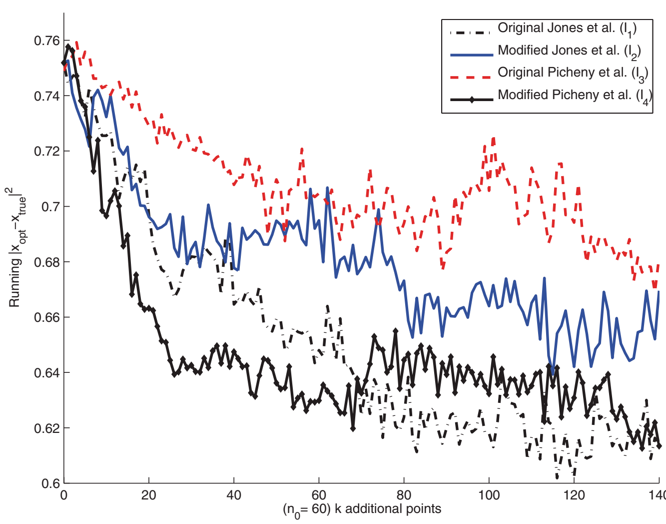
Two test functions with additive independent Gaussian noise are used for generating outputs from noisy simulators, that is, $\tilde{y}(x) = y(x) + \varepsilon$, where $\varepsilon \sim N(0, \tau^2\sigma^2)$. For consistency with Picheny et al., assume $\tau = 0.05$ and 0.2. In each case, I present the average running estimate of the global minimum ($f_{\min} = \min\{\hat{y}(x)\}$) and the corresponding minimizer of the simulator



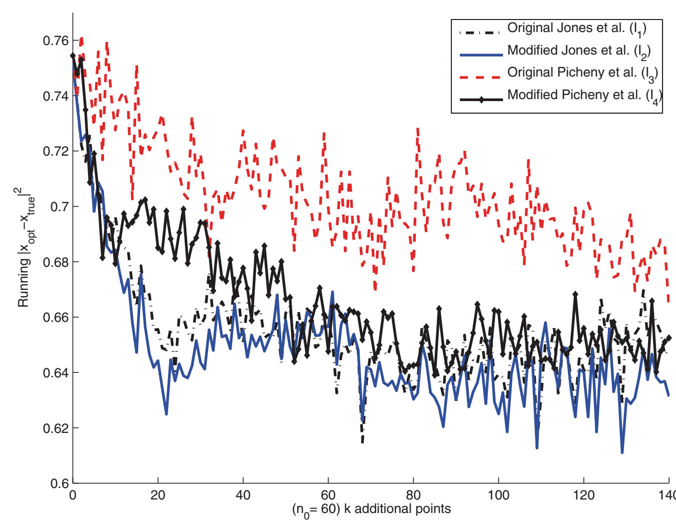
(a) Average global minimum ($\tau = 0.2$)



(b) Average global minimum ($\tau = 0.05$)



(c) Average minimizer ($\tau = 0.2$)



(d) Average minimizer ($\tau = 0.05$)

Figure 2. 6d simulator—Hartman function. Average of running global minimum and minimizer with $n_0 = 60$ and $n_{\text{new}} = 140$. The horizontal lines in (a) and (b) show the target global minimum. The online version of this figure is in color.

outputs based on the prediction set used for EI optimization. The estimated global minimum $f_{\min}$ can sometimes be smaller than the true minimum because $\hat{y}(x)$ can be lower than the true minimum. The results presented here are averaged over 100 simulations.

Example 1. Suppose the underlying deterministic simulator outputs are generated using the simple one-dimensional test function

$$y(x) = \frac{\sin(10\pi x)}{2x} + (x-1)^4, \quad x \in [0.5, 2.5].$$

The inputs are scaled to [0, 1] for implementation, and suppose the process variance is σ^2. Figure 1 presents the average running minimum and minimizer for the sequential design implementation with four improvement functions discussed in Section 3.

Figure 1 shows that I_1, I_2, and I_4 are somewhat comparable; however, I_4 is clearly the best criterion for minimizing the surrogate. The authors' proposed approach using I_3 exhibit significantly slower convergence. For the large noise variance case, minimizing the underlying deterministic simulator output is a challenging problem, and perhaps it is logical to focus more on the minimizer instead of the global minimum. Figure 1(c) shows that I_3-based sequential design would require significantly more follow-up trials than I_2- or I_4-based designs to reach the same accuracy. Figure 1(a) shows that the estimated global minima obtained from I_4 and I_1 can be lower than the true minimum; however, if run long enough, all estimates will converge to the true global minimum.

Example 2. We now consider the six-dimensional Hartman function in Picheny et al. (2013) as the deterministic component of a noisy simulator. Following Picheny et al., we used 60-point random maximin Latin hypercubes for choosing initial designs and added 140 new points one at a time using the improvement criteria discussed in Section 3. Figure 2 shows the average running best estimates of the global minimum and the Euclidean distance between the true and running best estimate of the location of the global minimum.

Figure 2 shows that all four criteria exhibit similar trend as in Example 1. The average $f_{\min}$ curves in Figure 2(a) and (b) indicate the need for more follow-up points to reach the global minimum, however the trend is clear. The I_1- and I_2-based designs are more conservative than I_4, and the proposed EQI method leads to the slowest convergence of the estimated global minimum and the minimizer. For the low noise case in Figure 2(d), I_1-, I_2-, and I_4-based designs show comparable performance in finding the location of the global minimum.

5. CONCLUDING REMARKS

Though my comments may appear to be critical, I really liked the innovative methodologies (especially the strategy of allocating the computational resources in evaluating follow-up trials) developed in Picheny et al. (2013).

The simulations in Section 4 suggest that the EQI criterion is developed for minimizing a quantity like the upper confidence limit and not to find the value (or location) of the global minimum of a deterministic simulator observed with noise. The simulation study presented here is only for a proof of concept, and more stable results can be obtained with extensive simulations. For instance, I used only $500d$-point random LHDs for optimizing EI criteria, which is almost surely insufficient for good optimization of EI in high dimensional input space.

The preliminary results in Section 4 indicate that I_1, I_2, and I_4 lead to faster convergence as compared to I_3 (the criterion proposed by Picheny et al.) for finding the true value and/or location of the global minimum of a computer simulator observed with noise. Interestingly, in Example 2, I_1 and I_4 exhibit faster convergence in high noise ($\tau = 0.2$) case [Figure 2(a) and (c)] as compared with that in low noise ($\tau = 0.05$) case [Figure 2(b) and (d)]. This counterintuitive result could be due to inaccurate optimization of EI criteria.

SUPPLEMENTARY MATERIALS

R codes: The zipped folder consists of Readme.txt and imp1.R to imp4.R. The detailed variable description and information on the saved outputs are provided in Readme.txt. The R codes imp1.R to imp4.R implement I_1- to I_4-based EI criterion, respectively.

ACKNOWLEDGMENTS

The author thanks Derek Bingham for many helpful comments. This work was supported by a Discovery grant from the Natural Sciences and Engineering Research Council of Canada.

ADDITIONAL REFERENCES

Chipman, H., Ranjan, P., and Wang, W. (2012), “Sequential Design for Computer Experiments With a Flexible Bayesian Additive Model,” *The Canadian Journal of Statistics*, 40, 663–678. [25,26]

Huang, P., Chen, D., and Voelkel, J. O. (1998), “Minimum-Aberration Two-level Split-Plot Designs,” *Technometrics*, 40, 314–326. [24]

Mockus, J., Tiesis, V., and Zilinskas, A. (1978), “The Application of Bayesian Methods for Seeking the Extremum,” *in Towards Global Optimisation 2*, eds. L. Dixon and G. Szego, Amsterdam, Elsevier, pp. 117–129. [25]

Picheny, V., Ginsbourger, D., Richet, Y., and Caplin, G. (2013), “Quantile-Based Optimization of Noisy Computer Experiments with Tunable Precision,” *Technometrics*, 55, 2–13. [24,25,27,28]

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989), “Design and Analysis of Computer Experiments,” *Statistical Science*, 4, 409–435. [25]

Schonlau, M., Welch, W., and Jones, D. (1998), “Global Versus Local Search in Constrained Optimization of Computer Models,” in *New Developments and Applications in Experimental Design* (vol. 34), pp. 11–25, Hayward, CA: Institute of Mathematical Statistics Lecture Notes. [25]