This is an interesting paper that addresses some important problems in computer experiments. Our discussion will focus on the first part of the paper, namely, modeling computer experiments with tunable precision.

Since the proposed EQI criterion depends on the estimation of the model parameters, we consider a general setting for the proposed model. The kriging meta models in Section 2.3 of Picheny et al. (2013) can be summarized as

\[ Y_i = y(x_i) + e_i, \]  

where \( y \) is a realization of a stationary Gaussian process with variance \( \sigma_1^2 \), \( e_i \) is \( N(0, \sigma_2^2/t_i) \) distributed representing the intrinsic error in stochastic simulation and the \( e_i \) are mutually independent and also independent of \( y \). In practice \( Y_i \) is observed and the \( t_i \) are known tuning variables (e.g., computational time or iteration number) and \( \sigma_1^2 \) and \( \sigma_2^2 \) are unknown variances. From the work of Bull (2011), if the variance of a Gaussian process model is fixed, the EI algorithm converges to the global optimal value; but the algorithm may not converge if the variance is estimated by the MLE. Therefore, we shall focus on the estimation of \( \sigma_2^2 \), which represents the intrinsic variance of the stochastic simulator. Since neither \( y(x_i) \) nor \( e_i \) is observed directly, there is a serious identifiability issue for estimating \( \sigma_2^2 \) when the sample size is small.
To address the identifiability problem, we need to examine how data are collected. One solution by Ankenman et al. (2010), which holds for stochastic simulation with one fidelity, is to choose replicated design points to identify $\sigma^2_2$. This allows $\sigma^2_1$ and $\sigma^2_2$ to be estimated separately. For the problem under consideration, this is not an efficient approach because the output accuracy can be controlled by the choice of the tuning variable $t$. In the remaining part of the discussion, we will present an efficient sampling plan that allows $\sigma^2_2$ to be estimated accurately. It will have the indirect benefit of improving the parameter estimation in the EQI criterion.

For stochastic simulations with a tunable variable, if the computer output is obtained at time $T$, then for any $T' < T$, the computer output at time $T'$ is available at no or minimal cost. Let $Y(x, t)$ be the “potential output” at point $x$ and time $t$. For given $x$, $Y(x, t)$ is a stochastic process in $t$. We can observe an output sequence $Y(x, t_i)$ over computational times $t_1 < \cdots < t_k$. This sequential sampling does not cost much computational time because the samples come from one realization run of $Y(x, \cdot)$. Thus we can use $Y(x, t_1), \ldots, Y(x, t_n)$ to estimate $\sigma^2_2$ in lieu of independent replications.

Nested space-filling designs (Qian, 2009; Qian, Tang and Wu, 2009; Qian, Ai and Wu, 2009) have been introduced for multi-fidelity deterministic simulations. Such designs need further development for stochastic simulations tuning precision.

In this new data structure, the $Y(x, t_i)$ sequence is from the same realization of $Y(x, \cdot)$. To incorporate this dependence, we propose the following Brownian motion model. First we change (1) to

$$Y(x, t) = y(x) + e_x(t),$$

(2)

where $e_x$ are mutually independent for different $x$ and $e_x(t) \sim N(0, \sigma^2_2/t)$.

For fixed $x$ we suggest a simple covariance structure for $e_x(\cdot)$ with independent increments, i.e., for any $t_0 < t_1 < t_2$, $e_x(t_0) - e_x(t_1)$ is independent of $e_x(t_1) - e_x(t_2)$. Define

$$B(t) = e_x(\sigma^2_2/t).$$

(3)

Note that $B(t) \sim N(0, t)$ and also has independent increments. Assume that $e_x(t)$ is continuous with respect to $t$, which implies that $B(\cdot)$ has continuous sample paths. Clearly, $B(\cdot)$ is a Brownian motion (Durrett, 2010).

We now consider $Y(\cdot, \cdot)$ in (2) as a stochastic process in two arguments. Since $y(\cdot)$ is stationary and $e_x(\cdot)$ are mutually independent for different $x$ values, the covariance structure for $Y(x, t)$ is

$$\text{Cov}(Y(x_1, t_1), Y(x_2, t_2)) = \sigma^2_1 K(x_1, x_2) + \sigma^2_2 I_{(x_1=x_2)} \min(1/t_1, 1/t_2),$$

(4)

where $I_{(x_1=x_2)}$ is the indicator function. Note that (4) gives a nonstationary kriging model, as $\text{Var}(Y(x, t))$ depends on $t$. If a smooth function is used instead of the indicator function in (4), the predictive mean of (2) has the interpolation property for deterministic computer
experiments with multiple fidelities like finite element analysis with different mesh densities (Tuo et al., 2012).

We now give some justification for the Brownian motion model. Suppose in model (I), \( e_i \) is given by an average of \( n \) i.i.d. random variables. Let

\[
E_n = \frac{1}{n} \sum_{i=1}^{n} \epsilon_i
\]

be the simulation error, where the \( \epsilon_i \) are i.i.d. sequence with \( E\epsilon_i = 0 \). Suppose sampling for \( \epsilon_i \) is time-consuming and the total computational time is proportional to \( n \). This is used in Monte Carlo integration as in equation (1) of the paper. Note that the independence assumption and the results below can be generalized by using a dependent sequence of \( \{\epsilon_i\} \) with an ergodic condition which covers broader situations including MCMC.

First consider a simple case with \( \epsilon_i \) following \( N(0, \sigma^2) \). It is easy to show that \((E_1, \ldots, E_n)\) is normally distributed with

\[
Cov(E_i, E_j) = \sigma^2 \min(1/i, 1/j),
\]

for \( 1 \leq i, j \leq n \), which is the second term on the right side of (1).

Even without the normality assumption, Theorem 1 shows that the limit process of \( \{E_n\} \) is a Brownian motion.

**Theorem 1.** Define

\[
W_n(t) = \frac{t}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i.
\]

Suppose \( E\epsilon_i = 0 \) and \( Var(\epsilon_i) = 1 \). Then on any compact subinterval of \((0, +\infty)\), as \( n \to \infty \)

\[
W_n \xrightarrow{d} W;
\]

where \( W \) is a Brownian motion.

**Proof.** Define the partial sum process

\[
W_n'(t) = \frac{1}{\sqrt{n}} \sum_{i=1}^{\lfloor nt \rfloor} \epsilon_i, \quad 0 \leq t \leq 1.
\]

Then for any \( 0 < t_0 \leq t \leq t_1 \),

\[
W_n(t) = \frac{t}{t_0} W_n'(t_0/t).
\]
Donsker’s Theorem (Durrett, 2010) implies that $W_n \xrightarrow{d} W$ as $n \to \infty$. By the continuous mapping theorem (van der Vaart, 1998),

$$W_n(t) \xrightarrow{d} \frac{t}{\sqrt{t_0}} W(t_0/t) \text{ on } [t_0, t_1] \text{ as } n \to \infty.$$ 

The scaling and inversion properties of Brownian motions imply that

$$\frac{t}{\sqrt{t_0}} W(t_0/t) \equiv W(t).$$

This completes the proof.

We now relate Theorem 1 to the Brownian motion model in (2)-(3). For a sample size $C$ that is obtained by the computer within a unit time, suppose the stochastic simulation can be approximated by a numerical integration process, i.e.,

$$e_x(t) \approx \frac{1}{[Ct]} \sum_{i=1}^{[Ct]} \sigma \epsilon_i.$$ 

As before, define $B(t)$ by (3). Then Theorem 1 implies that $\sqrt{C} B(t)$ converges to a Brownian motion as $C \to \infty$. Therefore, if the computational time for each level of fidelity increases, the resulting distribution will tend to a Brownian motion. Since the computational time of a typical stochastic simulation is often long, the true distribution of the output is close to the asymptotic distribution. This supports use of the Brownian motion model.

We now provide a simple example to illustrate the advantage of the proposed sampling plan. Suppose the true function of interest is $y(x) = \exp(-1.4x) \cos(3.5\pi x)$ (Santner et al., 2003, pp. 56-57). We compare two schemes for taking observations from $Y(x, t)$ in (2) with $\sigma^2 = 0.64$. The first scheme takes seven points $x_1, \ldots, x_7$ by drawing $x_1$ from $[0, 1/7]$ and defining $x_i = x_{i-1} + 1/7$ for $i = 2, \ldots, 7$. Assume the computational times $t_1, \ldots, t_7$ for the $x_i$ are 2, 6, 4, 2, 4, 6, 2, respectively. We replicate this scheme 1000 times, where in each time the maximum likelihood estimate $\hat{\sigma}^2$ for $\sigma^2$ is computed. Over the 1000 replications, the sample mean of $\hat{\sigma}^2$ is 1.044 with standard deviation 0.596. According to the proposed sampling plan, for each $x_i$ of the seven points $\{x_1, \ldots, x_7\}$ in the first scheme, values of $Y(x, t)$ are available for any positive integer $t \leq t_i$. The second scheme produces 26 data points, which give 19 ($= 26 - 7$) independent increments in the form $Y(x_i, j) - Y(x_i, j + 1)$. These independent increments give an unbiased estimator of $\sigma^2$ with standard error $\sigma^2 \sqrt{2/19} = 0.208$, which is much smaller than the standard deviation 0.596 of the first scheme.

In summary, we have proposed a sequential sampling plan for stochastic simulation with tunable precision, which allows more data to be collected for the same computational budget. We have also introduced a Brownian motion kriging model, which can be used to fit the data from the proposed sampling plan. Some mathematical justification of the proposed model is given. One remaining question is how to choose the tuning parameter sequence $t_1, \ldots, t_k$ as different fidelities.
Additional References


