

Comparison study of sampling methods for computer experiments using various performance measures

Inyong Cho¹ · Yongbin Lee² · Dongheum Ryu¹ · Dong-Hoon Choi³

Received: 28 October 2015 / Revised: 4 May 2016 / Accepted: 18 May 2016 / Published online: 30 May 2016
© Springer-Verlag Berlin Heidelberg 2016

Abstract This study compares the performance of popular sampling methods for computer experiments using various performance measures to compare them. It is well known that the sample points, in the design space located by a sampling method, determine the quality of the meta-model generated based on expensive computer experiment (or simulation) results obtained at sample (or training) points. Thus, it is very important to locate the sample points using a sampling method suitable for the system of interest to be approximated. However, there is still no clear guideline for selecting an appropriate sampling method for computer experiments. As such, a sampling method, the optimal Latin hypercube design (OLHD), has been popularly used, and quasi-random sequences and the centroidal Voronoi tessellation (CVT) have begun to be noticed recently. Some literature on the CVT asserted that the performance of the CVT was better than that of the LHD, but this assertion seems unfair because those studies only employed space-filling performance measures in favor of the CVT. In this research, we performed the comparison study among the popular sampling methods for computer experiments (CVT, OLHD, and three quasi-random sequences) with employing both space-filling properties and a projective property as performance measures to fairly compare them. We also compared the root mean square error

(RMSE) values of Kriging meta-models generated using the five sampling methods to evaluate their prediction performance. From the comparison results, we provided a guideline for selecting appropriate sampling methods for some systems of interest to be approximated.

Keywords Sampling method · Space-filling property · Projective property · Root mean square error (RMSE) · Optimal Latin hypercube design (OLHD) · Centroidal Voronoi tessellation (CVT)

1 Introduction

Computer experiments (or simulations) have been widely used to circumvent expensive and time-consuming physical experiments. However, an analysis using simulation models is still expensive and time-consuming because of their ever-increasing complexity (Crombecq et al. 2011; Viana et al. 2010; Gorissen et al. 2006). To reduce this computational burden, cheap meta-models, replacing the expensive simulation models, have been widely used for the past three decades (Jones 2001; Wang et al. 2001; Wang 2003; Wang and Simpson 2004; Wang and Shan 2007). A meta-model is generated based on expensive simulation results obtained at sample (or training) points (Sobester et al. 2005). Thus, the quality of the meta-model generated depends on the sample points, and it is very important to locate the sample points using a sampling method suitable for the system of interest to be approximated.

There are two ways of evaluating the performance of a sampling method: (1) without using simulation results and (2) with using simulation results. The representative performance measures, without using simulation results, are the space-filling property and the projective property (Xiong et al. 2009). The

✉ Dong-Hoon Choi
dhchoi@hanyang.ac.kr

¹ Graduate School of Mechanical Convergence Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea

² PIDOTECH Inc., HIT #312, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea

³ School of Mechanical Engineering, Hanyang University, 222 Wangsimni-ro, Seongdong-gu, Seoul, Korea

projective property measures how well sampled points are distributed when projected onto axes, while the space-filling property measure show evenly sampled points are spread in the design space. The representative performance measure with using simulation results is the root mean square error (RMSE) (Chai and Draxler 2014) values of meta-models generated using the sampled points to evaluate the prediction performance. The space-filling property, the projective property, and the RMSE values are employed to evaluate the performance of sampling methods for computer experiments.

Among sampling methods for computer experiments, the Latin hypercube design (LHD) is one of the most popular (Loh 1996), and quasi-random sequences and, recently, centroidal Voronoi tessellation (CVT) have begun to be noticed (Du et al. 1999). LHDs intrinsically have an excellent projective property because of their generation algorithm, but many of them may have an unsatisfactory space-filling property. To overcome this weak point, the optimal Latin hypercube design (OLHD) was proposed. When it selects samples, it considers not only the projective property but also the space-filling property by introducing an optimization criterion such as maximizing the minimum distance between samples (Husslage et al. 2011). Quasi-random sequences (also called “low discrepancy sequences”) consider the projective property only, and Halton, Hammersley (Wong et al. 1997), and Sobol sequences (Sobol 1967) are famous among quasi-random sequences. The CVT, based on a Voronoi tessellation, considers the space-filling property only.

Some comparison studies between the LHD and the CVT asserted that the CVT was found better than the LHD (Romero et al. 2006; Saka et al. 2007). We, however, think that the performance measures they used were inadequate because they only used the performance measures for the space-filling property that are favorable to the CVT. As mentioned in Goel et al. (2008), a sampling method taking a single performance measure into account may lead to small gains in that performance measure at the expense of large deteriorations in other performance measures. Moreover, they did not use the OLHD, but the LHD whose space-filling property is worse than that of the OLHD. Also, we could not find the literature on comparing the performance of popular sampling methods for computer experiments including the CVT, quasi-random sequences, and the OLHD. In this study, we perform the comparison study among the popular sampling methods for computer experiments (the CVT, the OLHD, and three quasi-random sequences) with employing both space-filling properties and a projective property as performance measures to fairly compare them. We employ two existing performance measures for the space-filling property (the coefficient of variance measure and the mesh ratio), and propose a performance measure for the projective property. We also compare the root mean square error (RMSE) values of Kriging

meta-models generated using the five sampling methods to evaluate their prediction performance.

The rest of the paper is organized as follows: Section 2 describes three kinds of performance measures used in this study (the space-filling properties, the projective property, and the prediction performance of the meta-model). In Section 3, we compare the performances of CVT, OLHD, and three quasi-random sequences (Halton, Hammersley, and Sobol sequences) using the performance measures described in Section 2. We summarize the comparison results and provide a guideline for selecting appropriate sampling methods for some systems of interest to be approximated in the final section (Section 4).

2 Performance measures

We employ two types of performance measures usually used for evaluating sampling methods for computer experiments. Also, to assess the ability of sampling methods in generating a meta-model with a good prediction performance, we employ the RMSE of a meta-model generated by each sampling method of interest as a performance measure of a sampling method for generating a meta-model with a good prediction performance.

2.1 Performance measures for space-filling property

We adopt two measures, COV measure (λ) and mesh ratio (γ), for evaluating the space-filling property of sampling methods (Gunzburger and Burkardt 2004). To assess the COV measure (λ), the minimum distance between the point \mathbf{z}_i and any point \mathbf{z}_j other than \mathbf{z}_i in the set of NEXP sample points ($\{\mathbf{z}_i\}_{i=1}^{NEXP}$) is first calculated as

$$\gamma_i = \min_{j \neq i} |\mathbf{z}_i - \mathbf{z}_j|. \quad (1)$$

Then, the COV measure (λ) is calculated by

$$\begin{aligned} \lambda &= \frac{1}{\bar{\gamma}} \left(\frac{1}{NEXP} \sum_{i=1}^{NEXP} (\gamma_i - \bar{\gamma})^2 \right)^{1/2} \\ &= \left(-1 + \frac{1}{\bar{\gamma} NEXP} \sum_{i=1}^{NEXP} \gamma_i^2 \right)^{1/2} \\ &= \left(NEXP \frac{\sum_{i=1}^{NEXP} \gamma_i^2}{\left(\sum_{i=1}^{NEXP} \gamma_i \right)^2} - 1 \right)^{1/2}, \end{aligned} \quad (2)$$

where

$$\bar{\gamma} = \frac{1}{NEXP} \sum_{i=1}^{NEXP} \gamma_i \tag{3}$$

The smaller the λ is, the better space-filling property the set of the sample points has. For a perfectly uniform case, $\gamma_1 = \gamma_2 = \dots = \gamma_{NEXP} = \bar{\gamma}$ so that $\lambda=0$.

The mesh ratio (γ) is defined as

$$\gamma = \frac{\max_{i=1, \dots, NEXP} \gamma_i}{\min_{i=1, \dots, NEXP} \gamma_i} \tag{4}$$

The closer to unity the γ is, the better space-filling property the set of the sample points has. For a perfectly uniform case, $\gamma_1 = \gamma_2 = \dots = \gamma_{NEXP}$ so that $\gamma = 1$.

2.2 Performance measure for projective property

We also evaluate the projective property measuring how well the points project onto design variable axes. The star discrepancy (Clerck 1986) is a typical measure to assess the projective property. It, however, requires a heavy computational burden, and thus we propose an alternative measure for evaluating the projective property, denoted as *P.P.*, as

$$P.P = \sum_{i=1}^{NDV} K_i, \tag{5}$$

where *NDV* denotes the number of design variables and

$$K_i = \sum_{j=1}^{NEXP-1} (I_i - G_{ij})^2. \tag{6}$$

Denoting the lower and upper bounds of the *i*th design variable as *XL_i* and *XU_i*, respectively, *I_i* in the above equation is defined as

$$I_i = \frac{XU_i - XL_i}{NEXP - 1}, \tag{7}$$

which is the projected distance onto the *i*th design variable axis between adjacent sample points when a set of sample points has the perfect projective property. In Eq. (5), *G_{ij}* is the projected distance between the *j*th sample point and the (*j* + 1) th sample point when the projected sample points (*j* = 1, 2, ..., *NEXP*) onto the *i*th design variable axis when the projected sample points are numbered in ascending order. Thus, *K_i* is a measure of deviation of the projective property onto the *i*th design variable axis from the perfect one, and the *P.P* in Eq. (4) is an aggregate measure of the projective property onto all design variable axes.

2.3 Performance measure for generating a meta-model with a good prediction performance

One of important purposes of applying sampling methods in the field of design and analysis of computer experiments

(DACE) is to build a meta-model for exploring the design space or finding an approximate optimum. In this study, to assess the ability of sampling methods in generating a meta-model with a good prediction performance, we employ the RMSE of a meta-model generated by each sampling method of interest, mathematically expressed as

$$RMSE = \sqrt{\frac{1}{NTEST} \sum_{i=1}^{NTEST} [y(x_i) - \bar{\gamma}(x_i)]^2}, \tag{8}$$

where *y(x_i)* is the real function value at *x_i*, $\bar{\gamma}(x_i)$ is the approximate function value at *x_i*, and *NTEST* is the number of test points. The meta-model adopted in this study is a Kriging model, one of the most popular models. The Kriging model was mathematically developed and established by Matheron (1963) based on Krige’s research (1951). Then, the Kriging model was applied to various fields of engineering since Sacks et al. (1989). In 1998, Simpson confirmed that the Kriging model had good prediction performance in systems with many design variables and high nonlinearity through comparative studies using several meta-models (Simpson et al. 1998). For RMSE evaluation, we employ 50**NDV* test points different from the sample points for building Kriging models. If the sample points are used as the test points for RMSE evaluation, the RMSE value of interpolation models such as Kriging models will be always zero, which is meaningless. The test points are generated using OLHD, and the same test points are used for RMSE evaluation of five sampling methods for each test problem.

To represent a variety of responses in optimization problems, we adopt nine mathematical functions whose types are valley-shaped, having many local minima, plate-shaped with steep ridges/drops, and bowl-shaped as listed in Table 1. Five of nine functions are scalable for which the number of design variables is set to 2, 4, 6, 8, and 10. Altogether, we use 29 test problems.

Table 1 Nine mathematical test functions

ID	Function name	Function type	Dimension
1	Branin function	Valley shape	2
2	Six-hump camelback function	Many local minima	2
3	Haupt function	Many local minima	2
4	Waving function	Many local minima	2
5	Rosenbrock function	Valley shape	2, 4, 6, 8, 10
6	Griewank function	Many local minima	2, 4, 6, 8, 10
7	Rastrigin function	Many local minima	2, 4, 6, 8, 10
8	Trid function	Bowl type	2, 4, 6, 8, 10
9	Styblinski-Tang function	Plate shaped & steep ridges/drops	2, 4, 6, 8, 10

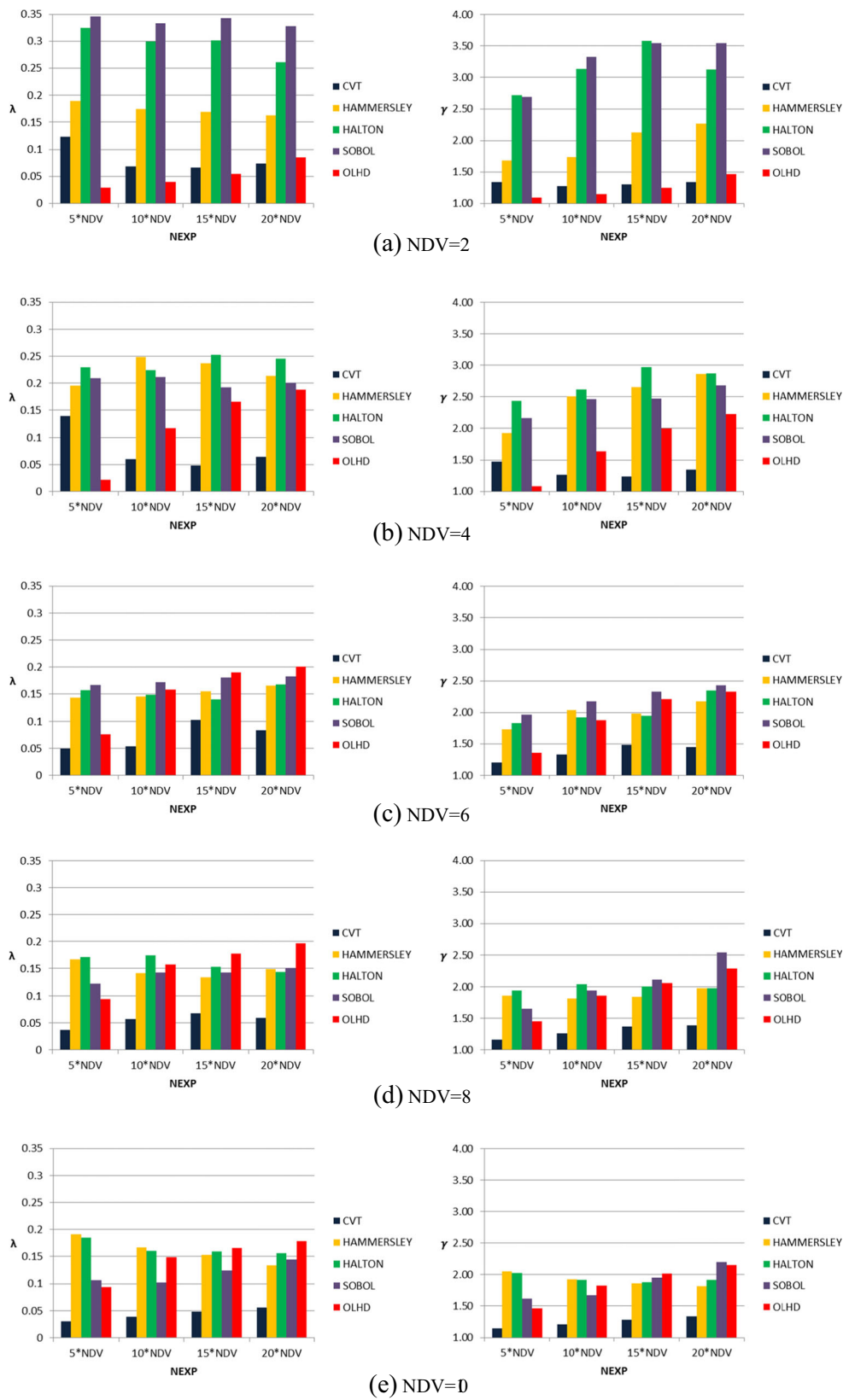


Fig. 1 The space-filling properties (λ , γ) of five sampling methods (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV

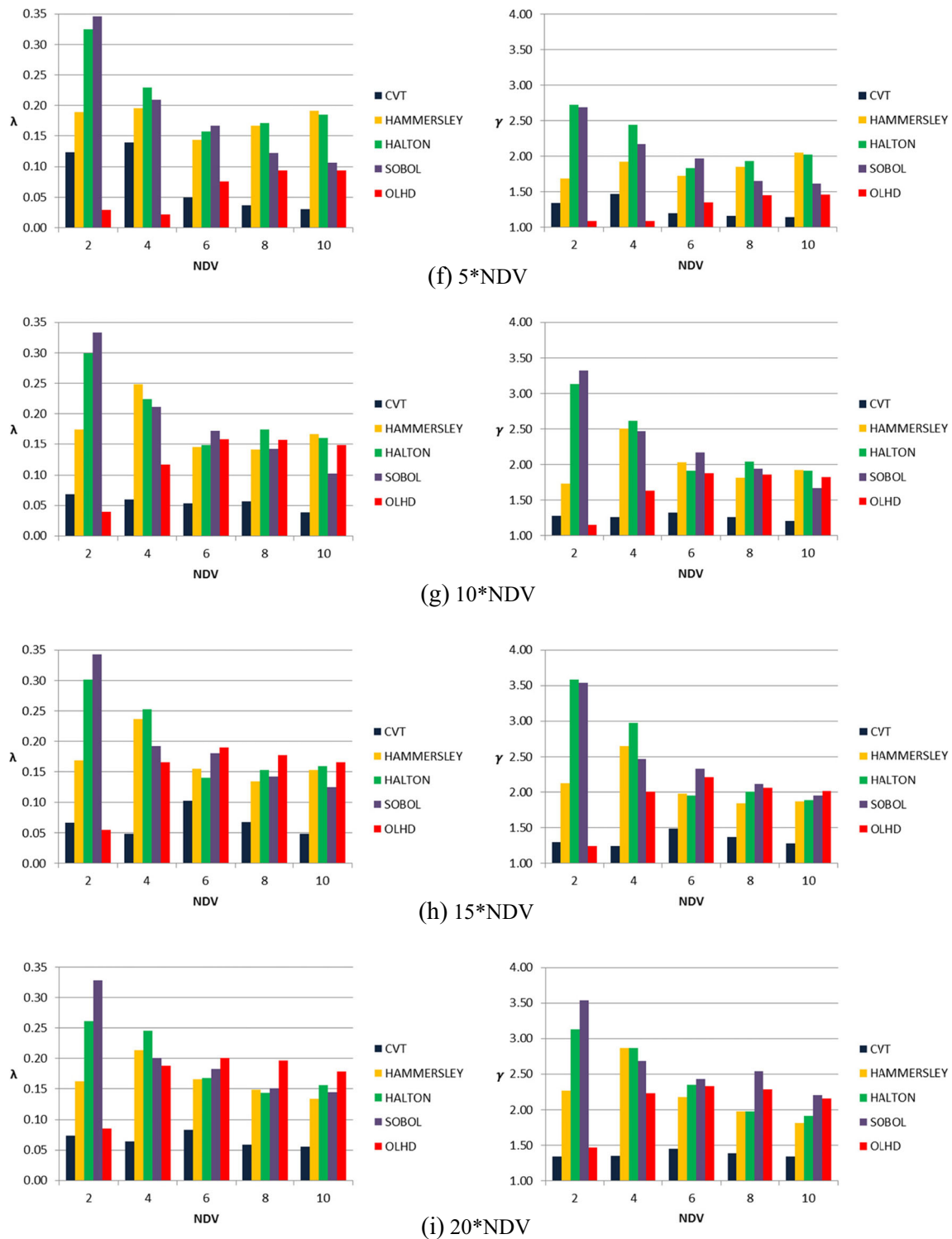


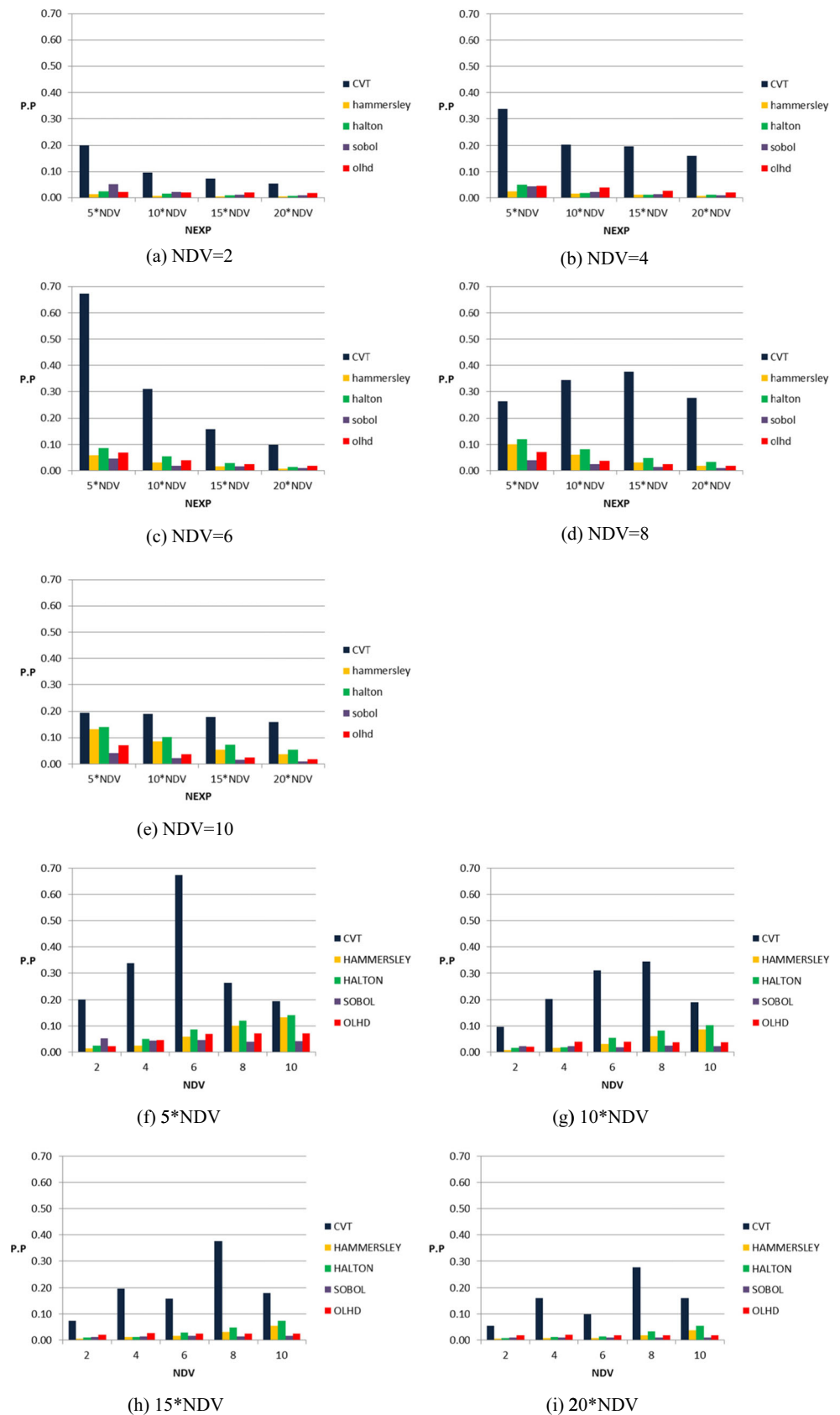
Fig. 1 (continued)

3 Comparison results

In this study, we compare performances of five sampling methods for computer experiment: CVT, OLHD, and three quasi-random sequences (Halton, Hammersley, and Sobol). The optimization criterion of the OLHD in this study is

maximizing the minimum distance between samples. To assess the performance measures for space-filling and projective properties described in Sections 2.1 and 2.2, respectively, the number of design variables (NDV) is varied as 2, 4, 6, 8, and 10 to see the effect of NDV on the performance measures, and the number of experimental points or sampling points (NEXP) is

Fig. 2 The projective property (*P.P.*) of five sampling methods (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV



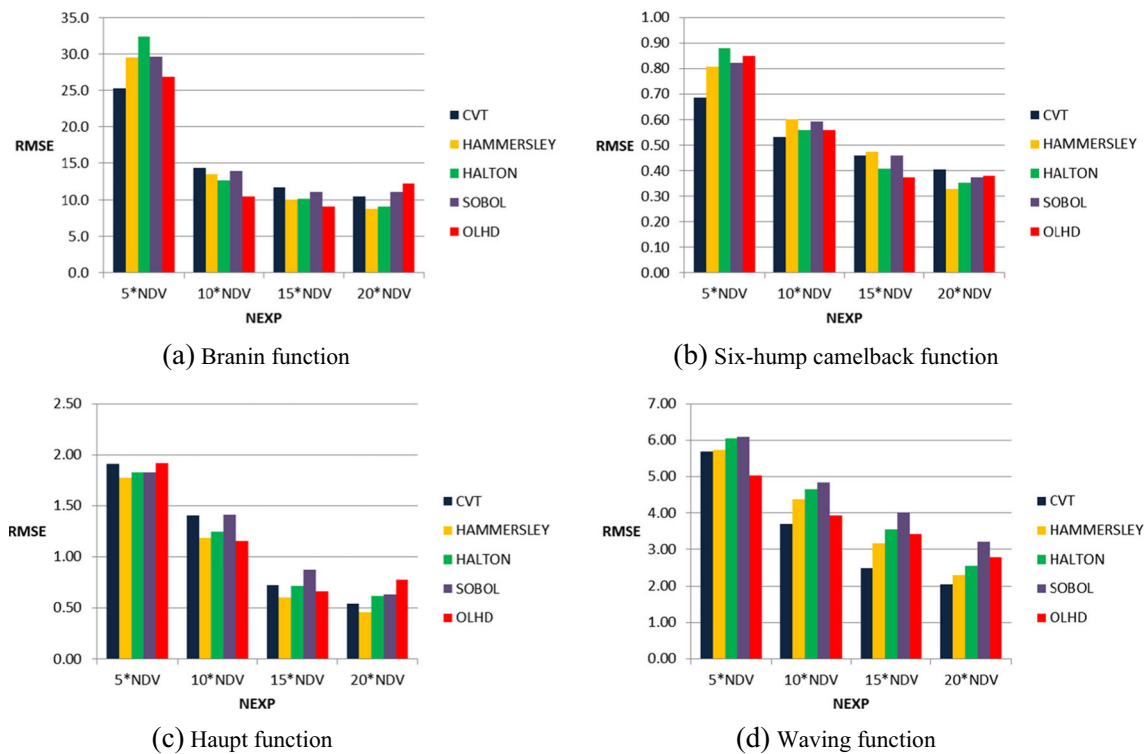


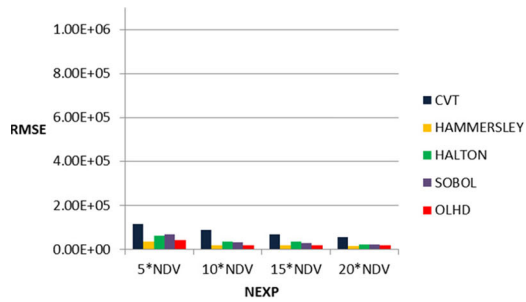
Fig. 3 The RMSE values of the four 2-D test functions (a) Branin function (b) Six-hump camelback function (c) Haupt function (d) Waving function

varied as 5*NDV, 10*NDV, 15*NDV, and 20*NDV to see the effect of NEXP on the performance measures. Thus, we have 20 test cases of sampling points for each sampling method. Also, for each test case, we generate 50 replications of sampling points, evaluate 50 corresponding performance measures, and use the average of 50 measure values as the performance measure value of each test case in order to mitigate the non-repeatability of the five sampling methods. Altogether, we generate 5,000 kinds of sampling points (5 kinds of NDVx4 kinds of NEXPx50 replicationsx5 kinds of sampling methods).

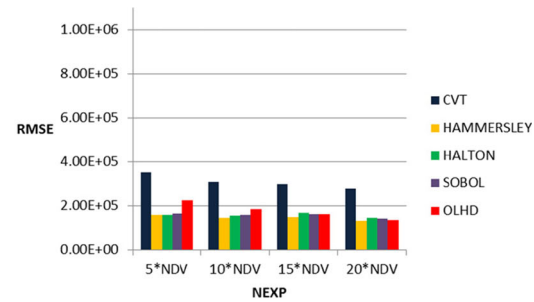
Figure 1a through e show space-filling properties, λ and γ , of five sampling methods with varying NEXP for different NDVs (2, 4, 6, 8, and 10), and Fig. 1f through i with varying NDV for different NEXPs (5*NDV, 10*NDV, 15*NDV, and 20*NDV). First of all, λ and γ show similar trends, which means that either of the two can be used as a measure for the space-filling property, as expected. Secondly, the space-filling property of OLHD is better than that of CVT when the total number of sampling points (NDV x NEXP) is small, but the former becomes worse than the latter when the total number of sampling points reaches around 30 and the difference in the space-filling performance gets larger as the total number of sampling points increase. The reason seems due to the difference of the generation mechanism between OLHD and CVT: OLHD primarily satisfies the projective property and then enhance the space-filling property by optimizing an optimality criterion while CVT mainly concerns the space-filling property. Furthermore, it becomes more difficult for OLHD to

improve the space-filling property without deteriorating its projective property as the total number of sampling points gets larger. Thirdly, the space-filling properties of quasi-random sequences, which only consider the projective property (or puts priority on the projective property) when they select points, are generally worse than those of CVT and OLHD.

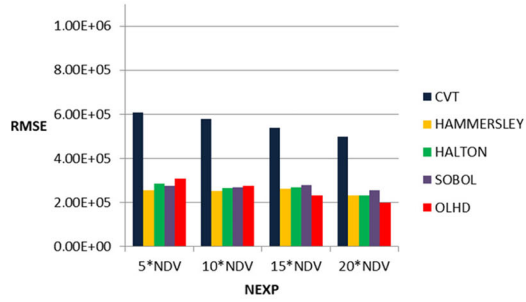
Figure 2a through e show the projective properties, $P.P.s$, of five sampling methods with varying NEXP for different NDVs (2, 4, 6, 8, and 10), and Fig. 2f through i with varying NDV for different NEXPs (5*NDV, 10*NDV, 15*NDV, and 20*NDV). First of all, the projective property of CVT is worst in all cases, which seems due to the difference in generation mechanisms of sampling methods. CVT only considers the space-filling property, and OLHD and quasi-random sequences primarily satisfy the projective property. Therefore, the projective property of CVT is always worse than those of OLHD and quasi-random sequences. Secondly, in general, the projective properties of OLHD and quasi-random sequences are comparable. Third, we can see that the projective property of Halton and Hammersley become worse than that of Sobol when NDV becomes large (larger than six). It is because Halton and Hammersley lose their structures in a high dimension (Krykova 2003). On the other hand, Sobol keeps its structure in a high dimension because of its *additional uniformity conditions*, known as *property A and property A'* (Sobol et al. 2011). Therefore, it does not lose its structure even in a high dimension. Thus, its projective property is best among all the other sampling methods when NDV gets large.



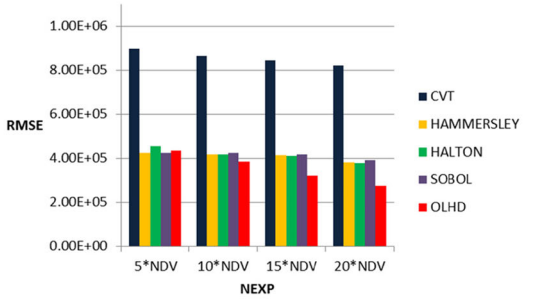
(a) NDV=2



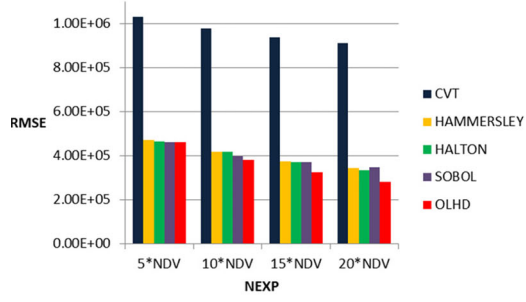
(b) NDV=4



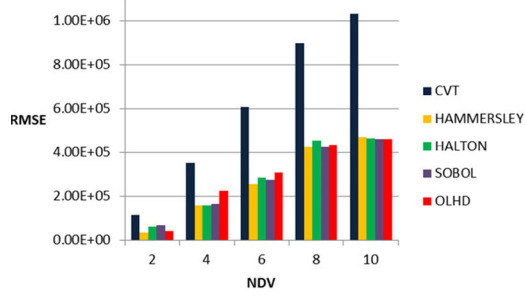
(c) NDV=6



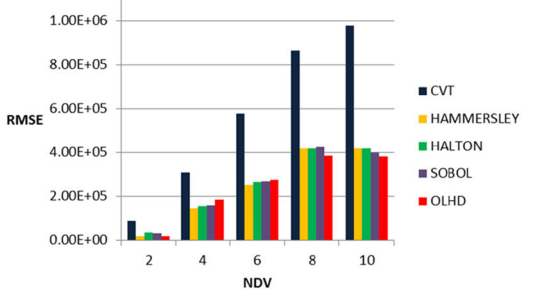
(d) NDV=8



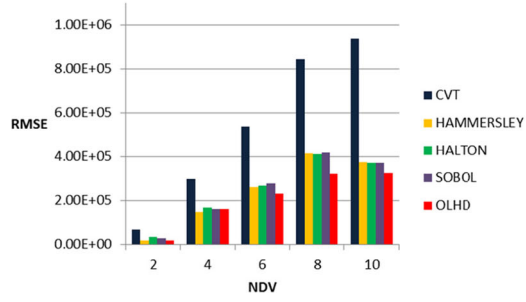
(e) NDV=10



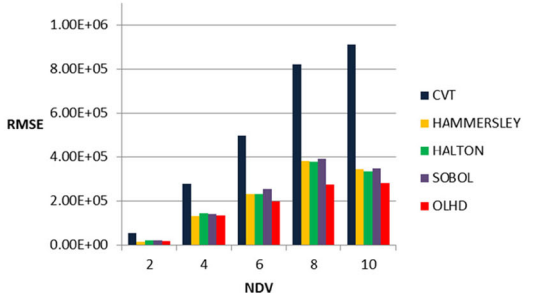
(f) 5*NDV



(g) 10*NDV



(h) 15*NDV

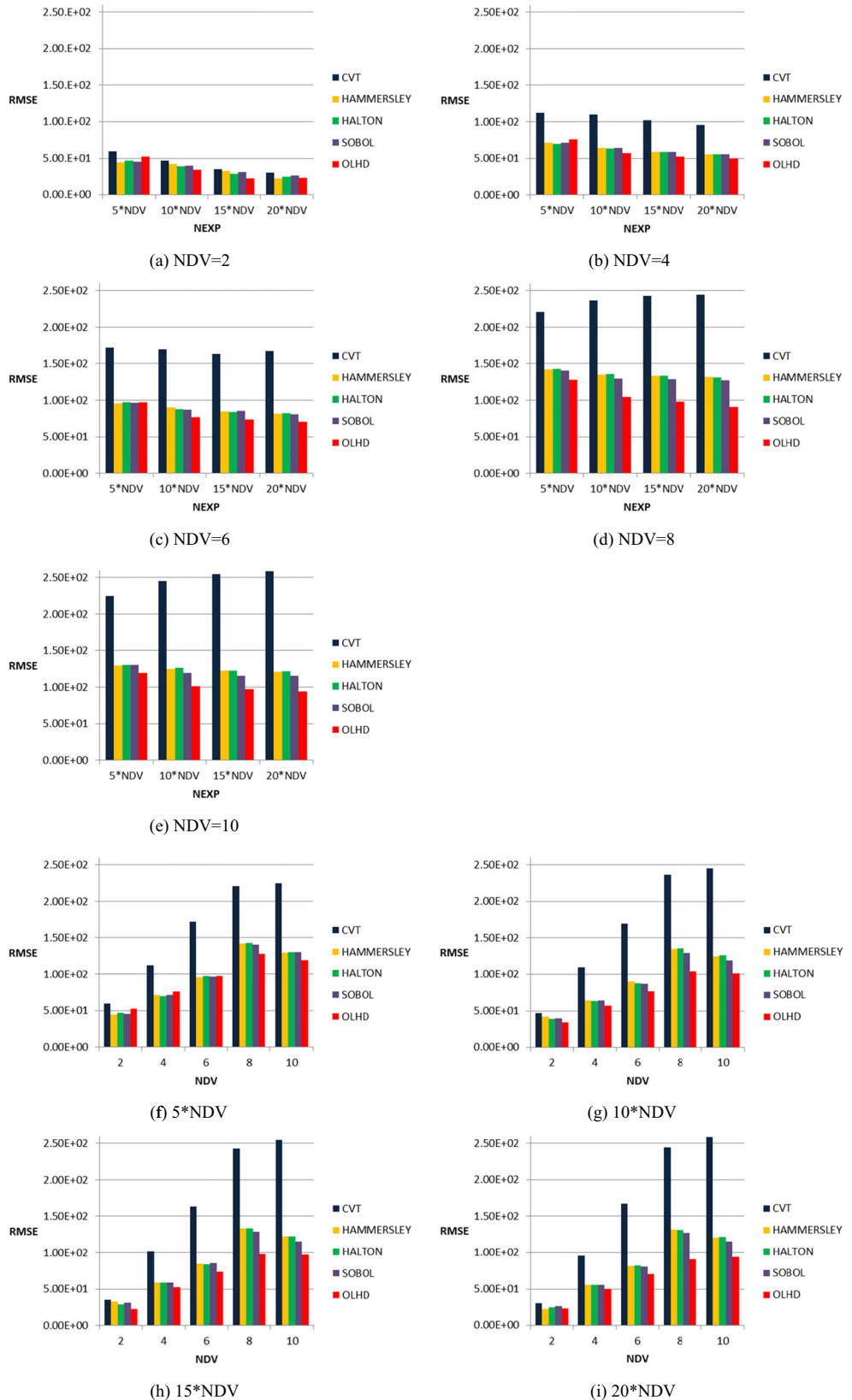


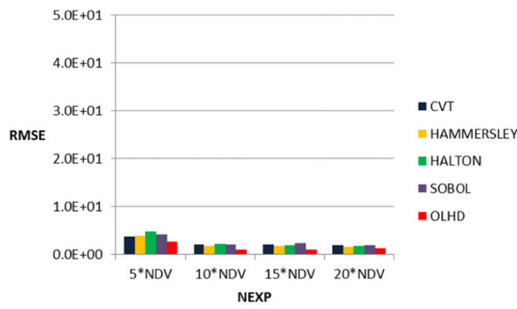
(i) 20*NDV

◀ **Fig. 4** The RMSE values of the Rosenbrock function (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV

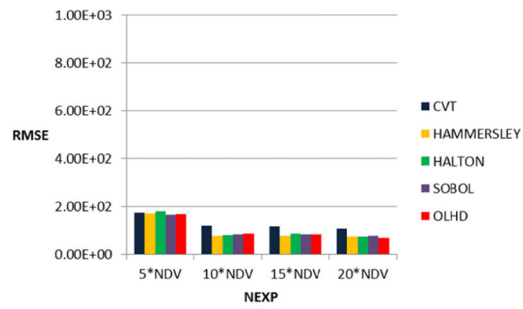
Figure 3 shows RMSE values of five sampling methods applied to the four 2-D test functions with varying NEXP. For all four test problems, the performance comparison among

Fig. 5 The RMSE values of the Styblinski-Tang function (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV

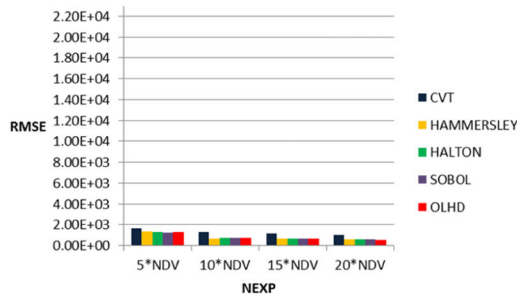




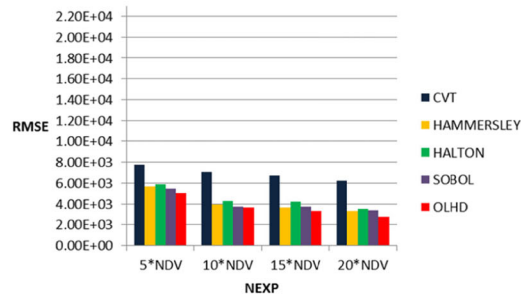
(a) NDV=2



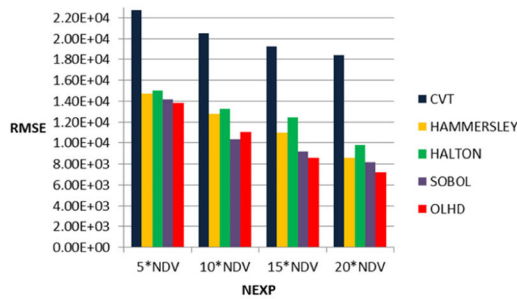
(b) NDV=4



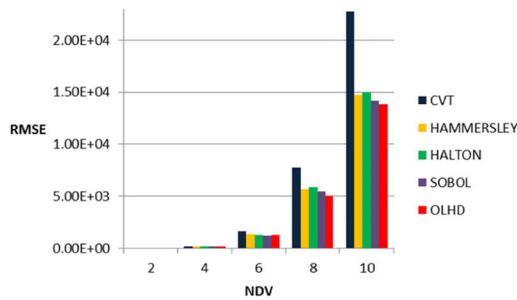
(c) NDV=6



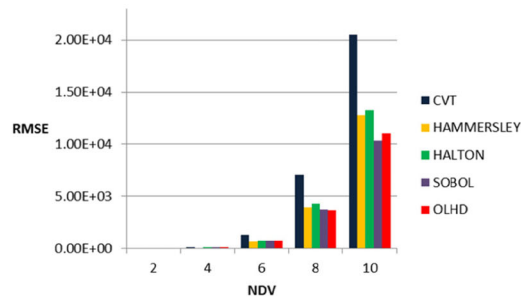
(d) NDV=8



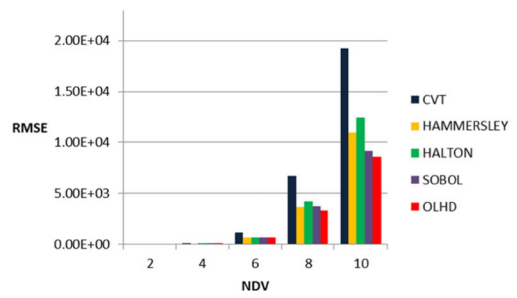
(e) NDV=10



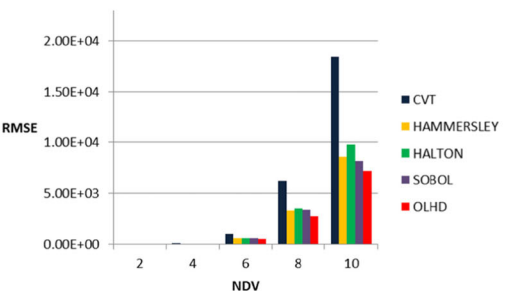
(f) 5*NDV



(g) 10*NDV



(h) 15*NDV



(i) 20*NDV

◀ **Fig. 6** The RMSE values of the Trid function (a) NDV=2 (b) NDV=4 (c) NDV=6 (d) NDV=8 (e) NDV=10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV

sampling methods does not show any noteworthy tendency, but confirms an expected result that the prediction performances of meta-models are generally improved when NEXP increases.

Figures 4, 5, and 6 show RMSE values of five sampling methods for the Rosenbrock function (valley-shaped), the Styblinski-Tang function (plate-shaped with steep ridges/drops), and the Trid function (bowl-shaped), respectively. In each figure, Figures (a) through (e) show the RMSE values of five sampling methods with varying NEXP for different NDVs (2, 4, 6, 8, and 10), and Figures (f) through (i) with varying NDV for different NEXPs (5*NDV, 10*NDV, 15*NDV, and 20*NDV). In all cases for all the three test problems, CVT is found to be far inferior to the other sampling methods while the other four sampling method show similar performance in general. To investigate the reason of the apparent inferiority of CVT with noting that function values are dramatically changing as approached to the boundary in the case of all the three test functions, we counted the number of sample points located in the outer 10 % region along each design variable. Figure 7 shows a 2-D example. The average number of sample points in the boundary region in 50 replications of the five sampling methods for different NDVs are listed in Table 2. As clearly shown in Table 2, CVT has no sample points in the boundary region (except for the case of NDV=2 with NEXP=40 which has almost no sample points in the boundary region). Thus, the Kriging model built using CVT sample points cannot well represent the rapid change of the function values in the boundary region, resulting in a far inferior performance of CVT. Based on this investigation, we recommend not to use CVT emphasizing just space-filling, but to use one of the other sampling methods emphasizing projective properties for functions with a rapid local change in value.

Figure 8a through e show the RMSE values of five sampling methods for the Griewank function with varying NEXP for different NDVs (2, 4, 6, 8, and 10), and Fig. 8f through i with varying NDV for different NEXPs (5*NDV, 10*NDV, 15*NDV, and 20*NDV). The Griewank function

has many local minima with equidistant intervals. From Fig. 8a through e, we can observe that CVT performs best for NDV=2. However, as NDV increases all five sampling methods show similar performances. The reason seems to be that the role of projective properties becomes important for producing a good Kriging model even though the sampling considering space-filling only can produce a good model when NDV is very small. From Fig. 8f through i, RMSE values are observed to clearly decrease as NDV increases. This behavior does not agree with our expectation of increased RMSE values with increased NDVs for fixed NEXPs because of increased complexity, and is opposite to those of all other test problems in this study. This counterintuitive behavior is due to the peculiar characteristics of the Griewank function. The Griewank function is a quadratic convex function superimposed by an oscillatory nonconvex function giving rise to the large number of local minima, and the region influenced by the oscillatory nonconvex function becomes narrower and narrower as NDV increases (Locatelli 2003). This means that, as the NDV increases, more and more region is represented by the quadratic convex function and thus the RMSE values become smaller and smaller.

Figure 9a through e show the RMSE values of five sampling methods for the Rastrigin function with varying NEXP for different NDVs (2, 4, 6, 8, and 10), and Fig. 9f through i with varying NDV for different NEXPs (5*NDV, 10*NDV, 15*NDV, and 20*NDV). The Rastrigin function has many local minima with equidistant intervals as the Griewank function does. Thus, the same behaviors can be observed as those of the Griewank function except for the counterintuitive behavior of decreased RMSE values of the Griewank function with increased NDVs. The RMSE values of the Rastrigin function increase with increased NDVs as expected. In general, we do not recommend to use of CVT but to use OLHD and quasi-random sequences for applications with many local minima.

4 Summary

In this study, we compared the popular sampling methods for computer experiments (the CVT, the OLHD, and three quasi-random sequences) by employing both space-filling

Fig. 7 The boundary region when the number of design variables is two

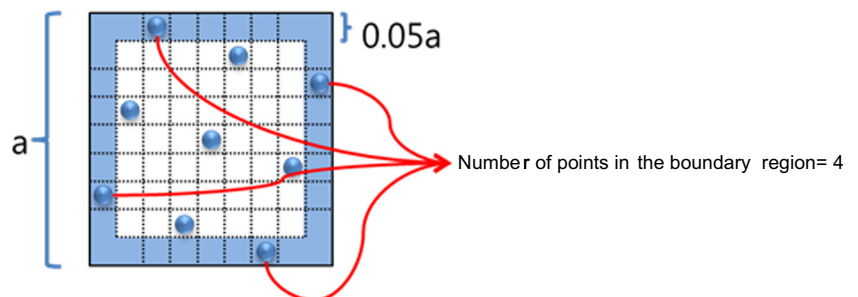


Table 2 The average number of sample points in the boundary region of the five sampling methods for different NDVs

NEXP	CVT	OLHD	HAMMERSLEY	HALTON	SOBOL
(a) NDV = 2					
10	0	3.96	1.76	1.66	2.08
20	0	4.86	4.78	3.76	3.96
30	0	7.02	5.72	5.48	5.94
40	0.02	7.70	8.28	7.22	7.84
(b) NDV = 4					
20	0	7.14	4.54	3.64	4.28
40	0	11.36	8.54	7.4	8.24
60	0	16.48	12.24	10.88	12
80	0	20.84	16.16	14.84	15.28
(c) NDV = 6					
30	0	12.84	5.8	5.48	5.94
60	0	19.94	12.32	11.62	11.62
90	0	28.06	17.16	16.62	16.36
120	0	37.38	23.76	22.66	21.44
(d) NDV = 8					
40	0	18.06	8.44	7.22	7.98
80	0	30.62	16.26	14.76	14.78
120	0	44.20	23.64	22.92	21.74
160	0	58.22	31.2	30.3	28.8
(e) NDV = 10					
50	0	25.22	9.48	9.44	10.16
100	0	42.08	20.02	18.84	18.02
150	0	63.52	28.42	27.92	27.08
200	0	84.32	39.14	38.3	36

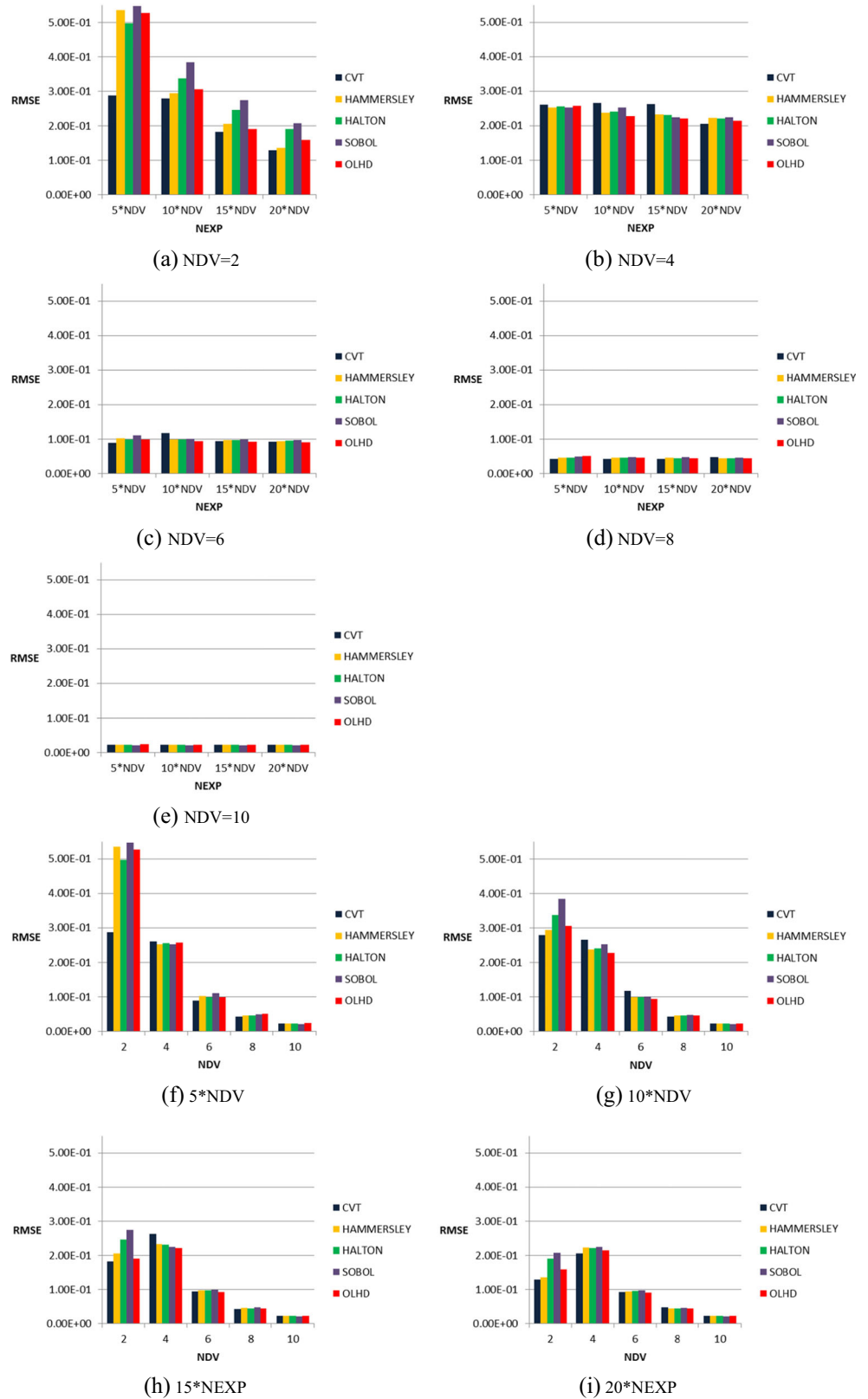
properties and a projective property as performance measures. We also compared the RMSE values of Kriging meta-models generated using the five sampling methods to evaluate their prediction performance. As test functions, we adopted nine mathematical functions whose types were valley-shaped, having many local minima, plate-shaped with steep ridges/drops, and bowl-shaped in order to represent a variety of responses in optimization problems. Five of nine functions were scalable for which the number of design variables was set to 2, 4, 6, 8, and 10. Altogether, we used 29 test problems. The results of our comparison study are summarized below.

1. The space-filling property of CVT was better than that of OLHD when the total number of sampling points reaches around 30 and the superiority of CVT over OLHD in the space-filling performance got higher as the total number of sampling points increased. The reason seemed due to the difference of generation mechanism between OLHD and CVT: OLHD primarily satisfied the projective property and then enhanced the space-filling property by

optimizing an optimality criterion while CVT focused on improving the space-filling property. The space-filling properties of quasi-random sequences, which only considered the projective property (or put priority on the projective property) when they selected points, were generally worse than those of CVT and OLHD.

2. The projective property of CVT was worst in all cases because CVT did not consider the projective property but only considered the space-filling property. The projective properties of OLHD and quasi-random sequences were generally comparable. Among quasi-random sequences, the projective property of Halton and Hammersley became worse than that of Sobol when the number of design variables (NDV) got large (larger than six) because Halton and Hammersley lost their structures in a high dimension. On the other hand, Sobol kept its structure in a high dimension because of its additional uniformity conditions. Thus, the projective property of Sobol was best among all the other sampling methods when NDV got large.
3. The RMSE values of the five sampling methods applied to the four 2-D test functions did not show any noteworthy tendency among the sampling methods. As expected, however, the prediction performances of Kriging models generated by all five sampling methods were generally improved for all four test functions as the number of experimental points (NEXP) increased.
4. Comparison of the RMSE values of five sampling methods for the Rosenbrock function (valley-shaped), the Styblinski-Tang function (plate-shaped with steep ridges/drops), and the Trid function (bowl-shaped) reveals that the prediction performance of CVT is found to be far inferior to those of the other sampling methods while the other four sampling method show similar performance in general. The reason was found due to the fact that CVT had far less average number of sample points in the boundary region compared to those of the other four sampling methods, and thus could not well represent the rapid local change of the three functions. For functions with a rapid local change in value, we recommend against use of CVT emphasizing just space-filling, rather to use one of the other sampling methods emphasizing projective properties.
5. Comparison of the RMSE values of five sampling methods for functions having many local minima with equidistant intervals such as the Griewank function and the Rastrigin function revealed that the prediction performance of CVT became worse than or similar to those of the other four sampling methods as NDV increased. The reason seemed that the role of projective properties

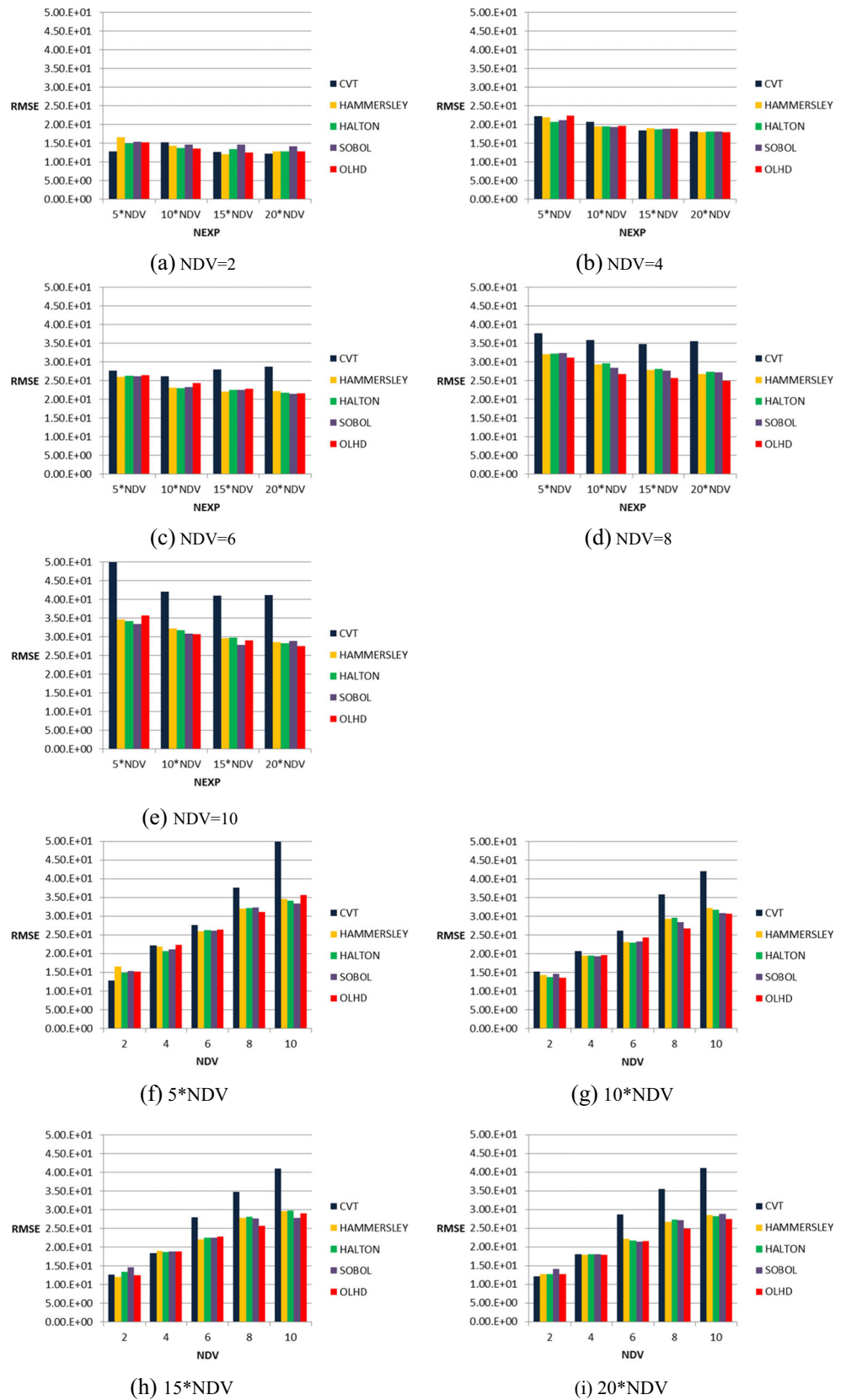
Fig. 8 The RMSE values of the Griewank function (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NEXP (i) 20*NEXP



became important for producing a good Kriging model even though the sampling considering space-filling only could produce a good model when NDV was very

small. In general, we do not recommend to use of CVT, but to use OLHD and quasi-random sequences for functions with many local minima.

Fig. 9 The RMSE values of the Rastrigin function (a) NDV = 2 (b) NDV = 4 (c) NDV = 6 (d) NDV = 8 (e) NDV = 10 (f) 5*NDV (g) 10*NDV (h) 15*NDV (i) 20*NDV



References

- Chai T, Draxler RR (2014) Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RSME in the literature. *Geosci Model Dev* 7:1247–1250
- Clerck LD (1986) A method for exact calculation of the Star discrepancy of plane sets applied to the sequences of Hammersley. *Mh Math* 101:261–278
- Crombecq K, Laermans E, Dhaene T (2011) Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. *Eur J Oper Res* 214:683–696
- Du Q, Faber V, Gunzburger M (1999) Centroidal Voronoi tessellations: applications and algorithms. *SIAM Rev* 41(4):637–676
- Goel T, Haftka RT, Shyy W, Watson LT (2008) Pitfalls of using a single criterion for selecting experimental designs. *Int J Numer Methods Eng* 75:127–155
- Grississen D, Crombecq K, Hendrickx W, Dhaene T (2006) Adaptive distributed metamodeling. *High Perform Comput Computational Sci-VECPAR* 4395:579–588
- Gunzburger MD, Burkardt JV (2004) Uniformity measures for point samples in hypercubes. <https://people.scs.fsu.edu/~burkardt/pdf/ptmeas.pdf>
- Husslage B, Rennen G, van Dam E, den Hertog D (2011) Space-filling Latin hypercube designs for computer experiments. *Optim Eng* 12(4):611–630
- Jones DR (2001) A taxonomy of global optimization methods based on response surfaces. *J Glob Optim* 21:345–383
- Krige DG (1951) A statistical approach to some basic mine valuation problems on the Witwatersrand. *J Chem Metal Mining Soc South Africa* 52(6):119–139
- Krykova I (2003) Evaluating of path-dependent securities with low discrepancy methods. Worcester polytechnic institute.
- Locatelli M (2003) A note on the Griewank test function. *J Glob Optim* 25:169–174
- Loh WL (1996) On Latin hypercube sampling. *Ann Stat* 24(5):2058–2080
- Matheron G (1963) Principles of geostatistics. *Econ Geol* 58(8):1246–1266
- Romero VJ, Burkardt JV, Gunzburger MD, Peterson JS (2006) Comparison of pure and “Latinized” centroidal Voronoi tessellation against various other statistical sampling methods. *Reliab Eng Syst Saf* 91:1266–1280
- Sacks J, Welch WJ, Mitchell TJ, Wynn HP (1989) Design and analysis of computer experiments. *Stat Sci* 4(4):409–435
- Saka Y, Gunzburger MD, Burkardt JV (2007) Latinized, improved LHS, and CVT point sets in hypercubes. *Int J Numer Anal Model* 4(3–4):729–743
- Simpson TW, Korte JJ, Mauery TM, Mistree F (1998) Comparison of response surface and kriging models for multidisciplinary design optimization. *AIAA* 98-4755:381–391
- Sobester A, Leary SJ, Keane AJ (2005) On the design of optimization strategies based on global response surface approximation models. *J Glob Optim* 33:31–59
- Sobol IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. *USSR Comput Math Math Phys* 7(4):86–112
- Sobol IM, Asotsky D, Kreinin A, Kucherenko S (2011) Construction and comparison of high-dimensional sobol’ generators. *Wilmott* 2011(56):64–79
- Viana FAC, Venter G, Balabanov V (2010) An algorithm for fast optimal Latin hypercube design of experiments. *Int J Numerical Method Eng* 82(2):135–156
- Wang GG (2003) Adaptive response surface method using inherited Latin hypercube designs. *J Mech Des* 125(2):210–220
- Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. *J Mech Des* 129:370–380
- Wang GG, Simpson T (2004) Fuzzy clustering based hierarchical metamodeling for design space reduction and optimization. *Eng Opt* 36(3):313–335
- Wang GG, Dong Z, Aitchison P (2001) Adaptive response surface method – a global optimization scheme for computation. *Eng Opt* 33(6):707–734
- Wong TT, Luk WS, Heng PA (1997) Sampling with Hammersley and Halton points. *J Graphics Tools* 2(2):9–24
- Xiong F, Xiong Y, Chen W, Yang S (2009) Optimizing Latin hypercube design for sequential sampling of computer experiments. *Eng Opt* 41(8):793–810