2010-01

Comparing computer experiments for fitting high-order polynomial metamodels

Johnson, Rachel T.; Montgomery, Douglas C.; Jones, Bradley; Parker, Peter T.

ABI/INFORM Global

Journal of Quality Technology, Jan. 2010, vo.42, no.1, pp.86-102.
https://hdl.handle.net/10945/39563

Comparing Computer Experiments for Fitting High-Order Polynomial Metamodels

RACHEL T. JOHNSON

Naval Postgraduate School, Operations Research Department, Monterey, California 93943, USA

DOUGLAS C. MONTGOMERY

Arizona State University, Tempe, Arizona 85287-5906, USA

BRADLEY JONES

SAS Institute, SAS Campus Drive, Cary, North Carolina 27513, USA

PETER A. PARKER

National Aeronautics and Space Administration, Hampton, Virginia 23681, USA

The use of simulation as a modeling and analysis tool is wide spread. Simulation is an enabling tool for experimenting virtually on a validated computer environment. Often the underlying function for a computer experiment result has too much curvature to be adequately modeled by a low-order polynomial. In such cases, finding an appropriate experimental design is not easy. We evaluate several computer experiments assuming the modeler is interested in fitting a high-order polynomial to the response data considering both optimal and space-filling designs. We also introduce a new class of hybrid designs that can be used for deterministic or stochastic simulation models.

Key Words: Optimal Design; Response Surface; Space-Filling Design.

COMPUTER simulation models are often used in place of or in conjunction with physical experiments. Simulation is an extraordinarily powerful tool that allows the study of a system virtually. If the computer-simulation model is a valid representation of the real system, the experimentation and analysis with the model can lead to results and conclusions that are valid for the real system. It is as desirable to experiment with computer simulation models as it is to experiment with physical systems. This requires the choice of an experimental design and analysis technique. The issues associated with the creation of design and analysis of computer experiments are sometimes different than those encountered in the physical domain.

Usually computer simulation models have no stochastic component so the result for any particular set of conditions is deterministic. Examples of deterministic computer-simulation models include

Dr. Johnson is an Assistant Professor in the Operations Research Department at the Naval Postgraduate School. Her email address is rtjohnso@nps.edu.

Dr. Montgomery is a Regents' Professor of Industrial Engineering and Statistics in the Department of Industrial, Systems and Operations Engineering. He is a Fellow of ASQ. His email address is doug.montgomery@asu.edu.

Dr. Bradley Jones is the Director of Research & Development for the JMP Division of the SAS Institute. His email address is Bradley.Jones@jmp.com.

Dr. Parker is a Senior Research Scientist at NASA Langley Research Center. He is a Senior Member of ASQ. His email address is peter.a.parker@nasa.gov.

finite-element analysis (FEA) models and circuit-simulation (or SPICE) models. Experimental designs intended for use on a deterministic model are often associated with several desirable properties that are not encountered in traditional design of experiments literature. First, the design should span or fill a large portion of the design region. Filling the design region is important because little is known about what portions of the region will provide the most informative or interesting results. Also, responses from computer experiments can be complex, and space-filling designs are robust in the sense that they allow a variety of models to be fit. Second, if factors are deemed insignificant or removed from the model, the resultant projected design should span or fill a large portion of the lower dimension design region. In physical experimentation or stochastic simulation, when factors are removed from the model after they are found to provide no contribution to the response, the projected design frequently results in replicated design points. For example, a 2^3 factorial design projected into two factors results in a replicated 2^2 design. In a computer simulation, this projection of the factorial design would result in a loss of half of the information because replicated points provide no new information in a deterministic setting. These properties have driven new research on experimental design for computer simulation. Designs that aim to fulfill these properties are called space-filling designs.

Computer experiments can be computationally expensive in terms of time required to run a single simulation—a single input. Therefore, surrogate models are used to mimic the input–output relationship in the form of a simpler mathematical expression that can be quickly computed. Surrogate models encompass a broad range of techniques ranging from parametric to nonparametric analysis. As previously stated, the response surface in a deterministic model can be very complex. Determining which model-fitting technique to use is dependent on several factors, such as the problem, the goal of the model, and the knowledge of the analyst. The use of high-order polynomials is only one of many techniques available as a surrogate modeling choice. Polynomial regression models of order two are commonly encountered; however, models of order greater than two can also be found in the literature. Allen et al. (2003) use second-order response models to fit test functions that mimic computer-simulation models. Fang et al. discuss the application of polynomial regression as a surrogate modeling technique. In Danehy et al. (2002) and Tedder et al. (2007), second- through sixth-order polynomials are used to fit flow-field data (such as temperature) from a combustion experiment. This paper presents a comparison of experimental designs for computer-simulation models when the expected form of the surrogate model is a high-order polynomial. Note that the sample size requirements for polynomials become large as the order of polynomial increases and the dimension of the design increases. In some cases, these sample sizes may not be feasible for extremely computationally intensive simulations.

We compare experimental designs based on their prediction variance with respect to the polynomial model. The results drive the motivation for a new class of hybrid designs that are a combination of space-filling and optimal designs.

Next a literature review of design comparison for computer simulations is presented, followed by a discussion of the method that will be used to compare the designs in this paper. The designs that will be compared are then presented, followed by the results, the introduction to a new class of hybrid designs, a case study that evaluates robustness of the design and model-selection criterion, and finally conclusions and future work.

Literature Review

Three articles that evaluate experimental designs for computer simulations are reviewed. Allen et al. (2003) compares combinations of experimental design classes with respect to second-order response surfaces and Kriging modeling methods. They find that there was inconclusive evidence as to which modeling method performed the best, due to the dependency on the experimental design used. They also observed that unavoidable bias errors constituted a large source of prediction error in regression modeling and estimation errors dominated prediction errors in Kriging modeling.

Hussain et al. (2002) present seven two-dimensional functions that were used to test two metamodels. The metamodels tested were a radial basis function, which was originally developed to fit irregular topographic contours of geographical data, and quadratic polynomial models. Factorial designs and Latin hypercube designs (LHDs) were used as the experimental designs. The paper concludes that the factorial design had the best performance with respect to the polynomial surfaces and the LHD has the best performance with respect to the radial-basis functions. Also, the radial-basis functions were shown to have better performance (in terms of model fit to the response surfaces generated by the test function and prediction capabilities) than the polynomial models.

Bursztyn and Steinberg (2004) develop a new method of design comparison based on a Bayesian interpretation of an alias matrix. They use this new alias sum-of-squares design-comparison method and three other comparison methods—entropy criterion, minimum-distance criterion, and integrated mean-square error (IMSE) criterion—to compare designs for computer simulations. In the paper, the comparisons are motivated by random field-regression models and low-degree polynomial approximations with some terms considered as bias. They compared Latin Hypercube designs, U-designs, lattice designs, rotation designs, and fractional factorial designs. The results demonstrate that the new alias sum-of-squares criteria developed in the paper tends to favor the rotational designs. The fractional factorial designs perform best in terms of the entropy and minimum-distance criteria. While this makes sense mathematically, the authors point out that these designs may not be the best choices due to their lack of good coverage and tendency to replicate points when projected. The IMSE criteria favored the space-filling designs with the best particular covariance model and demonstrated consistently low performance for the two-level fractional factorial designs.

None of the three papers investigates higher than second-order–polynomial terms. This paper evaluates experimental designs based on their prediction performance for second-order– through fifth-order–polynomial models. The next section presents a discussion of the comparison method used in this paper followed by a section containing the designs that will be used for the comparison.

Evaluation of Designs

The prediction variance is a standard criterion for comparing designs when modeling physical systems with a stochastic component. The scaled prediction variance (SPV) normalizes the prediction variance over the design region and is computed as

$$\frac{NV[\hat{y}(x_0)]}{\sigma^2} = Nx_0'(\mathbf{X}'\mathbf{X})^{-1}x_0, \qquad (1)$$

where $\mathbf{X}$ is the model matrix and x_0 is the point being evaluated. Zahran et al. (2003) introduce fraction of design space (FDS) plots that graph the empirical distribution function of SPV over the design region. FDS plots are used in the assessment of prediction capability for response surface designs. This is done by generating a large number of design points in the region and evaluating their scaled prediction variance based on Equation (1). These values are then sorted and plotted versus their order expressed as a proportion. Tables containing percentiles based on the scaled prediction variance over the regions and fraction of design (FDS) plots are used as methods for comparing designs for computer experiments where the expected response is a high-order polynomial.

Because deterministic computer experiments have no stochastic component, it is necessary to justify the use of scaled-prediction variance as a performance criterion. Suppose that a given computer experiment is adequately modeled using a polynomial fit. The difference between the observed and fitted values in a deterministic computer model, however, is not stochastic error, it is model bias. If the polynomial model adequately describes the response surface of the true underlying function, the model bias of the βs is negligible. The model bias of an individual prediction is also fairly small because the fit is adequate. Assume that the source of this bias is due to multiple high-order terms. Thus, deviations between the observed and predicted values will behave like the sum of a number of independent small quantities. Appealing to the central limit theorem, as the number of these quantities get large, these deviations will converge to the normal distribution. We then justify the prediction variance criterion as a measure of the sum of a large number of small biases.

Designs Used for Comparison

Numerous space-filling designs have been proposed in the last 30 years. We investigated sphere-packing, Latin-hypercube, uniform, and maximum-entropy designs. We compared the prediction performance of the space-filling designs to the I-optimal and D-optimal designs. These designs have low prediction variance with respect to polynomial models. The I-optimal designs minimize the average variance of prediction over the design region. While they do not satisfy desirable properties of computer simulation experiments—for example, projections of these designs result in replicate points—they can act as a baseline for good performance in terms of the prediction variance criterion. A description of the designs used follows.

The sphere-packing design, also known as the maximin design, maximizes the minimum distance between pairs of design points. This design was developed in Johnson et al. (1990). The maximin design maximizes the minimum intersite distance and is specified by

$$\max_{D} \min_{u,v \in D} d(u,v) = \min_{u,v \in D_0} d(u,v), \qquad (2)$$

where $d(u,v)$ is a distance that is greater than or equal to zero and D represents the design points. Examples of applications of the maximin designs can be found in Jank and Shmueli (2007), Liefvendahl and Stocki (2006), Chen et al. (2006), Roux et al. (2006), and Bursztyn and Steinberg (2006).

The Latin-hypercube design was developed by McKay et al. (1979). The Latin-hypercube design is defined in Fang et al. (2006) as, "A Latin Hypercube design (LHD) with n runs and s input variables, denoted by LHD(n, S), is an $n \times s$ matrix, in which each column is a random permutation of $\{1, 2, \ldots, n\}$." Examples of applications of LHDs can be found in Bayarri et al. (2007), Welch et al. (1992), Mease and Bingham (2006), Tyre et al. (2007), and Storlie and Helton (2008). The maximin criterion is used as a secondary criterion for creating the LHDs. All of the LHDs generated in this paper are maximin Latin-hypercube designs.

The uniform design was created by Fang (1980) and Wang and Fang (1981). The uniform design tries to generate a set of points in the design space to be uniformly scattered, as in the uniform distribution. The uniform design is a design created such that the discrepancy, a measure of uniformity, of the design is the smallest (the distribution closest to that of the uniform). Fang et al. (2006) detail several measures of discrepancy, where $F(x)$ is defined as a uniform distribution on C^s (the unit cube) and $F_{D_n}(x)$ is the empirical distribution of the design D_n. Thus,

$$F_{D_n}(x) = \frac{1}{n} \sum_{k=1}^{n} I\{x_{k_1} \le x_1, \ldots, x_{ks} \le x_s\}, \qquad (3)$$

where $x = (x_1, \ldots, x_s)$ and $I\{A\} = 1$ if A occurs, or 0 otherwise. In this paper, the software used to create the uniform design uses the centered L_2 discrepancy found in Hickernell (1998). The L_2 discrepancy can be treated as an objective function that can be minimized in continuous space. An example of the application of a uniform design is found in Bursztyn and Steinberg (2006).

The maximum-entropy design, developed in Shewry and Wynn (1987), uses entropy as the optimality criterion, where entropy is a measure of the amount of information contained in the distribution of a data set. In their paper, they show that the expected change in information is maximized by the design D that maximizes the entropy of the observed responses at the points in the design. If the data is assumed to be from a normal $(m, \sigma^2 R)$ distribution, where R is

$$R_{ij} = e^{\left(\sum_k \theta_k (x_{ik} - x_{ji})^2\right)}, \qquad (4)$$

which is the correlation of responses at two design points, then the design maximizes the determinant of R ($|R|$) (Sacks et al. (1989)). Note that this criterion requires the specification of the values in the θ_k vector, which, in practice, is not known prior to experimentation. To create maximum-entropy designs in this paper, we made the assumption that the values of θ are all equal. An application of this design can be found in Ko et al. (1995).

Optimal-design theory for experimental design emerged following World War II and was motivated by many authors. I-optimal designs, or integrated variance designs, minimize the average scaled prediction variance over the design region. A D-optimal design minimizes the generalized variance of the model coefficients. This is done by creating a design that maximizes $|\mathbf{X}'\mathbf{X}|$ (Myers and Montgomery (2002)). Examples of optimal designs can be found in Atkinson and Donev (1992). A specific application of the D-optimal design for stochastic computer simulation models can be found in Park et al. (2002).

Examples of each of the designs are presented in Figure 1. Note that the I-optimal and D-optimal designs are the same for 2 factors in 10 runs fit to a second-order polynomial. These designs are seen to only have nine points because they each have a replicated point. The I-optimal design replicates the center point and the D-optimal design replicates a corner point.

Results

Here we compare four types of space-filling designs (sphere packing, LHD, uniform, and maximum entropy) and two types of optimal designs (I-optimal and D-optimal) using percentiles of prediction variance over the design region and graphical evaluation via FDS plots (Zahran et al. (2003)).

In order to test the predictive capabilities of space-filling designs and optimal designs when fitting a high-order polynomial, we generated designs ranging from two to five factors and used second-order to

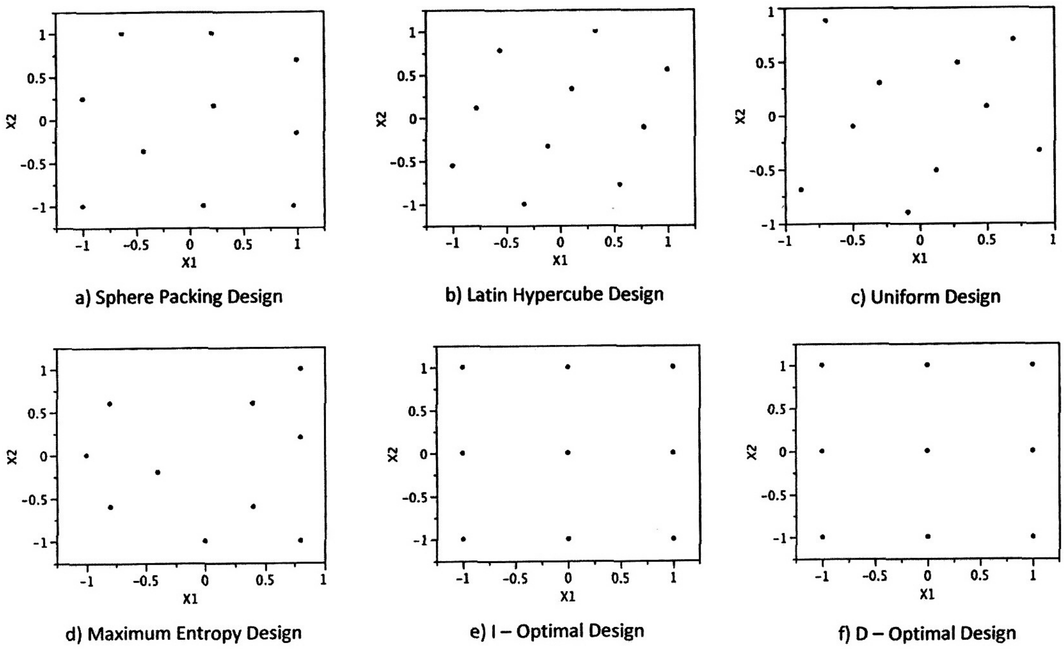


FIGURE 1. Examples of 2 Factors, 10-Run Designs for Each of the Designs Being Evaluated.

fifth-order polynomials to generate the **X** model matrix. To generate a sphere-packing, Latin-hypercube, or uniform space-filling design, no model specification is necessary, only the required number of points (sample size) is needed. To generate a maximum-entropy design, one must specify prior estimates of the values in the θ vector. In this paper, we assumed equal values of the θ's. To generate an optimal design, a model must be specified as well as the number of design points required. Table 1 illustrates the minimum number of design points needed to fit a given polynomial with two to five factors.

For all 16 cases (four separate factors levels and four increasing degrees of polynomial), four sets of designs were generated using an increasing sample size. One design contained a minimum number of design points, the second design contained the minimum design points plus two additional points, the third design contained the minimum design points plus four additional points, and finally the fourth design contained double the number of minimum points needed. Table 2 illustrates all of the designs generated with their respective number of runs.

TABLE 1. Minimum Number of Design Points Needed ($n = p$) to Fit a Given Polynomial for a Specified Number of Factors

	Order of polynomial			
Factors	2	3	4	5
2	6	10	15	21
3	10	20	35	56
4	15	35	70	126
5	21	56	126	252

For each of the designs illustrated in Table 2, we generate percentiles of prediction variance and FDS plots a uniform random sample of 10,000 points from the hypercube $[-1, 1]^p$. As an example, Figure 2 illustrates FDS plots for each of the designs evaluated for a second-order model in two factors and 10 runs. Figure 2 shows that the I-optimal design dominates the other by having the lowest prediction variance across 99.9% of the region. The I-optimal design is followed by the D-optimal and sphere-packing design, which have equivalent prediction variance performance in this example. The worst design in the example is the maximum-entropy design.

Tables 3 and 4 demonstrate the prediction variance percentiles for the second-order of the space-filling designs and the optimal designs, respectively. The space-filling designs are labeled as SP (sphere packing), LH (Latin hypercube), U (uniform), and

TABLE 2. Number of Runs Required for Each of the Designs Analyzed

Runs →	Order of polynomial															
	2nd order				3rd order				4th order				5th order			
Factors	p	$p+2$	$p+4$	$2p$	p	$p+2$	$p+4$	$2p$	p	$p+2$	$p+4$	$2p$	p	$p+2$	$p+4$	$2p$
2	6	8	10	12	10	12	14	20	15	17	19	30	21	23	25	42
3	10	12	14	20	20	22	24	40	35	37	39	70	56	58	60	112
4	15	17	19	30	35	37	39	70	70	72	74	140	126	128	130	252
5	21	23	25	42	56	58	60	112	126	128	130	252	252	254	256	504

ME (maximum entropy). The dark-gray boxes represent the percentiles corresponding to the minimum-prediction variance and the light-gray boxes corresponding to the percentiles with the maximum-prediction variance. Of the space-filling designs in Table 3, the sphere-packing design tends to have the best performance with respect to prediction variance for the second-order design. In 9 out of the 16 designs shown, the sphere-packing design has the lowest prediction variance for more than 50% of the design region. Table 3 also illustrates that the uniform design tends to have the worst performance when compared with the other space-filling designs.

Results from Table 4 indicate that the I-optimal design has the best performance in terms of prediction variance with respect to the second-order designs. We expected this because the I-optimal designs minimize the average prediction variance of a design with respect to the hypothesized model form. It is not surprising that, in general, the percentiles

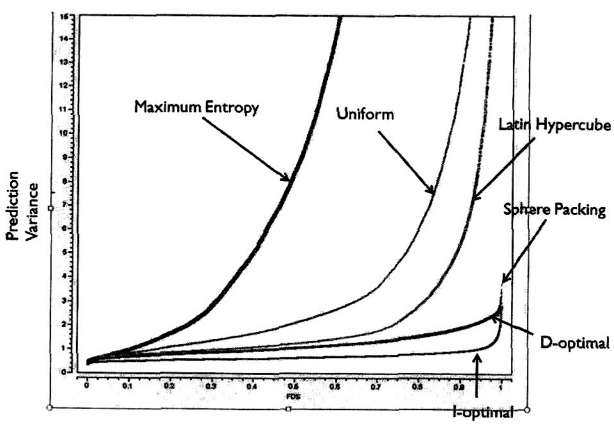


FIGURE 2. FDS Plot for Designs with 2 Variables and 10 Runs Fit to a Second-Order Polynomial Model.

of prediction variance for the I-optimal design would be lower than the competitive designs. In many of the designs shown in Table 4, the D-optimal design has a higher prediction variance at every percentile value recorded, with the exception of the maximum prediction variance. This result is also somewhat intuitive because the D-optimal design tends to spread points to the outside regions of the design, where the joint confidence regions for the unknown betas are influenced the most.

Similar results are demonstrated for the third-, fourth-, and fifth-order cases. The tables for all of these cases can be found in Appendices A, B, and C, which correspond to results from the third-, fourth-, and fifth-order polynomial cases, respectively. The best overall performance for the space-filling designs is not consistent as polynomial order increases. The sphere-packing designs generally perform better than the Latin hypercube. The uniform and maximum-entropy designs demonstrate the worst performance. As for the optimal designs, in almost all cases, the I-optimal design illustrates superior prediction variance properties to the D-optimal, by having the lowest overall prediction variance across most of the design space (the p-dimensional hypercube $[-1, 1]^p$). Additionally, when comparing all of the designs together, the I-optimal and D-optimal have a lower prediction variance than the space-filling designs.

A New Class of Hybrid Designs

From the results presented in the previous section, it is clear that space-filling designs are clearly inferior to optimal designs with respect to their prediction variance using a polynomial model of any order. We wished to determine if augmenting space-filling designs with points chosen specifically to reduce prediction variance with respect to a polynomial model would result in a satisfactory compromise.

TABLE 3. Prediction Variance Percentiles for Second-Order Models for Space-Filling Designs

2 Factors	n = p = 6 runs				n = p + 2 = 8 runs				n = p + 4 = 10 runs				n = 2p = 12 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	3.24	2.88	3.24	2.82	2.96	2.96	2.96	2.88	3.1	3.4	2.4	2.9	3	2.88	3.12	2.52
5%	3.3	3.36	4.32	3.18	3.12	3.12	3.36	3.12	3.2	3.5	2.7	3.2	3.12	3.12	3.24	2.88
25%	3.72	4.14	6.24	4.14	3.6	3.76	3.92	3.52	3.6	3.8	3.3	3.6	3.36	3.6	3.6	3.36
50%	4.38	5.16	9.3	5.82	3.92	4.72	5.12	4.08	4	4.5	5.7	4.1	3.72	4.32	4.56	3.72
75%	5.4	6.48	14.82	8.82	5.2	8.08	9.52	4.96	4.6	6.1	10.8	4.6	4.56	6.48	8.4	4.68
90%	6	9.66	31.32	13.74	6.64	13.04	15.44	6.32	5.4	10	15.8	5.4	6	9.84	13.2	6.12
95%	6.84	12.84	48.66	17.16	7.52	17.68	20.56	7.12	6.3	14.4	19.2	6	7.32	13.2	18.48	7.32
max	15.84	40.32	175.62	25.62	9.68	42.24	50.48	12.72	13.3	40.9	40.6	11.9	15.84	36.48	47.88	13.56

3 Factors	n = p = 10 runs				n = p + 2 = 12 runs				n = p + 4 = 14 runs				n = 2p = 20 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	5.6	4.7	4.8	4.4	5.04	4.44	3.48	4.68	3.92	4.9	4.34	4.62	4	4.4	3.8	4
5%	6.9	5.7	9.9	7.8	5.64	5.64	6	5.76	4.62	5.74	5.74	5.6	4.8	5.2	5	4.6
25%	8.9	7.3	19.6	17.8	6.6	6.84	9.6	7.44	5.18	6.58	8.26	7	5.2	6	6.6	5.2
50%	11.2	9.5	36.1	45.2	7.44	8.52	18.24	9.24	5.74	8.12	12.88	8.4	5.8	8	9.8	6
75%	15.3	15.1	89.7	93	8.88	12.84	37.08	13.92	7.14	12.18	22.54	10.92	7	13.2	16	8
90%	20.5	25.2	203.6	173.8	10.56	20.64	80.64	26.16	8.54	19.88	38.5	14.98	8.6	21.2	24.6	10.4
95%	23.4	34	311.7	236.7	12.36	28.2	120.24	38.04	9.52	25.48	53.06	20.02	9.8	28	31.6	12.8
max	87.9	154.5	1672.1	1313.3	28.92	112.08	319.92	131.4	16.38	67.76	207.48	84.14	20.8	82.6	85.8	38.4

4 Factors	n = p = 15 runs				n = p + 2 = 17 runs				n = p + 4 = 19 runs				n = 2p = 30 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	8.55	8.1	8.25	9.9	7.48	10.03	6.29	7.65	6.27	7.22	6.27	6.46	5.7	6.9	5.1	5.7
5%	88.05	12.75	14.7	54.6	13.6	13.26	13.6	13.26	8.55	8.93	9.69	9.88	6.6	7.8	6.6	7.5
25%	441.6	19.05	25.8	224.55	29.07	20.74	25.67	21.76	11.78	10.83	15.77	15.01	7.5	9.9	10.2	9
50%	995.55	27.9	41.55	484.95	66.13	37.06	50.66	36.04	17.1	13.49	25.46	21.47	8.7	13.8	16.5	11.4
75%	1925.7	43.5	69.6	902.4	138.21	75.65	100.3	57.12	27.93	20.52	43.7	31.54	10.5	21.6	26.4	17.1
90%	3381.75	67.2	132.45	2265.15	252.28	127.16	189.89	96.39	43.32	31.54	72.39	47.88	13.2	33.3	39.6	25.2
95%	4552.05	97.35	215.85	3914.25	332.18	163.37	289	152.83	55.86	40.09	100.89	64.79	15	42.3	49.8	32.1
max	9812.7	608.4	1067.7	23638.5	843.03	769.42	1250.18	895.22	106.97	137.37	636.88	305.33	28.8	125.4	133.5	94.8

5 Factors	n = p = 21 runs				n = p + 2 = 23 runs				n = p + 4 = 25 runs				n = 2p = 42 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	21.63	14.28	9.03	9.87	11.5	12.19	11.5	8.28	10.5	12.5	9	9.75	7.98	9.24	6.3	6.72
5%	133.35	27.72	30.87	23.52	17.48	17.25	22.08	15.41	15.25	17	17.75	17.75	9.66	11.76	9.24	8.82
25%	254.94	64.47	107.73	49.77	23.23	24.61	38.64	28.29	18.75	21.75	32	28	10.92	14.7	16.8	13.02
50%	360.57	135.45	300.09	96.6	31.74	35.88	60.03	52.9	22	28	56.5	43.75	12.18	19.74	26.88	19.32
75%	482.58	265.23	880.53	199.92	45.31	54.97	95.22	109.71	26.75	41	103.5	82.5	13.86	30.66	42.42	28.98
90%	607.95	441.84	2058.84	383.04	63.71	87.86	166.52	206.08	33.5	64.75	176.75	163.25	16.38	45.36	63	42
95%	683.97	605.22	3174.57	574.77	80.73	138	239.89	314.87	39.5	88.75	235.5	256.5	18.9	56.7	79.38	52.08
max	1048.74	3014.55	19244.4	2967.93	286.58	1148.39	1208.65	1596.89	75	328	736.75	1663.5	45.78	178.08	225.12	136.08

A small example demonstrates the idea. Consider Figure 3, which illustrates two designs, both containing two factors and 10 runs designs to fit a third-order polynomial (note these are saturated designs). The FDS plot for each of these designs is shown in Figure 4.

It is clear from this figure that the I-optimal design performs better because it has lower prediction variance over the entire curve. Can the performance of the LHD be improved by augmenting the design with I-optimal points? Figure 5 illustrates two designs each with two factors. The first design (Figure 5a) is an I-optimal design with 16 design points and the second (Figure 5b) is the LHD shown in Figure 3b augmented with 6 I-optimal points that were chosen to minimize the average prediction variance of the combined design given the preexisting 10 points. The FDS plot for these designs is shown in Figure 6.

The improved performance of the augmented LHD is clear from examination of Figure 6. Design augmentation is a standard procedure in physical experimentation. Building up information sequentially through design augmentation is efficient and economical. Montgomery (2006) points out that it is almost always preferable to run a fractional design, analyze the results, and then decide on the best set of runs to perform next. We believe that design augmentation can also be used in computer-simulation modeling. We find that design augmentation is especially effective when the initial analysis of a space-filling design indicates that a polynomial is adequate. The results illustrated in Figure 4 and Figure 6 demonstrate that this procedure can be effective in improving the prediction variance across the design region. This is important because the goal of a surrogate model is to closely approximate the computer model. We introduce augmentation of space-filling designs and refer to them as space–filling-hybrid designs.

The next section demonstrates the augmentation

TABLE 4. Prediction Variance Percentiles for Second-Order Models for Optimal Designs

2 Factors	n = p = 6 runs		n = p + 2 = 8 runs		n = p + 4 = 10 runs		n = 2p = 12 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	0.46	0.48	2.88	2.88	2.9	2.7	2.76	3.12
5%	0.5	0.67	3.12	3.12	2.9	3.3	3	3.48
25%	0.63	1.06	3.6	3.6	3.1	3.9	3.48	4.2
50%	0.9	1.62	4.16	4.16	3.4	4.3	3.84	4.8
75%	1.15	2.09	5.04	5.04	4	4.7	4.2	5.28
90%	1.35	2.48	6	6	4.8	5.2	5.04	5.88
95%	1.57	2.62	6.88	6.88	5.1	5.3	5.64	6.12
max	5.45	2.79	9.84	9.92	7.9	7.7	9	9.12

3 Factors	n = p = 10 runs		n = p + 2 = 12 runs		n = p + 4 = 14 runs		n = 2p = 20 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	27.6	28.8	3.72	4.32	4.34	4.34	3.8	5
5%	30	40.2	4.32	5.64	4.62	4.62	4	6
25%	37.8	63.6	4.8	6.96	5.04	5.04	4.2	7
50%	54	97.2	5.52	8.16	5.46	5.46	5	8
75%	69	125.4	7.08	9.36	6.58	6.44	6.2	9
90%	81	148.8	9.36	10.92	7.28	7.42	7.2	10
95%	94.2	157.2	12.12	12.12	7.7	7.7	8.2	10.6
max	327	167.4	30.48	17.16	9.8	10.22	13.2	12

4 Factors	n = p = 15 runs		n = p + 2 = 17 runs		n = p + 4 = 19 runs		n = 2p = 30 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	5.85	6.6	4.25	6.12	4.75	6.46	4.5	7.5
5%	8.25	8.55	5.44	8.67	5.51	8.36	4.5	9.9
25%	10.05	10.2	6.46	10.71	6.27	9.88	5.4	11.4
50%	11.1	11.7	8.33	12.58	7.79	11.21	6.9	12.6
75%	12.6	13.35	10.54	14.62	9.69	12.92	8.7	13.8
90%	14.55	15.75	12.92	16.32	11.97	14.25	10.8	15.3
95%	16.65	17.55	14.79	17.34	13.49	15.39	12.3	15.9
max	40.5	28.65	29.58	23.97	27.93	23.75	24.6	18.9

5 Factors	n = p = 21 runs		n = p + 2 = 23 runs		n = p + 4 = 25 runs		n = 2p = 42 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	7.77	6.93	6.44	8.05	6	9.25	5.04	10.08
5%	10.5	13.44	7.59	13.11	7.25	13	5.46	13.44
25%	12.81	17.01	9.66	15.64	9	15.5	7.14	15.54
50%	15.12	19.53	11.96	17.25	11.25	17.5	9.24	17.64
75%	18.27	22.68	14.95	19.09	13.5	19.5	11.34	19.32
90%	22.68	26.25	18.17	21.39	15.75	21.25	14.28	21
95%	26.46	28.56	20.24	22.77	17	22.25	15.54	22.26
max	70.98	40.74	41.17	30.36	24.25	28.5	25.62	27.3

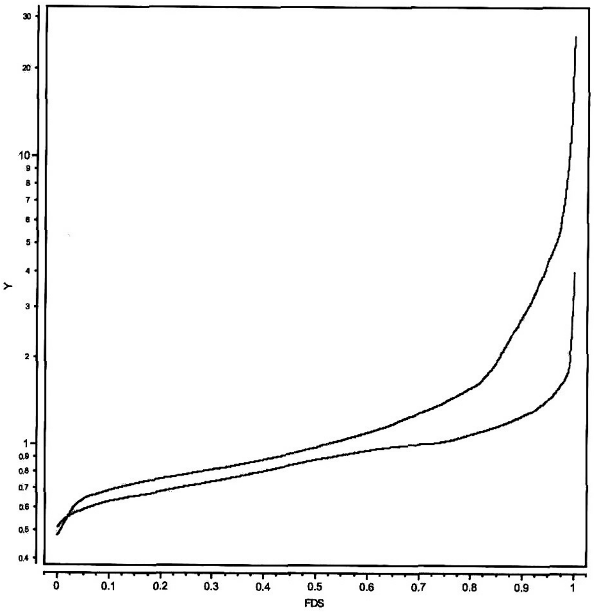
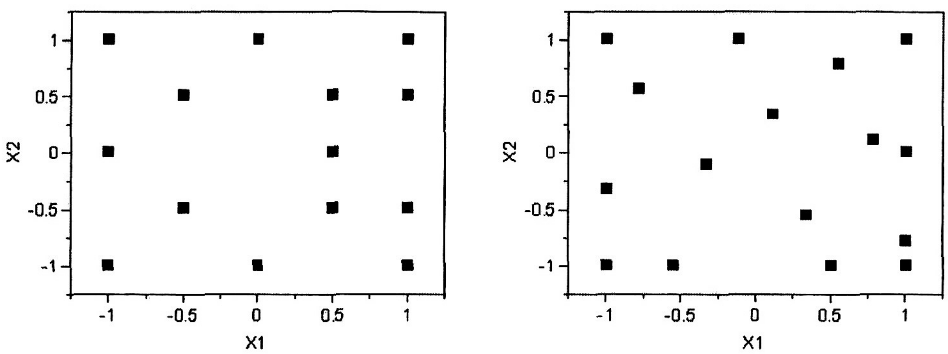


FIGURE 4. FDS Plots of the *I*-Optimal Design (Figure 3a) and LHD (Figure 3b). In this figure, the LHD is the design with the poorer prediction variance. Note: the log scale is used for the *y*-axis.

technique and investigates the robustness of the designs to the choice of surrogate model fitted. That is, we consider the performance of the designs with respect to not only the polynomial model, but also the Gaussian process (GASP model). Trade-offs between design and modeling choices are illustrated in the example through a study of the theoretical integrated prediction variances and empirical mean-square error.

Case Study

In this section, we present a case study relating to a NASA-sponsored air-breathing propulsion experiment. Theoretical prediction variance of the pure de-



a) 16 Run I-Optimal Design for a 3rd order polynomial Model

b) 16 Run Augmented Latin Hypercube Design for a 3rd order polynomial Model

FIGURE 3. Two Designs Each with 2 Factors and 10 Runs Designed for a Third-Order Polynomial. Figure 3a contains an *I*-optimal design and Figure 3b is an LHD.

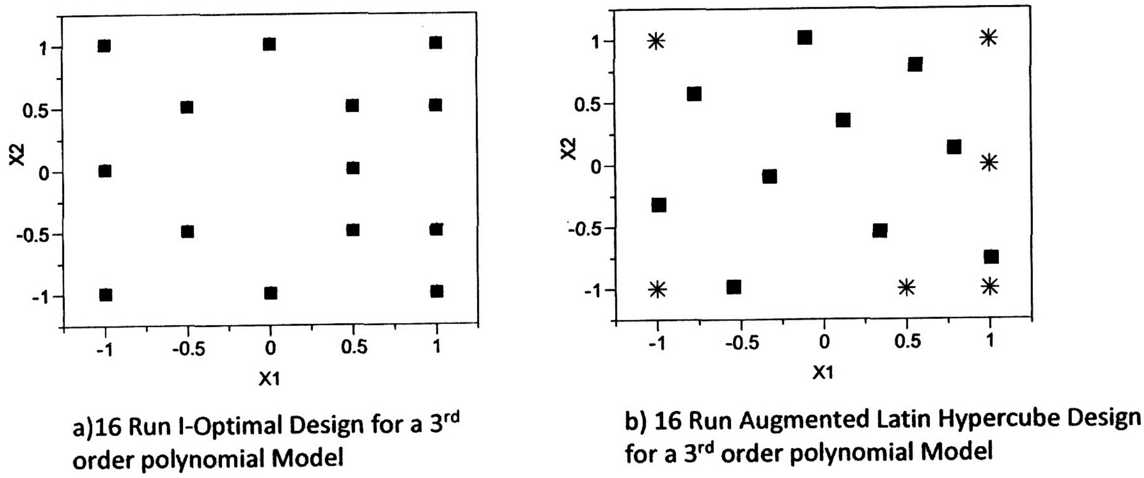


FIGURE 5. Two 16-Run 2-Factor Designs for a Third-Order Polynomial Model.

sign and an augmented hybrid design are considered. Additionally, empirical mean square error of the designs is calculated by using the surrogate model to predict points not in the original design. The simulation results are based on a computational fluid dynamics (CDF) model built to mimic the flow-field parameters within an open jet flame. Because the CFD is not available for commercial use, we created a mathematical model that predicts the CFD accurately to four decimal places. The physical experiment on which the CFD is based is described in Johnson et al. (2009). In the example in this section, we are interested in modeling response, oxygen, as a function of two input factors: x- and y-axis location. See Figure 7, which displays a surface profile plot of the function.

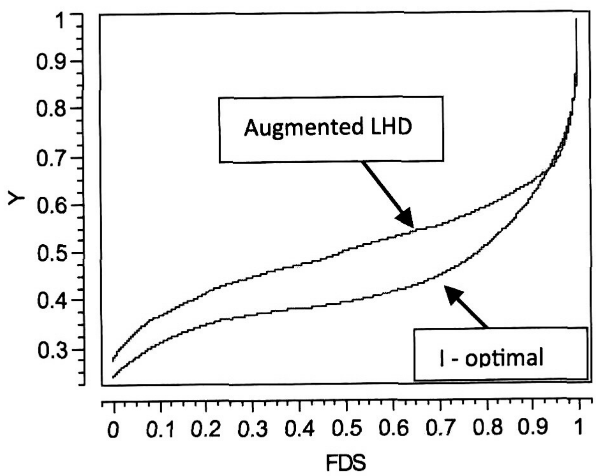


FIGURE 6. FDS Plots of Designs in Figure 5.

While the output response is expected to be modeled by a third-order polynomial model, we also consider the GASP model. Note that we expect a third-order model to provide a good fit to the data, but because we are using a mathematical model in place of the CFD, we know that the true underlying function is neither polynomial nor a Gaussian process. We compared three designs: the maximin Latin-hypercube design, the augmented maximin Latin-hypercube design, and the I-optimal design. The Latin-hypercube design was chosen because is has good performance with respect to the kriging model and it is a staple in the computer-simulation literature.

Augmentation Technique

The hybrid design strategy that we propose is based on the design-augmentation technique described below:

Step 1: Assume that the total design size is $N = n_1 + n_2$.

Start with n_1 points using some space-filling design. Collect data at each of the n_1 points and find and fit an adequate polynomial model.

Step 2: Augment the space-filling design with n_2 points using a polynomial variance criterion.

The degree of the polynomial and important terms can be influenced by the fit in step 1.

Theoretical Integrated Variance

In this subsection, we evaluate the theoretical performance of the designs in the case study. We com-

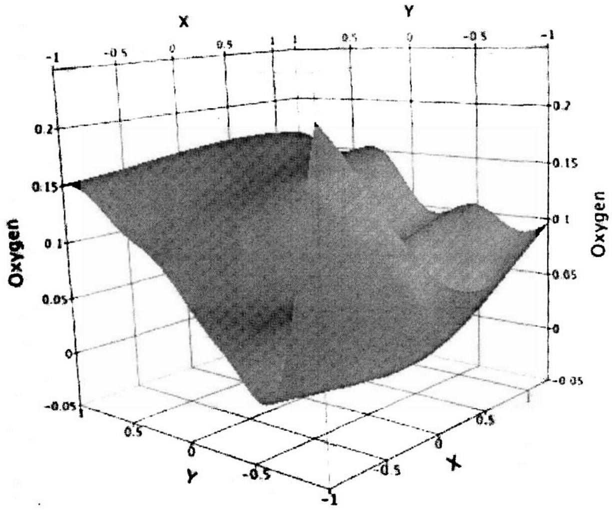


FIGURE 7. Oxygen as a Function of *x*- and *y*-Axis.

pare the designs with respect to the GASP integrated variance (IV) and the polynomial IV. The GASP prediction variance is given by the following equation:

$$\frac{\mathrm{Var}(\hat{y}(x))}{\sigma^2} = 1 - r'(x,\hat{\boldsymbol{\theta}})R^{-1}(\mathbf{X},\hat{\boldsymbol{\theta}})r(x,\hat{\boldsymbol{\theta}}) + \frac{(1 - 1'R^{-1}(\mathbf{X},\hat{\boldsymbol{\theta}})r(x,\hat{\boldsymbol{\theta}}))^2}{1'R^{-1}(\mathbf{X},\hat{\boldsymbol{\theta}})_1}. \quad (5)$$

The polynomial prediction variance equation was presented in Equation (1). Note that both the GASP IV and the polynomial IV can be computed by integrating the prediction variance equation over the design region, which is the $[-1, 1]$ hypercube. Three 18-run designs are compared. The first design is a maximin LHD, the next is an augmented maximin LHD generated with $n_1 = 12$ and $n_2 = 6$, and the third is an I-optimal design (optimal with respect to the third-order model in two variables). Table 5 presents the theoretical results. Note that the theoretical IV cannot directly be compared between the GASP model and the cubic polynomial model. The LHDs have the lowest expected IV with respect to

TABLE 5. Theoretical Integrated Prediction Variance for Each of Three Designs for the Gaussian Process (GASP) Model and the Cubic Polynomial

	LHD	Hybrid	I-optimal
GASP	0.009	0.0214	0.0407
Cubic	0.576	0.3999	0.365

TABLE 6. Empirical Root-Mean–Square Error for Each of Three Designs Fit Using Either the GASP Model or the Cubic Model

	LHD	Hybrid	I-optimal
GASP	0.01776	0.01514	0.02996
Cubic	0.01983	0.01794	0.02488

the GASP model and the I-optimal design has the lowest IV with respect to the cubic model. The hybrid design demonstrates a compromise between the pure LHD and the pure I-optimal design. By augmenting the LHD, we see an improved performance with respect to the cubic polynomial and a degradation in performance with respect to the GASP model.

Empirical Mean-Square Error

In addition to the theoretical performance of the designs, the empirical performance is also considered. Table 6 presents the results for empirical root mean-square error (ERMSE) of the design with respect to the model used to fit the design points. In this case study, the hybrid design outperformed the other design/surrogate model combinations. While all of the values are relatively small, the results demonstrate that the hybrid LHD has the lowest ERMSE for when the design is fit to both the GASP model and the cubic polynomial model.

Conclusions

The results presented in this paper give insight into how space-filling designs perform with respect to prediction variance properties for polynomial models. The designs are compared with optimal designs, which are designed with respect to criteria pertaining to the polynomial models. Of the space-filling designs, the sphere-packing design was generally the best choice of design for minimizing the SPV for a polynomial model, followed by the Latin-hypercube design. We also introduce a new class of hybrid designs. These designs consist of a space-filling design augmented with I-optimal points. These designs are shown to have much improved prediction variance with respect to polynomial models.

One of the benefits of deterministic computer-simulation models is the ability to build up a design sequentially, without concern for blocking or randomization. In such cases, the space-filling-hybrid design is an excellent choice. After running a preliminary set of runs, the experimenter has a better idea of what modeling strategy to use. At this point, the design can be augmented with a criterion that is optimal for that strategy.

While some might question the use of the space-filling design for polynomials at all, it is important to remember that, in advance of any experimentation, it is impossible to know whether a polynomial model of any order will prove to be adequate. Using a space-filling design for initial exploration makes sense.

The implications to the researcher are summarized in the following points: (1) space-filling designs do not perform as well as optimal designs with respect to a polynomial model, (2) of the space-filling designs, sphere-packing designs generally have the lowest prediction variance with respect to polynomials, and (3) augment space-filling designs with I-optimal points whenever initial modeling indicates that the computer-simulation model can be adequately approximated by a polynomial.

Acknowledgments

The authors gratefully acknowledge the support of the NASA Langley Research Center. Specifically, thanks to Phil Drummond, Andrew Cutler, Paul Danehy, Sara Tedder, and Daniel Bivolaru from NASA Langley Research Center and Gaetano Magnotti from the George Washington University for their interest, support, and application of statistical methods in the air-breathing propulsion research efforts. The authors would also like to thank the referee for the helpful revision suggestions and comments.

Appendix A
Prediction Variance for Third-Order Models

TABLE 7. Prediction Variance Percentiles for Third-Order Models for Space-Filling Designs

	n = p = 10 runs				n = p + 2 = 12 runs				n = p + 4 = 14 runs				n = 2p = 20 runs			
2 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	5.1	4.6	4.5	5.6	4.2	4.2	4.8	4.2	4.34	4.34	4.06	5.6	3.8	4.2	4	4.2
5%	7.6	6.5	8.1	30.2	5.76	5.64	7.2	6.24	5.6	5.6	5.6	6.02	4.8	5.2	5.2	5.2
25%	10.1	9.1	19.8	495.9	7.44	6.96	10.44	9.12	7.14	6.86	7.56	7.14	6.4	6.4	6.6	6.4
50%	15	12.5	39.9	2616	8.4	8.04	15.48	23.64	7.98	7.42	9.94	8.54	7	7.2	7.8	8.6
75%	28.5	20.2	83.2	8341	9.96	9.6	24.96	76.8	9.1	10.22	17.08	11.48	7.6	10	12.4	14.2
90%	43.6	33.7	203.3	18879.6	11.76	15.84	63.24	179.52	10.08	19.46	37.24	45.78	9.8	16.8	24.2	30.8
95%	50.1	50.3	394.8	47626.9	13.08	24.72	128.4	330.12	11.34	28.98	56.42	125.02	11.4	23.6	37.2	50
max	65.9	340.8	5308.2	177126	51.24	115.92	713.16	2940.6	38.08	89.04	392.84	1065.12	35.4	103.4	141	270

	n = p = 20 runs				n = p + 2 = 22 runs				n = p + 4 = 24 runs				n = 2p = 40 runs			
3 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	9	9.4	11.2	8	7.26	8.36	9.46	8.36	7.44	8.64	7.68	6.48	5.6	6	6.4	5.6
5%	24.6	18.2	42.8	14.2	13.42	14.74	16.28	15.84	14.88	14.4	15.12	10.8	9.2	10.4	10	8.8
25%	46.6	29.6	172.8	25.8	18.26	19.58	26.62	45.98	21.12	17.28	37.92	14.16	11.6	12.4	14.8	11.2
50%	89.4	51.8	470.8	52	23.1	26.84	46.2	185.9	26.88	21.36	93.6	16.8	13.2	14.8	20.8	12.8
75%	190.8	100	1277.6	139.8	30.8	47.08	101.42	794.2	34.32	32.16	231.12	21.84	15.2	24.8	41.6	15.2
90%	350.8	189	3722	355.2	41.14	120.34	275	2914.12	42	66.72	514.56	30.24	18	46	80	18.4
95%	447	446.4	6407.4	650	47.74	219.12	469.7	7767.76	46.8	109.92	960.96	37.92	20.4	66.4	124.8	21.2
max	1196.6	6293.6	62956.4	3099.2	83.16	1521.52	3682.58	121052	67.92	762.24	7399.92	116.16	56.4	314.8	593.6	71.6

	n = p = 35 runs				n = p + 2 = 37 runs				n = p + 4 = 39 runs				n = 2p = 70 runs			
4 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	12.25	21	16.1	22.4	15.91	14.43	16.65	17.02	12.87	12.87	9.75	19.5	9.1	9.8	7.7	7
5%	42.7	39.9	58.1	138.6	33.67	27.75	79.18	37	27.3	24.18	33.93	33.54	17.5	18.9	16.1	14
25%	98	65.45	165.9	958.65	51.8	39.59	333.74	71.04	37.44	34.32	128.7	56.55	21.7	22.4	24.5	18.9
50%	205.8	107.45	383.6	3703.35	80.29	61.42	1001.22	129.5	48.75	53.43	349.83	92.04	23.8	29.4	43.4	26.6
75%	444.85	211.75	980.7	10938.6	133.94	122.84	2942.24	265.66	65.13	101.79	886.86	175.11	26.6	52.5	82.6	40.6
90%	905.45	568.4	2238.6	26606	210.9	253.45	7209.82	529.1	85.41	211.38	2023.71	376.35	30.8	95.2	150.5	65.8
95%	1519.7	984.9	3641.75	47555.2	294.89	393.68	11912.5	789.21	99.84	336.57	3060.72	590.07	34.3	137.2	212.8	89.6
max	13384.7	18964.8	19146.4	230053	4460.35	2945.2	89726.9	6631.88	202.8	2906.67	14390.6	3320.85	92.4	910	1242.5	619.5

	n = p = 56 runs				n = p + 2 = 58 runs				n = p + 4 = 60 runs				n = 2p = 112 runs			
5 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	43.12	47.04	44.24	36.96	26.68	43.5	24.36	34.8	28.2	35.4	26.4	34.2	11.2	16.8	12.32	13.44
5%	179.76	141.12	304.08	110.88	81.78	112.52	111.94	107.3	60	71.4	108	69.6	23.52	30.24	25.76	25.76
25%	469.28	441.28	1090.32	265.44	154.28	348.58	322.48	280.14	88.8	135	301.8	145.2	30.24	36.96	40.32	36.96
50%	913.36	1202.88	2638.16	557.2	278.4	984.84	752.26	585.8	123	247.8	741.6	284.4	34.72	51.52	67.2	56
75%	1692.88	3108	6375.6	1298.64	531.86	2669.74	1803.8	1380.4	175.2	580.8	1824	621	40.32	91.84	125.44	97.44
90%	2687.44	7762.16	13788.9	2972.48	856.08	6679.86	3823.94	3678.94	242.4	1413.6	3898.2	1275	47.04	169.12	212.8	166.88
95%	3332.56	14556.1	21989.5	4776.8	1093.88	11118.6	5689.8	6637.52	294.6	2496	6031.2	1924.2	52.64	243.04	293.44	232.96
max	10851.1	179100	161255	39326.6	3665.6	101441	39424.9	55452.6	1304.4	29947.8	47083.2	9603	136.64	1407.84	1186.08	1361.92

TABLE 8. Prediction Variance Percentiles for Third-Order Models for Optimal Designs

	n = p = 10 runs		n = p + 2 = 12 runs		n = p + 4 = 14 runs		n = 2p = 20 runs	
2 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	5.1	5	4.56	4.56	4.48	4.62	3.8	5.2
5%	5.9	6.1	5.4	5.4	5.04	5.88	4.4	6.8
25%	7.1	8.7	6.6	6.72	6.02	7	5.4	8
50%	8.8	10.7	7.44	7.44	7.42	7.28	6.2	9
75%	10.3	13	8.04	8.04	8.12	7.7	7.6	10
90%	12.8	14.7	8.76	8.76	9.24	8.68	9.8	10.4
95%	14.3	15.4	9.6	9.6	10.22	9.52	11.2	10.8
max	40.1	18.7	11.28	11.4	13.16	12.46	17.2	14

	n = p = 20 runs		n = p + 2 = 22 runs		n = p + 4 = 24 runs		n = 2p = 40 runs	
3 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	8	9	8.14	10.34	6.48	8.64	5.2	10
5%	12.4	14.4	10.34	13.42	9.6	13.44	7.6	13.2
25%	14.8	17.8	12.32	17.82	12	16.32	9.2	15.2
50%	17	20.8	14.08	22.88	13.2	18.24	10.8	16.8
75%	21.2	26.2	16.94	31.68	15.36	20.64	14	18.4
90%	26	34	20.24	41.8	18.48	23.28	16.8	20
95%	29.2	39.6	23.1	47.96	20.4	24.96	18.8	21.2
max	66.6	57.2	65.12	61.16	51.6	30.24	42	26.4

	n = p = 35 runs		n = p + 2 = 37 runs		n = p + 4 = 39 runs		n = 2p = 70 runs	
4 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	12.25	18.55	13.32	14.8	10.14	13.65	6.3	16.8
5%	19.6	27.65	19.24	25.16	16.38	23.01	10.5	24.5
25%	23.8	34.65	23.68	31.45	20.28	29.25	14.7	28
50%	28	41.3	27.75	36.63	23.79	33.93	18.2	30.1
75%	34.65	50.75	32.93	42.55	28.47	40.95	23.1	32.9
90%	44.1	62.65	40.7	49.95	34.32	49.53	28	35
95%	52.15	70.7	48.47	55.13	39	54.6	30.8	36.4
max	131.95	105.7	193.51	82.88	97.89	93.6	57.4	43.4

	n = p = 56 runs		n = p + 2 = 58 runs		n = p + 4 = 60 runs		n = 2p = 112 runs	
5 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	21.28	35.28	14.5	34.8	13.8	28.2	7.84	26.88
5%	33.6	56.56	28.42	50.46	25.2	45	15.68	40.32
25%	42	70	36.54	63.8	33.6	55.2	21.28	45.92
50%	49.84	82.88	44.08	77.14	40.8	64.8	25.76	49.28
75%	62.16	100.24	54.52	92.8	51	74.4	32.48	52.64
90%	78.4	117.6	69.02	106.72	63.6	84.6	40.32	54.88
95%	90.72	127.12	80.04	114.84	73.8	92.4	44.8	57.12
max	276.08	198.24	262.16	176.32	205.2	145.2	96.32	68.32

Appendix B
Prediction Variance for Fourth-Order Models

TABLE 9. Prediction Variance Percentiles for Fourth-Order Models for Space-Filling Designs

2 Factors	n = p = 15 runs				n = p + 2 = 17 runs				n = p + 4 = 19 runs				n = 2p = 30 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	9.15	8.25	7.5	8.55	7.31	6.97	6.46	7.31	7.6	8.17	6.27	7.41	6.9	6.6	6.6	6.9
5%	13.5	9.9	19.5	10.95	8.16	9.35	9.52	8.67	8.93	9.12	9.5	7.98	7.5	7.5	7.5	7.5
25%	25.05	13.5	93	15.9	11.05	11.9	15.98	11.05	11.02	11.4	13.11	9.69	9.3	9	11.4	9
50%	43.5	18	381	24	14.45	15.47	23.12	13.6	13.68	14.06	16.72	12.35	11.4	11.4	15	10.5
75%	81.9	30.45	1459.2	49.8	17.68	20.91	53.38	18.53	18.24	18.24	23.37	14.82	13.8	15.3	23.7	13.2
90%	125.55	80.1	4535.1	101.25	23.29	40.46	162.69	24.14	23.75	32.3	66.31	18.05	15.6	27.9	63.3	17.4
95%	148.65	142.8	6958.5	133.8	30.77	100.13	284.58	29.92	27.74	57.19	131.67	19.57	17.4	44.1	112.5	19.5
max	723.9	1526.55	100142.3	246.9	88.74	1301.18	2613.41	92.14	81.89	702.62	1134.49	108.3	81.6	410.4	1499.4	82.5

3 Factors	n = p = 35 runs				n = p + 2 = 37 runs				n = p + 4 = 39 runs				n = 2p = 70 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	19.6	17.85	20.3	20.3	16.28	19.98	19.98	15.54	16.38	19.89	12.87	17.55	14	13.3	11.9	13.3
5%	43.75	37.1	46.2	42.35	25.9	50.69	172.79	25.53	23.4	24.96	36.27	22.62	16.1	15.4	16.8	15.4
25%	91.35	78.4	117.25	84.7	37.37	301.55	1515.15	34.04	33.15	33.54	97.11	31.2	21	22.4	25.2	19.6
50%	184.45	159.6	271.95	182.35	54.76	1152.92	7647.9	42.18	43.29	49.14	231.66	41.73	25.9	28.7	39.2	23.8
75%	345.45	455.7	760.9	386.05	92.5	3953.45	39744.29	57.35	61.23	100.62	703.56	57.33	30.8	46.2	75.6	29.4
90%	527.1	1537.2	2603.3	763.7	172.42	14236.12	162511.4	85.1	85.41	284.7	1904.76	75.27	35	93.8	151.9	35
95%	663.95	3165.4	5116.3	1199.45	238.28	37415.14	428477.02	103.6	107.25	531.57	3355.56	88.92	39.2	162.4	235.9	40.6
max	3997.7	142059.1	67863.95	9359	530.21	529365.66	11992464	430.68	376.74	13378.6	43967.8	301.47	111.3	2429	3818.5	93.8

4 Factors	n = p = 70 runs				n = p + 2 = 72 runs				n = p + 4 = 74 runs				n = 2p = 140 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	63.7	41.3	51.8	46.9	46.8	46.8	38.88	41.76	40.7	44.4	31.82	38.48	23.8	23.8	21	23.8
5%	283.5	98.7	219.8	249.2	113.76	97.92	207.36	123.12	71.78	84.36	111	71.78	30.8	30.8	33.6	29.4
25%	1005.9	293.3	788.9	864.5	285.84	171.36	794.88	285.84	111.74	161.32	380.36	139.12	39.2	44.8	49	39.2
50%	2943.5	833.7	2517.2	2114	748.8	287.28	2354.4	527.76	158.36	312.28	1046.36	247.16	47.6	60.2	81.2	50.4
75%	7675.5	2854.6	8088.5	5134.5	1990.8	632.16	7527.6	1047.6	232.36	689.68	2959.26	503.94	57.4	117.6	166.6	70
90%	15145.2	9442.3	23260.3	13273.4	4134.96	2129.04	20486.16	2332.08	330.04	2093.46	7069.96	1033.78	70	250.6	354.2	103.6
95%	20278.3	18734.1	42473.2	29230.6	6264	4611.6	38268	4203.36	433.64	4849.96	12917.4	1725.68	79.8	390.6	560	142.8
max	137232.9	246286.6	470064	820859	35667.4	136063.44	674393.04	53618.4	1669.4	159974	211978	41193.6	236.6	2839.2	7445.2	995.4

5 Factors	n = p = 126 runs				n = p + 2 = 128 runs				n = p + 4 = 130 runs				n = 2p = 252 runs			
	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	104.58	118.44	149.94	134.82	97.28	101.12	120.32	81.92	93.6	83.2	83.2	63.7	42.84	45.36	32.76	37.8
5%	1208.34	395.64	1116.36	496.44	272.64	245.76	1713.92	486.4	219.7	184.6	370.5	263.9	52.92	63	50.4	52.92
25%	11066.58	1253.7	4315.5	1549.8	608	519.68	10278.4	1569.28	384.8	319.8	1235	711.1	70.56	85.68	90.72	78.12
50%	45051.3	3763.62	12233.34	3693.06	1162.24	1107.2	36746.24	3988.48	568.1	552.5	3146	1604.2	85.68	115.92	158.76	123.48
75%	132699.4	12389.58	35964.18	9477.72	2241.28	3120.64	130310.4	11581.4	841.1	1353.3	8676.2	4035.2	103.3	229.32	320.04	231.84
90%	272734.6	36169.56	92985.48	23291.1	3747.84	8840.96	435159.04	32389.1	1192.1	3742.7	21361.6	9921.6	123.5	501.48	612.36	466.2
95%	385675.9	69117.3	159935.6	39027.2	5053.44	16505.6	859343.36	55932.2	1467.7	6696.3	34790.6	16422.9	136.1	778.68	904.68	715.68
max	2476063	1066477	2256729	873213	18741.8	371531.52	16379917	925300	11617	86529.3	542974	221402	473.8	7020.72	4394.88	6688.08

TABLE 10. Prediction Variance Percentiles for Fourth-Order Models for Optimal Designs

	n = p = 15 runs		n = p + 2 = 17 runs		n = p + 4 = 19 runs		n = 2p = 30 runs	
2 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	8.25	7.95	8.16	8.5	8.36	6.84	6.9	8.1
5%	10.35	10.95	9.18	10.88	9.12	11.21	7.2	11.1
25%	11.85	13.5	10.03	12.24	10.07	12.54	8.1	12.3
50%	12.9	15.3	11.39	13.43	11.4	13.3	9.9	13.2
75%	14.1	17.7	13.6	15.13	12.92	15.01	11.1	14.7
90%	15	20.85	15.98	18.87	14.25	17.1	14.4	15.9
95%	15.75	23.1	17.17	22.27	15.58	18.43	16.8	16.2
max	50.4	29.1	45.9	26.35	22.42	19.95	25.5	19.2

	n = p = 35 runs		n = p + 2 = 37 runs		n = p + 4 = 39 runs		n = 2p = 70 runs	
3 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	18.9	17.15	17.02	17.02	16.77	18.72	13.3	18.9
5%	23.8	25.2	19.98	24.79	18.33	23.4	14	27.3
25%	26.25	34.3	24.42	29.97	22.62	28.47	16.8	30.1
50%	28.35	46.55	27.75	37	25.74	31.98	20.3	31.5
75%	31.5	66.15	31.08	48.47	29.64	37.05	25.2	33.6
90%	36.05	78.4	34.78	58.83	35.88	44.07	30.8	35.7
95%	40.6	82.95	38.11	65.49	40.95	47.97	33.6	36.4
max	146.3	92.05	167.98	78.44	84.63	60.45	63.7	40.6

	n = p = 70 runs		n = p + 2 = 72 runs		n = p + 4 = 74 runs		n = 2p = 140 runs	
4 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	35	37.8	32.4	35.28	29.6	44.4	19.6	39.2
5%	44.1	65.1	38.16	58.32	35.52	56.98	22.4	54.6
25%	50.4	81.9	44.64	69.84	44.4	67.34	29.4	58.8
50%	56	97.3	51.84	79.92	51.06	77.7	36.4	61.6
75%	65.8	116.2	61.2	94.32	60.68	92.5	46.2	65.8
90%	81.9	140	74.16	112.32	72.52	108.78	56	70
95%	94.5	157.5	85.68	126	82.14	120.62	64.4	72.8
max	323.4	277.2	328.32	240.48	217.56	201.28	123.2	85.4

	n = p = 126 runs		n = p + 2 = 128 runs		n = p + 4 = 130 runs		n = 2p = 252 runs	
5 Factors	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	60.48	79.38	55.04	92.16	48.1	85.8	27.72	78.12
5%	75.6	128.52	67.84	130.56	63.7	136.5	32.76	103.32
25%	86.94	158.76	81.92	158.72	79.3	171.6	47.88	110.88
50%	100.8	191.52	97.28	183.04	93.6	206.7	60.48	115.92
75%	126	238.14	119.04	212.48	115.7	247	75.6	123.48
90%	158.76	293.58	149.76	245.76	145.6	288.6	95.76	128.52
95%	186.48	327.6	175.36	267.52	171.6	313.3	108.36	133.56
max	478.8	539.28	390.4	448	438.1	471.9	178.92	156.24

Appendix C
Prediction Variance for Fifth-Order Models

TABLE 11. Prediction Variance Percentiles for Fifth-Order Models for Space-Filling Designs

	n = p = 21 runs				n = p + 2 = 23 runs				n = p + 4 = 25 runs				n = 2p = 42 runs			
2 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	13.23	17.64	11.55	10.92	8.28	10.81	9.2	8.74	11	8.75	9	9.25	9.24	9.24	8.4	7.56
5%	149.52	879.69	23.1	18.69	16.1	13.34	18.86	11.96	13.5	12	13.5	12.5	11.34	10.92	10.5	11.34
25%	2946.3	19463.85	49.77	40.53	36.34	17.94	45.54	15.64	15.75	15.75	25.75	16.25	13.02	12.18	13.86	16.38
50%	15259.44	71401.26	129.36	90.3	101.43	28.75	111.32	23.23	20.25	25.25	39.5	22.25	15.12	16.38	18.9	23.1
75%	44937.9	187185.6	483.21	249.69	342.7	55.66	314.87	39.33	28.25	46.75	97.75	32.25	20.16	23.52	28.56	33.18
90%	118104.8	616205.52	4902.45	1752.87	857.67	194.12	1282.48	66.93	38.25	102.25	356	45.75	25.62	48.72	65.94	90.72
95%	179711.1	4170833.31	20081.67	4100.88	1116.19	420.21	4440.84	119.14	51.25	214	779.5	61.75	28.56	87.78	120.96	242.76
max	506326.2	248090098	593927.25	204561.21	7233.04	2288.96	141624.11	608.12	4466.75	1876.75	8435	165.5	147.42	1257.48	2284.38	4552.38

	n = p = 56 runs				n = p + 2 = 58 runs				n = p + 4 = 60 runs				n = 2p = 112 runs			
3 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	33.6	29.68	34.72	33.04	31.32	26.68	33.06	29.58	31.8	28.2	28.8	24.6	19.04	17.92	20.16	17.92
5%	124.32	82.32	200.48	94.64	51.62	62.06	157.18	51.04	46.2	44.4	62.4	45.6	25.76	24.64	28	25.76
25%	346.08	226.8	806.4	368.48	98.02	301.6	1064.88	89.32	70.8	78	229.8	64.2	31.36	31.36	36.96	31.36
50%	778.96	708.4	3236.8	1178.8	215.76	1610.08	4636.52	175.16	115.8	151.2	777	94.2	38.08	48.16	53.76	40.32
75%	1819.44	2647.12	16870	3411.52	583.48	7699.5	19263.54	408.32	193.2	408	2739.6	145.8	52.64	78.4	112	53.76
90%	3690.96	11039.28	68297.6	7162.4	1366.48	27347	78008.26	780.68	312	1397.4	8037.6	231.6	67.2	165.76	290.08	69.44
95%	5230.4	26107.2	143396.96	9900.8	2249.82	53730.04	175367.06	1050.96	423	3327	15192.6	316.2	75.04	319.2	545.44	80.64
max	25872.56	758089.36	2139024.7	107511.04	16493.5	3843412.3	13602403	4464.26	5176.8	67502.4	252365.4	2082.6	225.12	4992.96	13533	313.6

	n = p = 126 runs				n = p + 2 = 128 runs				n = p + 4 = 130 runs				n = 2p = 252 runs			
4 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	85.68	88.2	177.66	83.16	75.52	99.84	81.92	69.12	70.2	87.1	91	72.8	35.28	40.32	30.24	32.76
5%	315	355.32	2707.74	289.8	207.36	403.2	499.2	389.12	169	174.2	353.6	166.4	52.92	55.44	52.92	52.92
25%	892.08	1179.36	21297.78	887.04	444.16	1771.52	2795.52	1409.28	323.7	370.5	1471.6	384.8	68.04	80.64	85.68	70.56
50%	1906.38	3262.14	110243.7	2268	783.36	6224.64	9758.72	3680	538.2	813.8	4904.9	756.6	88.2	113.4	148.68	93.24
75%	4204.62	10032.12	542542.14	6311.34	1413.12	20897.28	36211.2	9402.88	946.4	2364.7	16775.2	1576.9	113.4	199.08	317.52	128.52
90%	8133.3	30134.16	2009925.5	16049.88	2288.64	87644.16	119863.04	23077.12	1567.8	9426.3	49907	3250	141.12	501.48	761.04	191.52
95%	11463.48	63653.94	4049973.9	28322.28	2993.92	223239.68	235362.56	39828.48	2063.1	23072.4	94069.3	5586.1	158.76	970.2	1333.08	277.2
max	45130.68	8455909.14	81253087	328589.1	17006.1	12150177	4290402.6	509726.72	10179	1083313	1012513	82513.6	771.12	13255.2	15276.2	4465.44

	n = p = 252 runs				n = p + 2 = 254 runs				n = p + 4 = 256 runs				n = 2p = 504 runs			
5 Factors	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME	SP	LH	U	ME
min	380.52	289.8	420.84	395.64	200.66	292.1	472.44	233.68	240.64	230.4	296.96	202.24	55.44	65.52	45.36	55.44
5%	2797.2	1990.8	7101.36	3757.32	1805.94	1442.72	4310.38	1856.74	1090.56	716.8	1856	906.24	100.8	110.88	100.8	100.8
25%	10216.08	6894.72	42139.44	17158.68	5471.16	4831.08	22176.74	9438.64	3038.72	1984	7257.6	3100.16	131.04	166.32	181.44	156.24
50%	26094.6	18529.56	155582.28	56130.48	14879.3	13883.64	78983.84	35257.74	6195.2	4574.72	21985.28	8714.24	171.36	241.92	327.6	252
75%	64355.76	57289.68	556479	202882.68	38150.8	43058.08	256971.8	131406.9	12815.4	13857.28	70108.16	26229.76	221.76	483.84	705.6	504
90%	147034.4	197159.76	1822716	662946.48	82245.2	138739.88	722591.9	409285.44	24673.3	49064.96	185216	70643.2	282.24	1219.68	1547.28	1038.24
95%	231847.6	397771.92	3565192.7	1299329.6	127836	288709.1	1322669.4	762142.24	36167.7	98309.12	312650.2	126855.7	322.56	2283.12	2484.72	1648.08
max	1445273	9203415.48	75758337	31106676	592958	8632370.3	18877778	23734133	220088	3693494	2595438	1904297	1249.92	27261.4	24590.2	20442.2

TABLE 12. Prediction Variance Percentiles for Fifth-Order Models for Optimal Designs

2 Factors	n = p = 21 runs		n = p + 2 = 23 runs		n = p + 4 = 25 runs		n = 2p = 42 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	12.81	12.18	12.19	12.65	11.75	13.25	9.66	11.76
5%	15.12	16.59	13.57	15.18	12.75	15.75	11.34	16.8
25%	17.22	21	15.64	17.71	14.5	17.75	12.18	18.06
50%	18.48	24.78	17.25	20.24	16	19.5	13.44	18.9
75%	20.37	32.76	19.32	23.92	19	21.25	16.8	19.74
90%	21.84	40.11	22.31	28.06	22.5	23.25	21	21
95%	23.52	44.31	24.15	30.82	24.75	24.5	24.78	21.84
max	98.7	49.98	38.87	36.8	69.5	31.25	38.22	26.04

3 Factors	n = p = 56 runs		n = p + 2 = 58 runs		n = p + 4 = 60 runs		n = 2p = 112 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	28	31.92	30.16	31.9	28.8	28.8	16.8	32.48
5%	40.88	48.72	37.12	42.92	36	45	23.52	43.68
25%	46.48	59.36	43.5	52.78	40.8	52.2	26.88	48.16
50%	50.4	71.12	48.14	59.16	45	58.2	33.6	50.4
75%	55.44	87.92	54.52	67.86	49.8	64.2	42.56	53.76
90%	64.96	112	63.22	77.14	57	71.4	51.52	56
95%	75.04	131.04	71.34	83.52	63.6	75.6	58.24	57.12
max	421.68	254.24	234.32	121.8	264	107.4	99.68	64.96

4 Factors	n = p = 126 runs		n = p + 2 = 128 runs		n = p + 4 = 130 runs		n = 2p = 252runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	64.26	86.94	60.16	84.48	59.8	81.9	27.72	70.56
5%	90.72	132.3	84.48	125.44	81.9	120.9	42.84	103.32
25%	105.84	165.06	98.56	154.88	94.9	152.1	52.92	113.4
50%	117.18	196.56	111.36	184.32	106.6	183.3	65.52	118.44
75%	137.34	241.92	133.12	220.16	123.5	224.9	85.68	126
90%	167.58	296.1	165.12	258.56	150.8	275.6	105.84	131.04
95%	194.04	332.64	189.44	281.6	171.6	308.1	120.96	133.56
max	921.06	643.86	742.4	483.84	553.8	553.8	219.24	148.68

5 Factors	n = p = 252 runs		n = p + 2 = 254 runs		n = p + 4 = 256 runs		n = 2p = 504 runs	
	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt	I Opt	D Opt
min	128.52	231.84	119.38	228.6	122.88	202.24	45.36	146.16
5%	186.48	345.24	175.26	347.98	171.52	317.44	70.56	211.68
25%	216.72	443.52	203.2	447.04	199.68	399.36	95.76	226.8
50%	254.52	529.2	241.3	528.32	230.4	465.92	120.96	236.88
75%	317.52	637.56	302.26	622.3	279.04	552.96	156.24	252
90%	410.76	750.96	391.16	716.28	355.84	640	196.56	262.08
95%	493.92	834.12	464.82	774.7	424.96	701.44	226.8	267.12
max	1479.24	1353.24	2369.82	1170.94	1648.64	1108.48	423.36	302.4

References

Allen, T. T.; Bernshteyn, M. A.; and Kabiri-Bamoradian, K. (2003). "Constructing Meta-Models for Computer Experiments". *Journal of Quality Technology* 35(3), pp. 264–274.

Atkinson, A. and Donev, A. (1992). *Optimum Experimental Designs*. Oxford: Oxford Science Publications.

Bayarri, M. J.; Berger, J. O.; Paulo, R.; Sacks, J.; Cafeo, J. A.; Cavendish, J.; Lin, C. H.; and Tu, J. (2007). "A Framework for Validation of Computer Models". *Technometrics* 49(2), pp. 138–154.

Bursztyn, D. and Steinberg, D. M. (2006). "Comparison of Designs for Computer Experiments". *Journal of Statistical Planning and Inference* 136, pp. 1103–1119.

Chen, V.; Tsui, K.-L.; Barton, R.; and Meckensheime, M. (2006). "A Review on Design, Modeling and Applications of Computer Experiments". *IEE Transactions* 38, pp. 273–291.

Danehy, P.; DeLoach, R.; and Cutler, A. (2002). "Application of Modern Design of Experiments to CARS Thermometry in a Supersonic Combustor." AIAA-2002-2914 40th AIAA Aerospace Sciences Meeting, 2002.

Fang, K. T. (1980). "The Uniform Design: Application of Number-Theoretic Methods in Experimental Design". *Acta Mathematicae Applicatae Sinica* 3, pp. 363–372.

Fang, K. T.; Li, R.; and Sudjianto, A. (2006). *Design and Modeling for Computer Experiments*. Boca Raton, FL: Taylor & Francis Group.

Hickernell, F. J. (1998). "A Generalized Discrepancy and

Quadrature Error Bound". *Mathematics of Computation* 67, pp. 299–322.

HUSSAIN, M. F.; BARTON, R. R.; and JOSHI, S. B. (2002). "Metamodeling: Radial Basis Functions, Versus Polynomials". *European Journal of Operational Research* 138, pp. 142–154.

JANK, W. and SHMUELI, G. (2007). "Modelling Concurrency of Events in On-Line Auctions via Spatiotemporal Semiparametric Models". *Applied Statistics* 56, pp. 1–27.

JOHNSON, M. E.; MOORE, L. M.; and YLVISAKER, D. (1990). "Minimax and Maximin Distance Design". *Journal for Statistical Planning and Inference* 26, pp. 131–148.

JOHNSON, R. T.; PARKER, P. A.; MONTGOMERY, D. C.; CUTLER, A. D.; DANEHY, P. M.; and RHEW, R. D. (2009). "Design Strategies for Response Surface Models for the Study of Supersonic Combustion". *Quality and Reliability Engineering International* 25, pp. 365–377.

KO, C.-W.; LEE, J.; and QUEYRANNE, M. (1995). "An Exact Algorithm for Maximum Entropy Sampling". *Operations Research* 43, pp. 684–691.

LIEFVENDAHL, M. and STOCKI, R. (2006). "A Study on Algorithms for Optimization of Latin Hypercubes". *Journal of Statistical Planning and Inference* 136, pp. 3231–3247.

MCKAY, N. D.; CONOVER, W. J.; and BECKMAN, R. J. (1979). "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code". *Technometrics* 21, pp. 239–245.

MEASE, D. and BINGHAM, D. (2006). "Latin Hyperrectangle Sampling for Computer Experiments". *Technometrics* 48(4), pp. 467–477.

MYERS, R. H. and MONTGOMERY, D. C. (2002). *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, 2nd edition. New York, NY: John Wiley.

PARK, S.; FOWLER, J. F.; MACKULAK, G. T.; KEATS, J. B.; and CARLYLE, W. M. (2002). "*D*-Optimal Sequential Experiments for Generating A Simulation-Based Cycle Time—Throughput Curve". *Operations Research* 50, pp. 981–990.

ROUX, W.; STANDER, N.; GUNTHER, F.; and MULLERSCHON, H. (2006). "Stochastic Analysis of Highly Non-Linear Structures". *International Journal for Numerical Methods in Engineering* 65, pp. 1221–1242.

SACKS, J.; WELCH, W. J.; MITCHELL, T. J.; and WYNN, H. P. (1989). "Design and Analysis of Computer Experiments". *Statistical Science* 4(4), pp. 409–423.

SHEWRY, M. C. and WYNN, H. P. (1987). "Maximum Entropy Sampling". *Journal of Applied Statistics* 14, pp. 898–914.

STORLIE, C. B. and HELTON, J. C. (2008). "Multiple Predictor Smoothing Methods for Sensitivity Analysis: Example Results". *Reliability Engineering & System Safety* 93, pp. 55–77.

TEDDER, S.; BIVOLARU, D.; DANEHY, P.; WEILK, M.; BEYRAU, F.; SEEGER, T.; and CUTLER, A. (2007). "Characterization of a Combined CARS and Interferometric Rayleigh Scattering System." AIAA-2007-871 45th AIAA Aerospace Sciences Meeting and Exhibit 2007.

TYRE, A.; KERR, G. D.; TENHUMBERG, B.; and BULL, M. (2007). "Identifying Mechanistic Models of Spatial Behaviour Using Pattern-Based Modelling: An Example from Lizard Home Ranges". *Ecological Modeling* 208, p. 307–316.

WANG, Y. and FANG, K. T. (1981). "A Note on Uniform Distribution and Experimental Design". *KeXue TongBao* 26, pp. 485–489.

WELCH, W. J.; BUCK, R. J.; SACKS, J.; WYNN, H. P.; MITCHELL, T. J.; and MORRIS, M. D. (1992). "Screening, Predicting, and Computer Experiments". *Technometrics* 34(1), pp. 15–25.

ZAHRAN, A.; ANDERSON-COOK, C. M.; and MYERS, R. H. (2003). "Fraction of Design Space to Assess the Prediction Capability of Response Surface Designs". *Journal of Quality Technology* 35, pp. 377–386.