Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes

Thomas M. CIOPPA

U.S. Army Training and Doctrine Command
Analysis Center
Fort Leavenworth, KS 66027
(*thomas.cioppa@us.army.mil*)

Thomas W. LUCAS

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943
(*twlucas@nps.edu*)

This article presents an algorithm for constructing orthogonal Latin hypercubes, given a fixed sample size, in more dimensions than previous approaches. In addition, we detail a method that dramatically improves the space-filling properties of the resultant Latin hypercubes at the expense of inducing small correlations between the columns in the design matrix. Although the designs are applicable to many situations, they were developed to provide Department of Defense analysts flexibility in fitting models when exploring high-dimensional computer simulations where there is considerable a priori uncertainty about the forms of the response surfaces.

KEY WORDS: Experimental design; Latin hypercube; Orthogonal; Simulation; Space-filling.

1. INTRODUCTION

Physical experimentation can be resource intensive in terms of material, time, and money. Therefore, the United States Department of Defense (DoD) often relies on simulation models in its decision making. Among other things, simulation models are used to help test war plans, decide what equipment to acquire, determine weapon mixes, and study doctrine and potential operational concepts (see, e.g., Appleget 1995; Loerch, Pudwill, and LaBarbera 1996; and the Defense Modeling and Simulation Office website: *www.dmso.mil*). Because there is a dearth of data with which to assess the veracity of many of these simulations, especially at the force-on-force level or when investigating future scenarios, they are often used in an exploratory manner. That is, the simulations are used to assist senior leadership in understanding and reasoning about extremely complex systems and processes (Bankes 1993).

The simulations that DoD analysts use are often quite large and almost unimaginably complex. Many contain thousands of input variables, a substantial number of which may be significant. Moreover, many of the input variables (e.g., with whom, where, and how a future conflict may take place) are uncertain. Furthermore, as Saeger and Hinch (2001) showed, the response surfaces may be highly nonlinear. In addition, due to the stochastic nature of combat, most of the simulations include pseudo-random events. Unfortunately, the complexity and uncertainty associated with these simulations makes using strong prior knowledge, such as the distributional form of the error term, unreliable. To efficiently explore these simulations, we want readily available experimental designs that allow us to screen a large number of input variables, fit commonly used main-effects models with nearly uncorrelated coefficient estimates, while providing flexibility to fit complex models (including nonparametric) on a modest number of dominant factors.

To address this goal, this article extends Ye's (1998a) algorithm for constructing orthogonal Latin hypercubes (OLHs) to allow more factors to be included within a fixed sample size. Unfortunately, the space-filling properties of these designs can be poor. A good *space-filling* design is one in which the design points are scattered throughout the experimental region

with minimal unsampled regions. To rectify this potential shortcoming, we present a computer-intensive algorithm to generate Latin hypercubes (LHs) that have good space-filling properties and are nearly orthogonal; for example, all correlations between the columns in the design matrix are in the interval $(-.03, .03)$. It takes extensive time to generate these designs using our algorithm; therefore, a catalogue of ready-to-use, nearly orthogonal and good space-filling designs for up to 22 factors in as few as 129 runs has been given by Cioppa (2002) and available for download at *http://harvest.nps.edu*. The methodology used to construct these designs can be applied to generate designs for more than 22 factors.

The article is organized as follows. Section 2 provides the background and frames the issues that our designs address. Section 3 details our extension of Ye's algorithm to allow more variables in a fixed-sample size OLH. But because these designs may have poor space-filling properties, we present an algorithm in Section 4 that provides dramatic improvement in the space-filling properties by slightly relaxing the orthogonality requirement. Section 5 discusses approaches that we have found useful in extending these designs to situations requiring different combinations of factors and sample sizes. Section 6 illustrates the use of the 22-variable design in an agent-based simulation study of a military peace enforcement scenario. Section 7 gives conclusions and suggests directions for future research.

2. BACKGROUND

In this section we define our notation, discuss building metamodels to better understand the relationship between simulation inputs and outputs, briefly review some designs frequently used in simulation experiments, and examine the measures that we use to assess the quality of our designs.

45

1. REPORT DATE **FEB 2007**	2. REPORT TYPE	3. DATES COVERED **00-00-2007 to 00-00-2007**
4. TITLE AND SUBTITLE **Efficient Nearly Orthogonal and Space-Filling Latin Hypercubes**		5a. CONTRACT NUMBER
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Naval Postgraduate School,Operations Research Department,Monterey,CA,93943**		8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)
12. DISTRIBUTION/AVAILABILITY STATEMENT **Approved for public release; distribution unlimited**		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT		
15. SUBJECT TERMS		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT **unclassified**	b. ABSTRACT **unclassified**	c. THIS PAGE **unclassified**	**Same as Report (SAR)**	**11**	

2.1 Meta-Models

Consider the situation in which a simulation model contains k continuous input variables that we wish to explore with n computational experiments over a rectangular region. Suppose that the model generates a vector of output responses denoted as $\mathbf{y}$. Let the ith input variable be denoted as x_i, with $\mathbf{X}$ representing the $n \times k$ input design matrix and y_j representing an output response from the simulation. To help us understand our simulation models, we often build meta-models to quantify the relationship between the input variables $(x_1, x_2, \ldots, x_k)$ and the output measures. A *meta-model* is a relatively simple function, g, compared with the original simulation, which is constructed given an experimental design and the corresponding responses. A good meta-model is one in which g makes parsimonious use of the variables available and the errors (i.e., the differences between the meta-model and the simulation output) are small.

One of the simplest and most commonly used meta-models is one in which g is a linear combination of the inputs, that is,

$$g(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i. \tag{1}$$

When estimating the coefficients in (1), the precision of the estimates can be adversely affected by collinearity among the input variables (Ryan 1997). If the columns in the design matrix between input variables x_i and x_j are uncorrelated, then the regression estimates of β_i and β_j in (1) are uncorrelated.

For many simulations, a linear meta-model may not sufficiently characterize the response surface. Unfortunately, it takes many more observations to estimate meta-models with curvilinear and interaction terms. For example, suppose that g includes quadratic and bilinear interaction effects, as well as the linear terms, that is,

$$g(\mathbf{x}) = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \sum_{i=k+1}^{2k} \beta_i x_i^2 + \sum_{i=1}^{k-1} \sum_{j>i}^{k} \beta_{i,j} x_i x_j. \tag{2}$$

To have sufficient degrees of freedom to estimate the coefficients in (2), the number n of simulation runs must satisfy $n \geq k + k + \binom{k}{2} + 1$. Therefore, n must grow on the order of k^2. More complicated meta-models require even larger n. However, in practice, typically only a small percentage of the input variables turn out to be significant. For example, in an empirical simulation screening experiment, using a deterministic ecological model, Bettonvil and Kleijnen (1997) found that only 15 of 281 factors considered had effect sizes in their linear with two-factor interactions meta-model above a certain threshold. Furthermore, when constructing meta-models, it is often reasonable to focus the higher-order terms by assuming that significant interactions and nonlinear terms likely include factors that tested significant when fitting a linear response.

2.2 Designs Commonly Used to Explore Simulations

Various designs are available for practitioners to use in simulation studies (Kleijnen, Sanchez, Lucas, and Cioppa 2005). The appropriateness of the various designs depends on many factors, including the number of variables that one wishes to

explore, the number of computational experiments that are feasible, the types of meta-models that one wishes to be able to analyze, whether iterations between the experiments and the analysis are possible, a priori assumptions on the response, and statistical criterion over which the analysts want to optimize. We are especially interested in cases, such as those often found in defense analysis, in which there may be multiple responses of interest and little a priori knowledge about the forms that the response function may take. Thus, we adopt the principle of Santner, Williams, and Notz (2003) for selecting designs that "allow one to fit a variety of models and provide information about all portions of the experimental region." Specifically, we desire to simultaneously be able to efficiently fit linear-effects models over many variables (often for screening purposes) and quite complex models on a handful of dominant factors—all within a constrained number of runs.

Perhaps the most used exploratory designs are factorial and fractional-factorial designs. Unfortunately, the number of runs necessary increases dramatically as the number of factors or levels increases. Highly fractionated (e.g., Plackett and Burman 1946) and supersaturated (Lin 1993) designs can mitigate the "curse of dimensionality"; however, they have extremely poor space-filling properties and severely limit the meta-models that an analyst can examine. Thus they are not sufficient in and of themselves for situations in which the response may be complex—for example, with substantial nonlinearities, higher-order interactions, and changepoints.

Designs commonly used in response surface methodology (see Meyers and Montgomery 2002), such as central composite designs (CCDs), are excellent designs for identifying quadratic and interaction terms. Furthermore, by using highly fractionated two-level designs in constructing a CCD, the number of runs required grows modestly (on the order of k^2) with the number variables. However, these designs do not have good space-filling properties when they are projected into the subspaces determined by a small number of input variables. For example, with only five levels per factor, CCDs lack the granularity provided by designs such as LHs, which limits their effectiveness with such exploratory techniques as regression trees.

Group screening (Dorfman 1943) and sequential bifurcation (Bettonvil 1995) designs are able to screen a numerous factors (e.g., >200) in a relatively few number of runs. However, these approaches require factor sparsity and assume a priori known directional effects of the factors. Chaloner and Verdinalli (1994) provided an excellent expository on Bayesian experimental design and noted that Bayesian statistics can be effective in determining a design and input parameter values. A deficiency of these designs is their reliance on expert opinion and a priori beliefs. This is particularly worrisome in complex high-dimensional problems for which little expert consensus exists. Srivastava's (1975) search linear models assume that the number of important factors is known a priori. This would be a rather rare phenomenon, especially in defense analyses. Frequency domain designs (Schruben 1986) are capable of identifying complex relationships between input factors and output responses but use nonterminating simulations, which are uncommon in defense analyses.

Many different criteria exist for constructing designs. Santner et al. (2003) provided a thorough description of entropy,

mean squared prediction error, and multiple optimization criteria. Johnson, Moore, and Ylvisaker (1990) described minimax and maximin distance designs. Potential drawbacks of these methods include the extensive time that may be needed to construct these designs based on the specified criterion, and the fact that orthogonality may not exist.

Two promising classes of designs are LH designs—particularly the orthogonal variants—and uniform designs. A brief description of each follows.

In LH sampling, the input variables are considered to be random variables with known distribution functions. For each input variable x_i, "all portions of its distribution [are] represented by input values" by dividing its range into "n strata of equal marginal probability $1/n$, and [sampling] once from each stratum" (McKay, Beckman, and Conover 1979). In practice, many analysts take a fixed value in each stratum (e.g., the median) rather than a random value or sample from a discrete uniform distribution (see Koehler and Owen 1996). For each x_i, the n sampled input values are assigned at random to the n cases, with all $n!$ possible permutations being equally likely. This determines the column in the design matrix for x_i and is done independently for each of the k input variables. Therefore, for each variable x_i, all of the n input values appear once and only once in the design. Also, for a given row in the design matrix, all of the n^k potential combinations of the input variable values have an equal chance of occurring. A great strength of basic LHs is that they are easy to generate for any k and n; indeed, many simulation software packages include them. When the input factors are given uniform distributions, LHs tend to have reasonable space-filling properties. In fact, when a discrete uniform is used on the input variables, the one-dimensional projections are optimum space-filling designs. Furthermore, when the number of input combinations (n) is sufficiently large with respect to the number of input variables (k), there likely will be small correlations between columns in the design matrix (see Owen 1994; Lucas, Sanchez, Brown, and Vinyard 2002). When n is not large with respect to k, somewhat restrictive OLHs do exist; we discuss this later.

Fang, Lin, Winker, and Zhang (2000a) defined a *uniform design* as one "that allocates experimental points [which are] uniformly scattered on the domain"; that is, they are good space-filling designs. Uniform designs can be particularly useful when identifying thresholds or fitting nonparametric surfaces (Fang and Wang 1994). However, these designs do not require orthogonality and can be difficult to obtain; that is, they are difficult to generate, and relatively few situations are cataloged.

2.3 Measures for Assessing Designs

Our objective is to generate and catalog designs that are (1) readily available, like LHs; (2) are orthogonal, like fractional factorial designs, or at least "nearly orthogonal," and (3) have good space-filling properties, like uniform designs. We use two measures to assess both space-filling and degree of nonorthoganility, allowing for enhanced discrimination when choosing among candidate designs. In particular, ties in one measure are often broken by the other measure.

An orthogonal design is desirable because it gives uncorrelated estimates of the coefficients in a linear regression model

and enhances the performance of many other procedures, such as classification and regression tree models (Kim and Loh 2003). Unless specifically stated otherwise, when we refer to a design matrix, we mean only the matrix composed of k columns (one for each unique input variable, i.e., not including functions of the input variables, like quadratics) and n rows (which specify the levels of values taken by the input variables).

Our first degree of nonorthogonality measure is the maximum$|\rho_{ij}|$, over all i, j such that $i \neq j$, with ρ_{ij} the pairwise correlation between columns x_i and x_j. We refer to this measure as the maximum pairwise correlation and denote it as ρ_{max}. The second measure of orthogonality is a condition number of $X^T X$, where X is the $n \times k$ design matrix. Condition numbers are commonly used in numerical linear algebra to examine the sensitivities of a linear system (Golub and Van Loan 1983). They also can reveal the degree of nonorthogonality for a candidate design matrix. The condition number that we use is defined as $cond(X^T X) = \psi_1 / \psi_n$, where ψ_1 and ψ_n are the largest and smallest eigenvalues of $X^T X$ after the columns of X are centered to sum to 0 and scaled to the range $[-1, 1]$. A $cond(X^T X)$ value of 1 indicates an orthogonal design matrix. A large condition number indicates that the candidate design matrix may be ill-conditioned. We seek a condition number as close to 1 as possible.

There is not a one-to-one correspondence between ρ_{max} and $cond(X^T X)$, but the condition number is related to the number of the pairs of columns that are correlated and to the magnitudes of the correlations. One measure, ρ_{max}, gives the worst correlation between design matrix columns, whereas the other measure, $cond(X^T X)$, assesses the overall degree of nonorthogonality of the design matrix. A design matrix will be classified as *nearly orthogonal* if it has a maximum pairwise correlation no greater than .03 and a condition number no greater than 1.13. Although these values are somewhat arbitrary, designs meeting them suffer minimal collinearity effects, and we show that good space-filling designs exist with this degree of nonorthogonality.

The two measures that we use to assess the space-filling of a design matrix are the modified L_2 discrepancy and the Euclidean maximin distance. In uniform design theory, *the L_∞ discrepancy*, equivalent to the Kolmogorov–Smirnov statistic, is usually used to assess the space filling of a design (Fang and Wang 1994). Fang et al. (2000a) stated that "this is probably the most commonly used measurement for discrepancy... and has been universally accepted in quasi–Monte Carlo methods and number theoretic methods." Unfortunately, as they noted, "one disadvantage of [this] measure is that it is expensive to compute." When the L_∞ discrepancy is too computationally burdensome, as for the designs considered herein, the *modified L_2 discrepancy* [ML_2; eq. (3), where designs are normalized to $[0, 1]$ in each dimension], is often used instead (Hickernell 1998; Fang, Ma, and Winker 2000b). This article uses ML_2 discrepancy to measure the space filling of a design, with smaller values preferred over large ones,

$$ML_2 = \left(\frac{4}{3}\right)^k - \frac{2^{1-k}}{n}\sum_{d=1}^{n}\prod_{i=1}^{k}(3 - x_{di}^2)$$

$$+ \frac{1}{n^2}\sum_{d=1}^{n}\sum_{j=1}^{n}\prod_{i=1}^{k}[2 - \max(x_{di}, x_{ji})]. \quad (3)$$

The second measure is the Euclidean maximin (Mm) distance (Johnson et al. 1990; Morris and Mitchell 1995). For a given design, define a distance list $\mathbf{d} = (d_1, d_2, \ldots, d_{[n(n-1)]/2})$, where the elements of $\mathbf{d}$ are the Euclidean distances between the n design points, ordered from smallest to largest. The *Euclidean Mm distance* is defined as d_1; a larger value is better. A large value of d_1 means that no two points are within d_1 of each other. Note that the Mm distances that we report here are for designs scaled to the domain $[-1, 1]^k$.

3. ORTHOGONAL LATIN HYPERCUBES WITH ADDITIONAL VARIABLES

The designs we develop in this article build on previous work in generating "good" LHs. The random elements in the construction of LHs means that any given realization may have poor design properties, for example, a high ρ_{max}. This is particularly likely when k and n are small. A common solution is to generate many random LHs and select the one with the best properties, such as the minimum ρ_{max}. Others (e.g., Florian 1992; Owen 1994) have developed algorithms that may reduce the off-diagonal correlations of a LH design matrix. Ye (1998a) went farther, deriving a method that generates OLHs (i.e., $\rho_{max} = 0$), when the number of runs is a power of 2 plus 1 (the plus 1 is the center point). Specifically, for any integer $m > 1$, his technique builds OLHs for k variables such that the number n of runs is related to k and m by

$$n = 2^m + 1 \quad \text{and} \quad k = 2m - 2. \quad (4)$$

One limitation of Ye's procedure is that too few factors can be varied in the design, especially for $n \geq 33$. Ye (1998b) noted that it is possible, given n in (4), "that more variables can be accommodated." Indeed, he listed a 17×8 OLH and provided the defining structure for a 33×9 OLH. However, Ye (1998b) remarked that for $m \geq 6$, "we are not able to [construct] an OLH... the maximum number of columns are still given by [our eq. (4)]."

In this section we extend Ye's procedure to construct OLH designs (for $k \leq 67$) that can accommodate more variables for a given sample size when $m \geq 5$. We conjecture that this is also true for $k > 67$. These orthogonal designs build directly from Ye's (1998a) OLH construction. Specifically, the design matrix is augmented with additional columns, thus permitting a greater number of variables in an orthogonal design matrix with the same number of runs. To show this, we first detail Ye's algorithm.

In developing the OLHs, Ye (1998a) constructed three matrices. One matrix, $\mathbf{M}$, has its columns composed of permutations of the ordinal values of the positive levels of the variables. The method assumes that there is an equal number of negative and positive levels for the variables. A second matrix, $\mathbf{S}$, attaches a sign to the levels in the design matrix; it is like a two-level factorial design matrix on $m - 1$ variables containing $m - 2$ interaction terms. All entries in $\mathbf{S}$ are ± 1, and the columns are orthogonal to one another. The third matrix, $\mathbf{T}$, is the elementwise (or Hadamard) product of $\mathbf{M}$ and $\mathbf{S}$. A mirror image of $\mathbf{T}$ and a row of 0's corresponding to the center point are then appended to the original $\mathbf{T}$ to create an OLH. Details now follow.

The dimensions of $\mathbf{M}$ are $q \times k$, with $q = (n - 1)/2$ being the number of positive levels of each variable. The first step in constructing $\mathbf{M}$ is to create a vector $\mathbf{e}$, any ordering of the first q natural numbers $(1, 2, \ldots, q)$. One column in $\mathbf{M}$ is $\mathbf{e}$. Given an initial $\mathbf{e}$, permutation matrices are used to generate the other columns of $\mathbf{M}$. Specifically, for $L = 1, 2, \ldots, m - 1$, create $q \times q$ permutation matrices $\mathbf{A}_L$ as follows. With $\mathbf{I}$ the 2×2 identity matrix and $\mathbf{R} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$, each $\mathbf{A}_L$ is constructed by the following:

$$\mathbf{A}_L = \underbrace{\mathbf{I} \otimes \cdots \otimes \mathbf{I}}_{m-1-L} \otimes \underbrace{\mathbf{R} \otimes \cdots \otimes \mathbf{R}}_{L},$$

where $\otimes$ denotes the Kronecker product. Additional permutation matrices are then created by multiplying distinct pairs of the permutation matrices $\mathbf{A}_1$ through $\mathbf{A}_{m-1}$ by one another.

In Ye's (1998a) algorithm, the $k = 2m - 2$ columns of $\mathbf{M}$ are composed of $\mathbf{e}$, $\mathbf{A}_i\mathbf{e}$ for $i = 1, 2, \ldots, m - 1$, and $\mathbf{A}_i\mathbf{A}_{m-1}\mathbf{e}$ for $i = 1, \ldots, m - 2$. The modification used in our construction is that the columns in $\mathbf{M}$ are, in order, the vector $\mathbf{e}$, the vectors $\mathbf{A}_i\mathbf{e}$, for $i = 1, 2, \ldots, m - 1$, as before, then the vectors $\mathbf{A}_i\mathbf{A}_j\mathbf{e}$, for all i and j such that $i = 1, \ldots, m - 2$ and, for each $i, j = i + 1, \ldots, m - 1$. Thus, Ye used only $m - 2$ of the $\binom{m-1}{2}$ possible pairwise combinations of the permutation matrices $\mathbf{A}_L$ in creating $\mathbf{M}$. Our construction of $\mathbf{M}$ uses all of the pairwise combinations of the $\mathbf{A}_L$ matrices. As a consequence, in our construction, not all permutations of $\mathbf{e}$ generate an OLH.

The number of variables that can be examined by using all pairwise combinations of the $\mathbf{A}_L$'s in $\mathbf{M}$ is as follows.

Fact 1. With n runs, where $n = 2^m + 1$ and m is an integer > 1, the maximum number of variables that can be examined in a LH, using all original and pairwise combinations of the $\mathbf{A}_L$ matrices, is $m + \binom{m-1}{2}$.

The proof is by construction. The vector $\mathbf{e}$ constitutes one variable. Each $\mathbf{A}_L$ yields a column in the design matrix, giving another $m - 1$. Finally, each of the $\binom{m-1}{2}$ pairwise combinations of the $\mathbf{A}_L$ matrices also corresponds to a column in the design matrix.

As an example, for $m = 4, n = 17$, and $k = 7$, with $\mathbf{e} = [1, 2, \ldots, 8]^T$, our corresponding $\mathbf{M}$ is as follows:

e	A_1e	A_2e	A_3e	A_1A_2e	A_1A_3e	A_2A_3e
1	2	4	8	3	7	5
2	1	3	7	4	8	6
3	4	2	6	1	5	7
4	3	1	5	2	6	8
5	6	8	4	7	3	1
6	5	7	3	8	4	2
7	8	6	2	5	1	3
8	7	5	1	6	2	4

For the matrices $\mathbf{M}$ and $\mathbf{S}$ to be conformable for elementwise products, Ye's $\mathbf{S}$ also must be given additional columns. In our construction, the first column of $\mathbf{S}$, labeled j, consists of $+1$, repeated q times, the next $m - 1$ columns of $\mathbf{S}$ are the columns used to estimate the main effects in a 2^{m-1} full factorial design matrix, and the remaining $\binom{m-1}{2}$ columns of $\mathbf{S}$ are the columns used to estimate pairwise interactions in a 2^{m-1} full factorial design matrix. The latter can be obtained by multiplying, element by element, the relevant pair of main effect columns.

Specifically, we extend Ye's approach as follows. For $k = 1, \ldots, m-1$, define the vector $\mathbf{a}_k$ as $\mathbf{a}_k = \mathbf{B}_1 \otimes \mathbf{B}_2 \otimes \cdots \otimes \mathbf{B}_{m-1}$, where $\mathbf{B}_{m-k} = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ and $\mathbf{B}_i = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ for $i \neq m-k$. The resulting $m + \binom{m-1}{2}$ columns of S are $\mathbf{j}$, $\mathbf{a}_i$ for $i = 1, \ldots, m-1$ and $\mathbf{a}_i \mathbf{a}_j$ for $i = 1, \ldots, m-2, j = i+1, \ldots, m-1$. In Ye's construction, the $\mathbf{a}_i \mathbf{a}_j$ columns are restricted to $i = 1$ and $j = 2, \ldots, m-1$.

Continuing the previous example, our corresponding S is as follows:

J	$\mathbf{a}_1$	$\mathbf{a}_2$	$\mathbf{a}_3$	$\mathbf{a}_1\mathbf{a}_2$	$\mathbf{a}_1\mathbf{a}_3$	$\mathbf{a}_2\mathbf{a}_3$
+1	−1	−1	−1	+1	+1	+1
+1	+1	−1	−1	−1	−1	+1
+1	−1	+1	−1	−1	+1	−1
+1	+1	+1	−1	+1	−1	−1
+1	−1	−1	+1	+1	+1	−1
+1	+1	−1	+1	−1	+1	−1
+1	−1	+1	+1	−1	−1	+1
+1	+1	+1	+1	+1	+1	+1

The matrix T is the elementwise product of M and S. The design matrix is completed by augmenting T with its mirror image and the center point, resulting in a 17×7 LH. We represent an OLH by O_k^n, where n represents the number of runs or experiments and k represents the number of variables. The resulting O_7^{17} design is as follows:

Run	Variable A	Variable B	Variable C	Variable D	Variable E	Variable F	Variable G
1	1	−2	−4	−8	3	7	5
2	2	1	−3	−7	−4	−8	6
3	3	−4	2	−6	−1	5	−7
4	4	3	1	−5	2	−6	−8
5	5	−6	−8	4	7	−3	−1
6	6	5	−7	3	−8	4	−2
7	7	−8	6	2	−5	−1	3
8	8	7	5	1	6	2	4
9	0	0	0	0	0	0	0
10	−1	2	4	8	−3	−7	−5
11	−2	−1	3	7	4	8	−6
12	−3	4	−2	6	1	−5	7
13	−4	−3	−1	5	−2	6	8
14	−5	6	8	−4	−7	3	1
15	−6	−5	7	−3	8	−4	2
16	−7	8	−6	−2	5	1	−3
17	−8	−7	−5	−1	−6	−2	−4

Whether the resulting $(2^m + 1) \times (m + \binom{m-1}{2})$ LH is orthogonal depends on the choice of e. In the case where $n = 17$ and $k = 7$, there are 8! possible permutations of e, resulting in 40,320 possible M matrices. A complete enumeration reveals 143 distinct O_7^{17} designs. Unfortunately, each of the 143 O_7^{17} designs has an Mm distance of 1.47902; thus if only this measure is used, then there is no space-filling distinction between the O_7^{17} designs. Therefore, we also consider the ML_2 discrepancies, which range from .151854 to .173952. The O_7^{17} design generated using $\mathbf{e} = [1, 2, \ldots, 8]^T$ has an ML_2 discrepancy of .173223 (almost, but not quite, the worst ML_2 discrepancy). The choice of e corresponding to the minimum (i.e., preferred) ML_2 discrepancy is $\mathbf{e} = [1, 2, 8, 4, 5, 6, 7, 3]^T$. The two-dimensional projections of the variables with the best space-filling O_7^{17}, as measured by ML_2 discrepancy, are shown in Figure 1.

For any e, Ye's OLH construction guarantees orthogonal designs for n and k as specified by (4). Using Proposition 1, OLHs can be constructed for the number of variables specified in Fact 1.

Proposition 1. If the matrices M, S, and T, constructed as described earlier, use $\mathbf{e} = [1, 2, \ldots, q]^T$, where q represents the number of positive levels, to generate a LH (for up to $m = 12$, i.e., $k = 67$ and $n = 4{,}097$), then the resulting LH is orthogonal.

The proof is by computational verification. That is, we have used this method to construct an OLH for all choices between 2 and 67 variables. It is conjectured that Proposition 1 applies for any (positive integer) value of m.

Table 1 compares the number of variables that can be examined using Ye's designs and the extended orthogonal designs. As with Ye's designs, these extended orthogonal designs have the elementwise square of each column orthogonal to all of the columns in the design matrix, and the elementwise product of every two columns orthogonal to all columns in the design matrix. Thus estimates of the linear effects [coefficients β_i, for $i = 1, \ldots, k$ in eq. (2)] are uncorrelated with the estimates of quadratic and bilinear interaction effects (coefficients β_i, for $i = k+1, \ldots, 2k$ and $\beta_{i,j}$ for $i = 1, \ldots, k-1, j = i+1, \ldots, k$ in eq. (2)]. However, in both our designs and Ye's designs, estimates of quadratic and interaction effects in (2) can be (and usually are) correlated with one another.

Table 1 shows that as the number of levels (n) doubles (less 1, for the center point), Ye's designs can accommodate exactly two more variables. In the new designs, as the number of levels doubles, the corresponding maximum number of variables increases by the previous m. This difference grows dramatically as the number of variables to be explored increases. For example, Ye's approach requires more then 16 million runs to build an OLH for 46 variables.

4. CONSTRUCTING AND CATALOGING NEARLY ORTHOGONAL LATIN HYPERCUBES

Using Proposition 1, $2^m + 1$ by $m + \binom{m-1}{2}$ OLHs exist for $k > 7$; however, their space-filling properties are often quite poor. By relaxing the requirement of orthogonality and considering LH designs that are orthogonal for at least one permutation of e, we can obtain nearly orthogonal designs that have dramatically better space-filling properties than the OLHs constructed using Proposition 1. Toward this end, we present a computationally intensive algorithm that produces nearly orthogonal Latin hypercube (NOLH) designs with improved space-filling properties.

The heuristic optimization used to generate the designs is described in detail later. Briefly, we screen millions of random LHs for good correlation properties. For the random LHs with the best correlation, Florian's (1992) correlation reduction method is iteratively applied. (the App. provides for details on Florian's method.) A large number of random LHs is screened because Florian's procedure provides only limited improvement in ρ_{max} and $\text{cond}(\mathbf{X}^T\mathbf{X})$. By extensive exploratory calculations, we have found that screening for ρ_{max} and $\text{cond}(\mathbf{X}^T\mathbf{X})$ speeds the process and enhances the nonorthogonality measures of the final design matrix. Of those candidate designs that satisfy the near-orthogonality constraints of a maximum pairwise correlation no greater than .03 and a condition number no greater than 1.13, the design with the smallest rank sum of the
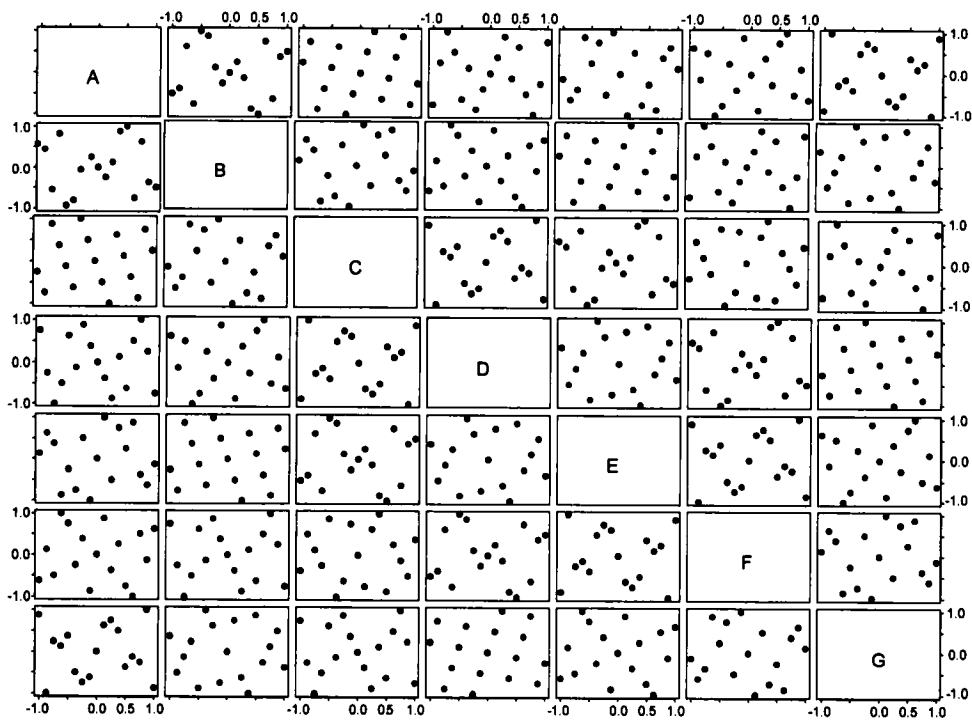
Figure 1. Two-Dimensional Projections of the Best Space-Filling O_7^{17} Design With All of the Variables Scaled to Range From -1 to 1.

Euclidean Mm distance and ML_2 is selected and cataloged. This optimization problem is represented as

$$\text{minimize} \quad f(\text{Mm}, ML_2),$$

$$\text{subject to} \quad \rho_{\max} \leq .03 \tag{5}$$

$$\text{cond}(\mathbf{X}^T\mathbf{X}) \leq 1.13.$$

The constraints in (5) require a nearly orthogonal design. The objective function, f, that we minimize is the rank sum of the two space-filling properties for the designs that meet the constraints.

Experimental designs with near orthogonality are denoted by N_k^n, where N represents near orthogonality, n is the number of runs or experiments, and k is the number of variables. Our algorithm for finding NOLH experimental designs comprises the following steps:

Step 1. Determine the number of variables ($k > 7$) required. If the number of variables is other than 11, 16, 22, or, more generally, $(m + \binom{m-1}{2})$, then round the required number of variables up to the nearest of these numbers.

Step 2. Establish a maximum threshold pairwise correlation value and a maximum threshold condition number. Based on extensive experimentation, we use $\rho_{\max} = .05, .17, .16$ and cond$(\mathbf{X}^T\mathbf{X}) = 1.15, 2.4, 2.8$ for $k = 11, 16,$ and 22.

Step 3. Using a randomly permuted $\mathbf{e}$, construct a design matrix as described in Section 3.

Step 4. Calculate the pairwise correlations and the condition number of the candidate matrix.

Step 5. If either value in Step 4 exceeds the thresholds in Step 2, then discard the design and return to Step 3 to regenerate with another randomly permuted $\mathbf{e}$. Otherwise, keep the design and proceed to Step 6. Repeat Steps 3–5 until a desired number of candidate designs are found. If not enough are found, relax the criteria in Step 2 and begin again. We have found that 15 candidate designs works well for Steps 6–8.

Step 6. Subject each of the candidate designs to repeated applications of Florian's method to decrease the maximum pairwise correlation and condition numbers. Stop when no further improvement is noted.

Table 1. A Comparison Illustrating the Increased Number of Variables That Can Be Examined by Extending Ye's (1998a, 1998b) Construction Algorithm for OLHs

Total number of levels for each variable (n)	m	Maximum number of variables (k) for Ye's (1998a) OLH method	Number of variables (k) in OLH identified by Ye (1998b)	Maximum number of variables (k) by extending Ye's OLH
17	4	6	8	7
33	5	8	9	11
65	6	10	NA	16
129	7	12	NA	22
257	8	14	NA	29
513	9	16	NA	37
1,025	10	18	NA	46

Step 7. Calculate the Mm distance and ML_2 discrepancy for each of the Step 6 designs. Rank the designs according to some combination of these measures. We chose the design with the minimum rank sum over the two measures.

Step 8. If a number of variables other than 7, 11, 16, 22, or $(m + \binom{m-1}{2})$ is required, then construct each possible subset having the appropriate number of columns from the Step 7 design and calculate the Mm distance and ML_2 discrepancy. Choose the design with the best combination of these two measures.

By applying these steps, the best N_{11}^{33}, N_{16}^{65}, and N_{22}^{129} designs (as of this writing) have been identified; the complete designs have been given by Cioppa (2002). For each of these cases, we identified 15 designs that satisfy our definition of near-orthogonality. The one with the best space filling—determined as the minimum rank sum of Mm distance and ML_2 discrepancy—was selected as the best NOLH by the procedure proscribed earlier. Table 2 compares the best NOLH design to the same-sized OLH and an average LH designs. The average LH, denoted "mean LH(k, n)," is empirically determined by generating 1,000 random LHs for the specified n and k and averaging the measures. Table 2 shows that the best NOLH has much better space-filling than the OLH, at the expense of only a slight departure from orthogonality. Moreover, we see that in the ranges of n and k examined, the NOLHs are vastly superior to a random discrete uniform LHs. Finally, the best NOLH has better space-filling properties than an average random LH, which is superior to the OLH.

Except for the O_7^{17} design, there is no guarantee that the N_k^n designs generated from this algorithm are optimal among designs created using the foregoing algorithm (i.e., over all permutations of e). However, these designs are nearly orthogonal and typically have much better space-filling properties than the orthogonal or random LHs. The designs are readily available; recommended designs for from 2 to 22 variables and 17 to 129 runs have been given by Cioppa (2002). An easy-to-use spreadsheet with this family of designs also has been made available by Lucas and Sanchez (2005). This algorithm has many ad hoc features in this algorithm. Many other approaches were tried: the one described above produces good designs in a reasonable amount of time.

5. EXPANDING THE SET OF READILY AVAILABLE NEAR–ORTHOGONAL LATIN HYPERCUBE DESIGNS

We developed the designs constructed in Sections 3 and 4 for specific n and k combinations. In this section we illustrate how effective the designs (obtained using Step 8 from the previous section) with fewer variables can be, and show a useful way to build off of the existing designs if taking more runs is feasible.

5.1 Designs With Fewer Variables

As mentioned in Step 8 of the algorithm (see Sec. 4), we may require a number of variables other than 7, 11, 16, 22, and so on. Of course, when we delete columns, the new (smaller) design is a subset of a design with nearly orthogonal columns. Therefore, collinearity will not be an issue, and consequently, we focus on space-filling properties when selecting which columns to delete. Rather than construct a design by other methods, we assume that a design obtained by eliminating columns from the algorithmically developed (Steps 1–7) design will result in a design with good space-filling properties. One test of this approach on the goodness of our design's space-filling properties is found for designs with 2 variables and 17 levels. Specifically, Table 3 compares three designs: (1) the O_2^{17} design, derived by taking the two columns with the top space-filling properties from the best O_7^{17} design (see Fig. 1); (2) the published optimal uniform design of Fang and Wang (1994); and (3) the design with the best Mm distance measure, taken from Morris and Mitchell (1995).

The space-filling measures of the O_2^{17} design are nearly equal to those of the designs with optimum space-filling properties. Indeed, the O_2^{17} has an ML_2 discrepancy only 15% higher than that of the optimal uniform design and an Mm distance about 3% lower than that of the best maximin design. In fact, the O_2^{17} design has a substantially better Mm distance than the uniform design. It is noteworthy that all of these designs have zero or minimal correlations. Our experience is that when n is large relative to k, designs with good space-filling properties often are close to being orthogonal.

The situation analyzed in Table 3, with $n = 17$ and $k = 2$, is not a situation in which a sophisticated design is typically needed. Indeed, a two factor, four-level full-factorial design, augmented with a center point, is orthogonal and has reasonable space-filling properties (with an Mm distance of .471 and an ML_2 discrepancy of .0295). This example was selected for comparison because there are very few combinations of n and k for which optimum space-filling designs are readily available, especially for large n and k. This is the only condition that we have found in which we can make a direct comparison between

Table 2. Comparing NOLHs to OLHs and Discrete Uniform LHs With Respect to Our Space-Filling and Nonorthogonality Measures

Design	Max pairwise correlation	Condition number	Mm distribution	ML_2
O_{11}^{33}	0	1	1.671	.95
Best N_{11}^{33}	.0234	1.123	1.758	.73
Mean LHC(11, 33)	.4401	8.671	1.295	1.01
O_{16}^{65}	0	1	1.794	7.98
Best N_{16}^{65}	.0219	1.103	2.035	4.46
Mean LHC(16, 65)	.3194	6.103	1.647	5.37
O_{22}^{129}	0	1	1.789	96.6
Best N_{22}^{129}	.0015	1.036	2.265	37.8
Mean LHC(22, 129)	.2332	4.073	1.899	59.8

Table 3. Comparison of the Best O_2^{17}, Uniform, and Mm Distance Designs for the 17-Run, 2-Variable Case

	Maximum pairwise correlation	Condition number	Mm distribution	ML_2
$(O)_2^{17}$ design	0	1	.515	.0025
Uniform design	0	1	.279	.0022
Best Mm distance design	.0588	1.125	.530	.0024

Table 4. A Comparison of Space-Filling and Nonorthogonality Measures for Nine-Level Designs With (about) 33 Levels

Design	Max pairwise correlation	Condition number	Mm distribution	ML_2
Best N_9^{33}	.023	1.100	1.512	.229
Mean LHC(9, 33)	.413	6.085	1.030	.347
Ye best Mm dist O_9^{33}	0	1	1.720	.215
L_{27} Taguchi array	0	1	2.236	2.806

our designs with fewer variables and published optimal Mm distance and uniform designs.

A more interesting comparison is between the suggested N_9^{33} from Cioppa (2002) and similar-sized alternative designs that we might consider. Table 4 shows the nonorthogonality and space-filling measures of the recommended N_9^{33} against the mean of 1,000 9-factor, 33-level random LHs, Ye's (2005) "33 × 9 maximin distance OLH," and a 9-factor (each with 3 levels), 27-run Taguchi array generated by using the JMP Statistical Discovery software (SAS Institute Inc. 2004). We see that the reduced variable N_9^{33} is greatly preferred to a typical random LH. It also performs quite well, although, of course, not quite as good as Ye's maximin distance OLH. A traditional three-level design, such as the L_{27} produced by JMP, has a sparse interior and thus has an ML_2 discrepancy substantially worse than those of any of the LHs and uniform designs. Finally, although not listed in the table (because the full design is not available), Fang (2005) gave the ML_2 discrepancies for uniform designs with up to 29 factors and up to 30 runs. The 9-factor, 30-run design has an ML_2 discrepancy of .180.

5.2 Generating Additional Design Points

It may be desirable to add design points to the original design matrix so as to improve the design's space-filling properties and maintain orthogonality (in the case of seven or fewer variables) or near orthogonality (for more than seven variables). The additional points can be used to test how well a meta-model fit to the original data predicts the new observations. It is possible to add the second-best design from the foregoing process to the original design, then the third-best design, and so on. A much simpler alternative permutes columns of the "best design" and appends them to the "best design." Permuting the columns of a design matrix does not affect its space-filling measures or degree of nonorthogonality. Thus this approach reuses all of the effort that went into generating the best space-filling NOLH and requires that only the best design be cataloged. The center point run is redundant and thus is not repeated. Therefore, if n was the number of runs in the initial design matrix, then the appended design adds $n - 1$ runs. Proposition 2 gives the encouraging result that the maximum pairwise correlation in the appended design is less than or equal to what it is in the original design. Note that because points are added to the original region without an increase in dimensionality, the ML_2 discrepancy usually decreases, and the Mm distance is nonincreasing.

Proposition 2. By permuting the columns of the original NOLH containing n runs and appending these columns to the original NOLH, the number of runs is increased to $(2n - 1)$, and the maximum pairwise correlation is nonincreasing.

Proof. The correlation $r(\mathbf{v}, \mathbf{w})$ between two columns, $\mathbf{v}$ and $\mathbf{w}$, in a design matrix is defined as $r(\mathbf{v}, \mathbf{w}) = [\sum(v_i - \bar{v})(w_i - \bar{w})]/\sqrt{[\sum(v_i - \bar{v})^2 \sum(w_i - \bar{w})^2]}$. Furthermore, without loss of generality, we consider the absolute value of $r(\mathbf{v}, \mathbf{w})$ in determining the maximum pairwise correlation. For a sample size of n, the values in the columns of our LHs take the integer values from $(-n + 1)/2$ to $(n - 1)/2$. Thus, for any column $\mathbf{v}$, $\bar{v} = 0$ and $\sum v_i^2 = [(n - 1)n(n + 1)]/12$. Therefore, for any two columns of $\mathbf{v}$ and $\mathbf{w}$, $r(\mathbf{v}, \mathbf{w}) = \sum v_i w_i / [(n - 1)n(n + 1)/12]$. Now assume that the columns of the design matrix are permuted, and that we append the permuted matrix to the bottom of the initial design matrix to create a new, expanded design matrix. The new columns consist of $n + (n - 1)$ entries. Suppose that columns $\mathbf{x}$ and $\mathbf{y}$ are appended to $\mathbf{v}$ and $\mathbf{w}$. Then the correlation between the two columns is $r_{new}(\mathbf{v} : \mathbf{x}, \mathbf{w} : \mathbf{y}) = [\sum v_i w_i + \sum x_i y_i]/[(n - 1)n(n + 1)/6]$. Note that the denominator of $r_{new}(\mathbf{v} : \mathbf{x}, \mathbf{w} : \mathbf{y})$ is twice that of $r(\mathbf{v}, \mathbf{w})$. Without loss of generality, suppose that maximum pairwise correlation is greater than or equal to the negative of the minimum pairwise correlation. Moreover, suppose that $r(\mathbf{v}, \mathbf{w}) = \rho_{max}$; then $r(\mathbf{x}, \mathbf{y}) \leq r(\mathbf{v}, \mathbf{w})$. Therefore, $r_{new}(\mathbf{v} : \mathbf{x}, \mathbf{w} : \mathbf{y}) \leq r(\mathbf{v}, \mathbf{w})$.

Because the original experimental design is nearly orthogonal, the maximum pairwise correlation value and condition number are generally improved only marginally. Thus, when selecting columns to permute, it seems wise to emphasize space-filling properties. In the O_7^{17} design, an exhaustive enumeration of the 7! column permutations is possible. In finding the best permutation of columns to be appended, the rank sum of the Mm distance and the ML_2 discrepancy are used in the same way as was done above when seeking columns to delete. Exhaustive enumerations of the column permutations for the N_{11}^{33}, N_{16}^{65}, and N_{22}^{129} designs are not feasible. One possibility is to sample randomly from the possible permutations, rank-order the resulting designs for their Mm distances and ML_2 discrepancies, and choose the permutation design with the smallest rank sum. To do this more efficiently, we use a heuristic to narrow the possible permutations for the random sampling. The objective of the heuristic is to identify variables that perform well and poorly on space-filling properties, so poor performers can be appended to good performers.

This is done as follows. The ML_2 discrepancy is calculated for each combination of three variables [e.g., in the N_{11}^{33} design, there are $\binom{11}{3} = 165$ combinations]. The ML_2 discrepancies are then rank-ordered from highest (i.e., worst space-filling) to lowest (i.e., best space-filling). The number of times that each variable appears in a combination with a high ML_2 discrepancy (in the N_{11}^{33} design, this is the upper 82 of the 165 measures) is compared with the number of times that each variable appears in a combination having a low ML_2 discrepancy (e.g., in the N_{11}^{33} design, this is the smallest 82 measures). Under the assumption of randomness that a variable has the same likelihood of appearing in either the upper half or lower half of these combinations, an exact binomial test (see Conover 1999) at the .10 significance level is performed to identify variables that are more likely to appear in the better combinations and those that are more likely to appear in the poorer combinations. For example, if a variable appears in the upper half 94 times and in the lower half 70 times, then the associated p value from the exact

binomial test is .072. This variable is designated as a variable more likely to appear in the upper half (worst space-filling). The best-performing variables are then restricted to being appended to the poorest-performing variables (e.g., the aforementioned variable).

Although other heuristics are possible, this one allows us to quickly find additional design points that improve on both the design's near-orthogonality measures and space-filling properties. A significance level of .10 is chosen (over, say .05) to permit identification of a greater number of variables as good and poor performers. This heuristic has identified good (although not necessarily globally optimal) permutations, whereas random sampling has not found a better permutation in a much greater number of attempts. Cioppa (2002) gave suggested permutations of the columns of the design matrices to append to the best O_7^{17}, N_{11}^{33}, N_{16}^{65}, and N_{22}^{129} designs.

6. EXPLORATORY ANALYSIS OF A PEACE ENFORCEMENT SCENARIO

These new designs were developed so that analysts could conduct computer experiments and explore the output flexibly. This section summarizes one such exploration (see Cioppa 2002).

Recent years have brought an increased emphasis on using military forces for operations other than war, such as peace enforcement. The United States Army Field Manual 100-23 (Department of the Army 1994) describes *peace enforcement* as "the application of military force or the threat of its use, normally pursuant to international authorization, to compel compliance with generally accepted resolutions or sanctions. The purpose of peace enforcement is to maintain or restore peace and support diplomatic efforts to reach a long-term political settlement." Operations of this nature are becoming common for the military. Furthermore, many questions exist about doctrine and tactics for units conducting peace enforcement operations.

To generate hypotheses about light-infantry platoon-level peace enforcement tactics, we explored a scenario of such a platoon clearing an area to facilitate United Nations (UN) food distribution and military convoy operations using the agent-based simulation MANA (Lauren and Stephen 2001). This scenario is a challenging one, because the Blue force (a U.S. Army light-infantry platoon) is subjected to a series of encounters with a hostile indigenous Red force and an originally nonhostile Yellow force that turns hostile as the scenario progresses. This scenario has been deemed doctrinally correct and plausible by the U.S. Army Infantry Simulation Center at Fort Benning, Georgia. To examine the effects of unit cohesion and agent personalities, 22 variables (labeled A–V) were identified for experimentation. These 22 variables were selected from among many available variables, and their levels were chosen based on the author's military experience and judgment and on the results of hundreds of small, interactive experiments. The chosen inputs control the entities' movement capabilities and personalities. Personalities refer to the agents' propensities to move toward or away from scenario objects, such as friendly agents, hostile agents, or a positional goal. These variables affect the simulated units' speed, cohesiveness, and aggressiveness. (See Cioppa 2002 for more details on MANA and the scenario.)

As is often the case with DoD exploratory analysis, the model is not being used to predict potential outcomes. Indeed, due to a lack of data, the model cannot be empirically validated. Rather, the model is used to help us devise new ideas or assess the consequences of certain assumptions. Potential insights gleaned from such exploration usually need to be tested elsewhere, perhaps with field experiments.

Because 22 continuous factors were chosen, the best space-filling N_{22}^{129} design was used. The output on which we focus here is the exchange ratio (ER), the number of Red agents attritted divided by the number of Blue agents attritted; higher is considered better. Due to the high variability of ER for given input values (e.g., coefficient of variation typically >.40), the availability of substantial computing assets, and our interest in examining output variability, for each of the 129 input combinations, 100 iterations were conducted using different random seeds; this gave 12,900 individual runs. A regression equation representing the MANA results was found interactively through trial and error, using visualization and forward and backward stepwise selection (to include quadratic terms and two-variable interactions). The Akaike information criterion (Akaike 1974), sum of squares, and residual plots were the primary measures used to build the model. Additional design points (12,800) were then generated, as specified in Section 4.4, as a way to cross-validate the hypothesized regression equation. This is in keeping with the Department of Defense's policy of model–experiment–model (Piplani, Mercer, and Roop 1994). Here the model is our meta-model. After building the meta-model, we assessed how well it predicts the experiments (in MANA) at new input combinations. This process was continued until the model was deemed "good enough." The results shown in Figure 2 indicate that the differences between the predicted responses and the actual responses were acceptably low in the opinion of military subject matter experts at the U.S. Army Infantry Simulation Center (Cioppa 2002). Thus the first and second experiments are combined, and a regression analysis was executed on the 257 input variable combinations and 25,700 total simulation runs.
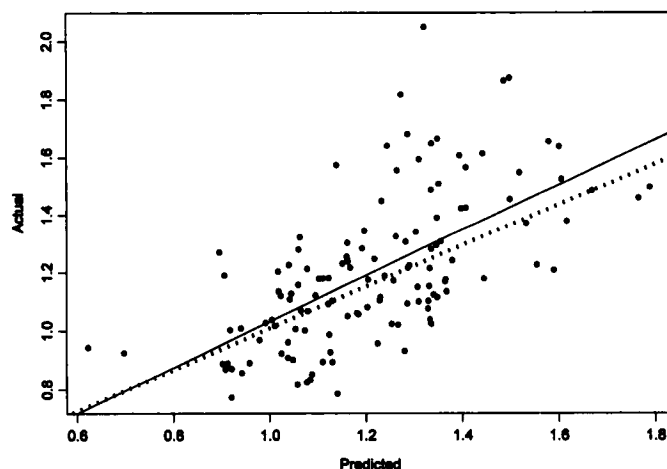

Figure 2. Predicted Values versus Actual Values (second experiment) With Least Squares Fitted Line (—) and Weighted Least Squares Line (·····) for the Mean ERs.

The resulting model, shown in (6), has an R^2 of .67. The fitted exchange ratio as a function of the 22 independent variables is

$$
\begin{aligned}
ER = {} & 1.890 + (1.928e{-}007)U^2 + (.000457)B + (.000736)E \\
& + (.00237)F + (.00568)G + (.000826)P - (.00898)U \\
& - (.00327)V - (4.866e{-}006)BU - (3.021e{-}005)GU \\
& - (2.688e{-}005)FV + (1.378e{-}005)IJ \\
& + (2.225e{-}006)BN. \tag{6}
\end{aligned}
$$

We see that the fitted model has quadratic and interaction terms, including an interaction in which neither variable shows up as a main effect. Guided by the regression, and many visual playbacks, several working hypotheses based on this scenario can be gleaned, including the following:

- Speed of execution (U) and precision of movement (B) are critical to Blue's success.
- When the Blue elements are in contact with the threat, they should consider moving toward other friendly elements (variables E, F, G)—that is, massing its forces.
- When Blue elements have taken casualties and are in contact, continuing the operation instead of ceasing, though simultaneously trying to move toward other friendly units, may be advantageous over the course of the battle (variables I, J, N, P).

It is important to emphasize that in this exploratory analysis, we are primarily trying to identify the factors that have a significant effect on the exchange ratio, along with the directions of those effects and which, if any, factors interact. We considered a quadratic to be sufficient for these purposes. However, for purposes of prediction, a second-order model may be inadequate; for example, a second-order response surface has at most one extremum. Unless one had reason to believe that the actual computer model has this property, perhaps a more flexible model (e.g., spatial models; see Cressie 1993) would be more appropriate.

An unexpected thing that came out of the exploration was the discovery of a software bug over a relatively narrow range of one of MANA's parameters. In particular, the input controlling the agent's movement range (U) was found to not work properly at values of 101–114. Note that this input takes integer values between 0 and 200. These thresholds were easily identified by regression trees. Identifying this bug, and specifying the range of the problem, would have been essentially impossible with a low-level factorial or central composite design. This discovery was reported to the developers of the model and illustrates these designs' potential in verifying models as well as using them. This example illustrates a critical advantage of using the N_{22}^{129} design for exploration over a replicated central composite or other competing design. Specifically, the N_{22}^{129} design provides better resolution on critical thresholds or changepoints. As such, we have found that they are particularly suitable for generating data to be analyzed with regression trees. (See Cioppa 2002 for more on this analysis.)

Although the foregoing exploration emphasized regression and visualization, the data are amenable to analysis by a host of methods. For example, using the N_{22}^{129} design, Ipekci (2002) applied neural networks, regression trees, multiple additive regression trees, and Bayes nets to explore the relationships among the input settings and the outputs.

7. CONCLUSIONS AND AREAS FOR FURTHER RESEARCH

In the DoD and elsewhere, we are increasingly reliant on computer simulations in which there is a vast space of possible computational experiments. In selecting a design, there is a rich literature from which we can draw. Many of the commonly used designs were originally developed for agricultural, industrial, or laboratory experiments. Unfortunately, most of these designs were developed for situations involving a modest number of factors and runs in which strong assumptions are made a priori about the response (e.g., low-order polynomial) and error (e.g., homoscedastic normal error). These situations are often not applicable to computer exploration. Thus we desire readily available designs that allow analysts to explore how well a diverse set of meta-models captures the relationships between many input variables of a simulation and one or more output variables. Toward that end, we have presented an algorithm that generates NOLHs with good space-filling properties. These designs allow an analyst to examine many factors by fitting models with main, quadratic, and interaction effects with nearly uncorrelated estimates of the regression coefficients for the linear effects terms. Having identified the dominant variables, analysts have considerable flexibility in fitting meta-models to them. Furthermore, we have cataloged and implemented in a spreadsheet a set of these designs for from 2 to 22 factors in as few as 129 input combinations, so they are readily available (Cioppa 2002; Lucas and Sanchez 2005; and *http://harvest.nps.edu*).

Two areas related to this research are particularly worthy of exploration. The first area concerns design matrices with a large number of both continuous quantitative and qualitative variables. Currently, when a variable contains fewer levels than runs, the levels are used more than once. For example, if one factor has two levels, say high and low, then all of the positive values in the appropriate column of our NOLH are set to high and the negative values are set to low. The appropriate column is chosen by searching over all columns and selecting the one that provides the best performance with respect to our near-orthogonality and space-filling measures. This method works reasonably well when there are only a handful of qualitative factors (see Cioppa and Brown 2003); a thorough examination over a variety of cases is necessary. The second area of continuing research concerns sequencing, combining, and crossing the proposed designs with full-factorial, fractional factorial, or group screening designs.

ACKNOWLEDGMENTS

APPENDIX: CORRELATION REDUCTION METHOD

In brief, Florian's (1992) method is as follows. For each column of a design matrix **X**, each element is replaced with its

rank within the column. This $n \times k$ matrix is denoted by $\mathbf{W}$. Let $\mathbf{C}$, a $k \times k$ matrix, be the rank correlation matrix of $\mathbf{W}$. If each pair of columns in $\mathbf{W}$ is uncorrelated, then $\mathbf{C}$ is the $k \times k$ identity matrix $\mathbf{I}$. Only those realizations of $\mathbf{W}$ for which $\mathbf{C}$ are positive definite are considered. The basic idea is to transform $\mathbf{W}$ into a set of uncorrelated variates. A Cholesky factorization scheme is used (because $\mathbf{C}$ is positive definite) to determine a lower-triangular matrix, $\mathbf{Q}$, which is $k \times k$. Then let $\mathbf{D} = \mathbf{Q}^{-1}$ and $\mathbf{C} = \mathbf{Q}\mathbf{Q}^T$ such that $\mathbf{D}$ has the property $\mathbf{D}\mathbf{C}\mathbf{D}^T = \mathbf{I}$. The original $\mathbf{W}$ is then transformed into a new matrix by $\mathbf{W}_B = \mathbf{W}\mathbf{D}^T$. Because the elements of the matrix $\mathbf{W}_B$ are not necessarily integral, the elements in each column are replaced by their rank.

Iman and Conover (1980) proved that the difference between corresponding elements in the correlation matrix of $\mathbf{W}_B$ and $\mathbf{I}$ is lower than the analogous difference in $\mathbf{W}$ and $\mathbf{I}$. Because the elements of $\mathbf{W}_B$ are replaced by ranks, this process can be repeated until there is no further decrease in the maximum pairwise correlation or condition number. When applying this procedure iteratively, it is quite common for the maximum pairwise correlation to not change, but the condition number to decrease. Thus, if the procedure uses only the maximum pairwise correlation value, then this iteration process may stop too soon, even though a better design matrix may exist.

REFERENCES

Akaike, H. (1974), "A New Look at the Statistical Model Identification," *IEEE Transactions on Automatic Control*, AC-19, 716–723.

Appleget, J. A. (1995), "The Combat Simulation of Desert Storm With Applications for Contingency Operations," *Naval Research Logistics*, 42, 691–713.

Bankes, S. (1993), "Exploratory Modeling for Policy Analysis," *Operations Research*, 41, 435–449.

Bettonvil, B. (1995), "Factor Screening by Sequential Bifurcation," *Communications in Statistics: Simulation*, 24, 165–185.

Bettonvil, B., and Kleijnen, J. P. C. (1997), "Searching for Important Factors in Simulation Models With Many Factors: Sequential Bifurcation," *European Journal of Operational Research*, 96, 180–194.

Chaloner, K., and Verdinelli, I. (1994), "Bayesian Experimental Design: A Review," Technical Report 599, Carnegie Mellon University, Dept. of Statistics.

Cioppa, T. M. (2002), "Efficient Nearly Orthogonal and Space-Filling Designs for High-Dimensional Complex Models," unpublished doctoral dissertation, Naval Postgraduate School, Monterey, CA; available at *http://library.nps. navy.mil/uhtbin/cgisirsi/Sun+Jul+13+19:36:20+PDT+2003/0/520/02sep_ Cioppa_PhD.pdf*. Accessed May 25, 2004.

Cioppa, T., and Brown, L. (2003), "Agent-Based Models as an Exploratory Analysis Approach to Combat Simulations," technical report, U.S. Army Training and Doctrine Command Analysis Center, Monterey, CA.

Conover, W. J. (1999), *Practical Nonparametric Statistics*, New York: Wiley.

Cressie, N. (1993), *Statistics for Spatial Data*, New York: Wiley.

Defense Modeling and Simulation Office, Joint Staff, available at *www. dmso.mil*. Accessed May 25, 2004.

Department of the Army (1994), *Field Manual 100-23, Peace Operations*, Washington, DC.

Dorfman, R. (1943), "The Detection of Defective Numbers of Large Populations," *The Annals of Mathematical Statistics*, 14, 436–440.

Fang, K. T. (2005), available at *http://www.math.hkbu.edu.hk/~ktfang/*. Accessed September 20, 2005.

Fang, K. T., Lin, D. K. J., Winker, P., and Zhang, Y. (2000a), "Uniform Design: Theory and Application," *Technometrics*, 42, 237–248.

Fang, K. T., Ma, C., and Winker, P. (2000b), "Centered L_2-Discrepancy of Random Sampling and Latin Hypercube Design, and Construction of Uniform Designs," *Mathematics of Computation*, 71, 275–296.

Fang, K. T., and Wang, Y. (1994), *Number-Theoretic Methods in Statistics*, London: Chapman & Hall.

Florian, A. (1992), "An Efficient Sampling Scheme: Updated Latin Hypercube Sampling," *Probabilistic Engineering Mechanics*, 7, 123–130.

Golub, G. H., and Van Loan, C. F. (1983), *Matrix Computations*, Baltimore: Johns Hopkins University Press.

Hickernell, F. J. (1998), "A Generalized Discrepancy and Quadrature Error Bound," *Mathematics of Computation*, 67, 299–322.

Iman, R. L., and Conover, W. J. (1980), "Small-Sample Sensitivity Analysis Techniques for Computer Models, With an Application to Risk Assessment," *Communications in Statistics—Theory and Methods*, A9, 1749–1845.

Ipekci, A. I. (2002), "How Agent-Based Models Can Be Utilized to Explore and Exploit Non-Linearity and Intangibles Inherent in Guerrilla Warfare," unpublished masters thesis, Naval Postgraduate School, Monterey, CA.

Johnson, M., Moore, L., and Ylvisaker, D. (1990), "Minimax and Maximin Distance Designs," *Journal of Statistical Planning and Inference*, 26, 131–148.

Kim, H., and Loh, W. (2003), "Classification Trees With Bivariate Linear Discriminant Node Models," *Computational and Graphical Statistics*, 12, 512–530.

Kleijnen, J. P. C., Sanchez, S. M., Lucas, T. W., and Cioppa, T. M. (2005), "A User's Guide to the Brave New World of Designing Simulation Experiments," *INFORMS Journal on Computing*, 17, 263–289.

Koehler, J. R., and Owen, A. B. (1996), "Computer Experiments," in *Handbook of Statistics*, Vol. 13, eds. S. Ghosh and C. R. Rao, Amsterdam: Elsevier, pp. 261–308.

Lauren, M. K., and Stephen, R. T. (2001), *Map Aware Non-Uniform Automata Version 1.0 Users Manual*, Auckland, New Zealand: New Zealand Defence Technology Agency.

Lin, D. K. J. (1993), "A New Class of Supersaturated Designs," *Technometrics*, 35, 28–31.

Loerch, A. G., Pudwill, R. A., and LaBarbera, L. (1996), "Army Program Value Added Analysis 98-03 (VAA 98-03)," Vol. I, Report CAA-SR-96-1, U.S. Army Concepts Analysis Agency.

Lucas, T. W., and Sanchez, S. M. (2005), available at *http://diana.cs.nps. navy.mil/SeedLab/*. Accessed September 20, 2005.

Lucas, T. W., Sanchez, S. M., Brown, L., and Vinyard, W. (2002), "Better Designs for High-Dimensional Explorations of Distillations," in *Maneuver Warfare Science 2002*, eds. G. Horne and S. Johnson, Quantico, VA: USMC Project Albert, pp. 17–45.

McKay, M. D., Beckman, R. J., and Conover, W. J. (1979), "A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output From a Computer Code," *Technometrics*, 21, 239–245.

Meyers, R. H., and Montgomery, D. C. (2002), *Response Surface Methodology: Process and Product Optimization Using Designed Experiments*, New York: Wiley.

Morris, M. D., and Mitchell, T. J. (1995), "Exploratory Designs for Computational Experiments," *Journal of Statistical Planning and Inference*, 43, 381–402.

Owen, A. B. (1994), "Controlling Correlations in Latin Hypercube Samples," *Journal of the American Statistical Association*, 89, 1517–1522.

Piplani, L. K., Mercer, J. G., and Roop, R. O. (1994), *Systems Acquisition Manager's Guide for the Use of Models and Simulations*, Fort Belvoir, VA: Defense Systems Management College Press.

Plackett, R. L., and Burman, J. P. (1946), "Design of Optimal Multifactorial Experiments," *Biometrika*, 23, 305–325.

Ryan, T. P. (1997), *Modern Regression Methods*, New York: Wiley.

Saeger, K. J., and Hinch, J. H. (2001), "Understanding Instability in a Complex Deterministic Combat Simulation," *Military Operations Research*, 6, 43–55.

Santner, T. J., Williams, B. J., and Notz, W. (2003), *The Design and Analysis of Computer Experiments*, New York: Springer-Verlag.

SAS Institute Inc. (2004), *JMP User's Manual*, Cary, NC: Author.

Schruben, L. W. (1986), "Simulation Optimization Using Frequency Domain Methods," in *Proceedings of the 1986 Winter Simulation Conference*, pp. 396–369.

Srivastava, J. N. (1975), "Design for Searching Non-Negligible Effects," in *A Survey of Statistical Design and Linear Models*, Amsterdam: North Holland, pp. 507–519.

Ye, K. Q. (2005), available at *http://www.ams.sunysb.edu/~kye/olh.html*. Accessed September 20, 2005.

——— (1998a), "Orthogonal Column Latin Hypercubes and Their Application in Computer Experiments," *Journal of the American Statistical Association*, 93, 1430–1439.

——— (1998b), "On the Structure of Orthogonal Latin Hypercubes," technical report, University of Michigan, Department of Statistics, Ann Arbor, MI.