Proceedings of the 2013 Winter Simulation Conference
R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, eds.

SIMULATION SCREENING EXPERIMENTS USING LASSO-OPTIMAL SUPERSATURATED DESIGN AND ANALYSIS: A MARITIME OPERATIONS APPLICATION

Dadi Xing
Hong Wan

School of Industrial Engineering
Purdue University
West Lafayette, IN 47907, USA

Michael Yu Zhu

Department of Statistics
Purdue University
West Lafayette, IN 47907, USA

Susan M. Sanchez
Turgut Kaymal

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

ABSTRACT

Screening methods are beneficial for studies involving simulations that have a large number of variables where a relatively small (but unknown) subset is important. In this paper, we show how a newly proposed Lasso-optimal screening design and analysis method can be useful for efficiently conducting simulation screening experiments. Our approach uses new criteria for generating supersaturated designs, and a new algorithm for selecting the optimal tuning parameters for Lasso model selection. We generate a 24x69 Lasso optimal supersaturated design, illustrate its potential with a numerical evaluation, and apply it to an agent-based simulation of maritime escort operations in the Strait of Gibraltar. This application is part of a larger project that seeks to leverage simulation models during the ship design process, and so construct ships that are both cost effective and operationally effective. The supersaturated screening design has already proved beneficial for model verification and validation.

1 INTRODUCTION

Complex systems such as large-scale computer simulation models typically involve a large number of factors. When investigating such systems, screening experiments are often used to sift through these factors and identify a subgroup that most significantly influences the response of interest. A secondary experiment can efficiently allocate the remaining computational resources to those important factors, and this two-phase process can therefore greatly expand the ability of analysts and decision makers to gain insights about a complicated system with a reasonable computing budget.

The focus of the paper is on using supersaturated designs for factor screening. A supersaturated design is a fractional factorial design with more factors than experimental runs (Booth and Cox 1962, Lin 1993, Wu 1993). Supersaturated designs share the same basic principle (i.e., the sparsity of effects principle) as screening experiments, and are specifically attractive for high dimensional scenarios since they require far less runs than the number of factors. On the other hand, supersaturated designs do not have sufficient degrees of freedom to simultaneously estimate all main effects for all factors. Therefore, different analysis

978-1-4799-2076-1/13/$31.00 ©2013 IEEE

methods are needed than those used for non-saturated designs. Recently, L_1-penalty-based methods such as Lasso have been developed and shown great potential for analyzing high dimensional data (Osborne et al. 2000, Efron et al. 2004). Our latest work (Xing et al. 2013a,b) discusses a new approach to utilize Lasso for factor screening and proposes new searching criteria to construct Lasso-optimal supersaturated design. In this paper, we review the method and generate an optimal supersaturated design to explore an agent-based simulation model of marine escort operations. The agent-based simulation model involves a naval escort mission scenario in an anti-surface warfare environment. It is based on a 2002 Morocco incident, where terrorists planned to conduct suicide attacks on vessels passing through the Strait of Gibraltar. The screening experiments take 24 runs to study 65 factors, and our proposed Lasso method picks up to 11 important factors.

The paper is organized as follows. Section 2 focuses on the methodology used in this paper. We review the Lasso method and its application in factor screening, discuss the probability of correct model selection using Lasso, propose a new criterion to generate Lasso-optimal supersaturated designs, and discuss the determination of the Lasso tuning parameter. In Section 3, we describe a simulation model of maritime escort operations, and study the model with a 24 by 65 Lasso-optimal supersaturated design. We conclude with a brief discussion.

2 METHODOLOGY

We begin by briefly describing the Lasso-optimal screening design and analysis method used in the two paper to analyze the simulation model. For more details, please refer to Xing et al. (2013b) and Xing et al. (2013a).

2.1 LASSO INTRODUCTION

The least absolute shrinkage and selection operator, or Lasso in short, refers to a group of methods that use an L_1 penalty to shrink parameter estimates and perform automatic variable selection (Tibshirani 1996). Lasso does both continuous shrinkage and automatic variable selection simultaneously, and has shown great potential for analyzing high-dimensional data.

Consider a linear regression model

$$Y = X\beta + \varepsilon,$$

where $Y = (Y_1, Y_2, \ldots, Y_n)'$ is the n dimensional vector of observed responses, X is the $n \times p$ design matrix, $\beta = (\beta_1, \beta_2, \ldots, \beta_p)'$ is a p-dimensional vector of the unknown regression coefficients, and $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)'$ is the vector of random errors from a multivariate normal distribution $N(0, \sigma^2 I_n)$. Denote the columns of X as $(X_1, X_2, \ldots, X_p)$; these represent the p independent variables. For $i = 1, 2, \ldots p$, X_i is said to be a true (important) variable if $\beta_i \neq 0$; otherwise, X_i is said to be an untrue (unimportant) variable. The Lasso estimate of β, denoted by $\hat{\beta}^{\lambda}$, is the solution of the following L_1-penalized least squares problem:

$$\hat{\beta}^{\lambda} = \frac{1}{2} \arg\min_{B \in \mathbb{R}^p} \left\{ \sum_{i=1}^{n} \left(Y_i - \sum_{j=1}^{p} B_j X_{ij} \right)^2 + \lambda \sum_{i=1}^{n} |B_i| \right\}, \tag{1}$$

where $\sum_{i=1}^{p} |B_i|$ is the L_1-norm penalty function of B, and λ is the tuning parameter that controls the amount of penalty. Note that we can assume, without loss of generality for factor screening, that the Y's have been centered with mean zero so the intercept is also equal to zero (Tibshirani 1996). For ease in discussion, let $A = \{i : \beta_i \neq 0\}$ and $C = \{i : \beta_i = 0\}$ be the collection of the true and untrue variables indexes, respectively.

Given a specific $\lambda > 0$, there are many fast algorithms to solve Lasso, such as LARS (Efron et al. 2004), Coordinate Descent (Friedman et al. 2007), and others. In this paper, we use the LARS algorithm because of its demonstrated efficiency for large-scale data. For a fixed λ, define $\hat{A}^{\lambda} = \{j : \hat{\beta}_j^{\lambda} \neq 0\}$, which is referred as the collection of selected variables (those estimated to be important factors) at λ. Denote

$\hat{C}^{\lambda}$ as the complement of $\hat{A}^{\lambda}$. When λ increases, more $\hat{\beta}_i^{\lambda}$'s will become zero due to the heavier impact of the penalty term in Equation (1). The path of $\hat{\beta}^{\lambda}$ as λ changes is referred to as the *solution path* of Lasso (Tibshirani 1996), or simply the *Lasso path*. Different methods have previously been proposed for choosing the appropriate tuning parameter λ which, in turn, controls how many variables will be selected.

2.2 PROBABILITY OF SIGN-CONSISTENT SELECTION VIA LASSO AND LASSO-OPTIMAL SUPERSATURATED DESIGNS

For any $\lambda > 0$, $\hat{\beta}^{\lambda}$ will always be a biased estimator of the true parameter vector β. There are two ways to evaluate the Lasso estimate $\hat{\beta}^{\lambda}$. One is to see whether the non-zero items in $\hat{\beta}^{\lambda}$ match the true important factors, which we call variable selection consistency; and another is to measure the difference between the estimated coefficients and β. Here we focus on the first one, which is consistent with our factor screening goal.

There are two types of variable selection consistency for Lasso discussed in the literature. When $\hat{A}^{\lambda} = A$, that is, all the true variables are selected, Lasso is said to have *ordinary consistency* in variable selection at λ. Similarly, Lasso is said to have *sign consistency* in variable selection at λ if the $p \times 1$ vectors of true and estimated signs of the coefficients are equal. Mathematically, define the function $sgn(B)$ to operate element-by-element on vector B so that $sgn(B_i) = B_i/|B_i|$ takes on values in $\{-1, 0, 1\}$. Then sign consistency means that $sgn(\hat{\beta}^{\lambda}) = sgn(\beta)$. Notice that the sign consistency of Lasso implies ordinary consistency, but the converse is not generally true.

Let q denote the number of true variables. Let β_A be the subvector of β that corresponds to the true variables, let X_A be the submatrix of X with columns in A, and let X_C be the complement of X_A. From these, define $P = X_A(X_A'X_A)^{-1}X_A'$, $R = X_C'X_A(X_A'X_A)^{-1}\text{sgn}(\beta_A)$, and $D = \frac{\lambda}{2}(X_A'X_A)^{-1}\text{sgn}(\beta_A)$.

The following theorem by Xing et al. (2013b) gives the probability that Lasso achieves sign consistency for a given sample of size n, design matrix X, and penalty parameter λ.

Theorem 1. *Let W be a (p-q)-dimensional multivariate $N(R, 4X_C'(I-P)X_C\sigma^2/\lambda^2)$, and let V be a q-dimensional multivariate $N(D, (X_A'X_A)^{-1}\sigma^2)$. The probability that Lasso achieves sign consistent selection, which is denoted by* P(SCS)*, is:*

$$\text{P(SCS)} = \text{P}(\mathscr{E}_1)\text{P}(\mathscr{E}_2), \tag{2}$$

where $\mathscr{E}_1$ and $\mathscr{E}_2$ are the events

$$\mathscr{E}_1 = \{-1 \leq \text{W} \leq 1\} \text{ and}$$

$$\mathscr{E}_2 = \bigcap_{i \in A} \{\beta_i \notin [\min(0, \text{V}_i),\ \max(0, \text{V}_i)]\}.$$

With fixed A, β, λ and σ^2, the probabilities of $\mathscr{E}_1$ and $\mathscr{E}_2$ depend on the following four terms: R and D, which are the mean vectors of W and V, along with $4X_C'(I-P)X_C\sigma^2/\lambda^2$ and $\frac{\lambda}{2}(X_A'X_A)^{-1}\sigma^2$, which are the variances of W and V, respectively. If we can optimize P(SCS), we will optimize the performance of Lasso for factor screening. There are two challenges, however. First, the direct evaluation and optimization of $\text{P}(\mathscr{E}_1)$ and $\text{P}(\mathscr{E}_2)$ with respect to β, λ, and X require high-dimensional integration, which becomes computational prohibitive as the dimensions increase. Second, these probabilities depend on unknown parameters.

To address the first challenge, Xing et al. (2013b) proposed both a lower bound method and Monte Carlo simulation method to approximate $\text{P}(\mathscr{E}_1)$ and $\text{P}(\mathscr{E}_2)$ efficiently in high dimensional cases. Notationally, let D_i $(i = 1, \ldots, q)$ denote the ith element of D, and let R_j $(j, \ldots, p-q)$ denote the jth element of R. Let $(\sigma_{W_1}^2, \sigma_{W_2}^2, \ldots, \sigma_{W_{p-q}}^2)'$ be the vector of diagonal elements of $(4X_C'(I-P)X_C\sigma^2)$, and let $(\sigma_{V_1}^2, \sigma_{V_2}^2, \ldots, \sigma_{V_q}^2)'$ be the vector of the diagonal elements of $((X_A'X_A)^{-1}\sigma^2)$. Finally, let $I(\cdot)$ be an indicator such that $I(x) = 1$ if x is true, and $I(x) = 0$, otherwise. The lower bound method is based on the following proposition from Xing et al. (2013b):

Proposition 1. *A lower bound for* $\mathrm{P}(\mathscr{E}_1)$ *is*

$$\mathrm{P}(\mathscr{E}_1)_{low} = \max\left(0, 1 - \frac{1}{\lambda\sqrt{2\pi}}\sum_{j=1}^{p-q}\sigma_{uj}\left(\mathbf{I}(|R_j|<1)\,Q_1(j) + \mathbf{I}(|R_j|>1)Q_2(j)\right)\right) \tag{3}$$

where

$$Q_1(j) = \frac{1}{(1+R_j)}exp\left\{\frac{-\lambda^2(1+R_j)^2}{2\sigma^2_{W_j}}\right\} + \frac{1}{(1-R_j)}exp\left\{\frac{-\lambda^2(1-R_j)^2}{2\sigma^2_{W_j}}\right\},$$

and

$$Q_2(j) = \frac{1}{\sqrt{2\pi}(1+R_j)}exp\left\{\frac{-\lambda^2(1+R_j)^2}{2\sigma^2_{W_j}}\right\} + \frac{1}{\lambda} - \frac{(R_j-1)}{\sqrt{2\pi}\left((1-R_j)^2+\frac{\sigma^2_{W_j}}{\lambda^2}\right)}exp\left\{\frac{-\lambda^2(1-R_j)^2}{2\sigma^2_{W_j}}\right\}.$$

Also, a lower bound for $P(\mathscr{E}_2)$ *is*

$$\mathrm{P}(\mathscr{E}_2)_{low} = \max\left(0, 1 - \sqrt{\frac{2}{\pi}}\sum_{i=1}^{q}\sigma_{V_i}\left[\frac{\beta_i}{|\beta_i|}\left(\frac{1}{2\beta_i-\lambda D_i}\right)exp\left\{\frac{-(2\beta_i-\lambda D_i)^2}{8\sigma^2_{vi}}\right\}\right]\right). \tag{4}$$

Then, a lower bound for P(SCS) *is given by* $\mathrm{P(SCS)}_{low} = \mathrm{P}(\mathscr{E}_1)_{low}\mathrm{P}(\mathscr{E}_2)_{low}$.

The lower bound approximation above can be calculated very efficiently compared to P(SCS). The simulation results show that when the total number of variables $p < 25$, the lower bound offers a satisfactory approximation for the probability of correct selection. For higher order cases, however, the lower bound deviates substantially from the true value and the approximation of $\mathrm{P}(\mathscr{E}_1)$ and $\mathrm{P}(\mathscr{E}_2)$ may approach zero. In these cases, we use a straightforward Monte Carlo sampling method instead of the lower bounds in (3) and (4). In summary, the algorithm sequentially and independently generates sample vectors from W and V, both of which are multivariate normal random variables, to obtain direct estimates $\mathrm{P}(\mathscr{E}_1)_{Monte}$ and $\mathrm{P}(\mathscr{E}_2)_{Monte}$, respectively. The algorithm stops when the estimation converges. For more details, please refer to Xing et al. (2013b).

The second challenge of directly optimizing P(SCS) is that the probability depends on unknown parameters, including the number of true factors, the magnitudes and signs of β_A, and the variance σ^2. These parameter values are not available in the screening stage—if they were, there would be no need for a screening experiment. One approach to solve this problem is to focus only on the influence of the design structure to the probability of correct selection. Specifically, Zhao and Yu (2006) prove that if $|R| < 1$, then $\mathrm{P(SCS)} = 1$ asymptotically. This is called the irrepresentable condition. From the P(SCS) expression in Theorem 1, it is easy to see that R dominates $\mathrm{P}(\mathscr{E}_1)$. When n goes to infinity, both $\mathrm{P}(\mathscr{E}_1)$ and $\mathrm{P}(\mathscr{E}_2)$ will converge to 1 (independent of $|R|$) if $|R| < 1$, which implies the irrepresentable condition discussed in Zhao and Yu (2006). This motivates us to propose a standard for optimal design that is based on the residual sum of squares RSS as a function of q_{max}, the maximum number of potentially important factors defined by the user. For ease of notation we will drop the subscript and simply use q. Let A be a submatrix of selected variables (important factors) at λ, and define

$$\mathrm{RSS}(A) = trace(X_C'X_A(X_A'X_A)^{-2}X_A'X_C).$$

Since we do not have prior knowledge of which q factors are important, all possible subsets of q factors need to be considered when constructing the optimal design. Therefore, we evaluate the suitability of design X for selecting q factors as follows: we compute

$$\overline{\mathrm{RSS}}_q(X) = \binom{p}{q}^{-1}\sum_{A\in\mathscr{A}}\mathrm{RSS}(A),$$

where $\mathscr{A}$ is the collection of all possible subsets of q important factors among the p factors. For fixed p and n, we refer to a design that minimizes $\overline{\mathrm{RSS}}_q(X)$ over all possible designs with specified q as an optimal $\overline{\mathrm{RSS}}_q$ design.

The design is generated by a partial gradient column exchange heuristic algorithm discussed in Xing et al. (2013a). The intuition of the algorithm is to find the partial gradient of the objective function with respect to each column in X, and use the gradient information to guide the column exchange by moving the least favorable column in the desired direction. When the progress stagnates, fresh columns will be introduced. Due to the limitations of the space, we omit the details of the algorithm here and direct the interested readers to Xing, Wan, and Zhu (2013a). We have run extensive numerical evaluation to compare the RSS-optimal supersaturated designs with other supersaturated designs proposed in the literature. The RSS-optimal designs do have higher probability of correct model selection compared to other designs when Lasso is used as the analysis method.

2.3 MODEL SELECTION USING LASSO: DETERMINE THE TUNING PARAMETER λ

With the optimal design, we collect observations from simulation model and apply the LARS algorithm proposed by (Efron et al. 2004) to analyze the data. One issue that has not yet been discussed in this paper is the selection of λ, the tuning parameter of Lasso. This matters because the magnitude of λ determines how many coefficients will be non-zero, therefore the amount of screening that will occur. We recently proposed an algorithm based on self-voting using ordinary least squares estimation, called SVOE, to select the optimal λ. We briefly summarize it here. For more details, please refer to Xing et al. (2013b).

To gain insight into the relationship between λ and $\hat{\beta}^{\lambda}$, consider the following generic example with 16 variables. The underlying true model is $Y = X\beta + \varepsilon$, where $\beta = (2,0,2,0,0,0,0,0,2,0,0,0,0,0,0,0)'$. Note that the true (important) variables in this model are factors 1, 3 and 9. Our X is the 12×16 supersaturated design given in Table 4 of Li and Wu (1997), and we let $\varepsilon \sim N(0, \sigma^2 I_{16})$. We generate one response independently for each design point using Monte Carlo simulation and apply the LARS algorithm repetitively with λ changing from 70 to 0. This gives us the Lasso solution path of Figure 1. Here the x-axis is the λ; and the y-axis is the estimated β value. When $\lambda \geq 70$, all estimated coefficients are 0. As λ decreases, factors are added to the model one by one.

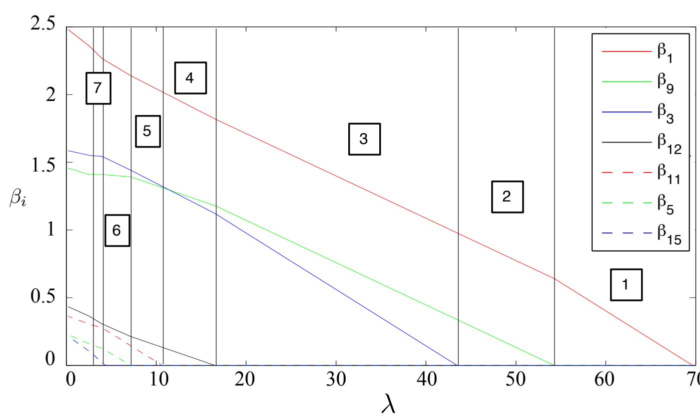

Figure 1: Lasso path for a generic 12×16 supersaturated design.

The seven lines in Figure 1 represent the sequential changes of $\hat{\beta}_i$ for the first seven factors selected into the model; we stop at seven factors to avoid crowding the figure, but twelve factors are selected at $\lambda = 0$. The vertical lines split the figure into different regions, and the numbers in squares specify how many factors are selected within the regions. For example, in Region 3, where λ is between 10.7 and 16.5, three factors are included in the model (factors 1, 3, and 9). Thus, for each λ value, there is a corresponding

$\hat{\beta}^{\lambda}$; if we treat this $\hat{\beta}^{\lambda}$ as if it is the true β and plug it into Equation (2) in Theorem 1, we can solve for a corresponding $\hat{\lambda}_{opt}$. We can repeat this process until the λ and its counterpart $\hat{\lambda}_{opt}$ match, and treat this converged λ as the true λ_{opt}.

We call our heuristic SVOE, for self-voting based on OLS estimation. It is based on the self-voting principle discussed in Gu (1992). The numerical evaluation in Xing et al. (2013b) shows that this heuristic works very well except in some rare cases. When there are multiple self-voting λ_{opt}, the user can determine which model they would like to pick. Compared with other parameter tuning methods, for example, CV (Devijver and Kittler 1982), AIC (Akaike 1974), BIC (Schwarz 1978), that focus on finding the balance between goodness of fit and overfitting by adding penalties for selecting too many factors into the model, our approach focuses on directly optimizing the probability of the selecting the true model. Numerical evaluation shows that our SVOE method selects the best λ the majority of the time (Xing et al. 2013b), and we apply it in the next section.

3 MARITIME ESCORT OPERATIONS SIMULATION

To illustrate the potential benefits of this screening procedure, we apply it to study an agent-based model of maritime escort operations. A brief summary of the context follows. For further details, see Kaymal (2013): his thesis is part of a multinational, multi-year effort to leverage modeling and simulation in order to better understand trade-offs between cost and operational effectiveness for surface ships, and so improve the ship design process.

3.1 MOTIVATION

In the past two decades, there has been a significant shift in navy missions towards operations other than war. Counter-piracy, search and rescue, maritime interdiction, maritime patrol, naval escort operations, and humanitarian efforts are the main focus of most fleets today, but the vessels that are currently being used in such operations were mainly built for other purposes. For instance, on 17 August 2009, the North Atlantic Council approved "Operation Ocean Shield" to fight piracy in the Gulf of Aden. Among six surface ships that were assigned in January-June 2012 rotation of this NATO mission, one was a destroyer and three were frigates. These are sophisticated warships which are capable of anti-surface warfare, anti-air warfare, and anti-submarine warfare. Although these sophisticated multi-mission capable fleets are able to achieve good results in expeditionary warfare against a strong enemy (Murphy 2007), their full capabilities will probably be used in less than 1% of their total life time. Frigates and destroyers are also expensive to build and operate.

Smaller combatants are much cheaper and better suited for the more common modern naval operations due to their flexibility. Therefore, many nations are reshaping their fleets to meet emerging operational demands, by starting to build multi-mission capable combatants such as Offshore Patrol Vessels (OPVs) that are smaller and less sophisticated than frigates and destroyers. These nations are also developing new tactics and counter measures to better deal with the new threats. The main purpose of OPVs is maintaining maritime security inshore and offshore, and they can be deployed globally (Kimber and Booth 2010). Being cheaper yet flexible, OPVs are also great options for those countries that either do not really need or cannot afford sophisticated naval combatants.

Unfortunately, the ship design and the acquisition process has not kept pace with the rapid technological advances of the past few decades (Ryan and Jons 2004). Historical cost estimation models, such as those based on cost per ton, often fail to adequately capture new technology-driven requirements. In the past, operational requirements were often not considered until late in the ship design process, and then might be based on limited input by small panels of subject matter experts. Cost effectiveness and operational effectiveness are both important, and it is extremely hard to achieve both when using a traditional ship design process. Advances in modeling and simulation, engineering design software, and massive amounts of data now allow for a much broader and richer range of ship specifications and capabilities to be explored.

Moreover, utilizing simulation and analytical models to build decision making tools will ensure collaboration between warfighters and engineers in the early stages of the process. Exploiting technology is paramount for accomplishing the navy objectives and increasing the effectiveness for both cost and operations (Mizine et al. 2012).

3.2 SIMULATION SCENARIO DESCTIPTION

Even though the term ”maritime terrorism” has been in the literature for more than a few decades, it was not really spelled out clearly until a series of incidents started in 2000. In January 2000, terrorists tried to attack USS The Sullivans (DDG-68) in Yemen. They failed because the terrorist boat sank right before the attack. Following this event, the terrorists attacked USS Cole (DDG-67) with a suicide boat, killing 17 of the crew members, in October 2000. Almost two years after these incidents, in October 2002, a boat with explosives hit the French oil tanker Limburg which was close to Yemen coastal waters (Luft and Korin 2004). These are just a few examples of how, with just a small boat, terrorists can cause damage to multi-million/billion dollar ships, and—more importantly—kill innocent people. Many other terrorist activities which were discovered in their planning phases and prevented. These incidents show that maritime terrorism is a serious problem, and that surface combatants must be ready to fight terrorists at sea.

In June 2002 a group of terrorists, who were planning an attack on two merchant vessels in the Strait of Gibraltar, were caught by Moroccan officials (Maggio 2008). The fact that the officials arrested those terrorists does not mean that similar plans will not be put into practice by other groups. If a terrorist group fills a boat with explosives and approach a ship in Strait of Gibraltar, it might be really hard to identify that boat as a terrorist boat since many vessels are in the strait at any point in time.

Our specific scenario involves a naval escort mission scenario in an anti-surface warfare environment. It is based on the 2002 Morocco incident, and implemented in an agent-based simulation modeling platform called MANA, designed by the Operational Analysis personnel of the New Zealand Defence Technology Agency (DTA) (Lauren and Stephen 2002). The key attributes of MANA which make it a useful tool for military applications are the situational awareness of the agents, advanced communication capabilities within squads and with other agents, the interaction of entities with friends and foes, and the user friendly design of the program. There are four main sets of parameters that form agent behaviors in MANA:

- Personality weightings determine the willingness of agents to perform a particular action;
- Movement constraints modify the basic personality weightings of the agents;
- Intrinsic capabilities determine the physical characteristics of the agents such as sensors, weapons, or fuel level; and
- Movement characteristic adjustments ensure that agent actions change in different terrain conditions and different situations.

Just recently, DTA released MANA-V. The “V” stands for both vector and five. In this version, the programmers replaced the cell-based movement of previous versions with vector based movement. This allows building larger battlefield regions with panning and zooming options, and allows defining users to define distances, and the attributes such as speed and range, in real-world units rather than pixel-based units (McIntosh 2009).

There are six types of agents in the scenario: the high value unit (HVU) being escorted, the OPV, the OPV’s helicopter, terrorist boats, known vessels, and unknown vessels. A screenshot from MANA appears in Figure 1; note that the sizes of the agents are not to scale, but are magnified for easier visualization by the user. The allegiance of the HVU, OPV, and the helicopter is “friend,” the allegiance of the terrorist boats is “hostile,” and the allegiance of the known and unknown vessels is “neutral.” The OPV’s mission is to escort the HVU through the Strait of Gibraltar and protect the HVU from attacks that can occur in the passage. The OPV has several guns, ranging from high caliber guns to machine guns. It also has a helicopter landing platform. The main role of the helicopter is to detect and classify unknown vessels. Its high speed and maneuverability give the friendly forces an advantage against the hostiles. The helicopter can also be equipped with a machine gun so that, if necessary, it can start firing before the hostile vessels come too close to the HVU.

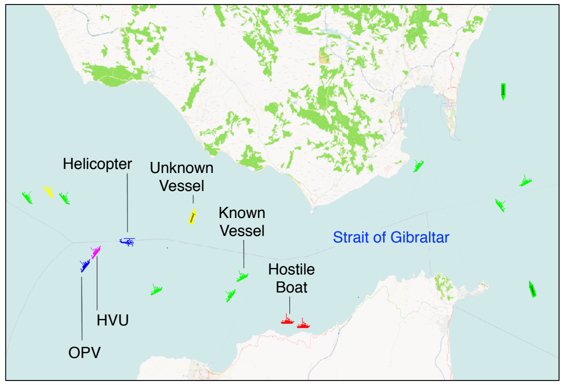

Figure 2: MANA screenshot, with six types of agents in the Strait of Gibraltar (from Kaymal 2013).

The terrorist boats are loaded with explosives, and their purpose is to get close enough to the HVU to perform a suicide attack. The only target of the hostile boats is the HVU. They do not attack the OPV, but in some cases they may try to evade it. One of the critical properties of the terrorist boats is that they are initially acquired as unknown vessels in the friendly force radar systems until they are classified as enemy. Therefore, either the helicopter classifies them as enemy, or they get closer to the OPV or HVU, and the friendly ships classify them depending on the range.

The known ships and the unknown ships are both neutral and they do not pose a threat to friendly assets. The only difference between the two is that while the OPV can instantly classify the known neutral ships using an Automatic Identification System (AIS) device, unknown ships cannot be classified when they are initially detected. This occurs for several reasons: some ships may be too small to carry an AIS device, others may have a device that is not working or turned off. This represents the type of marine traffic often found in the Strait of Gibraltar and other similar situations. It complicates matters for the OPV, since the OPV and its helicopter may be unable to fire on hostile boats if neutral ships are close by. Moreover, the OPV and helicopter may need to spend a substantial amount of time and effort identifying unknown ships, even though the vast majority are neutral. These distractions might give terrorists a better chance to approach the HVU without being classified as threats.

The simulation begins with the HVU and OPV approaching the strait. It halts the instant that either the HVU successfully navigates the strait, or it is attacked by a hostile ship.

3.3 SIMULATION EXPERIMENTS

We conducted multiple sets of experiments in the process of creating and exploring this simulation model. Those for exploring the trade-offs used nearly orthogonal and balanced mixed designs (Vieira Jr et al. 2011, Vieira Jr. et al. 2013), which are space-filling designs capable of handling a mix of continuous, discrete, and qualitative factors. The initial experiment involved 100 replications of a NOAB design for 38 factors, including one that represented the time step used in the scenario. This directly affects the time it takes to complete the simulation runs, but the time step must be small to avoid model artifacts.

Once the choice of time step was settled and some other minor changes were made, the final space-filling experiment involved 200 replications of a NOAB for 35 factors. Consolidating these results into a single file for analysis resulted in 153,600 simulation runs. Kaymal (2013) uses a range of statistical techniques, including regression, logistic regression, and partition trees, to model the relationship between the factors and key responses. These models also are used to suggest factor ranges and combinations that make efficient use of resources but achieve high operational effectiveness.

In this paper, we focus on just one of the many potential output measures. Our response Y is the proportion of time that the HVU successfully navigates the strait. We use a 24×69 Lasso-optimal design capable of examining up to 69 factors, and identifying up to 24 important ones. The design is available on request. We augmented Kaymal's list of 35 factors with an additional 30 that had not initially been explored. To maintain the spirit of a screening experiment, we made very tiny changes in factor settings for personality weights and communication characteristics (e.g., ± 2 on a scale of -100 to 100). In addition to serving as an example for the Lasso screening approach, we viewed this as an opportunity to beta-test some aspects of MANA that were not used in Kaymal's experiments. (Over the years, the developers have added many features to MANA in response to requests by NPS students. In turn, the beta-testing by NPS students of MANA's features has been useful for verification and validation efforts.)

To test the performance of the selected design, we ran a quick numerical evaluation of the selected design. Consider a linear model with 65 factors and $q = 5, 10, 15,$ and 20 respectively. The important factors are randomly selected and their effects are set at 3. The other factors all have an effect of 0. The white noise follows a standard normal distribution. We generate one set of responses of the 24×69 Lasso-optimal design for each case, and the results are summarized in Table 1. All factors that are both truly important and selected by Lasso are shown in boldface. Note that for the $q = 10$ case, there are multiple optimal λs, therefore multiple models.

Table 1: Numerical Evaluation of 24×69 Lasso-optimal design

q	Important Factors	SVOE Lasso Variable Selection
5	**4**,**5**,**32**,**50**,**56**	(**56**,**50**,**4**,**32**,**5**)
10	**3**,**4**,**6**,**17**,23,**25**,26,**31**,**37**,63	(**37**,**4**,**31**),(**37**,**4**,**31**,**3**),(**37**,**4**,**31**,**3**,59,25,**17**,**6**)
15	**5**,8,9,20,25,**26**,29,**33**,**37**,44,45,**46**,47,60,63	(3,**37**,39,**46**,**33**,**5**,52,24,48,58,**26**)
20	2,8,9,12,21,**26**,32,34,**39**,40, 46,47,**52**,53,55,57,**58**,62,64,65	(**39**,38,**52**,**26**,**58**)

Ideally, the SVOE Lasso method would select all of the important factors and none of the unimportant ones. It does this for our $q = 5$ example. As q increases, the number of selected factors remains roughly constant, which means there is a decrease in the proportion of correctly identified factors. Some unimportant factors also show up, although false positives are less of a concern in screening experiments because they can be eliminated in subsequent stages. Clearly, a much more extensive empirical investigation would be needed to provide a true picture of the design's screening capabilities, both in terms of false positives and false negatives. Nonetheless, it is promising to see that when the number of important factors is small or moderate ($q = 5$ or 10), the SVOE Lasso method selects a majority or all of the important factors with less than half number of runs required by saturated design. This implies that for practical screening problems where a small set of factors can drive the response, SVOE Lasso may work well.

With these promising (though limited) results giving us some confidence of the design, we ran 1000 replications to obtain sample proportions Y_i for each of the 24 design points. This took over 5.25 days of CPU time on our computing cluster. We found that seven of the 24 design points aborted before all replications were complete. After defining a new response $\tilde{Y} \equiv \mathrm{I}(\text{all reps complete})$, we quickly identified which factor was causing the problem. After confirming our findings by running MANA in its GUI mode, we removed this factor from our list and substituted another. This points out a clear benefit of using a screening procedure. We were able to identify a bug in just 5.25 CPU days, rather than the 112 CPU days required for 1000 replications of our 512 design point NOAB, or the 4.9 CPU years required for 1000 replications of an 8196 design point resolution V fractional factorial (Sanchez and Sanchez 2005).

Our second screening experiment ran to completion. There were two potential choices for λ_{opt}. The smallest selects two of the first 35 factors and three of the second 30, the largest selects six of the first 35 factors and five of the second 30. The Lasso path for $40 \leq \lambda \leq 220$ appears in Figure 3.

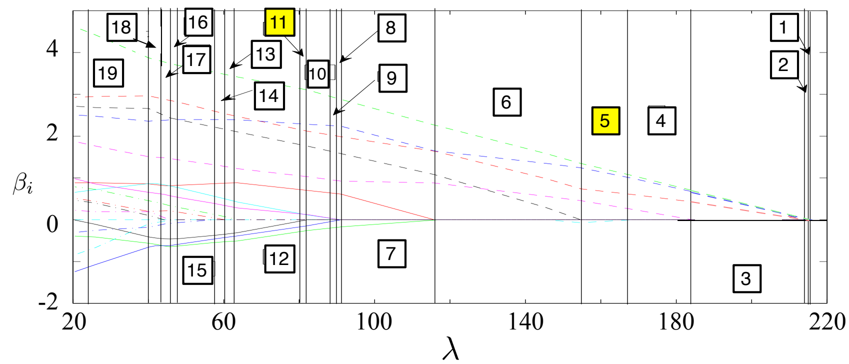

Figure 3: Lasso path for the 24 x 65 design, with the number of selected terms shown for each region.

These results were somewhat surprising, since we had anticipated that all the important factors would be among the first 35. Further experimentation is underway to determine an appropriate baseline for the study. Whatever these results, they will be beneficial for several stakeholders. As mentioned earlier, the MANA developers have been adding features to expand its capabilities, and they are always interested in results of large-scale beta tests that might indicate either model bugs ("artifacts"), or else modeling concepts that would benefit from additional clarification in the documentation. Those creating scenarios can use screening experiments early in the model-building process to aid in model verification and help set suitable factor ranges. The international consortium looking at the model-based ship design project will benefit from having information about the robustness of the findings—we advocate that it is better to experiment on a large number of factors and seek a simplified model than to limit the factors from the outset (Sanchez et al. 2012). Finally, by examining simulation models that are being used for decision-making, we will continue to identify challenges that would benefit from further methodological research.

4 DISCUSSION

In this paper, we provided a brief overview of a newly proposed Lasso method for factor screening, and a new search criterion for constructing Lasso-optimal supersaturated designs. The new method specifies one or more optimal tuning parameters for Lasso, which determines the model selection result. The criterion is based on the probability of correct variable selection of Lasso. Compared to the past work in supersaturated designs, our work uses the analysis method to direct the experimental design stage, and the preliminary numerical evaluation shows very promising results. The method is specifically intended for the initial study of complex systems with many potential factors.

The example in the paper involves an agent-based simulation of maritime escort operation. The insights we gain in this study may continue to aid in model verification and validation efforts—we have already identified one model bug by using the small Lasso-optimal screening design. The broader implications are also important: the operational insights from Kaymal's investigation (and related studies) may, in the near future, determine how nations decide to structure their naval forces.

In summary, preliminary results on the supersaturated designs are promising, both for artificially generated response surfaces, and for a more complex, agent-based simulation. Supersaturated screening designs may be of even greater interest for larger models: many simulations addressing defense or homeland security issues involve thousands of potential factors.

ACKNOWLEDGMENTS

Susan M. Sanchez was supported in part by grants from the Office of Naval Research and the U.S. Army Survivability and Lethality Analysis Directorate.

REFERENCES

Akaike, H. 1974. "A New Look at the Statistical Model Identification". *IEEE Transactions on Automatic Control* 19 (6): 716–723.

Booth, K. H. V., and D. R. Cox. 1962. "Some Systematic Supersaturated Designs". *Technometrics* 4 (4): 489–495.

Devijver, P. A., and J. Kittler. 1982. *Pattern Recognition: A Statistical Approach*. London: Prentice-Hall.

Efron, B., T. Hastie, I. Johnstone, and R. Tibshirani. 2004. "Least Angle Regression". *Annals of Statistics* 32 (2): 407–499.

Friedman, J. H., T. Hastie, H. Hofling, and R. Tibshirani. 2007. "Pathwise Coordinate Optimization". *The Annals of Applied Statistics* 1 (2): 302–332.

Gu, C. 1992. "Cross-Validating Non-Gaussian Data". *Journal of Computational and Graphical Statistics* 1 (2): 169–179.

Kaymal, T. 2013, June. "Assessing the Operational Effectiveness of a Small Surface Combat Ship in an Anti-Surface Warfare Environment". Master's thesis, Naval Postgraduate School, Monterey, California.

Kimber, A., and J. Booth. 2010. "Novel Technologies for Enhancing the Performance of Constabulary and Security Orientated Ships". In *Proceedings of the 2010 RINA Warship Conference*: London: The Royal Institute of Naval Architects.

Lauren, M., and R. Stephen. 2002. "Map-Aware Non-uniform Automata (MANA)—A New Zealand Approach to Scenario Modelling". *Journal of Battlefield Technology* 5 (1): 27–31.

Li, W. W., and C. F. J. Wu. 1997. "Columnwise-Pairwise Algorithms with Applications to the Construction of Supersaturated Designs". *Technometrics* 39 (2): 171–179.

Lin, D. K. J. 1993. "A New Class of Supersaturated Designs". *Technometrics* 35 (1): 28–31.

Luft, G., and A. Korin. 2004. "Terrorism Goes to Sea". *Foreign Affairs* 83 (6): 61–67.

Maggio, E. 2008. *Private Security in the 21st Century: Concepts and Applications*. Sudbury, Massachusetts: Jones and Bartlett Publishers.

McIntosh, G. C. 2009. *MANA–V (Map Aware Non-uniform Automata–Vector) Supplementary Manual*. New Zealand: Defence Technology Agency.

Mizine, I., B. Wintersteen, and S. Wynn. 2012. "A Multi-Level Hierarchical System Approach to Ship Concept Formulation Tools". *Naval Engineers Journal* 124 (3): 93–120.

Osborne, M., B. Presnell, and B. Turlach. 2000. "A New Approach to Variable Selection in Least Squares Problems". *IMA Journal of Numerical Analysis* 20 (3): 389–403.

Ryan, J. C., and O. P. Jons. 2004. "Improving the Ship Design, Acquisition and Construction Process". *Naval Engineers Journal* 104 (2): 39–57.

Sanchez, S. M., T. W. Lucas, P. J. Sanchez, H. Wan, and C. Nannini. 2012. "Designing large scale simulation experiments, with applications to defense and homeland security". In *Design and Analysis of Experiments Volume III: Special Designs and Applications*, edited by K. Hinkelman, 413–441. New York: John Wiley & Sons.

Sanchez, S. M., and P. J. Sanchez. 2005. "Very Large Fractional Factorial and Central Composite Designs". *ACM Transactions on Modeling & Computer Simulation* 15 (4): 362–377.

Schwarz, G. E. 1978. "Estimating the Dimension of a Model". *IEEE Transactions on Automatic Control* 6 (2): 461–464.

Tibshirani, R. 1996. "Regression Shrinkage and Selection via the Lasso". *Journal of the Royal Statististical Society, Series B* 58 (1): 267–288.

Vieira Jr, H., S. M. Sanchez, K. H. Kienitz, and M. C. N. Belderrain. 2011. "Improved efficient, nearly orthogonal, nearly balanced mixed designs". In *Proceedings of the 2011 Winter Simulation Conference*, edited by S. Jain, R. R. Creasey, J. Himmelspach, K. P. White, and M. Fu, 3600–3611: Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Vieira Jr., H., S. M. Sanchez, K. H. Kienitz, and M. C. N. Belderrain. 2013. "Efficient, Nearly Orthogonal-and-Balanced, Mixed Designs: An Effective Way to Conduct Trade-off Analyses via Simulation". Working paper, Department of Operations Research, Naval Postgraduate School, Monterey, CA.

Wu, C. F. J. 1993. "Construction of Supersaturated Designs Through Partially Aliased Interactions". *Biometrika* 80 (3): 661–669.

Xing, D., H. Wan, and M. Y. Zhu. 2013a. "Optimal Supersaturated Design via Lasso". Working paper, School of Industrial Engineering, Purdue University.

Xing, D., H. Wan, and M. Y. Zhu. 2013b. "Proper Lambda Chosen via Lasso". Working paper, School of Industrial Engineering, Purdue University.

Zhao, P., and B. Yu. 2006. "On Model Selection Consistency of Lasso". *Journal of Machine Learning Research* 7:2541–2563.

AUTHOR BIOGRAPHIES

DADI XING is a Ph.D. student in the School of Industrial Engineering at Purdue University. He is co-advised by Prof. Hong Wan and Prof. Michael Yu Zhu. His research focus on the experiment design and analysis based on the penalized variable selection methods and their application in complex system simulation. His email address is dxing@purdue.edu and his web address is http://web.ics.purdue.edu/~dxing.

HONG WAN is an Associate Professor in the School of Industrial Engineering at Purdue University. She received her Ph.D. in Industrial Engineering and Management Sciences from Northwestern University. Her research interests include design and analysis of large-scale simulation experiments, simulation analysis of manufacturing and healthcare systems; quality management in supply chain, and quality control and applied statistics. She has taught a variety of courses and is a member of INFORMS and ASA. She currently serves as the associate editor of ACM Transactions on Modeling and Computer Simulation. Her email address is hwan@purdue.edu and her web address is http://web.ics.purdue.edu/~hwan.

MICHAEL YU ZHU is a Professor in the Statistics Department at Purdue University. He received his Ph.D. in Statistics from the University of Michigan. His research interests include dimension reduction and variable selection in high dimensional data analysis, design and analysis of experiments, and Bioinformatics. His email and web addresses are yuzhu@purdue.edu and http://www.stat.purdue.edu/~yuzhu, respectively.

SUSAN M. SANCHEZ is a Professor in Operations Research at the Naval Postgraduate School, and Co-Director of the Simulation Experiments & Efficient Design (SEED) Center for Data Farming. She also holds a joint appointment in the Graduate School of Business & Public Policy. She has a B.S. in Industrial & Operations Engineering from the University of Michigan, and a Ph.D. in Operations Research from Cornell. She has been active in various service capacities within the simulation community over many years, including the WSC Board of Directors and NATO Modeling & Simulation Group's *MSG-088: Data Farming in Support of NATO*, and is currently the Chair of the ASA Section on Statistics in Defense and National Security. Her web page is http://faculty.nps.edu/smsanche/ and her email is ssanchez@nps.edu.

TURGUT KAYMAL is a Surface Warfare Officer in the Turkish Navy. He received his B.S. in Computer Science from the Turkish Naval Academy in Istanbul, Turkey. Upon graduation he attended officer training school, and then was assigned to Turkish frigate TCG KEMALREIS as an electronics officer. He participated in many military and humanitarian operations on board Turkish frigates TCG KEMALREIS and TCG SALIHREIS. He is working on his M.S. in Operations Research at the Naval Postgraduate School. His e-mail is turgutkaymal@gmail.com.