Sensitivity Analysis Using Design of Experiments in Ballistic Missile Defense

Jacqueline K. Telford
The Johns Hopkins University Applied Physics Laboratory
11100 Johns Hopkins Road, Laurel, Maryland 20723-6099

A sensitivity analysis discovers the functional relationship between a response variable (Y) and many possible explanatory variables (X_1, X_2, … , X_k). The first step is a screening experiment to determine which explanatory variables need to be included in the function. Since only two levels are used, any non-linearity of the effects cannot be detected. The second step is to fit a response surface model by a second-degree polynomial of the important X's found in the screening experiment. Three types of designs were used in this study: central composite design, three-level fractional factorial design, and D-optimal design. These three types of designs were evaluated using the same number of additional experiments to determine the efficiency in estimating the response surface. These design of experiment techniques have been successfully applied to several Ballistic Missile Defense sensitivity studies to maximize the amount of information in a minimum number of computer simulation runs.

Keywords: Sensitivity analysis, Design of experiments, Simulations

1.0 Introduction. The basic situation is that of needing to evaluate some process with input variables called factors and with measured output variables called responses. This process could be a complex computer simulation model or a manufacturing process with raw materials and temperature and pressure settings as the inputs and a product being produced. If the input variables to the process are varied, the outputs will vary. The question is which input variables (factors) are causing the majority of the variability in the output (responses), in other words, which factors are the "drivers." It is desirable to answer the question of where the variability is coming from (also known as "sensitivities") with a minimum expenditure of resources.

Experimental design is an effective tool for maximizing the amount of information gained from a study while minimizing the amount of data to be collected, which, in this case, is minimizing the number of computer runs. Factorial experimental designs investigate the effects of many different factors in a single study, instead of conducting many separate studies, each varying one factor at a time. Factorial designs allow estimation of the sensitivity to each factor and also to combinations of two or more factors at a time.

This paper describes a process for identifying Ballistic Missile Defense (BMD) Family of Systems (FoS) needs using experimental design techniques and shows some of the findings from a first implementation of the methodology. The sensitivity analysis proceeds in two steps: a screening experiment to determine the main drivers and a response surface experiment to determine the shape of the effects (linear or curved). The following sections describe the methodology, data, modeling assumptions, and some results from the study.

2.0 Screening Design Methodology. Many factors are screened in a sensitivity analysis to determine which are the main drivers of system performance. However, as the number of factors increases, the total number of combinations increases geometrically. For this reason, studies employing experimental design should use a method such as the Fractional Factorial Method which produces high confidence in sensitivity results using very small fractions, that is, a small subset of the total number of combinations, in this case as small as 1 in 200 hundred billion.

2.1 *General Concepts of Experimental Design.* In experimental design, certain terminology is used. The controllable input variables to the experiment (in this case, the simulation program) are the factors. The performance measures output from the experiment are called responses. The polynomial equation that is frequently used to model the response variable (Y) as a function of the input factors (X's) is :

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} \beta_{ij} X_i X_j + \sum_{\substack{i=1 \\ i \neq j \neq k}}^{p} \sum_{j=1}^{p} \sum_{k=1}^{p} \beta_{ijk} X_i X_j X_k + \cdots \qquad (1)$$

where β_0 represents the overall mean response

β_i represents the main effects for each factor (i = 1, 2, ... , p)

β_{ij} represents the two-way interaction between the ith and jth factors

β_{ijk} represents the three-way interaction between the ith, jth, and kth factors.

Usually, two values of the X's (called levels) are used in the experiment for each factor, denoted by "high" and "low" and coded +1 and –1. The use of only two levels implies that the effects are monotonic on the response variable, but not necessarily linear. At least three levels of the factors would be required to detect curvature. Interaction is present when the effect of a factor on the response variable depends on the level of another factor. Graphically, this can be seen as two non-parallel lines when plotting the response means from the four combinations of high and low levels of the two factors. The β_{ij}'s account for the two-way interactions and it is desirable to be able to estimate these effects. Two-way interaction can be thought of as the correction to (or lack of) perfect additivity of the factor effects.

Experimental designs can be categorized by their resolution level. The higher the resolution level (as in optics, being able to separately resolve objects), the more terms in the regression analysis equation that can be estimated. Confounding is the opposite of resolution. Confounding occurs when only the summation of several effects can be estimated, not the effects separately. Confounded results require that additional experiments be conducted to untangle the results, to clearly identify which factor combinations are drivers and which are not (Daniel, 1962). Resolution levels are usually denoted by Roman numerals (III, IV, and V are the most commonly used). The effects in the regression analysis equation are not confounded if the sum of their "ways" is less than the resolution level of the design. In order to have all of the two-way interactions unconfounded from each other, the resolution level needs to be at least V. However, in the resolution level V, the main effects ("one-ways") are confounded with some four-way interactions and the two-way interactions are confounded with some three-way interactions. Therefore, it is usual to assume that most of the three-way and higher order interactions are negligible. The three-way and higher order interactions are not estimated separately, but their combinations are used to estimate the precision of the estimates and to compute the confidence intervals on the effects.

Table 1. Resolution Levels and Confounding Patterns

Resolution Level	**Confounding Patterns**
II	Main effects confounded with themselves
III	Main effects not confounded with themselves but confounded with two-way interactions
IV	Main effects not confounded with two-way interactions, but two-ways confounded with themselves
V	Main effects and two-way interactions not confounded except with higher order interactions

Factorial designs collect data at the vertices of a cube in p-dimensions (p is the number of factors being studied). If data is collected from all the vertices, the design is a full factorial, requiring 2^p runs and no confounding occurs. Fractional factorial designs collect data from a specific subset of all possible vetrices and require 2^{p-q} runs. Fractional factorial designs can determine which factors and their combinations have significant effects on the response variable. Fractional factorial designs yield sets of unconfounded effects (main, two-way, three-way, . . .), depending on the resolution level of the design. The minimum number of runs needed for Resolution IV and V designs for different numbers of factors are shown in Tables 2 and 3. Law and Kelton (2000) specifically "recommend that only designs of resolution V or higher be considered," page 639. However, as the number of factors increase, it may not be feasible to perform the Resolution V design. Since the significant two-way interactions are most likely combinations of the significant main effects, a Resolution IV design can be used, especially if the factors have monotonic effects on the response variable. A second, smaller Resolution V design using only the significant main effects (as determined from the Resolution IV experiment) can be performed to determine if there are any significant two-way interactions. Fractional factorial designs have been suggested for use in computer simulations in Jacoby and Harrison (1962), Hunter and Naylor (1970), Kleijnen (1975), and Biles (1979).

Table 2. Resolution IV Designs: All Main Effects Free of Two-way Interactions

Number of Factors	Minimum Number of Runs
1	2
2	4
3 - 4	8
5 - 8	16
9 - 16	32
17 - 32	64
33 - 64	128
65 - 128	256
129 - 256	512

Table 3. Resolution V Designs: All Main Effects and Two-way Interactions Unconfounded

Number of Factors	Minimum Number of Runs
1	2
2	4
3	8
4 - 5	16
6 - 7	32
8	64
9 - 11	128
12 - 17	256
18 - 22	512
23 - 31	1,024
32 - 40	2,048
41 - 54	4,096

If a factorial design is used in a screening experiment of many factors, there is no need to replicate the same combinations of factors. Factorial designs, including fractional factorials, have essentially built-in replication. More design points are preferable to replicating the same points. An experimental design is a matrix of +1's and –1's with one column for each factor, and one row for each set of factor combinations, called a design point, labeled “Run” in Table 4.

Table 4. Notational Experimental Design Matrix

Run	Factor 1	Factor 2	Factor 3	.	.	.	Factor 47
1	–1	–1	–1				–1
2	+1	–1	–1				+1
3	–1	+1	–1				+1
.							.
.							.
.							.
xx	+1	+1	+1	.	.	.	–1

The statistical textbooks such as those by Box, Hunter, and Hunter (1978) and Montgomery (2000) describe the concepts of this section in more detail. Law and Kelton (2000) also discuss factorial designs; however, Law and Kelton’s analysis requires replicating the experimental design, is not efficient, and is therefore not recommended. The analysis of factorial data found in statistics texts and in statistical software packages is preferred. The traditional statistical analysis of factorial designs has the feature that there is equal precision on the main effects and on the interactions being estimated. The confidence intervals on the main effects and on the interactions are of equal length and are as small as possible. Confidence intervals on the main effects will be shown in Figure 2.

2.2 *Survey of Other Approaches to Experimental Design.* There are many approaches to and types of experimental designs. The typical engineering experiment involves varying one factor at a time, while holding constant the other factors such as to the standard operating conditions. The shortcomings of this type of design are that there is no information at other combinations of operation conditions or on possible two-way interactions. Another type of design is the Random Balance design, but this design randomly confounds the effects that one is trying to study. The Random Balance design might be only a resolution II design. A Plackett-Burman design is used when only a very limited number of runs can be performed (for example in Grier et al., 1999), but confounds the two-way interactions with the main effects and is only valid if there are no two-way interactions. A Blocked design reduces variability by testing for effects on groups of similar experimental units, but is not normally needed in a simulation study. A Latin hypercube design employs blocking in multiple directions and, similarly to the Plackett-Burman design, assumes no multi-way interactions. Plackett-Burman and Latin hypercube are resolution III designs

2.3 *Application to Simulation Model.* Each row of the design matrix specifies the particular combination of high or low values of each of the factors to be run. In this case, there are 47 factors, so there is a high value (denoted by +1) or low value (denoted by –1) for each of the factors specified for each simulation run, which corresponds to one row of the design matrix. For the very large experiments illustrated in this paper, the experimental designs were generated and the data from the simulation runs were analyzed using the SAS® software, version 7. The exact factor combinations specified by the experimental design must be run to achieve the desired resolution level.

The steps to perform the analysis are described here. First, each of the input factors specified as +1 or –1 in the design matrix are converted to the actual engineering values and then the simulation is run for each design point. The response variables or Measures of Effectiveness (MOEs) are associated with each of the design points and the sensitivity results for each MOE are calculated using regression analysis (Equation 1). The sensitivity results are the change in MOE caused by changing each main effect (factor) from the "low" to "high" value, along with the 95% confidence limit. The two-way interactions and their confidence limits are also estimated if they can be separated, depending on the resolution level of the design. It should be noted that in executing this process, a number of scripts have been written to run the Extended Air Defense Simulation (EADSIM) automatically rather than via the Graphical User Interface (GUI). Figure 1 contains a pictorial representation of the analysis process using the EADSIM program. The UNIX script shown is an example with a single processor machine operation. There are other scripts required to employ multiple processor machines.

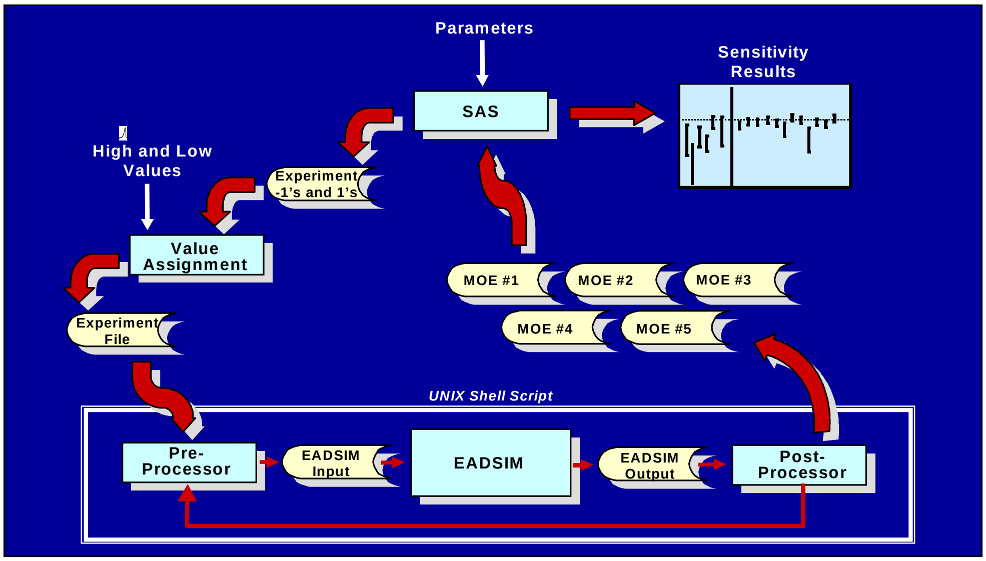

Figure 1. Process for Running the Sensitivity Analysis

We used a Fractional Factorial experimental design and EADSIM to screen 47 factors for their relative importance in far-term (i.e., 2010 timeframe) Northeast Asia (NEA) and Southwest Asia (SWA) scenarios over the first 10 days of the war. A three-tiered defense system was employed for both scenarios, including an Airborne Laser (ABL), a Ground-Based (GB) Upper Tier, and a Lower Tier comprised of both Ground-Based and Sea-Based (SB) systems.

The primary MOE for the study was FoS Protection Effectiveness and the secondary MOEs were inventory usage for each of the defensive weapon systems. We defined FoS Protection Effectiveness as the number of threats negated divided by the total number of incoming threats over the course of a scenario.

The sensitivities to the 47 factors were calculated from the MOEs at the completion of 10 days of warfare, but, as a check, the sensitivities at times less than 10 days were computed and the results were found to be comparable. This is important because it means the sensitivity results are roughly the same no matter what time is selected, and indicates the robustness of the method to variations in the scenario length.

Table 5 shows the 47 factors that were screened in the study. We selected these factors by doing a functional decomposition of the engagement process for each defensive weapon system (e.g., a radar must detect, track, discriminate, and assess the success of intercept attempts) and then by accuracy, reliability, and timeline factors associated with each of those functions.

Table 5. Factors to be Screened

Threat RCS	GB Lower Tier 2 Reaction Time
Satellite Cueing System Prob of Detection	GB Lower Tier 2 Pk
Satellite Cueing System Network Delay	GB Lower Tier 2 Vbo
Satellite Cueing System Accuracy	SB Lower Tier Time to Acquire Track
Satellite Cueing System Time to Form Track	SB Lower Tier Time to Discriminate
GB Upper Tier Time to Acquire Track	SB Lower Tier Time to Commit
GB Upper Tier Time to Discriminate	SB Lower Tier Time to Kill Assessment
GB Upper Tier Time to Commit	SB Lower Tier Prob of Correct Discrimination
GB Upper Tier Time to Kill Assessment	SB Lower Tier Prob of Kill Assessment
GB Upper Tier Prob of Correct Discrimination	SB Lower Tier Launch Reliability
GB Upper Tier Prob of Kill Assessment	SB Lower Tier Reaction Time
GB Upper Tier Launch Reliability	SB Lower Tier Pk
GB Upper Tier Reaction Time	SB Lower Tier Vbo
GB Upper Tier Pk	Network Delay
GB Upper Tier Vbo	Lower Tier Minimum Intercept Altitude
GB Lower Tier Time to Acquire Track	Upper Tier Minimum Intercept Altitude
GB Lower Tier Time to Discriminate	ABL Reaction Time
GB Lower Tier Time to Commit	ABL Beam Spread
GB Lower Tier Prob of Correct Discrimination	ABL Atmospheric Attenuation (j-95%)
GB Lower Tier 1 Launch Reliability	GB Upper Tier Downtime
GB Lower Tier 1 Reaction Time	GB Lower Tier Downtime
GB Lower Tier 1 Pk	SB Lower Tier Downtime
GB Lower Tier 1 Vbo	ABL Downtime
GB Lower Tier 2 Launch Reliability	

As in any factorial experimental design study, we selected “high” and “low” values for all factors to be screened and, in a full factorial design, we would have run EADSIM for all possible combinations of the high and low values. We selected high and low values in this study to cover a large range of operating conditions for each factor. Our goal was to assess FoS sensitivities resulting from large variations in the 47 factors. For example, we varied the Probability of Kill (Pk) for all weapon over a range of 30% between the low and high values. If no sensitivity for Pk is indicated in the screening analysis, we can say with reasonable confidence that Pk is not a driver of FoS. Follow-on response-surface analysis is warranted for those factors flagged as being drivers in the screening analysis to identify possible “knees in the curve” in FoS performance in response to smaller changes in those factors.

We conducted the NEA and SWA screening experiments to find the main factor (i.e., linear effects) and two-way interactions for the 47 factors. We assumed all three-way and higher interactions were insignificant. The number of required experiments (i.e., EADSIM runs) was driven by the number of factors, the precision needed to resolve technology drivers from underlying randomness in the problem, and the need for “unconfounded” results. A few of the factors in Table 5 were selected for reasons unrelated to technology issues. Weapon system downtimes are the best example of this. Firing units in this study experienced downtimes that were varied over a 20% range of the total scenario time. Future sensitivity studies could vary firing unit downtimes from 0% to 100%, effectively turning entire weapon systems on and off, to explore force structure and architectural issues.

We initially conducted 512 EADSIM runs to screen the sensitivities of the 47 factors in the NEA scenario. This is a Resolution IV design and resolves all the 47 main factors but leaves confounded most of the 1,081 possible two-way interactions. After analyzing results from the initial 512 runs, 17 additional, separate experimental designs were needed (for a total of 352 additional EADSIM runs) to resolve the confounding in the two-way interactions for FoS Protection Effectiveness.

We learned from the NEA screening study that more runs are warranted in the initial experiment to reduce or eliminate the number of additional experiments needed to untangle the results. The time saved by not having to untangle results is well worth the additional computer runtime. Thus, for the SWA screening study, we conducted 4,096 EADSIM runs to find the main factors and two-way interactions for the 47 factors, all unconfounded. This was a Resolution V design. An added benefit of conducting more experiments is that smaller error estimates are obtained (approximately one third less), meaning that the relative importance of the performance drivers can be identified with higher certainty.

Running EADSIM 4,096 times for the SWA analysis, each with a 25- to 40-minute runtime, was a formidable challenge. To make this feasible, we conducted the study as part of a cooperative effort with analysts at the Ballistic Missile Defense Simulation Support Center (BMD SSC) located at the Joint National Test Facility (JNTF) in Colorado Springs, Colorado. A majority of the 4,096 EADSIM runs for the SWA analysis were run on multi-processor computers operated by the BMD SSC.

Figure 2 illustrates the main factor sensitivities to the 47 factors for both NEA and SWA. The colored dots in Figure 2 represent the sensitivity to each factor and the error bars around the colored dots are 95% confidence bounds for the results. The y-axis is the difference in the average Protection Effectiveness for a factor between the “high” and “low” values. Factors are flagged as being FoS performance drivers if the 95% confidence bounds do not include zero as a probable result. Factors shown in Red in Figure 2 were found to be performance drivers in both the NEA and SWA scenarios. Factors shown in Blue were found to be drivers in NEA only, and factors shown in Green were found to be drivers in SWA only. Factors that were not found to be drivers in either scenario are shown in Grey.

The sensitivities in Figure 2 are ranked according to the relative importance of the factors in the NEA scenario. The Red factors all appear at or near the top of the figure, indicating that the same factors that are most important in the NEA scenario tend to be also the most important factors in the SWA scenario. The differences in the sensitivities between the two scenarios result from geometric and laydown differences inherent to those theaters.

If the initial experiment was designed to screen a large number of factors in a Resolution IV design, many two-way and three-way interactions are confounded, that is, only linear combinations of the two-way interactions can be estimated from the differences of means. When the data is reanalyzed for a smaller number of factors (for example, five factors), often all of the two-way and sometimes all of the three-way interactions among the five factors are unconfounded. However, the same combination of differences of means may estimate more than one of two-way and three-way interactions among the five factors. In this case, some of the interactions are confounded. Only by running an additional experiment to examine the specific five factors and their two-way and three-way interactions can be effects be separately estimated (resolved).

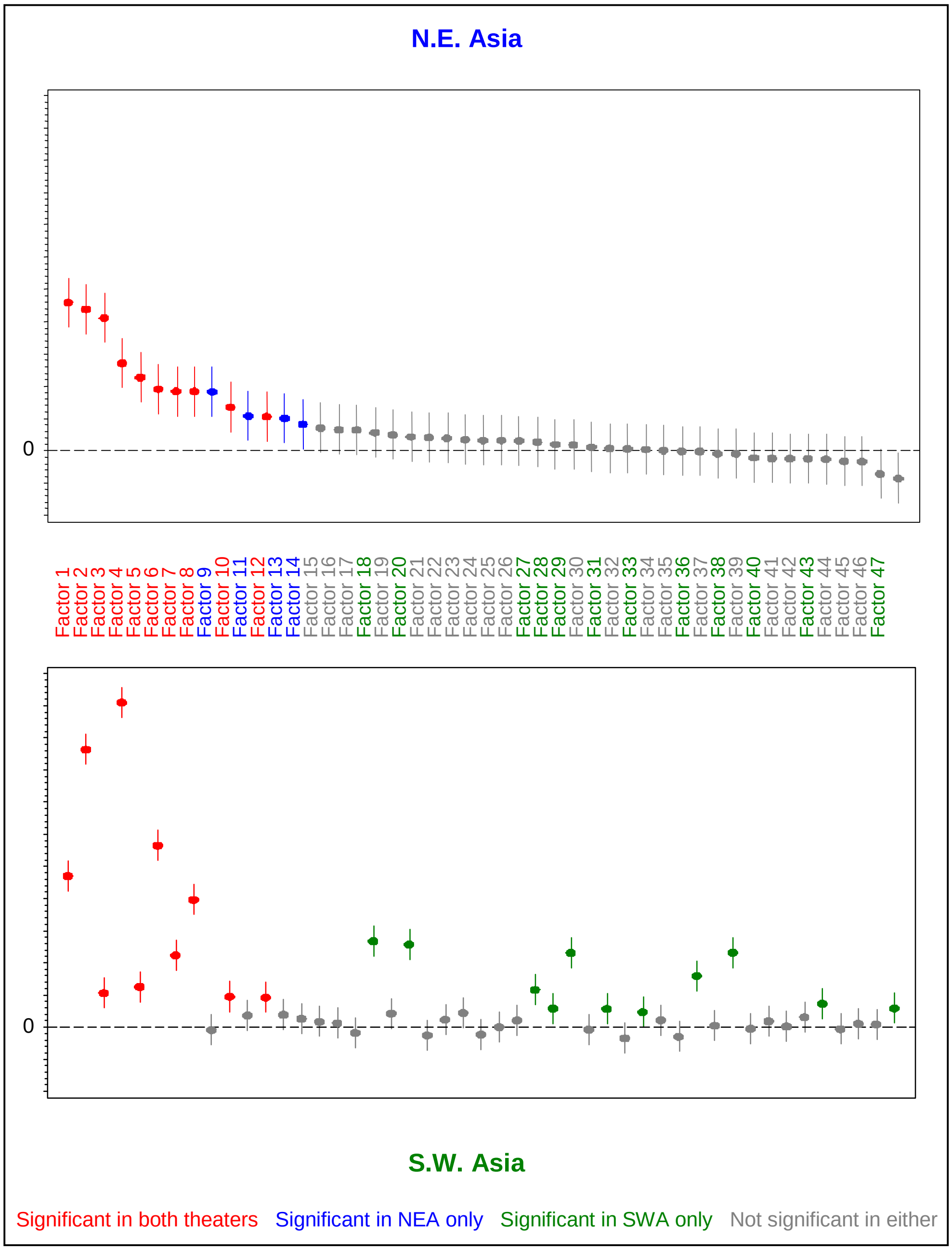

Figure 2. Protection Effectiveness: 47 Main Effects and 95% Confidence Limits

An example of a significant interaction effect can be seen in Figure 3, as the two lines in the interaction graph are not parallel. The increase in Protection Effectiveness from improving Factor 6 (denoted as F6 in the graph) is large if Factor 9 is at the low level, but essentially zero if Factor 9 is at its high level. (Factor 6 and Factor 9 are not the sixth and ninth values listed in Table 5.)

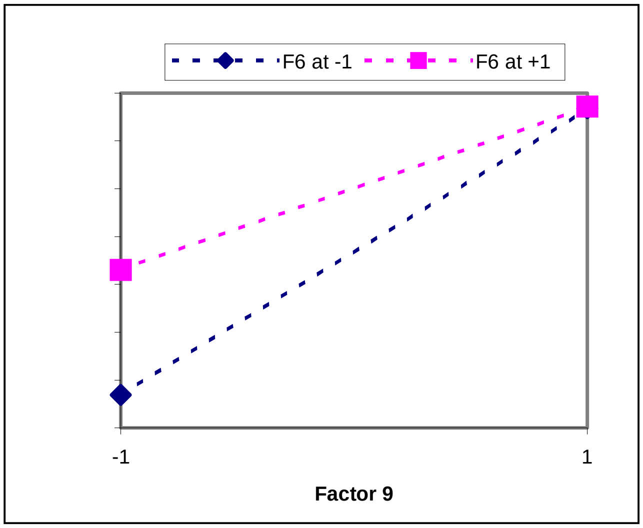

Figure 3. Protection Effectiveness: Two-way Interaction Between Factors 6 and 9 from the Screening Experiment

3.0 Response Design Methodology. Once the screening experiment has been performed and the important factors have been determined, the next step is to perform a response surface experiment. The polynomial equation that is frequently used to model the response surface is a quadratic model with cross-product terms is:

$$Y = \beta_0 + \sum_{i=1}^{p} \beta_i X_i + \sum_{\substack{i=1 \\ i \neq j}}^{p} \sum_{j=1}^{p} \beta_{ij} X_i X_j + \sum_{i=1}^{p} \beta_{ii} X_i^2 \qquad (2)$$

where β_0 represents the overall mean response

β_i represents the main effects for each factor (i = 1, 2, ... , p)

β_{ij} represents the two-way interaction between the ith and jth factors

β_{ii} represents the quadratic effect for the ith factor.

In order to fit the fully second-degree polynomial in Equation (2), more than two levels for the X's variables are needed, usually three, that is, a "medium" as well as a "high" and "low" and these are coded +1, 0, and –1. The use of three levels can model possible curvature. A total of 3^k computer simulations are needed to take observations at all the possible combinations of the three levels of the k variables. If 2^k computer simulations is large, then 3^k computer simulations is much larger. This is the value of conducting the initial screening study to reduce k to a smaller number. Even so, 3^k computer simulations may still be prohibitively large.

There are three types of experimental designs that are commonly used for response surfaces: the central composite design, the three-level fractional factorial design, and the "optimal" designs. For the central composite design, the design points are the augmentation of the two-level fractional factorial with points on the faces of the hypercube (or further out if a rotatable design is desired) and at the center of the design space. For the three-level fractional factorial design, the design points are a subset of all the possible 3^p points in the design space. For the "optimal" design, the design points are selected by a statistical criterion such as minimizing the uncertainty on the estimated effects, the determinant of X′X, where X is the design matrix, which are called D-optimal designs. An "optimal" design is useful if too many points are required by a fractional factorial design or there is an irregular design space. Response surface methods are discussed in more detail in Box and Draper (1987).

The minimum number of runs needed for Resolution V designs for different numbers of factors are shown in Table 6. From the screening design, there are 11 main effects that were statistically significant and have at least a 1% effect on Protection Effectiveness. For 11 factors, Table 6 shows that a minimum number of 243 runs are needed. To provide a fair comparison among the three types of response surface designs, 243 new runs were made for each of the three types of designs. The central composite design is 10 replicates for each of the 22 faces of the hypercube plus 23 replicates of the center of the cube. The "optimal" design also contained 243 points.

Table 6. Three-level Resolution V Designs: All Main Effects and Two-way Interactions Unconfounded

Number of Factors	Minimum Number of Runs
1	3
2	9
3	27
4 – 5	$81 = 3^4$
6 – 11	$243 = 3^5$
12 - 14	$729 = 3^6$
15 – 21	$2{,}187 = 3^7$
22 – 32	$6{,}561 = 3^8$

The comparison of the three designs are shown in Tables 7 and 8. The standard deviations of the effects are generally minimized by the three-level fractional factorial design, with quadratic effect term having nearly twice as large standard errors for the central composite and "optimal" designs as compared with the fractional factorial design (Table 7). The number of statistically significant effects with effects estimated to be larger than 1% is generally largest for the three-level fractional factorial design, with fewer quadratic effects found by the central composite and "optimal" designs (Table 8). Therefore, the three-level fractional factorial seems to be the best design for estimating quadratic effects, which is the reason for the response surface experiment.

Table 7. Comparison of Three-level Resolution V Designs: Statistical Measures

Statistical Measure	3^{11-6} Fractional Factorial	Central Composite: Only Star and Center Points	D-Optimal
Det[(X′X)]	50	8 (no cross-products)	59
Standard Error:			
Main Effects	.0016	.0030	.0015
Two-way Interactions	.0019	--	.0016
Quadratic Effects	.0027	.0041	.0051

Table 8. Comparison of Three-level Resolution V Designs: Number of Significant Effects

Number of Significant Effects	3^{11-6} Fractional Factorial	Central Composite: Only Star and Center Points	D-Optimal
Main Effects	11	11	8
Two-way Interactions	7	5	8
Quadratic Effects	6	4	2

Examples of a significant quadratic main effect and a significant two-way interaction for the three-level fractional factorial response surface experiment are shown in Figure 4. Factor 6 and Factor 9 are not the sixth and ninth factors listed as in Table 5. Factor 6 is denoted as F6 in Figure 4.

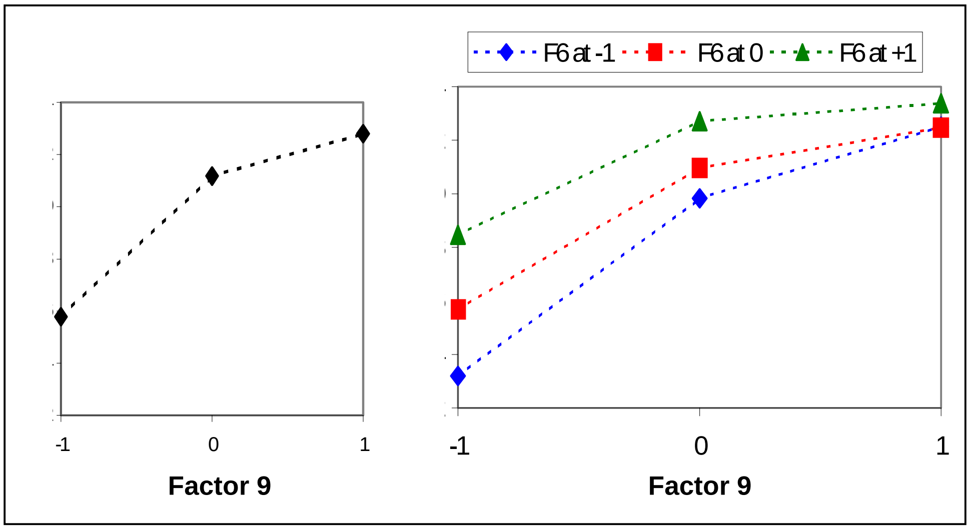

Figure 4. Protection Effectiveness: Quadratic Main Effect and Two-way Interaction Between Factors 6 and 9 from the Response Surface Experiment

The fitted model for Protection Effectiveness with quadratic and cross-product terms using the 3^{11-6} fractional factorial response surface experiment is as follows. The size of the effects are actually twice as large as the coefficients on the X terms since X has a range of 2 (from –1 to +1).

$$\text{P.E.} = .938 + .035X_9 + .026X_{11} + .017X_5 + .016X_2 + .015X_6 + .014X_1 + .012X_7 + .011X_4$$

$$+ .007X_3 + .006X_8 - .011X_6X_9 - .007X_8X_9 - .007X_2X_5 - .006X_5X_7 - .005X_3X_9$$

$$- .005X_5X_6 - .005X_1X_5 - .019X_9^2 - .011X_5^2 - .009X_{11}^2 - .008X_4^2 - .006X_3^2 - .006X_2^2$$

Not only were there two theaters examined (NEA and SWA) but also at four force levels. All of the preceding analysis was conducted at a Force Level 4, which is comparable to the Desert Storm level of logistics support prior to the operation. Force Level 1 is a rapid response with no prior warning, and Force Levels 2 and 3 are intermediate between Force Levels 1 and 4. The response surfaces for the four force levels are shown in Figure 5. The individual graphs are the response surfaces for Factors 9 and 11, the two largest main effects for Force Level 4. There is very noticeable curvature for Factor 9, especially at the lowest two force levels. As the force level increases, Protection Effectiveness increases. The different color bands are 5% increments in Protection Effectiveness, with Red being between 65% and 70% and Orange being between 90% and 95%. Therefore, the response surfaces flatten out and raise up as the force level increases, and correspondingly Protection Effectiveness improves and is less sensitive to changes in the factors. As the force level increases, there are more assets of the same type, so the reliance on the performance of any individual asset diminishes.

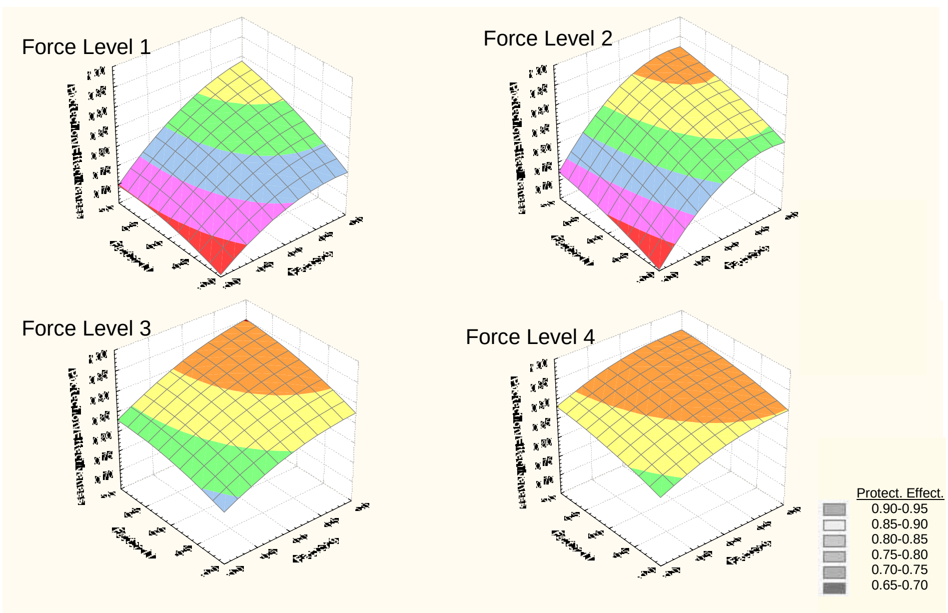

Figure 5. Protection Effectiveness Response Surfaces at Four Force Levels

4.0 Recommendations. The recommended experimental designs for the two steps in a sensitivity analysis are as follows.

1. Screening experiment: Use a two-level fractional factional design. If the number of factors is less than 32, use a Resolution V design. (If you can run more than 1,024 design points, the number of factors can be increased above 32 and Resolution V design can be used). Otherwise, use a Resolution IV design. To obtain some information on curvature, collect data at the center of design. Only one measurement at the center point is needed if the process is deteministic; if the process is stochastic, replicates are needed at the center point (10, 25, or 50 times, depending on the variability of the process). Even if curvature is indicated by an appropriate test, the factor causing the curvature cannot be identified.

2. Response Surface experiment: Use a Resolution V three-level fractional factorial design.

SAS is a registered trademark of the SAS Institute, Inc., Cary, NC in the U.S. and other countries.

References

Biles, W. E. (1979), "Experimental Design in Computer Simulation," in *Proceedings of the 1979 Winter Simulation Conference*, 3-9, Washington, DC: Institute of Electronics Engineers.

Box, G.E.P. and N. R. Draper. (1987), *Empirical Model Building and Response Surfaces*, New York: Wiley.

Box, G. E. P., W. G. Hunter and J. S. Hunter. (1978), *Statistics for Experimenters: An Introduction to Design, Data Analysis, and Model Building*, New York: Wiley.

Daniel, C. (1962), "Sequences of Fractional Replicates in the 2^{p-q} Series," *Journal of the American Statistical Association*, **57**:403-429.

Grier, J. B., T. G. Bailey and J. A. Jackson. (1999), "Response Surface Modeling of Campaign Objectives Using Factorial Analysis," *Military Operations Research*, **4**:2, 61-70.

Hunter, J. S. and T. H. Naylor. (1970), "Experimental Designs for Computer Simulation Experiments," *Management Science*, **16**:422-434.

Jacoby, J. E. and S. Harrison. (1962), "Multi-variate Experimentation and Simulation Models," *Naval Research Logistics Quarterly*, **9:**121-135.

Kleijnen, J. P. C. (1975), "Screening Designs for Poly-Factor Experimentation," *Technometrics*, **7:**487-493.

Law, A. M. and W. D. Kelton. (2000), *Simulation Modeling and Analysis (Third Edition)*, New York: McGraw-Hill.

Montgomery, D. C. (2000), *Design and Analysis of Experiments (Fifth Edition)*, New York: Wiley.

Weber, D. C. and J. H. Skillings. (2000), *A First Course in the Design of Experiments: A Linear Models Approach*, Florida: CRC Press.