A Modified Response Surface Methodology for Knowledge Discovery with Simulations

Donald E. Brown
University of Virginia, Charlottesville, VA 22903

Jeffrey B. Schamburg
University of Virginia, Charlottesville, VA 22903

Abstract: *The response surface methodology (RSM) provides an iterative process for learning that involves the sequential use of experimental design, empirical model building, and analysis of the developed models. This work provides a modified response surface methodology (MRSM) that can be applied to more complex simulation studies. These problems involve a larger number of input variables, multiple measures of performance, and complex systems relationships. Furthermore, the MRSM considers simulation analyses that have multiple objectives. These may include: 1) determining near optimal solutions; 2) understanding tradeoffs; and 3) translating the findings into generalizations. In many applications, one may begin the analysis with relatively little understanding of the variable relationships in the systems under study. The MRSM capitalizes on the underlying learning philosophy of the traditional RSM while benefiting from other knowledge discovery concepts and data mining techniques. We describe the general concepts behind the MRSM and a sequential procedure for analysis of a broad range of simulation applications. To show the steps of the MRSM, we provide examples based on an application involving combat agent-based simulation models.*

1. Introduction

The traditional response surface methodology (RSM) is an iterative process involving experimental design, empirical model building, and analysis of the developed model. This iterative process of learning through RSM is roughly formalized by [Box, 1987] and [Khuri, 1987] and consists of the repeated use of the steps, "conjecture, design, experiment, and analysis." With this approach, one must first conjecture where to conduct the experiments and the form of the model which may be used to represent the system over a given portion of the solution space. The next step is to design a "suitable experiment to test, estimate, and develop a current conjectured model." Finally, one runs the experiment and then conducts the analysis. The analysis leads to "verification of the postulated model and the working out of its consequences, or to the forming of a new or modified conjecture," [Box, 1987], and [Khuri, 1987]. The applications of the traditional RSM usually have one of three possible purposes: 1) To map a response surface over a particular region of interest; 2) To optimize the response; 3) To select operating conditions to achieve desired specifications, [Box, 1987] and [Myers, 2002].

We present a modified response surface methodology (MRSM) that extends the purpose of the traditional RSM (TRSM). The MRSM considers problems where one desires to develop generalizations about the system under study. For example, in many problems, the purpose of such an approach could be to gain an increased understanding of the tradeoffs among the various factors under study. In these problems one might desire to develop operational procedures based on the analysis of the system and the developed model. Typically, the investigator wants these generalizations to be robust to the system environment and to certain key factors in the study. Also, many of these analysis problems have multiple purposes. For example, we may want to find near optimal solutions at the same time we are trying to make generalizations in some given region of interest. In each case, the desired understanding goes beyond that which is usually obtained through TRSM. See [Schamburg, 1995].

Although TRSM has been applied to computer simulation problems and other physical systems, the most extensive applications of TRSM are in the "industrial world, particularly in situations where several input variables potentially influence some performance measure of the product or process, [Myers, 2002]." The MRSM considers applications involving complex adaptive systems simulations, or agent-based simulation models, and other complex systems problems. In many of these cases, one may begin the analysis with relatively little understanding of the variable relationships in the system under study. These problems are typified by having a large number of variables, multiple performance measures of interest, and complex stochastic relationships.

As an example of this, a complex adaptive systems simulation is used in this research for the analysis of future potential military technologies. Complex adaptive systems simulations offer potential improvements to complement current military analysis techniques. These simulations provide the ability to evaluate procedural related factors while simultaneously evaluating military technological and information related factors. In evaluating performance in these simulations, we want to understand the tradeoffs among the procedural and technological factors at the same time we are attempting to determine near optimal and robust settings for these factors. Additionally, we want to understand these relationships well enough to provide insights to procedures (rules of thumb) that can help take advantage of the benefits of the new military technologies.

TRSM requires the sequential use of experimental design, analysis of variance (ANOVA), and empirical model building through least squares. The developed MRSM includes concepts from the TRSM while capitalizing on other statistical learning and data mining concepts. For example, TRSM usually considers factorial, central composite, and Box-Behnken experimental designs. The MRSM may consider a wider range of variables by using variations of Latin hypercube designs. TRSM usually focuses on analysis of first and second order empirical models developed through the use of least squares. Our application of the MRSM relies more on the use of classification and regression trees (CART) and empirical models with restricted cubic splines (RCS). As iterations of the methodology progress, sequential experimentation, CART, and developed RCS functions allow us to get a clearer understanding of the system as the analysis solution space is reduced.

2. The Modified Response Surface Methodology

A. General Overview. The MRSM expands the RSM conjecture-design-experiment-analysis concept to *search*-conjecture-design-experiment-analysis-*generalize*. Because new technologies are being considered and because we want to understand a wide range of procedural implications, the initial set of experiments should examine a broad portion of the solution space. Searching the solution space in this way can help in refining the issues for analysis. It may result in several interesting directions for following iterations of the methodology. Additionally, it can give a better understanding of the system, the environment, and the solution space under study. Thus, the MRSM philosophy begins with "search."

Through the use of the conjecture-design-experiment-analysis process, the systems engineer gains knowledge of the system under study. Use of this methodology should result in generalizations that can be used to help develop operational procedures. The required level of understanding, in the end, is somewhat greater than might be expected from the TRSM. Therefore, the MRSM philosophy ends with "generalize." See Figure 1 below.

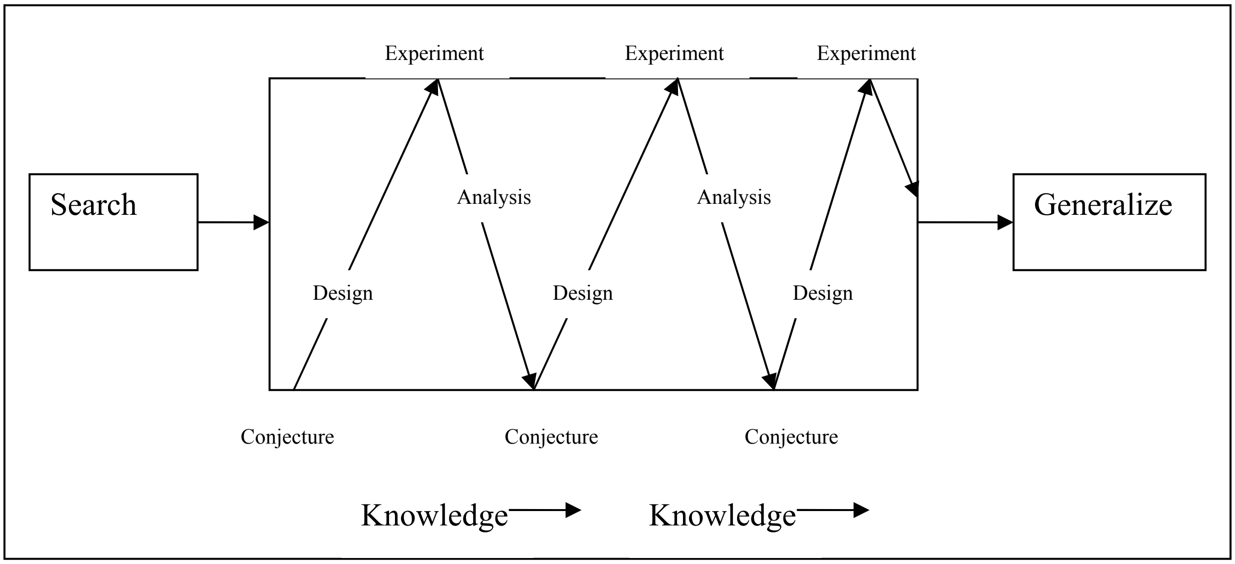

Figure 1: The Learning Process of the Modified Response Surface Methodology of Search-Conjecture-Design-Experiment-Analysis-Generalize (The Conjecture-Design-Experiment-Analysis concept is from [Box, 1987] and [Khuri, 1987]).

Using the fundamental aspects of RSM, agent-based modeling, knowledge discovery, data mining, and the concepts above, the intent of this methodology has four parts. First, we desire to gain an increased understanding of the system and how the various factors relate. Second, we wish to determine the tradeoffs and "optimal" or near "optimal" settings of the control factors. Third, we desire to translate the findings into meaningful implications for procedures. Finally, we wish to use this improved understanding to help direct higher resolution experimentation.

B. Steps of the Modified Response Surface Methodology. In an iterative way, this methodology is intended to complement existing simulation and analysis techniques. As part of a larger analysis process, this methodology receives input from existing experimentation and analysis. Likewise, the methodology provides information to improve existing simulations and analysis. The methodology includes 14 iterative steps. See figure 2. The primary focus of our research is on steps 6 through 14. These steps are described in what follows.

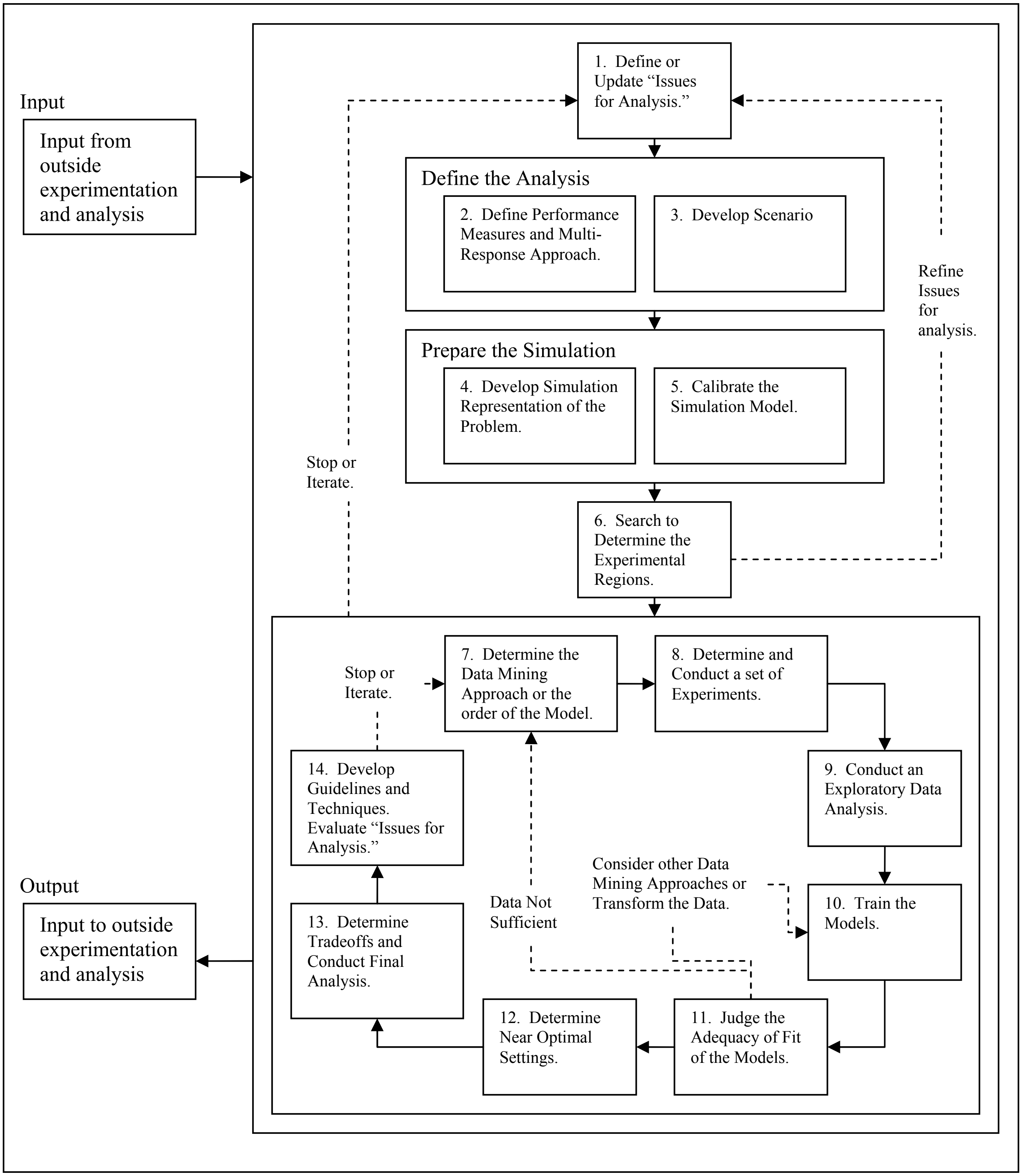

Figure 2: The Modified Response Surface Methodology.

Step 1. Define the Issues for Analysis. We first determine the issues and concepts upon which the study will focus. For example, we may be interested in finding the best combination of factors; learning how they affect the response(s); how they interact; which are most important; and which procedures are most favorable. In our application, we are interested in investigating new small-unit information technologies. As an example, some of the specific issues that we want to address in our research include:

a. How do sensor and information capabilities affect the outcome?
b. What technological parameters are the most important and what improvement do they provide?
c. What information is most helpful to soldiers and leaders in determining their actions?
d. What are the implications for squad tactical procedures? What movement rules should be used for the employment of friendly forces and the unmanned sensors?

Step 2. Define Performance Metrics and the Multiple Response Surface Approach. Here, metrics are selected to help address the issues for analysis. Most systems engineering problems involve multiple performance metrics. Our performance measures are directly related to the squad's survivability, the squad's lethality, seizing and controlling terrain, and sensor survivability. Therefore, our most important measures of performance include the number of enemy casualties, the number of friendly casualties, weather or not the squad seized the objective, and weather or not the sensor is destroyed.

Several authors provide techniques for dealing with multiple response surfaces. See [Myers, 2002], [Castillo, 1999], [Li, 2002], [Ben-Gal, 2001], [Rees, 1983], [Montgomery, 1977], [Veral, 2001], and [Yang, 2002].

For our application, in part, the multi-response approach is related to the tactical mission that will be described in the scenario. Therefore, we use an approach described in [Ben-Gal, 2001] that suggests the use of normalized performance measures to develop a "desirability function." In our application we develop a desirability function that is directly related to the tactical mission. We call this function the "mission response" function. Additionally, we also look at the performance measures individually. The "mission response" function is given by:

$$MR = \sum_{i=1}^{k} v_i z_i$$

where v_i is the value of the i[th] normalized response, z_i. The v_i can be developed so that $\sum_{i=1}^{k} v_i = 1$ and $0 \geq v_i \geq 1$ for all v_i. With our mission response function, larger is better.

Step 3. Develop the Scenario. This step requires further specification of the analysis. Here we determine the environment and the hierarchical level of the system that we want to investigate. In our application, this requires specification of the terrain, the type of operation, the friendly force, and the enemy scenario.

Our current analysis involves a small unit infantry attack on a defended enemy position. This scenario considers the use of potential future information technologies for the squad. These include: an unmanned aerial sensor, squad situational awareness software, squad call for fire capabilities, and squad communications. The mission is to destroy the enemy and to seize the key position.

Step 4. Develop the Simulation Representation of the Problem. Here we determine specific parameter values for the subsystems in the study. If possible, reliable, external data should be used to help determine parameter values where necessary. In addition, subject matter input can be used. For example, in our military problem we use the Map Aware Non-uniform Automata (MANA) simulation tool. MANA is a combat agent-based simulation under development for the New Zealand's Defence Technology Agency (DTA), [Lauren, 2002]. Initially we are interested in 44 different variables that relate to the information systems and maneuver that help address the issues for analysis.

Based on [Lauren, 2002], the following gives example definitions of some of our variables. We also indicate the ranges of values that are of interest in our study.

a. SLMSR - The Squad Leader's Maximum Sensor Range: This is the maximum range at which the squad leader can detect enemy forces. This range is of particular interest in the study because we give the squad leader direct communication to indirect fire assets. Variable range: 50 to 3000 meters.
b. SLCD - The Squad Leader's Communications Delay: This is the time to call for fire and the time to communicate to friendly forces. It represents a time delay in processing and communicating information. Variable range: 0 - 300 seconds.
c. SLSD - The Squad Leader's Speed: This is the squad leader's movement speed. Variable range: 2 to 5 miles per hour.
d. ATMDE – The Assault Team's Minimum Distance from Alive Enemy Agents: This variable limits the distance to which friendly entities will approach alive enemy agents. The value entered is the minimum distance that the agents try to maintain. Variable range is 0 to 2000 meters.
e. SMSR – The Sensor Maximum Sensor Range: In this scenario, the squad has an organic aerial sensor. This is the maximum range at which the sensor can detect enemy forces. Variable Range is 100 to 5000 meters.
f. STMSR – The Support Team's Maximum Sensor Range: Similar to the squad leader's maximum sensor range, this is the maximum range at which the support team members can detect enemy forces. Variable range: 50 to 3000 meters.

Weapons ranges and probability of hit, for example, are treated as constants in our study. For each weapon, however, we provide MANA with a table of probabilities for appropriate ranges that are less than the maximum weapon range. Probabilities associated with the initial locations, shooting, hitting, injuring, and killing opposing forces, and movement sequences add to the highly stochastic nature of our combat simulation. However, we believe that the uncertainties associated with these aspects of the problem are reflective of the uncertainties that we have about the true nature of combat. In turn, these uncertainties are reflective in our results for certain aspects of the analysis. Nevertheless, the MRSM provides a way to improve understanding of this highly stochastic system.

Step 5. Calibrate the Simulation Model. Once the simulation is set up, we run the simulation and compare the outputs to trusted or live data if available. The distribution of outputs, as a function of the inputs should be similar to what we find in the trusted or live data. Based on the results, the simulation parameters can be scaled and then adjusted so that the output is similar to the live data. This can be a complex process in itself. It may require experimental designs, experimental runs, and statistical analysis (from both the calibrated model and the trusted model) to be used in an iterative manner. Additionally, calibration may be required at several levels of the simulation model. For example, agent movement and communication may need to be calibrated first. Later phases of the calibration process may focus on overall outcomes such as casualties.

Step 6. Search to Determine the Experimental Regions. In this step we are interested in looking at a large portion of the solution space to gain a general understanding of the system. To accomplish this step of the analysis, we construct and run experimental designs that allow us to look at many factors over a broad range of values. Multiple experimental designs used in a sequential manner may be appropriate for this step. Considering the multiple measures of performance, we then conduct exploratory data analysis to determine interesting experimental regions. Based on statistical significance and relative importance with respect to the issues for analysis, some factors will be selected for further study and some factors may be screened out[1].

Alternatively, instead of continuing with the current issues for analysis, this step may result in findings that cause us to alter and refine the current issues for analysis and/or the current scenario. For example, other interesting issues may exist or some select subsystems may be more or less interesting than expected.

In our application, our initial 44 factors of interest are those that relate to the information technologies and those that provide implications for tactical procedures. Because we are interested in gaining a general understanding of the entire system, we require many observations at many different combinations of levels for the 44 factors. Here, our data is developed by running and combining multiple experimental designs with multiple replications. These experimental designs are described in step 8 below. To select the subsequent experimental regions, we use "Spearman's rho" rank correlation, statistics related to empirical models, graphical techniques, and classification and regression trees (CART).

Spearman's rho rank correlation uses the ordinary R-squared value from "predicting the rank of Y based on the rank of X and the square of the rank of X." This helps to detect the "linear and the nonlinear marginal relationships," [Harrell, 2001]. Generally, this gives us an initial look at which variables might be the most important in our analysis. We use Spearman's rho rank correlation to give us an understanding of the relative importance of the variables with respect to the response(s). For our example, figure 3 shows the Spearman rho rank correlation for friendly casualties. The graph shows that the variable, SLMSR might be the most important variable in determining friendly casualties. Additionally, STDWP might be the least important variable in determining friendly casualties.

---

[1]An alternative approach for this step could involve heuristic search methods. However, in our study, we have been more interested in learning (as illustrated in figure 1) through the use of experimental designs and analysis.

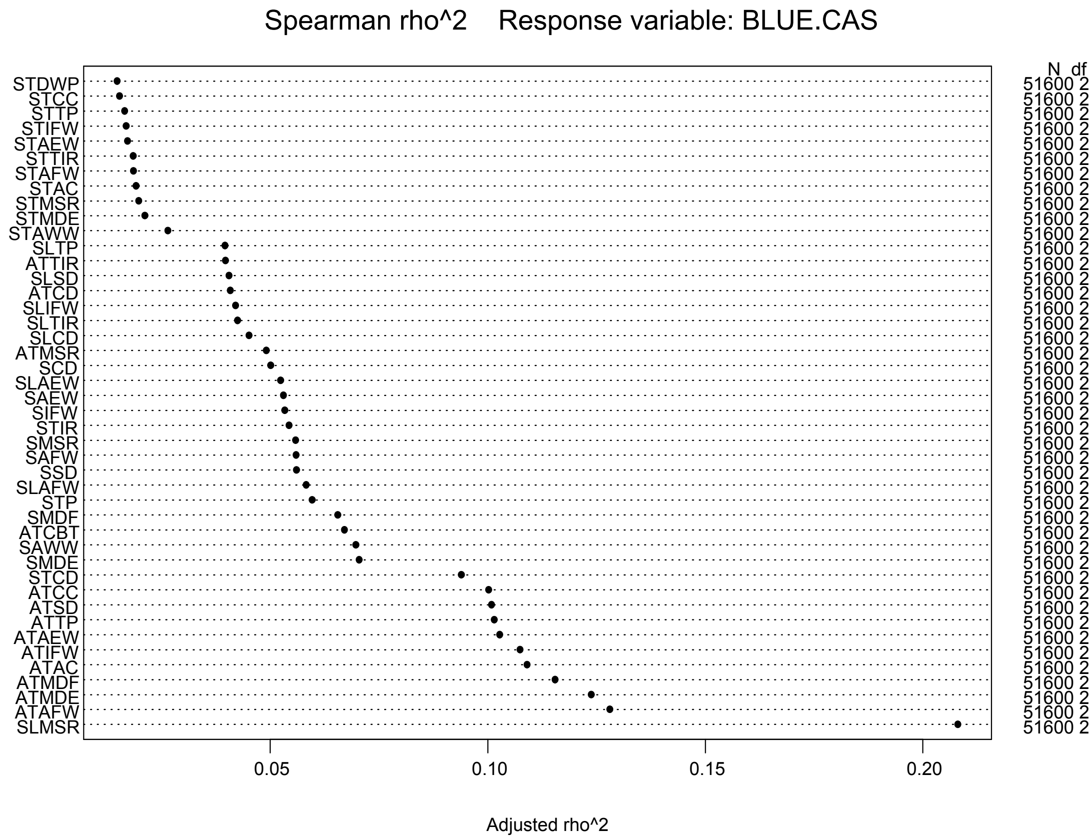

Figure 3: Spearman rho^2 for friendly casualties.

Next, we develop empirical models through the use of least squares. The resulting empirical models are selected through use of an iterative sequence of analysis of variance (ANOVA), t- statistics, and stepwise regression techniques. With 44 variables under study, we consider 1st order interactions, squared terms, and restricted cubic splines (RCS) in our model development. Considering up to more than 1000 potential terms, we screen groups of terms until about 200 or fewer terms remain in the model. Ultimately, we reduce our models until only about 175 (or fewer) of the most important terms remain. Given the response under study, the resulting models are then used to help identify the important terms in the analysis. To improve understanding of the resulting relationships, we use response surface and multi-dimensional graphical techniques. This is further described in following sections.

Finally, we use CART to help further understand these relationships and to help determine the next search region. With CART, we find good solutions through observation of the results at the terminal nodes. Then, we simply work our way up the tree to define the regions in which those solutions lie. In our application, figure 4 shows an example reduced regression tree for the mission response function. The best average solution of .7242 is found at the terminal node where the variable STAWW > 29. From this terminal node, working back through the branches of the tree, this average solution is found when SLMSR > 147, STMDE > 1, SMSR > 249, SLAFW > -57, STMSR > 64, and STAWW > 29. This information, plus information from CART analysis based on the other individual responses of interest, is used to help select subsequent experimental regions. We describe a method for developing CART in following sections.

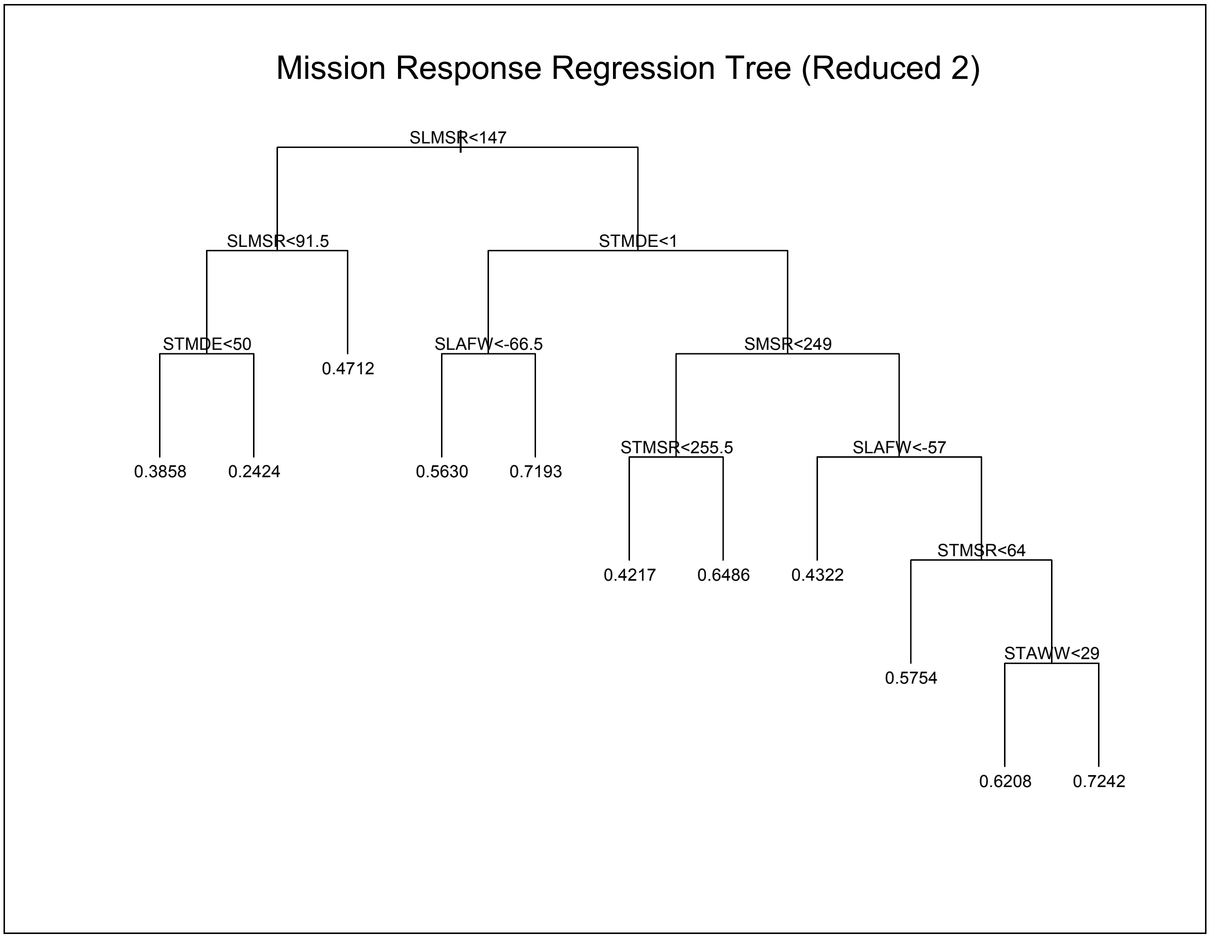

Figure 4: Regression Tree for the Mission Response Function.

Step 7. Determine the Data Mining Approach or the Order of the Model. This step requires that we next conjecture as to the form of the model which may be used to represent the system over a given portion of the solution space. Appropriately determining the order (or type) of the model in this step leads to an appropriate set of experiments in the next step. Different data mining approaches may have different training data needs. For example, higher order function approximations, developed through use of least squares, require three or more different levels for each factor. On the other hand, for first order function approximations, only two levels may be required for each factor.

In addition to the common RSM models, we consider models developed with RCS's, logistic regression models, multi-dimension graphical techniques, and CART. For development of this methodology we have found that experimental designs based on [Cioppa, 2002] allow us to use data mining analysis methods that go beyond the common RSM analysis methods. Some advantages of these designs are touched upon in step 8 below.

Step 8. Determine and Conduct a set of Experiments. Here, we determine and conduct a set of experiments that will yield the measures of performance. In our application, the experiments are conducted through use of the computer simulation program. This step includes determining which variables and what levels of these variables should be considered for an analysis. In part, this is directed by the issues for analysis and then refined through the techniques in step 6.

Limiting the reasonable number of variables that may be investigated in the analysis, TRSM considers factorial, central composite, and Box-Behnken experimental designs. The MRSM considers a wider range of variables by using variations of Latin hypercube designs. In our application, we use Cioppa's "nearly orthogonal Latin hypercube designs," [Cioppa, 2002]. Cioppa's designs allow us to look at up to 22 variables at 129 levels. While maintaining "near orthogonality," these designs also have good space-filling properties, allowing us to get a better representation of the solution-space. With these designs, we can get a better understanding of the variable relationships with the response(s) under-study over a broader portion of the solution space.

In some cases, we have used a modification of the nearly orthogonal Latin hypercube designs because some of our experimentation included one or more variables that have fewer than 129 discrete levels. In these cases, our approach has been to round the design specified level to the nearest discrete level allowed for the variable. While this method may not be optimal for some parametric statistical analysis techniques, we have found this method to provide the necessary data for our example data mining approaches. Additionally, this method does not increase the complexity of the experimentation and it does not require more than the 129 runs specified by the original design. Furthermore, with one dataset, we are able to use multiple data mining approaches while looking at multiple continuous and discrete responses. See [Cioppa, 2002].

Step 9. Conduct an Exploratory Data Analysis. The exploratory data analysis is used to determine which factors and interactions are most important and how they affect the response. The exploratory data analysis, in part, addresses some of the issues and concerns brought forth at the beginning of the process. It also helps determine which terms may be most important in training the models. Here we again use the Spearman rho rank correlation, descriptive statistics, and graphical techniques to get a better understanding of the data with respect to the responses and the issues under study. For our example, figure 5 shows how the mean of one variable on the horizontal axis (stratified by six groups of another variable) affects the number of friendly casualties on the vertical axis. The trend lines for each frame are developed through use of nonparametric smoothers. They show how the relationship changes with respect to the different stratified groups. The variable on the horizontal axis relates to one of our information technology variables while the six stratification groups relate to one of our maneuver variables[2].

[2] The nonparametric smoother is based on Cleveland's moving linear regression smoother "loess." See [Cleveland, 1979] and [Harrell, 2001].

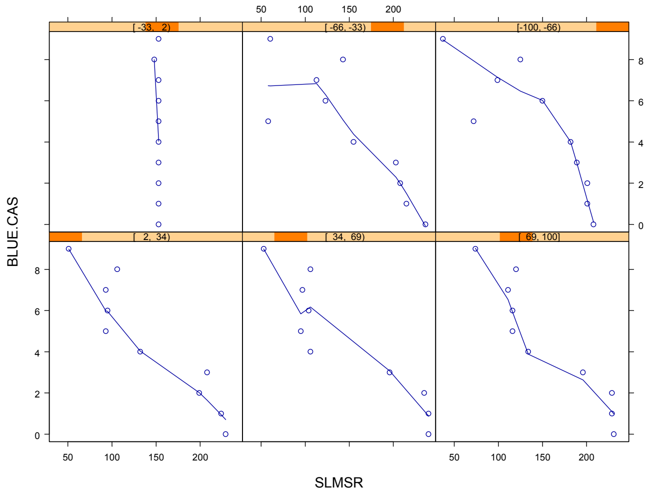

Figure 5: loess trend lines for the relationship between the mean of an information technology related variable to friendly casualties. The six frames show how the relationship changes for six groups of a maneuver variable.

Step 10. Train the Models. This step requires the determination and fitting of appropriate mathematical models from which to analyze the relationships between the input variables and the response variables. As touched on above, we use least squares regression, logistic regression, and CART techniques to fit the models. As one example, we develop regression trees through methods described by [Hastie, 2001]. See [Hastie, 2001] and [Harrell, 2001] for more detail. The general method follows:

a. Find the predictor, X_a, that gives the best possible binary split for partitioning the data into two sub-regions:

$$SR_1(a,s) = \{X \setminus X_a \leq s\} \text{ and } SR_2(a,s) = \{X \setminus X_a > s\}$$

- by finding the variable a in X at the split point s that gives:

$$\min_{a,s}\left[\min_{t_1} \sum_{x_i \in SR_1(a,s)} (y_i - t_1)^2 + \min_{t_1} \sum_{x_i \in SR_2(a,s)} (y_i - t_2)^2\right]$$

  where y_i is the response.

- For any choice a and s, the inner minimization is solved by:

$$\hat{t}_1 = ave(y_i \setminus x_i \in SR_1(a,s)) \text{ and } \hat{t}_2 = ave(y_i \setminus x_i \in SR_2(a,s))$$

b. Continue splitting previously formed subsets until some minimum node size is reached. [Hastie, 2001] says 5 for example.

c. Prune the tree backward using "cost-complexity" pruning. This may be accomplished several times using different tuning parameters (AIC, BIC, and others). Alternatively, inspection of deviance versus size or misclassifications versus size may be used.

d. Pruned trees may be compared by:

- number of total nodes or number of terminal nodes
- classification error
- deviance

In our study, we prune trees until unacceptable error or deviance is incurred. Commonly, this has resulted in trees with about 10 to 15 terminal nodes. Furthermore, trees reduced to this size are efficient; provide easy interpretation; and a good general understanding of the important relationships in the analysis. This is because there are fewer nodes to examine and the result at each node is based on a larger average number of observations. For example, we have 6450 observations from one set of experiments. For larger trees of 100 or more terminal nodes, this dataset would result in an average of 64.5 observations per terminal node. For a smaller tree with only 10 terminal nodes, there would be an average of 645 observations per terminal node. As necessary, larger trees can also be used to provide more detailed information about regions of interest or "near-optimal" solutions.

As another example, through the use of least squares, we develop models with RCS terms. These models can be more flexible than second order polynomials and can be used to better represent non-linear relationships while providing models that are linear in the tails of the relationships. An empirical model with RCS's is formed by dividing the ranges of the factors into intervals and developing a piecewise function. The endpoints of the intervals are called knots. As an example, a function with one factor X and j knots would have j + 2 terms, counting the intercept. These functions have the form:

$$f(X) = \beta_0 + \beta_1 X + \beta_2 (X - k_1)_+^3 + \beta_3 (X - k_2)_+^3 ... + \beta_{j+1} (X - k_j)_+^3$$

where the β_i are the estimated coefficients, $(z)_+ = \begin{cases} z, & if \quad z > 0 \\ 0 & if \quad z \leq 0 \end{cases}$,and the k_j represent the knots. Because the RCS's are forced to be linear in the tails, only j coefficients (counting the intercept) have to be calculated. The remaining 2 coefficients are found as functions of the other terms in the model. Normally, j = 3, 4, or 5. See [Harrell, 2001]. [Stone and Koo, 1985] and [Harrell, 2001] describe other concepts and advantages of RCS's.

Step 11. <u>Judge the Adequacy of Fit of the Models</u>. The models can be judged through use of statistical analysis, analysis of the mean square error, residual analysis techniques, and through the performance on test data. Models developed through the use of least squares require validation of the usual assumptions. First, the least squares regression requires that the residuals have constant variance and zero mean. The statistics associated with the analysis of variance also require that the residuals be normally distributed. The residual analysis includes the analysis of several residual plots to verify the required assumptions. The normality assumption may be checked through use of a normal probability plot and histogram of the residuals. The residuals should also be plotted against the predicted response values to determine if the variance is relatively constant. Finally, to check the variability in the residuals at each factor level, the residuals are plotted against each factor. Although the method of least squares and the analysis of variance are fairly robust to these assumptions, larger violations may require data transformations or weighting. See [Schamburg, 1995]. Overall, unsatisfactory model performance may result in one of the following:

a. If the model fails particular tests in this step, we may attempt to try a different transformation of the data or a different data mining approach and then return to step 10.

b. If the model does not satisfactorily predict the response, return to step 7 above, make adjustments to the experiment and go through the sequence again to improve the model.

In our application, we have been restrictive in the number of terms allowed in the final model. Because we consider a large number of variables in this methodology, we try to restrict our final models to about 175 or fewer terms. This allows only the most important relationships to remain in the model. We developed our final empirical models through use of an iterative sequence of analysis of variance (ANOVA), t- statistics, and stepwise techniques.

Because of the highly stochastic nature of the combat agent-based simulation in our study, in our case, R^2 values greater than .40 indicated relatively good model performance. In our application, we found that, per degrees of freedom, models with the RCS's typically outperform second order polynomials in terms of R^2 (or deviance in the case of our logistic regression models). This was especially true when comparing the larger models of 175 or more terms. See tables 1 through 5 below that show the results of models for all 5 of our responses of interest. In table 1, for example, model 1 is a 2nd order polynomial developed through use of stepwise regression techniques[3]. Model 1 was developed using Akaike's information criterion (AIC) for model selection[4]. Model 1 has 286 degrees of freedom and resulted in an R^2 value of .523. Model 3, a model with RCS terms and only 181 degrees of freedom, also resulted in an R^2 value of .523. Model 3 was developed using more restrictive criteria for model selection.

	R^2	D.F.
Model 1 (AIC selected 2nd order polynomial)	.523	286
Model 2 (Reduced 2nd order polynomial)	.508	140
Model 3 (Model 2 w/ RCS Terms in place of 2nd order terms)	.523	181
Model 4 (Reduced w/ RCS Terms)	.514	141

Table 1: Model Comparison: Developed linear regression models for Friendly Casualties.

[3] In each table, we start by developing the first model, Model 1, using Akaike's information criterion (AIC) for model selection. Subsequent models were selected using the more restrictive Schwarz's Bayesian Information Criterion (BIC). See [Schwarz, 1978] and [Harrell, 2001].

[4] For information about AIC see [Atkinson, 1980], [Harrell, 2001], and [Hastie, 2001].

	R^2	D.F.
Model 1 (AIC selected 2nd order polynomial)	.487	291
Model 2 (Reduced 2nd order polynomial)	.477	175
Model 3 (Model 2 w/ RCS Terms in place of 2nd order terms)	.495	217
Model 4 (Reduced w/ RCS Terms)	.487	174

Table 2: Model Comparison: Developed linear regression models for Enemy Casualties.

	R^2	D.F.
Model 1 (AIC selected 2nd order polynomial)	.450	261
Model 2 (Reduced 2nd order polynomial)	.443	174
Model 3 (Model 2 w/ RCS Terms in place of 2nd order terms)	.457	199
Model 4 (Reduced w/ RCS Terms)	.438	159

Table 3: Model Comparison: Developed Models for the Mission Response Function.

	Residual Deviance	% of Null Deviance Accounted for in Model	D.F.
Null (intercept only)	62323.7	0%	1
Model 1 (AIC selected 2nd order polynomial)	6233.4	90.0%	134
Model 2 (Reduced 2nd order polynomial)	6295.2	89.9%	122
Model 3 (Model 2 w/ RCS Terms)	6676.0	89.3%	95
Model 4 (Reduced w/ RCS Terms)	7125.0	88.6%	78

Table 4: Model Comparison: Developed logistic regression Models for the sensor survivability.

	Residual Deviance	% of Null Deviance Accounted for in Model	D.F.
Null (intercept only)	71060.7	0%	1
Model 1 (AIC selected 2nd order polynomial)	49293.5	30.6%	221
Model 2 (Reduced 2nd order polynomial)	49963.6	29.7%	160
Model 3 (Model 2 w/ RCS Terms)	47901.3	31.7%	212
Model 4 (Reduced w/ RCS Terms)	49443.0	30.4%	145

Table 5: Model Comparison: Developed logistic regression models for Seizing the Objective.

Step 12. Determine Near Optimal Settings. Here mathematical programming or heuristic search methods are used to determine "optimal" or "near-optimal" settings of the factors if required in the study. In our application, we have accomplished this step by exploiting the special structure of outputs of tree-based methods such as CART. Here, we observe terminal nodes that give the best solutions and we work back up the tree to determine the variable values for those solutions. Using this special structure, we start with the regression tree for the mission response function (i.e. the desirability function). Based on the prioritization of the other responses of interest, we repeat this process with the other trees, in order, until we have input related to all of the important variables in the study. In cases where we have conflicting input from the competing responses, preference is given to the trees related to the highest priority responses. We desire robust solutions. Therefore, we first use trees that are reduced to fewer terminal nodes (about 10 terminal nodes in our case). Terminal nodes that provide “good” solutions with many observations (say 1,000 or more in our case) are preferred to terminal nodes that provide the “best” observed solution but have only a few observations (say 50 in our case).

As another potential alternative approach, the CART optimal solution can then be used to reduce the size of the developed statistical models. It may be especially helpful in reducing the combinatorial nature and size of the solution space in the models with many RCS terms. Here, the analyst can "fathom" many of the branches of the model with RCS terms by reducing the range of variable values. In agreement with the knot locations, some of the RCS terms can be dropped from the model and others can be treated as normal cubed terms. For accurate empirical models, further improvement might be beneficial. Other methods involve using the CART solution as a good start point for optimization of the empirical model. From the start point, one can use heuristic techniques on the empirical model in an effort to find "better" solutions.

As an example partial solution in our application, we found the factor constraints in table 6 through use of the CART. In our case, we started with the tree for the mission response function. Then, in order, we used information from the blue casualties tree, the red casualties tree, the seize the objective tree, and the sensor survivability tree. Within the tree constrained region for these “good solutions,” we selected three specific solutions that are partially defined by the factor settings in table 6. The variables were previously described in step 4, “Develop the Simulation Representation of the Problem.” For an example comparison, table 7 shows the mean performance of the entire experimental dataset, the data within the tree-based constraints, and the three example solutions. The solution performance comparison is based on 50 observations of each solution. Through the use of box plots, figure 6 shows the distributions of these observations and indicates that solution 2 seems to outperform solutions 1 and 3 in terms of blue casualties, red casualties, and mission response value. Using the nonparametric Kruskal-Wallis test to compare the differences of these distributions, we found that there was a significant statistical difference at the .001 significance level based on these three responses. There was less difference in the solutions in terms of sensor survivability and seizing the objective. Further iterations of the methodology may be used to help identify a reduced region of the solution space for good solutions. Furthermore, subsequent iterations may result in improved solution performance.

		Tree-based Constraints for “Good Solutions”	Solution 1	Solution 2	Solution 3
1	SLMSR	> 1895 meters	2000 meters	3000 meters	3000 meters
2	SLCD	< 102 seconds	45 seconds	15 seconds	0 seconds
3	SLSD	> 4.2 miles/hour	4.4 miles/hour	4.4 miles/hour	5 miles/hour
4	ATMDE	< 315 meters	250 meters	250 meters	250 meters
	ATMDE	> 10 meters			
5	STMSR	> 915 meters	1500 meters	2500 meters	3000 meters
6	SMSR	> 2610 meters	3000 meters	3500 meters	4950 meters
	SMSR	< 4950 meters			

Table 6: Factor settings for the example partial solution found through use of CART.

	Entire Experimental Region Data Set	Tree-based Constraints for "Good Solutions"	Solution 1	Solution 2	Solution 3
Blue Kills (Survivability)	3.70	1.92	1.14	0.68	1.74
Red Kills (Lethality)	26.90	33.30	34.32	37.54	36.40
Seize Objective (Yes= 1)	0.56	0.97	0.98	1.00	0.96
Sensor Destroyed (Yes = 1)	0.71	0.08	0.02	0.06	0.06
Mission Response Value	0.59	0.84	0.90	0.95	0.89

Table 7: Example mean performance for the entire experimental region data set and the partial solutions indicated in Table 6.

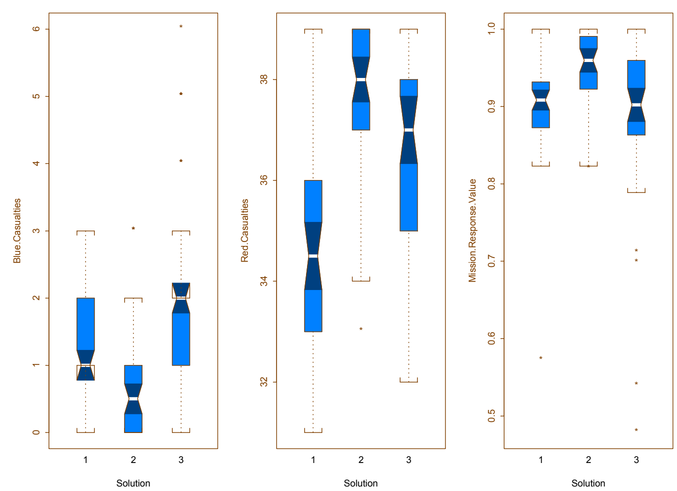

Figure 6: Box plots for comparison of the distributions of solution performance. Each graph represents 50 observations of each solution.

Step 13. <u>Determine Tradeoffs and Conduct Final Analysis</u>. Of the factors considered, we determine which of these factors or combinations of factors have the greatest effect on the responses. The sensitivity analysis includes an analysis of the tradeoffs among these factors. We desire robust solutions. That is: desired solutions are those in which slight deviations from the solution would still result in a relatively good response.

Analysis of a graphical representation of the empirical models and the CART is one of the final procedures in this methodology. A graphical representation of the response surface may be provided through the use of contour diagrams and response surface plots. Analysis of these graphs shows the relationship between two variables and the response when the values of all other variables are held constant. The contour diagram is a two dimensional representation of constant response contours over the ranges of values for two variables. The response surface plot is a three dimensional representation of the response surface over the ranges of values for two variables. For any pair of variables, multiple graphs may be analyzed by changing the values of the other variables. These graphical representations of the response surface allow one to investigate and analyze the interactions, relationships, and tradeoffs between a response variable and pairs of input variables.

For our application, figure 7 shows a two dimensional response surface plot for the most important factor found in the analysis. This graph is developed by using the reduced empirical model with RCS terms and setting the other factors to their median values. SLMSR was the most important factor in determining the mission response value, the number of friendly casualties, and the number of enemy casualties. Increasing SLMSR had a strong favorable affect. Its affect on the mission response value is shown in figure 7.

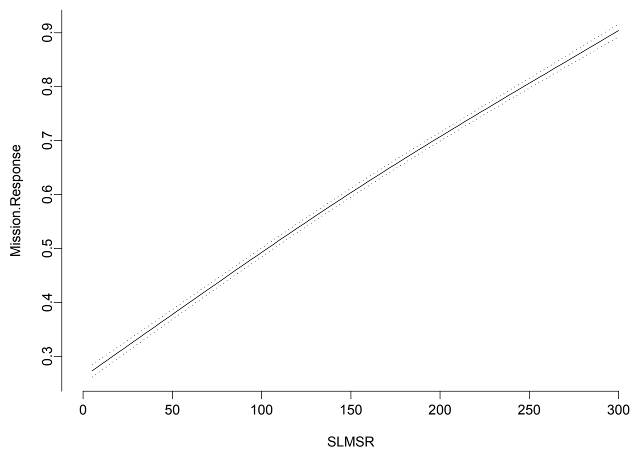

Figure 7: Two dimensional response surface plot for the most important factor found in the analysis. SLMSR is a factor that represents one of our combat technologies and is on the horizontal axis. The Mission Response is on the vertical axis.

For our application, figure 8 shows a three dimensional response surface plot for two important factors found in the analysis. SLTP and STTP represent two of our information related variables. The interaction of SLTP and STTP was found to have a highly significant affect on the Mission Response. The best Mission Response values were found where SLTP is relatively high while STTP is relatively low and where SLTP is low while STTP is high. Additionally, good Mission Response values were found where SLTP and STTP were both about the same near their mid-range values (about 300 for SLTP and STTP as observed on the graph). Furthermore, when both SLTP and STTP were high or when both were low, we expect a negative affect on the Mission Response.

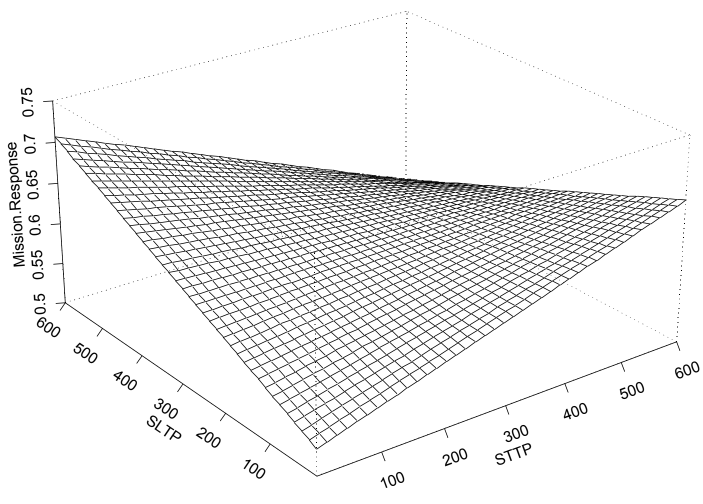

Figure 8: Three dimensional response surface plot for two important factors found in the analysis. SLTP and STTP represent two of our information related variables. Again, the Mission Response is on the Y-axis.

Step 14. <u>Develop guidelines and techniques and evaluate issues for analysis</u>. The final phase of this methodology includes drawing conclusions from the analysis techniques described in the above steps. Through the analysis and conclusions found in the steps above, we attempt to make generalizations that will be beneficial in developing procedures for the system under study. In this step we attempt to address the key issues and summarize the most important findings in our analysis. Although this step is specified for this part of the methodology, the process of evaluating the issues for analysis is actually on-going throughout the entire methodology. That is, at each step of the methodology, we may be able to draw some conclusions. At this step, we reflect on the other portions of the analysis and attempt to summarize the most important points. In cases where generalizations can not be made, the results are reported and the process may be iterated. Table 8 provides an example partial summary of our current findings related to our infantry squad attack scenario. These findings help address the issues that were identified in step 1 of the methodology. More specific findings related to how the factor settings affect the response may be observed through use of the CART(s).

a. The squad leader's maximum observation range is the most important factor in 4 out of 5 responses. This is because the squad leader has the ability to communicate directly with indirect fire assets. This direct communication capability coupled with increased observation range greatly improves squad performance.
b. Overall, the most important technological factors were SLMSR, SLCD, STMSR, and SMSR. In general, increasing SLMSR and decreasing SLCD improves squad performance. Increasing STMSR also seems to improve squad performance. Higher values of SMSR, between about 3500 and 4950 meters, depending on other factor settings, resulted in better performance. However, increasing SMSR beyond 4950 meters, in some cases, may degrade performance.
c. The support team and the assault team should use stand-off to increase the mission performance. For the support team, this was shown to have a strong impact on lethality and seizing the objective. For the assault team, being overly aggressive or assaulting too fast can reduce survivability. In turn, this decreases lethality as well. As an example we found that in most cases, the assault team should try to maintain about 250 meters between itself and live enemy forces.
d. Information for the assault team related to the location of enemy, friendly, and injured friendly forces is important in determining survivability. On the other hand, similar information for the support team and the squad leader is more important in determining lethality.

Table 8: Example partial summary of current findings related to the infantry squad attack scenario.

3. Conclusion

In this paper we have described a general methodology for analysis of simulation optimization problems involving many factors (we looked at 44, for example) and multiple responses of interest. Initially, we develop the focus of the analysis so that the investigation is pointed at answering the important issues. Through use of the methodology, we have found good solutions and have made generalizations that can ultimately be used to provide input to decision making about military technology selection and appropriate tactical procedures.

The presented MRSM is intended to provide a general approach to analysis involving complex systems simulations. This analysis process is intended to be iterative. The iterative approach helps verify the conclusions drawn from previous phases of the process. It additionally leads to increased understanding of the complex relationships involved in large systems. The study should be set up so that one may gain information from the analysis of a given scenario that may be beneficial in the analysis of upcoming scenarios.

The methodology is also intended to be flexible. The steps and tools described above should be adapted to the problem and the issues at hand. While the TRSM provides a useful approach to more specific analysis situations involving fewer factors, the MRSM broadens TRSM concepts by incorporating other data mining techniques. Our examples focus more on using modified Latin hypercube experimental designs, CART, RCS functions, and graphical techniques in an iterative way and this seems to be appropriate for a broad range of analysis cases. We selected these techniques for ease of interpretation and for the natural way they allow one to gain increased understanding over sequentially reduced experimental regions of interest. However, with the use of experimental designs that allow for a large number of variables (more than 10 for example), a broader range of data mining techniques may be appropriate for other analysis situations. For example, in analysis where the resulting empirical model is more important, neural networks or support vector machines may be more appropriate because of their predictive potential, [Hastie, 2001]. However, interpretability may become an issue with the use of these techniques. If so, a combination of neural networks and CART, for example, may be appropriate. Nonetheless, the selection of techniques used in the MRSM should be dependent on the analysis situation and the analysis objectives.

4. References

(1) Atkinson, A. C., "A Note on the Generalized Information Criterion for Choice of a Model," Biometrika, Vol. 67, pages 413 to 418, 1980.

(2) Ben-Gal, Irad, and Joseph Bukchin, "The Ergonomic Design of Workstations using Virtual Manufacturing and Response Surface Methodology," IIE Transactions, Vol. 34, pages 375 to 391, 2002.

(3) Box, George E. P., and Norman R. Draper, Empirical Model-Building and Response Surfaces, John Wiley and Sons, New York, 1987.

(4) Brown, Donald E., C. L. Pittard, and Andrew R. Spillane, "ASSET: A Simulation Test Bed for Data Association Algorithms," Computers and Operations Research, Vol. 19, No. 6, pages 479 to 493, 1992.

(5) Cioppa, Thomas M., Efficient Nearly Orthogonal and Space-Filling Experimental Designs for High-Dimensional Complex Models, Dissertation, Naval Postgraduate School, Monterey, CA, 2002.

(6) Cornell, John, Experiments with Mixtures: Designs, Models, and the Analysis of Mixture Data, 3rd Edition, John Wiley And Sons, Inc., New York, 2002.

(7) Del Castillo, Enrique, and Shu-Kai Fan, "Optimization of Dual Response Systems: A Comprehensive Procedure for Degenerate and Nondegenerate Problems," European Journal of Operational Research, Vol. 112, 1999.

(8) Fang, Kai-Tai, Dennis K. J. Lin, Peter Winker, and Yong, Zhang, "Uniform Design: Theory and Application," Technometrics, Vol. 42, No. 3, Aug 2000.

(9) Fayyad, U. M., G. Piatetsky-Shapior, and P. Smyth, "From Data Mining to Knowledge Discovery," in Advances in Knowledge Discovery and Data Mining, pages 1-34, AAAI/MIT Press, 1996.

(10) Fayyad, U. M., G. G. Grinstein, and A. Wierse, Information Visualization in Data Mining and Knowledge Discovery, Morgan Kaufmann Publishers, New York, 2002.

(11) Greenwood, Allen G., Loren Paul Rees, and Fernando C. Siochi, "An Investigation of the Behavior of Simulation Response Surfaces," European Journal of Operational Research, Vol. 110, pages 282 – 313, 1998.

(12)Han, Jiawei, and Micheline Kamber, Data Mining Concepts and Techniques, Morgan Kauifmann Publishers, New York, 2001.

(13)Harrell, F.E., Regression Modeling Strategies, Springer-Verlag Inc. , New York, 2001.

(14)Hastie, T., Tibshirani, R., and Friedman, J. The Elements of Statistical Learning, Springer-Verlag Inc., New York, 2001.

(15)Khuri, Andre I., and John A. Cornell, Response Surfaces: Designs and Analyses, Marcel Dekker, Inc., New York, 1987.

(16)Lauren, Michael K., Roger T. Stephen, and Mark A. Anderson, MANA: Map Aware Non-uniform Automata Version 2.0 Users Manual, New Zealand Defence Technology Agency, New Zealand, July 2002.

(17)Li, Z., R. V. Grandhi, and R. Shivpuri, "Optimum Design of the Heat-Transfer Coefficient during Gas Quenching using the Response Surface Method," International Journal of Machine Tools and Manufacture, Vol. 42, pages 549 to 558, 2002.

(18)Lucas, Thomas W., Susan M. Sanchez, Lloyd Brown, and William Vinyard, "Better Designs for High-Dimensional Explorations of Distillations," Maneuver Warfare Science, 2002.

(19)Montgomery, Douglas C., and Vernon M. Bettencourt, Jr., "Multiple Response Surface Methods in Computer Simulation", Simulation, pages 113 – 121, October 1977.

(20)Myers, Raymond H., and Douglas C. Montgomery, Response Surface Methdology: Process and Product Optimization Using Designed Experiments, 2nd Edition, John Wiley and Sons, Inc., New York, 2002.

(21)Rees, Loren P., Edward R. Clayton, and Bernard W. Taylor, III, "Solving Multiple Response Simulation Models Using Modified Response Surface Methodology within a Lexicographic Goal Programming Framework," IIE Transactions, March 1995.

(22)Schamburg, Jeffrey B., Deployment Planning and Analysis for Time Difference of Arrival and Differential Doppler Location Finding Assets, Thesis, University of Virginia, Charlottesville, Virginia, 1995.

(23)Scherer, William T., and Christopher W. Zobel, "Simulation-based Policy Generation for Large-scale Markov Decision Processes," IEEE Transactions on Systems, Man, and Cybernetics – Part A: Systems and Humans, Vol. 31, No. 6, page 609, Nov 2001.

(24)Veral, Emre A., "Computer Simulation of Due-date Setting in Multi-Machine Job Shops," Computers and Industrial Engineering, Vol. 42, pages 77 to 94, 2001.

(25)Yang, T. and L. Tseng, "Solving a Multi-Objective Simulation Model using a Hybrid Response Sruface Method and Lexicographical Goal Programming Approach; A Case Study on Integrated Circuit Ink-Marking Machines," Journal of the Operational Research Society, Vol. 53, No. 2, pages 211 to 222, 2002.