NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

EXPERIMENTS IN ERROR PROPAGATION WITHIN HIERARCHAL COMBAT MODELS

by

Russell G. Pav

September 2015

Thesis Advisor: Thomas W. Lucas
Second Reader: Jeffrey E. Kline

Approved for public release; distribution is unlimited

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE

Form Approved OMB No. 0704–0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instruction, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

1. AGENCY USE ONLY ***(Leave blank)***	**2. REPORT DATE** September 2015	**3. REPORT TYPE AND DATES COVERED** Master's Thesis

4. TITLE AND SUBTITLE
EXPERIMENTS IN ERROR PROPAGATION WITHIN HIERARCHAL COMBAT MODELS

5. FUNDING NUMBERS

6. AUTHOR(S) Pav, Russell G.

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School
Monterey, CA 93943-5000

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING /MONITORING AGENCY NAME(S) AND ADDRESS(ES)
N/A

10. SPONSORING/MONITORING AGENCY REPORT NUMBER

11. SUPPLEMENTARY NOTES The views expressed in this thesis are those of the author and do not reflect the official policy or position of the Department of Defense or the U.S. Government. IRB Protocol number ____N/A____.

12a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited

12b. DISTRIBUTION CODE

13. ABSTRACT (maximum 200 words)

The Office of the Chief of Naval Operations (OPNAV) uses a hierarchy of simulation models as part of scenario-based planning to help decide which new platforms to procure and how to employ them. Simulation is used at every level of the acquisition process, from platform design to tactics to force structure. In hierarchal combat modeling, the mean output of lower-level, higher-resolution models are used as inputs to higher-level, lower-resolution models. The goal of this process is to inform military commanders how design changes in new platforms will affect tactical performance, and how changes in tactical performance can enhance campaign effectiveness.

This thesis uses a hierarchal modeling structure to examine whether including the distributions of mission model inputs instead of just the mean can affect campaign model results. A mission model of a one-on-one submarine battle is developed to determine the mean time to kill (MTTK) for the belligerents. The MTTK is sampled in a variety of ways, including just the mean, and used to calculate the attrition coefficients for a stochastic Lanchester campaign model that contains 18 Blue and 25 Red submarines. The outputs of the campaign models are analyzed statistically. The results indicate that the sampling methodology has a significant impact on the mean probability Blue wins the campaign and the mean amount of losses Blue takes when it wins. In addition, sampling methodology has a significant effect on the standard deviation for the probability Blue wins and the amount of losses Blue expects to take when it wins. These results also have practical significance: estimates of Blue's average odds of winning range from 0.58 to 0.94, while estimates of average losses range from 4.69 to 8.31. Hierarchal combat models must adopt methods for including the entire distribution of lower-level model outcomes in order to better represent risk.

14. SUBJECT TERMS Campaign analysis, hierarchal combat models, simulation, error propagation, anti-submarine warfare.			**15. NUMBER OF PAGES** 83
			16. PRICE CODE
17. SECURITY CLASSIFICATION OF REPORT Unclassified	**18. SECURITY CLASSIFICATION OF THIS PAGE** Unclassified	**19. SECURITY CLASSIFICATION OF ABSTRACT** Unclassified	**20. LIMITATION OF ABSTRACT** UU

NSN 7540–01-280-5500 Standard Form 298 (Rev. 2–89)
Prescribed by ANSI Std. 239–18

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release; distribution is unlimited

EXPERIMENTS IN ERROR PROPAGATION WITHIN HIERARCHAL COMBAT MODELS

Russell G. Pav
Lieutenant, United States Navy
B.S., State University of New York College at Old Westbury, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

NAVAL POSTGRADUATE SCHOOL
September 2015

Author: Russell G. Pav

Approved by: Thomas W. Lucas
Thesis Advisor

Jeffrey E. Kline
Second Reader

Patricia A. Jacobs
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

The Office of the Chief of Naval Operations (OPNAV) uses a hierarchy of simulation models as part of scenario-based planning to help decide which new platforms to procure and how to employ them. Simulation is used at every level of the acquisition process, from platform design to tactics to force structure. In hierarchal combat modeling, the mean output of lower-level, higher-resolution models are used as inputs to higher-level, lower-resolution models. The goal of this process is to inform military commanders how design changes in new platforms will affect tactical performance, and how changes in tactical performance will enhance campaign effectiveness.

This thesis uses a hierarchal modeling structure to examine whether including the distributions of mission model inputs instead of just the mean can affect campaign model results. A mission model of a one-on-one submarine battle is developed to determine the mean time to kill (MTTK) for the belligerents. The MTTK is sampled in a variety of ways, including just the mean, and used to calculate the attrition coefficients for a stochastic Lanchester campaign model that contains 18 Blue and 25 Red submarines. The outputs of the campaign models are analyzed statistically. The results indicate that the sampling methodology has a significant impact on the mean probability Blue wins the campaign and the mean number of losses Blue takes when it wins. In addition, sampling methodology has a significant effect on the standard deviation for the probability Blue wins and the amount of losses Blue expects to take when it wins. These results also have practical significance: estimates of Blue's average odds of winning range from 0.58 to 0.94, while estimates of average losses range from 4.69 to 8.31. Hierarchal combat models must adopt methods for including the entire distribution of lower-level model outcomes in order to better represent risk.

THIS PAGE INTENTIONALLY LEFT BLANK

THESIS DISCLAIMER

The reader is cautioned that the computer programs presented in this research may not have been exercised for all cases of interest. While every effort has been made, within the time available, to ensure that the programs are free of computational and logical errors, they cannot be considered validated. Any application of these programs without additional verification is at the risk of the user.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

LIST OF TABLES

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

ASW	Anti-Submarine Warfare
ASUW	Anti-Surface Warfare
ANOVA	Analysis of Variance
BIC	Bayesian Information Criterion
CNO	Chief of Naval Operations
DOE	Design of Experiments
DOD	Department of Defense
MANA	Map Aware Non-uniform Automata
MOE	Measure of Effectiveness
MOP	Measure of Performance
MTTK	Mean Time to Kill
MTTD	Mean Time to Detect
NOLH	Nearly Orthogonal Latin Hypercube
N81	Office of the Chief of Naval Operations Plans and Assessments Department
N97	Office of the Chief of Naval Operations Undersea Warfare Department
N98	Office of the Chief of Naval Operations Navy Air Warfare Department
OPNAV	Office of the Chief of Naval Operations
OR	Operations Research
R5FF	Resolution 5 Fractional Factorial
SSN	Nuclear Fast Attack Submarine
SSK	Diesel Attack Submarine

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The Office of the Chief of Naval Operations (OPNAV) uses a hierarchy of simulation models as part of scenario-based planning to help decide which new platforms to procure and how to employ them. Simulation is used at every level of the acquisition process, from platform design to tactics to force structure. In hierarchal combat modeling, the mean output of lower-level, higher-resolution models are used as inputs to higher-level, lower-resolution models. The goal of this process is to inform military commanders how design changes in new platforms will affect tactical performance, and how changes in tactical performance can enhance campaign effectiveness. By using hierarchal simulation models, the Navy can gain insight into questions such as "how will better sensors affect the outcome of a blue water battle?" Senior leadership can then use this information to determine the best investments to achieve and sustain warfare dominance within a particular budget.

This work explores how error propagates through hierarchal simulation models at the mission and campaign levels to quantify the degree of inaccuracy between the two methods using a "ground up" approach. First, it develops a mission-level model for one on one submarine combat in Map Aware Non-uniform Automata (MANA) simulation, an agent-based simulation that can model the different behavioral postures of submarines. The measures of performance (MOP) for this model are based on open source operating characteristics of submarines. The measures of effectiveness are mean time to kill (MTTK) and the average probability each side wins. The result is excluded if no kill is made. Uncertainty in the MOPs is obtained with two different designs of experiments (DOEs), nearly orthogonal Latin hypercube (NOLH) and resolution V fractional factorial (R5FF), to determine how the mission model experimental design affects the MOE.

Next, the work constructs a stochastic Lanchester campaign model. The attrition coefficients are determined by multiplying the reciprocal of a randomly sampled MTTK value by the average probability of winning. Sampling is used to account for the variance in the distributions of MANA output metrics. Several types of sampling are explored: sampling one side's MTTK in isolation with the other's mean value, sampling both sides,

constructing a NOLH DOE using the min and max value, constructing a NOLH DOE using a range that excludes outliers, and using both sides' means. In order to isolate the effects of the sampling method, the Blue and Red units are held constant at 18 and 25, respectively.

The analysis finds that there is a statistically significant difference in average Blue MTTK for the NOLH (μ = 15.98, σ = 7.53) and R5FF (μ = 18.33, σ = 12.04) data sets. In addition, there is a significant difference between the variance of Blue MTTK for the NOLH and R5FF designs, according to the Levene test. The effect of the experimental design used on Red MTTK and winning percentage is not statistically significant. Therefore, the analysis continues distinguishing both data sets. It uses "FF" to denote campaign simulations that sampled the R5FF data set, and "NOLH" to denote campaign simulations that sampled the NOLH data set.

The analysis fits a one-way analysis of variance (ANOVA) model to determine the effect of sampling methodology on campaign MOEs. The results indicate that the sampling methodology has a significant correlation with the probability Blue wins the campaign and the amount of losses Blue takes when it wins. In addition, sampling methodology has a significant effect on the standard deviation for the probability Blue wins and the amount of losses Blue expects to take when it wins.

These results also have practical significance in assessing the risk to an operational commander. The graphs in Figure 1 and Figure 2 illustrate this significance graphically. In Figure 1, the estimated odds that Blue wins the campaign vary dramatically based upon the sampling method. This is because the different sampling methods produce input variables with different means and variances, leading to different campaign simulation outputs. Similarly, Figure 2 displays the chances that Blue loses a certain amount of submarines given it wins the battle. Again, the risk changes significantly based upon the method used to construct the hierarchal simulation.

This study demonstrates that the effect of accounting for the distribution in random input variables has a significant impact on campaign model results. Further research should be conducted to determine which method provides the best estimate of the output MOE.

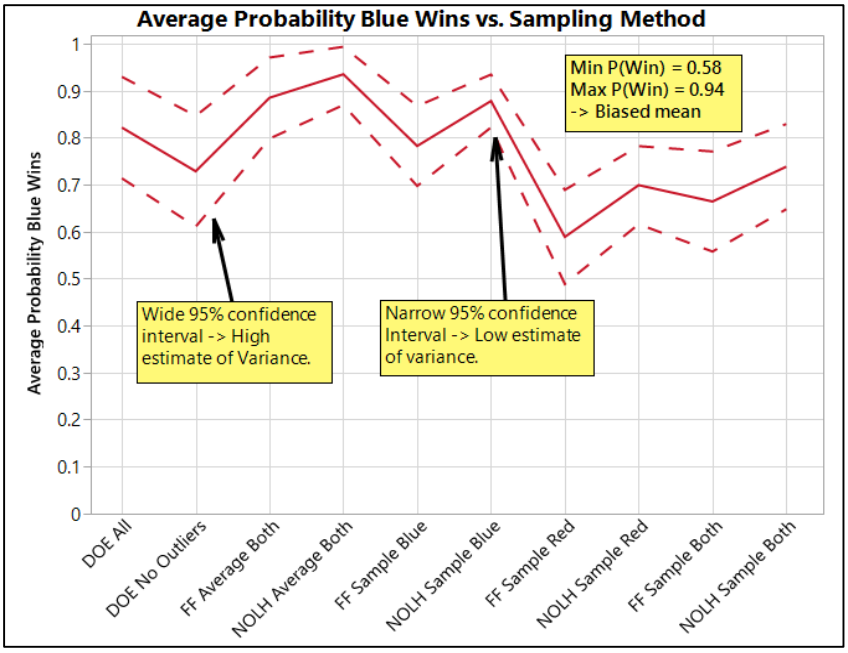


Figure 1. Graph of the average probability Blue wins versus sampling methodology.

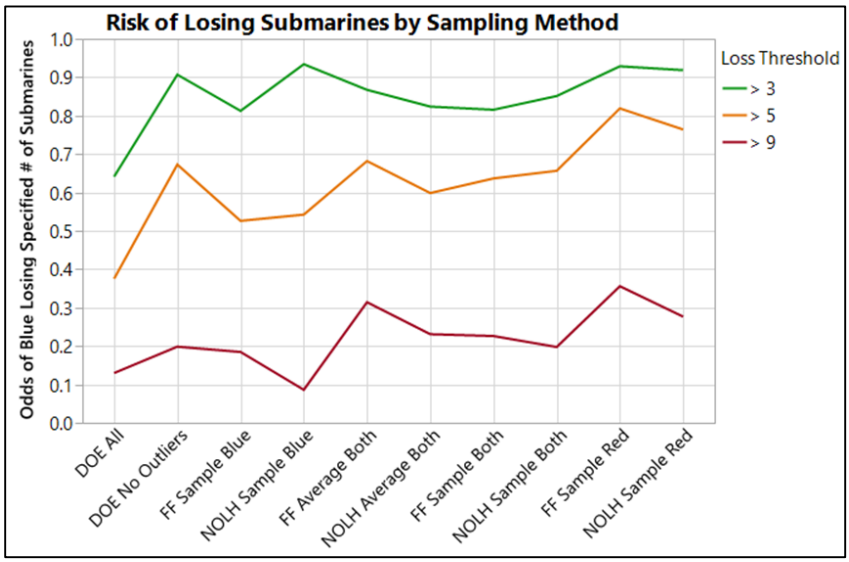


Figure 2. Graph of estimated probability for losing > 3, > 5, and > 9 submarines in this campaign.

THIS PAGE INTENTIONALLY LEFT BLANK

ACKNOWLEDGMENTS

First, I would like to thank my thesis advisors, Dr. Thomas W. Lucas and Professor Jeffrey Kline, for their mentorship and guidance while completing this thesis. In addition, I would like to thank Mary McDonald for her assistance in creating the mission model and conducting simulation experiments.

Next, I would like to thank the Naval Postgraduate School Department of Operations Research staff members for their expertise and patience. The professors at this institution have provided some of the best academic instruction I have ever received and made my education at Naval Postgraduate School extremely rewarding.

Finally, I would like to thank my wonderful wife, Debbie. Her support, encouragement, and understanding were key to my success at NPS and the completion of my thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

A. BACKGROUND: WHY SIMULATION?

The Office of the Chief of Naval Operations (OPNAV) uses a hierarchy of simulation models as part of scenario based planning to help decide which new platforms to procure and how to employ them. As shown in Figure 1, which displays a pyramid representation of the hierarchal combat modeling process, simulation is used at every level of the acquisition process, from platform design to tactics to force structure. The outputs of models on the lower levels of the pyramid are used as inputs to models on the next level.

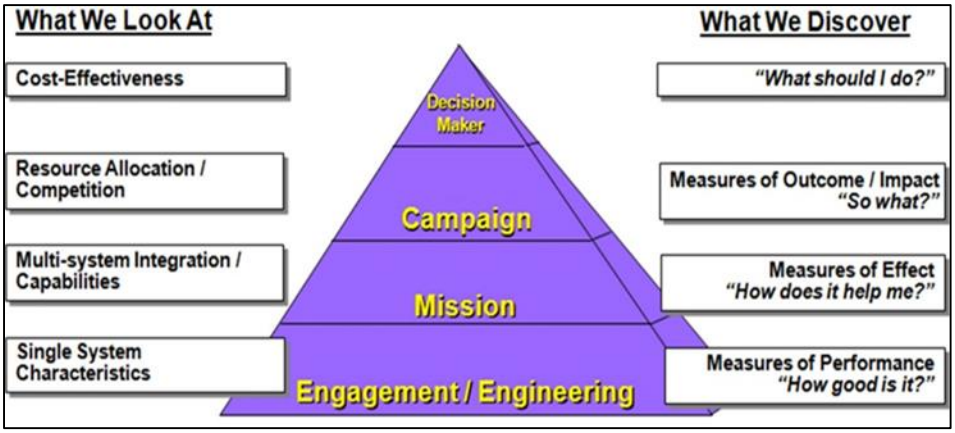


Figure 1. Hierarchal combat model process (from Cappellini 2011).

The use of simulation at each level is useful for several reasons. First, engineering simulations are useful because testing of military equipment, particularly destructive testing, is costly. The unit cost estimate of a single F/A-18 strike fighter is $57 million (U.S. Navy 2009), while the unit cost of a single Mk 48 Heavyweight Torpedo is between $2-3 million (*Defense Industry Daily* Staff 2014). These high costs prevent the Navy from conducting large-scale destructive tests and live fire exercises to determine their effectiveness. Simulation helps ensure that the design is sound before proceeding to building prototypes and conducting the limited number of feasible live tests.

Second, the Navy cannot conduct live-fire exercises to test tactics because doing so would not only result in the loss of costly equipment, but also potentially sacrifice lives. Consequently, mission-level combat simulations are useful because they allow the Navy to evaluate specific platform effectiveness and assess tactical doctrine. The Navy uses the results of these simulations to develop tactical publications and to help train its operators to employ ships, aircraft, and submarines in ways that give them the best chance of success.

Finally, campaign models help shape the force structure of the Navy as a whole. The acquisition of major combatant platforms occurs over decades. For example, the Navy initially drafted plans for the Virginia Class submarine in 1991. General Dynamics delivered the first ship of its class, *USS Virginia* (SSN 774), to the Navy 13 years later in October 2004. The Navy plans to procure Virginia class submarines until 2043, and plans to operate them until 2060 (Osborn 2014). A submarine designed in 1991 must be capable to combat a threat in 2060, almost 70 years later. As a result, the Navy must anticipate the nature of future conflicts and develop flexible platforms that can adapt to emerging threats. Campaign models allow the Navy to analyze the outcome of potential future conflicts given a particular force structure. They even allow the Navy to analyze the effects of capabilities and weapons platforms that are in early development for both the U.S. and other nations. The outcomes of these models identify capability gaps and help develop the focus of future platform acquisitions.

The goal of this process is to inform military commanders how design changes in new platforms may affect tactical performance, and how changes in tactical performance can enhance campaign effectiveness. By using hierarchal simulation models, the Navy can gain insight into questions such as “how will better sensors affect the outcome of a blue water battle?” Senior leadership can then use this information to determine the best investments to achieve and sustain warfare dominance within a particular budget.

B. LITERATURE REVIEW

1. Intelligent Experimental Design

The use of combat simulations within the Department of Defense (DOD) presents additional challenges. These simulations are complex, computationally intensive programs that take a long time to run. In addition, the simulations analyze dozens of variables that each has multiple levels. In general, to fully explore an experiment with m levels and k variables requires m^k runs multiplied by the number of replications needed (Sanchez et al. 2012). This is called a full factorial design. To illustrate how simulations can quickly grow to become infeasible, one replication of an experiment that examines the interaction of all combinations of 30 variables at two levels, such as low and high, requires over 10^9 runs. If each run took just one second, then a single experiment would take over 34 years. To do the replications required to obtain output that can be analyzed with statistical techniques would take multiple lifetimes (Sanchez et al. 20120). Additionally, this design is not capable of analyzing for non-linearity in the effects because it only samples the end-points. In the world of military simulation, an experiment with 30 factors is relatively small, and the ability to run them in one second is usually not possible.

There are several techniques available to reduce the amount of simulation runs required to obtain useful analysis. This thesis selects two techniques for comparison: resolution V fractional factorial (R5FF) and nearly orthogonal Latin Hypercubes (NOLH). The fractional factorial design reduces the amount of runs needed by some factor of two by assuming higher level interactions are negligible (Sanchez and Sanchez 2005). The NOLH design reduces the amount of runs by calculating an uncorrelated matrix of design points for all the factors under consideration (Cioppa and Lucas 2007). After the experiment is completed in either design, an analyst can fit a regression meta-model to the results to determine which factors are statistically significant and their relationship to the output (Cioppa and Lucas 2007). By employing thoughtful experimental design and fitting meta-models to the output, analysts can reduce the amount of simulation runs required to obtain statistically useful output while analyzing a broader range of factors.

2. Simulation Optimization

As mentioned previously, a primary goal of the military simulation process is to help senior military and political leadership decide where to allocate resources for future force structure. Once the relationship between input factors and the output measures of effectiveness (MOE) is understood, the next step is to determine how to create the best possible military force structure. This concept is called simulation optimization, and provides a more efficient means of determining the optimal factor settings than a brute-force approach of iteratively running a simulation until the analyst stumbles upon the answer. One method of achieving this goal is to develop a linear optimization algorithm for the meta-model to determine the optimal factor settings (Osorio and Chong 2015). These programs can then be solved using commercial linear program solvers.

Another method for optimizing meta-models uses the statistical software JMP, developed by SAS. JMP has a "prediction profiler" in its regression tools that allows analysts to vary inputs via slider bars and observe the change in output. This offers a simpler approach that is more user-friendly to analysts because it does not require one to develop and program an optimization algorithm. Therefore, this is the method utilized by OPNAV N98, Air Warfare and The Maritime Dominance Branch of Mission Engineering Analysis at Patuxent River (Pax River) when modeling future platforms.

3. Hierarchal Meta-Modeling

Analysts at N98 and Pax River take the meta-modeling optimization process one step further by using hierarchal meta-modeling. This develops a series of regression equations that models the hierarchal combat simulation pyramid displayed in Figure 1. Each level of the pyramid has its own meta-model developed based upon the results of their respective simulation outputs. The hierarchal meta-model is a recursive equation where the regression equation for each lower-level model serves as the independent variables for the higher level models. This allows analysts to estimate the effect of engineering changes to platforms, such as P-8 radar range, on campaign anti-submarine warfare (ASW) effectiveness using JMP's prediction profiler. This hierarchal meta-

modeling process attempts to eliminate the need to re-run a chain of complex simulations, which takes up to a year to accomplish.

C. RESEARCH PROBLEM: MEASURING UNCERTAINTY

Unfortunately, each time an analyst performs a linear regression there is uncertainty in the output. The hierarchal meta-modeling process has no known way to propagate uncertainty; each factor in the campaign meta-model is based upon a point estimate for the mean in the lower-level meta-models. As a result, the variance represented in the campaign hierarchal meta-model does not accurately represent the variance in the actual output from the campaign simulation (Lucas 2000), (Davis, Exploratory Analysis Enabled by Multiresolution, Multiperspective Modeling 2001), (Cappellini 2011). In addition, it is possible that this method introduces bias into estimating the mean for measures of effectiveness from a campaign level model. Thus, the analyst employing this technique cannot accurately quantify the risk.

D. SCOPE OF THESIS

This work explores how error propagates through hierarchal simulation models at the mission and campaign levels to quantify the degree of inaccuracy between the two methods. It develops a mission-level model for one on one submarine combat in Map Aware Non-uniform Automata (MANA), an agent based simulation that can model the different postures of submarines. It feeds the results from MANA into stochastic Lanchester campaign models by using different sampling techniques. Next, the resulting campaign measures of effectiveness (MOEs), average winning percentage, W, and average Blue attrition when Blue wins, will be compared using statistical analysis. Finally, the study will determine whether the differences have real-world significance in determining the optimal force mix and present a risk assessment for commanders.

THIS PAGE INTENTIONALLY LEFT BLANK

II. MODELING TOOLS

A. MANA

Map Aware Non-uniform Automata (MANA) was developed by the New Zealand Defense Technology Agency (DTA) (McIntosh 2009). It is an agent-based simulation software package that employs a time-stepped, stochastic, mission-level modeling environment. It creates a modeling environment that facilitates an abstraction of a scenario that captures the essence of the physical and behavioral aspects, but avoids unnecessary details. Its intended use of providing a quick-turn ability to explore a wide range of possible outcomes is ideal for conducting the extensive statistical analysis that is explored in this thesis.

There are several features in MANA that make it an ideal choice of mission-level model for this analysis. As an agent-based simulation, it employs entities of any size that share common physical and behavioral characteristics. The physical characteristics include, but are not limited to, sensor capabilities, weapon effectiveness, speed, and fuel capacity. This feature allows the user to quickly and easily create units that have multiple sensors and weapons using engineering design specifications. The behavioral characteristics allow the user to define the rules for how a unit behaves and interacts with other units, such as search patterns, rules of engagement, and target prioritization. Since the agents in this simulation are submarines that have multiple sensors, weapons, and distinct search patterns, the use of an agent-based simulation like MANA that can incorporate all of these attributes allows for modeling of the mission scenario. Figure 2 and Figure 3 display screenshots of the sensor and weapons input, respectively.

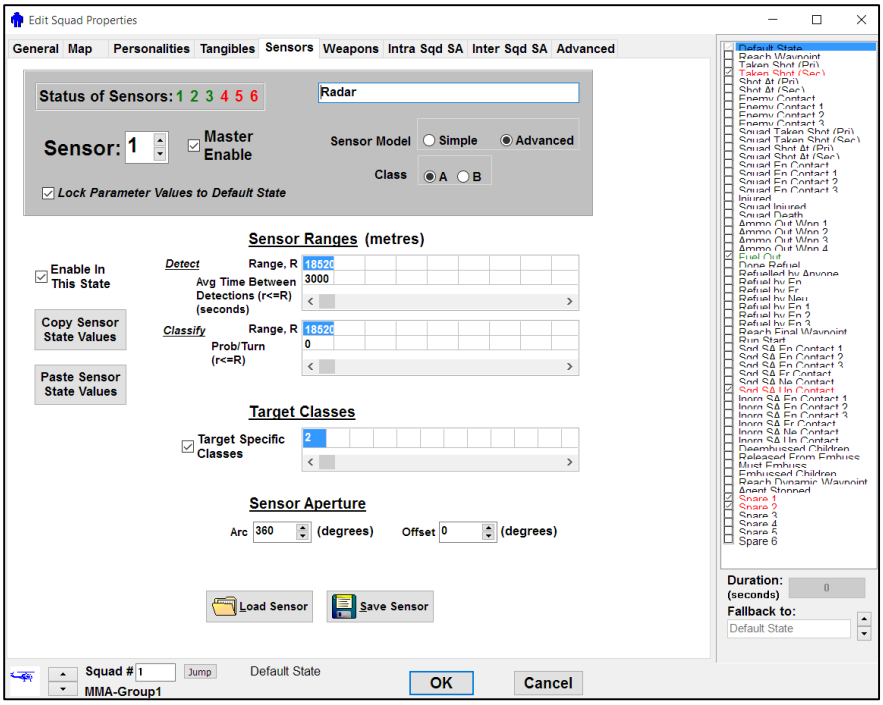


Figure 2. MANA sensor characteristics screenshot.

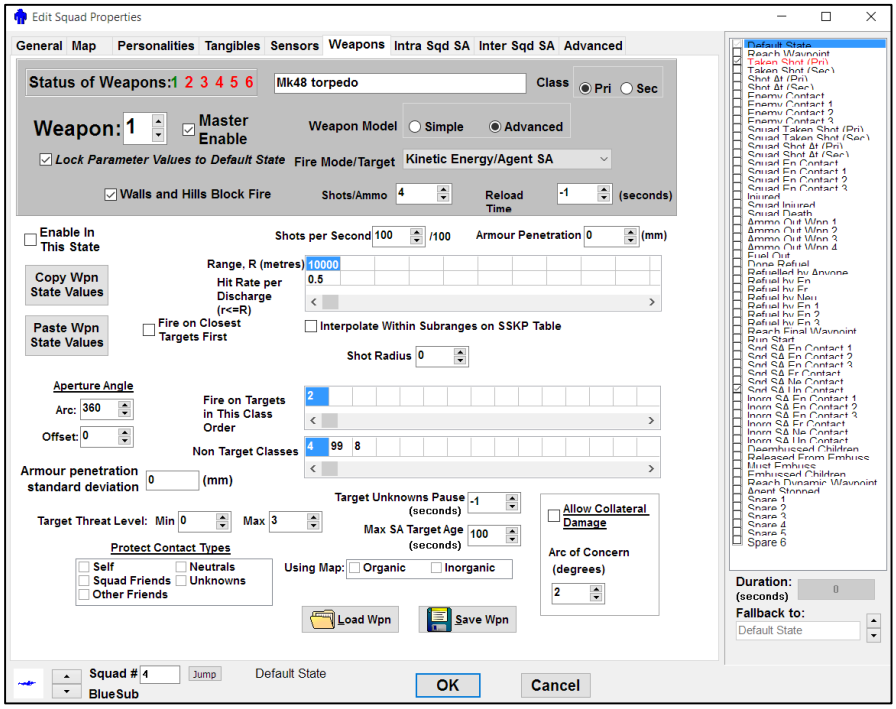


Figure 3. MANA weapons characteristics screenshot.

However, MANA does have some limitations that are applicable to this analysis. First, the agents in MANA cannot exercise fire control over wire-guided munitions. Guided weapons such as submarine launched heavyweight torpedoes must be modeled to either hit on a straight path or miss. In addition, MANA cannot force an agent to shoot a torpedo down a bearing, which it would do when conducting a 'snapshot' against an incoming torpedo from an unknown source. The work-around for this is to create a dummy agent that travels with the weapon, which the target submarine can use to fire upon. Finally, there is no specific shooter-to-target probability of kill. The effectiveness is a function only of the weapon employed. In the context of ASW, the unit-on-unit effectiveness can be adjusted by modifying the sensor probability of detections.

B. CAMPAIGN MODELS

The analysis divides the campaign into discrete sub-campaigns for modeling. Researchers for the Rand Corporation employed this methodology in a naval campaign analysis in the Mediterranean Sea (Kelley 1974). The underlying assumption is that because military ships, aircraft, and vehicles are built to do a specific set of missions, units will primarily engage a particular type of enemy force. For example, Blue SSNs tasked with anti-submarine warfare (ASW) clearing will only engage Red submarines and will not participate in anti-surface warfare (ASUW) combat. However, the underlying MOE for the overall campaign is the force exchange ratio. This is what enables analysts to divide the campaign to be analyzed into distinct phases with the appropriate units and then aggregate the results. Using this technique, this thesis focuses on the error propagation within the Blue submarine ASW operation sub-campaign.

1. Python 2.7 and JMP version 12

The stochastic Lanchester campaign models are coded using Python 2.7, an open source language available at http://www.python.org. Python is selected because it is an open-source coding language with several free analytical packages with extensive user documentation. In addition, the Python coding language is very readable, so researchers interested in follow-on work can more easily implement the scripts developed for this research.

The statistical analysis is conducted using JMP version 12. JMP is a windows based statistical software package that allows analysts to visualize data manipulation and to employ advanced analytical techniques without the need for programming. It is selected for this thesis primarily because DOD analysts and engineers employ the software suite to perform hierarchal meta-modeling with regression techniques. Its most useful feature for this task is the dynamic prediction profiler, shown in Figure 4. In Figure 4, each factor is constructed from the residuals of a meta-model built from results of its respective simulation. The residuals are input into the regression equation for the campaign MOE. This feature allows analysts to alter factors and see the effect on the output real-time. In addition, it links the effect of changing engineering specifications to campaign effectiveness even though engineering specifications cannot be input directly into campaign simulations. In this way, the prediction profiler provides a coarse but fast optimization technique that can be employed without the need for commercial solvers.

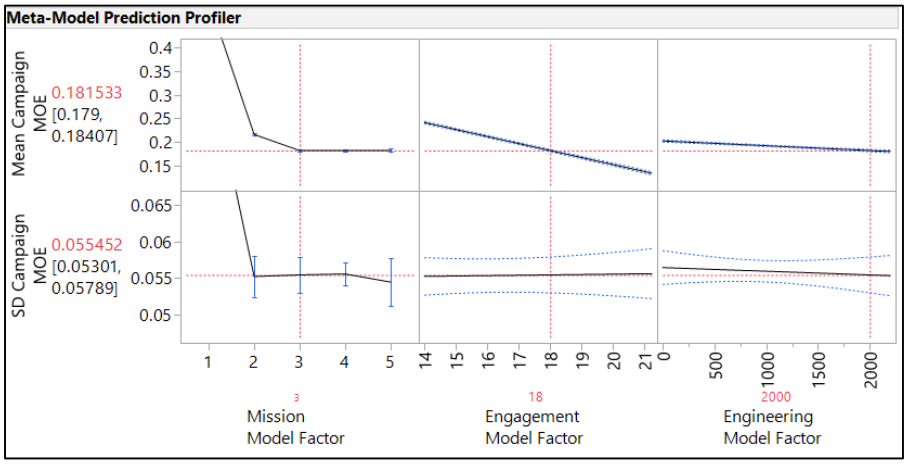


Figure 4. Prediction Profiler Example

2. Lanchester Models

Frederick Lanchester developed a series of differential equations to mathematically model combat over time. Each force's attrition is dependent upon two factors; the amount of enemy forces and the enemy's effectiveness in battle. He developed two general models: the linear case, which modern campaign analysts use to model forces employing area fire, and square case, which modern campaign analysts use

to describe forces employing aimed fire (Lanchester 1916). The campaign simulations in this research utilize a stochastic extension of the Lanchester Square Law. Future research can explore the scenario using the Lanchester Linear Law.

In the stochastic version of Lanchester's Squared Law, the time to next casualty is drawn from an exponential distribution with rate λ. In the case of aimed fire, the rate is proportional to the amount of units that Blue (denoted by *x*) or Red (denoted by *y*) has and the attrition coefficients for Blue (a) and Red (b). Therefore:

$$\begin{aligned} \lambda_{Blue} &= ay \\ \lambda_{Red} &= bx \end{aligned} \tag{2.1}$$

which gives the following expression for the determining the expected time to next casualty, $\mathrm{E}[\mathrm{X} \mid x, y]$ and the probability Blue suffers the casualty, $\mathrm{P}[\mathrm{X} \mid x, y]$:

$$\mathrm{E}[\mathrm{T} \mid x, y] = \frac{1}{\lambda_{Blue} + \lambda_{Red}} \tag{2.2}$$

$$\mathrm{P}[\mathrm{X} \mid x, y] = \frac{\lambda_{Blue}}{\lambda_{Blue} + \lambda_{Red}} \tag{2.3}$$

The time to next casualty is a random exponential variable with rate equal to $1 / \mathrm{E}[\mathrm{X} \mid x, y]$. Equation 2.3 is compared to a random uniform [0, 1] variable to determine which side suffers the casualty. If $\mathrm{P}[\mathrm{X} \mid x, y]$ is less than or equal to the random uniform variable, then Blue suffers a casualty; otherwise, Red suffers a casualty. These equations are employed in an event based simulation until all units are destroyed or until a time limit is reached.

THIS PAGE IS INTENTIONALLY LEFT BLANK

III. SCENARIO AND MODEL DEVELOPMENT

A. SCENARIO

This analysis considers a potential future naval conflict between an enemy nation, herein referred to as Red, and the United States, herein referred to as Blue. The Red nation has established a submarine blockade, either off a coast or around an island, using a fleet of 25 diesel submarines (SSKs). Red sinks any merchants or warships that enter the area, and Blue seeks to clear the blockade using its nuclear fast attack submarines (SSNs).

This scenario is useful because it provides a background to conduct analysis to answer the following questions:

(1) What is the best way to conduct ASW clearing operations?

(2) Do different modeling approaches produce different answers?

This thesis focuses on question (2) by conducting a thorough analysis of error propagation between the campaign and mission level model.

B. GENERAL MODELING APPROACH

This scenario has many parameters that are unknown to the analyst. These parameters, or factors, fall under three broad categories: those that are within control of Red, those that are within the control of Blue, and those that are beyond the control of either force. The analysis ultimately seeks to determine the most prominent factors that affect the outcome of the campaign within Blue and Red's control. In addition, the analysis seeks to determine whether the method of choosing the parameters in the hierarchal simulation changes the measures of effectiveness (MOEs). Armed with that insight, operational commanders can construct a force structure and concept of operations that can maximize the odds of success in such a scenario.

Because it is infeasible to conduct live experiments for this scenario, the analysis employs stochastic simulations of the ASW clearing campaign to clear the SSK blockade. Stochastic simulations are selected because they can account for the uncertainties

associated with military operations. The analysis starts with a mission-level model using known unit performance characteristics as the measures of performance (MOP) to estimate Blue and Red submarine effectiveness in one-on-one combat by obtaining a mean time to kill (MTTK) and probability of kill (P_K) values as the MOEs. These mission level MOEs are then sampled and fed into a campaign-level model, the stochastic Lanchester simulation, as MOPs with multiple Red and Blue submarines to obtain distributions for the probability Blue wins (P_W) and Blue attrition when blue wins over a two week campaign as campaign MOEs. This relationship is summarized in Table 1.

Table 1. Summary of hierarchal model construction. The MOEs from the mission model are sampled using various methods and input as MOPs into the campaign model.

Level of Model	**Model Tool**	**MOPs**	**MOEs**
Mission	MANA	Submarine operating characteristics constructed from a R5FF and NOLH DOE (see Table 2 and 3).	1. P_K 2. MTTK
Campaign	Stochastic Lanchester Square Law	1. Units (held constant) 2. mean P_K 3. Sampled MTTK	1. P_W 2. E[Blue Losses| Win]

The first campaign MOE provides the odds of success, while the second campaign MOE quantifies the risk of the campaign. These MOEs can be compared to other ASW clearing options, such as using maritime patrol aircraft (MPA), in future analysis. This work focuses on an in-depth analysis using Blue submarines. The average blue attrition on the whole is not considered because when blue loses, we know that it loses all submarines committed to theater. This occurs because the campaign model assumes that the campaign continues until one side is annihilated. The campaign output is compared by mission model MOE sampling method and mission model MOP design of experiments to determine if the methods of constructing a hierarchal combat simulation are statistically or practically different.

C. ASW CLEARING OPERATIONS

1. Mission Model

The mission model is developed in Map Aware Non-uniform Automata (MANA). In the mission model, a one versus one battle between a red and blue submarine is programmed. In this model, a Blue submarine approaches a 60 nm x 60 nm datum to search for a single Red submarine. A wartime scenario is assumed, so any positive detection by one submarine will result in the unit firing on the detected unit. The scenario runs for two weeks of simulated time and records the winning unit and the time to kill as the MOEs. If no kill occurs then that data is excluded from the analysis.

There are several unknown performance characteristics in the mission-level model. To handle the uncertainty, the study employs the design of experiment (DOE) techniques discussed in section I.B.1. Two techniques are employed to compare whether they eventually lead to the same campaign model results: a Resolution 5 Fractional Factorial (R5FF) and Nearly Orthogonal Latin Hypercube (NOLH) design. The R5FF allows the exploration of factors at their end-points and second-order interactions in output analysis (Sanchez and Sanchez 2005). The NOLH design provides a space-filling design to examine the entire range of each factor in output analysis (Sanchez, Lucas, et al. 2012). The NOLH design is more efficient in the number of required samples, so the purpose of including both is to examine whether NOLH designs ultimately sacrifice accuracy for efficiency.

The performance characteristics varied for the Blue and Red submarine are summarized in Table 2 and Table 3. These characteristics are derived from open source information and varied to account for environmental and performance uncertainty beyond each force's control. The NOLH experiment had 257 design points, each repeated for 40 replications, for a total of 10,280 runs. The R5FF experiment had 512 design points, each repeated for 40 replications, for a total of 20,480 runs. Given an average run time per replication of 5 minutes and 128 processors, the R5FF experiment took approximately 27 hours and the NOLH experiment took approximately 13 hours. The runs would have taken over six months on a single processor without the computing cluster.

Table 2. Blue submarine design factors used in the DOE for the MANA model.

Factor	Description	Unit	Low	High
Experimental Design	Used to track whether the simulation was constructed using a NOLH or R5FF DOE.	-	-	-
Blue.Sub.Patrol.Spd	The speed that a Blue submarine moves when searching for Red submarines. Variation accounts for environmental and ship acoustic characteristics.	Knots	8	12
Blue.Sub.Evade.Spd	Average speed that a Blue submarine moves while evading a torpedo. Variation accounts for reaction time.	Knots	22	28
Blue.Sub.Mk48.Pkill	Probability that a torpedo fired from a Blue submarine hits a Red submarine. Variation accounts for firing solution precision, enemy counter-measures, and environmental acoustic characteristics.	-	0.5	0.7
Blue.Sub.TA.Rng	Maximum range that a Blue submarine can detect a Red submarine using the towed array. Variation accounts for varying acoustic characteristics in the environment and Red submarines.	Yards	15,000	25,000
Blue.Sub.TA.MTTD	Average time it takes a Blue submarine to recognize a detection once one occurs. Accounts for environmental acoustic conditions and operator training.	Hours	0.17	0.33
Blue.Sub.Active.Sonar.Rng	Maximum range that a Blue submarine can detect a Red submarine using the active sonar. Variation accounts for varying acoustic characteristics in the environment and Red submarines.	Yards	4,000	6,000
Blue.Sub.Active.Ploc	The probability a Blue submarine detects a Red submarine on active sonar given the Red Submarine is within maximum range. Variation accounts for environmental acoustic conditions and operator training.	-	0.6	0.8

Table 3. Red submarine factors used in the DOE for the MANA model.

Factor	Description	Unit	Low	High
Red.SSK.Evade.Spd	Average speed that a Red submarine moves while evading a torpedo. Variation accounts for reaction time.	Knots	9	12
Red.SSK.Det.Range	Maximum range that a Red submarine can detect a Blue submarine using the towed array. Variation accounts for varying acoustic characteristics in the environment and Red submarines.	Yards	5,000	10,000
Red.SSK.MTTD	Average time it takes a Red submarine to recognize a detection once one occurs. Accounts for environmental acoustic conditions and operator training.	Hours	0.17	0.33
Red.SSK.Pkill	Probability that a torpedo fired from a Red submarine hits a Blue submarine. Variation accounts for firing solution precision, enemy counter-measures, and environmental acoustic characteristics.	-	0.3	0.6
Red.SSK.Avg.Time.Bet.Snorkel	Average time a Red submarine spends submerged before it must snorkel to recharge the battery. Varied to account for variations in electrical loading and to examine the effect of better battery technology.	Hours	24	96
Red.SSK.Time.in.Snorkel	Amount of time that a Red submarine spends snorkeling. Varied to account for different battery usage upon snorkeling.	Hours	1	3

2. Campaign Model

The campaign model uses the stochastic Lanchester simulation detailed in section II.B.2. Use of this model assumes that submarines employ aimed fire in a homogeneous battle. In reality, submarines would engage in a series of one versus one battles, so use of more detailed official DOD models that account for this behavior in future work could produce results that are more accurate.

In the campaign model, there are 18 Blue submarines and 25 Red submarines. This is based upon an estimate of material readiness for each force, with the U.S. having

approximately half of its submarines stationed on each coast and a pessimistic 66–70% readiness. The attrition coefficient is calculated by multiplying the average P_K from the MANA simulation by the reciprocal of the sampled MTTK, since a MTTK is only produced when that submarine gets a kill. The number of submarines is held constant because the analysis seeks to isolate the effect of sampling method for the MTTK on error propagation.

IV. ANALYSIS

This thesis is interested in determining whether the method in constructing a hierarchal model affects the campaign model measures of effectiveness (MOE). First, the MANA output is analyzed to determine which parameters most affect a submarine's mean time to kill (MTTK). In addition, the MANA output will determine P[Blue Wins] in a one-on-one battle. Next, several methods are employed to sample the MTTK distribution used to calculate the attrition coefficients in a stochastic Lanchester campaign to determine the distribution of Blue's probability of winning the campaign, denoted W to distinguish it from the mission MOE, and the amount of Blue attrition when Blue wins, denoted A. Finally, the results of the stochastic Lanchester campaign are analyzed to determine if there is any statistical or practical significance between the sampling methodology and campaign MOE. The analysis uses a typical convention that μ_X is the sample mean, σ_X is the sample standard deviation, E[X] is the expectation, and V[X] is the variance of the random variable X.

A. MANA SCENARIO

1. Comparing Design of Experiments

The MANA simulation of a one-on-one battle between a Red and Blue submarine produces two output data sets—one from the resolution V fractional factorial design (R5FF) and one from the nearly orthogonal Latin hypercube design (NOLH). The two DOEs specify different parameter values for the simulation. From herein, the data sets will be referred to by the design of experiments that produced them. Recall that the MOEs for the mission-level model are the P[Blue Wins], which occurs when a Red submarine is killed, and MTTK given a kill occurs for each force, so these output metrics are the subject of further analysis. First, the study uses JMP's summary table feature to produce the μ_{MTTK} by each MANA design point, which contain 40 replications. This summary is performed because the analysis is primarily interested in the mean and variance of MTTK across a variety of favorable and unfavorable conditions, and not the stochastic effects within a single design point that has a very particular set of conditions.

The distributions for both the NOLH and R5FF output data are displayed in Figure 5 and Figure 6. In both instances, $\mu_{P[Blue\ Wins]}$ is around .75, which means the model is reflecting the intended combat superiority of Blue submarines. The shape of these distributions and their summary statistics differ by experimental design. Surprisingly, the Blue μ_{MTTK} in the R5FF data set is higher than Red μ_{MTTK} despite the fact that μ_W is greater than 0.5, although this result is not statistically significant. One would expect Blue's μ_{MTTK} to be lower than Red's because in a one on one submarine battle, the first unit to detect the other has a decided advantage. Finally, note that the reason Red's N is not always equal to the number of design points is because there are some design points where Red never wins, which results in no MTTK value generated for that design point.

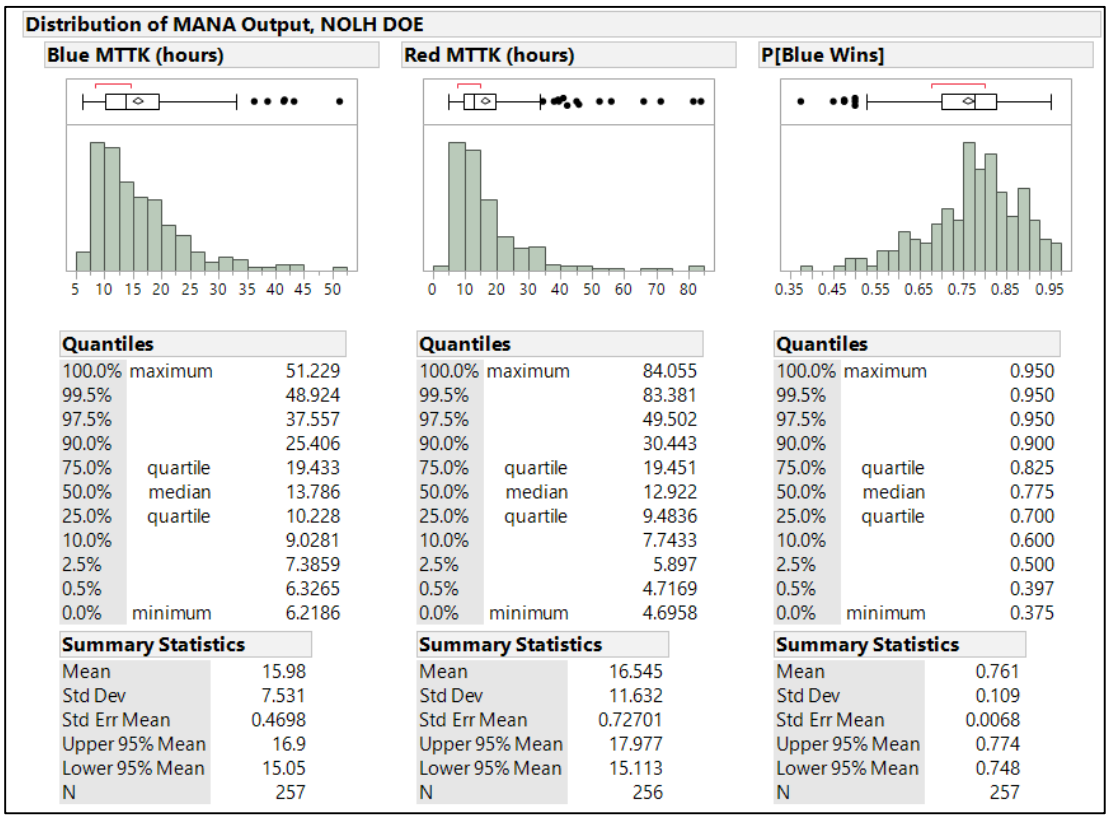


Figure 5. Distributions for Blue MTTK, Red MTTK, and P[Blue Wins] from the MANA experiment constructed with a NOLH DOE. Blue MTTK denotes the average amount of time a Blue submarine requires to kill a Red submarine, and Red MTTK denotes the average amount of time required for a Red submarine to kill a Blue submarine.

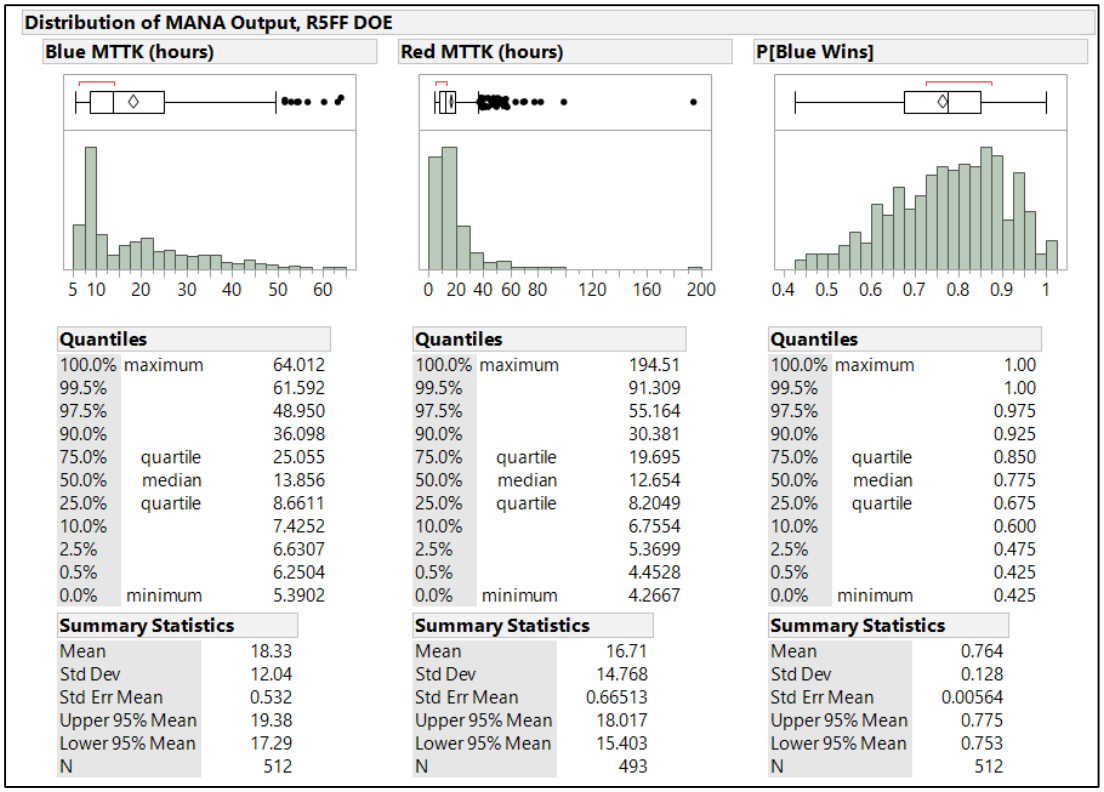


Figure 6. Distribution of Blue MTTK, Red MTTK, and P[Blue Wins] from the MANA experiment constructed with a R5FF DOE.

Next, the analysis uses a student's t-test and a Levene test to determine if the means and variances of Blue MTTK, Red MTTK, and P[Blue Wins] by Experimental Design are statistically different (Wackerly, Mendenhall III and Schaeffer 2008). Although the data is not normally distributed, the t-test is robust to non-normality (Wackerly, Mendenhall III and Schaeffer 2008). There is only a significant difference in average Blue MTTK for the NOLH (μ = 15.98, σ = 7.53) and R5FF (μ = 18.33, σ = 12.04) data sets; t_{767} = 1.96, p = .0021. In addition, there is a significant difference in variance of Blue MTTK for the NOLH and R5FF design according to the Levene test; F = 67.96, p = 0.0001. The effect of experimental design on Red MTTK and P[Blue Wins] is not statistically significant. At this stage of the analysis, the differences among the mean and standard deviations of MTTKs and P[Blue Wins] are not practically significant; under three hours is a negligible difference in time in a submarine campaign that takes weeks, and there is only a 0.003 difference in P[Blue Wins]. This confirms the

efficiency of the NOLH design does not produce different results in most cases; the same results are achieved with approximately half the runs.

However, because the difference in Blue μ_{MTTK} is statistically significant by experimental design, and because the shapes of the distributions vary, the analysis will sample from both data sets to determine the attrition coefficient for the Lanchester models and make a detailed comparison of each experiment. Even though these data sets are not practically significant, the statistically different results could magnify themselves to become practically significant in the campaign output analysis.

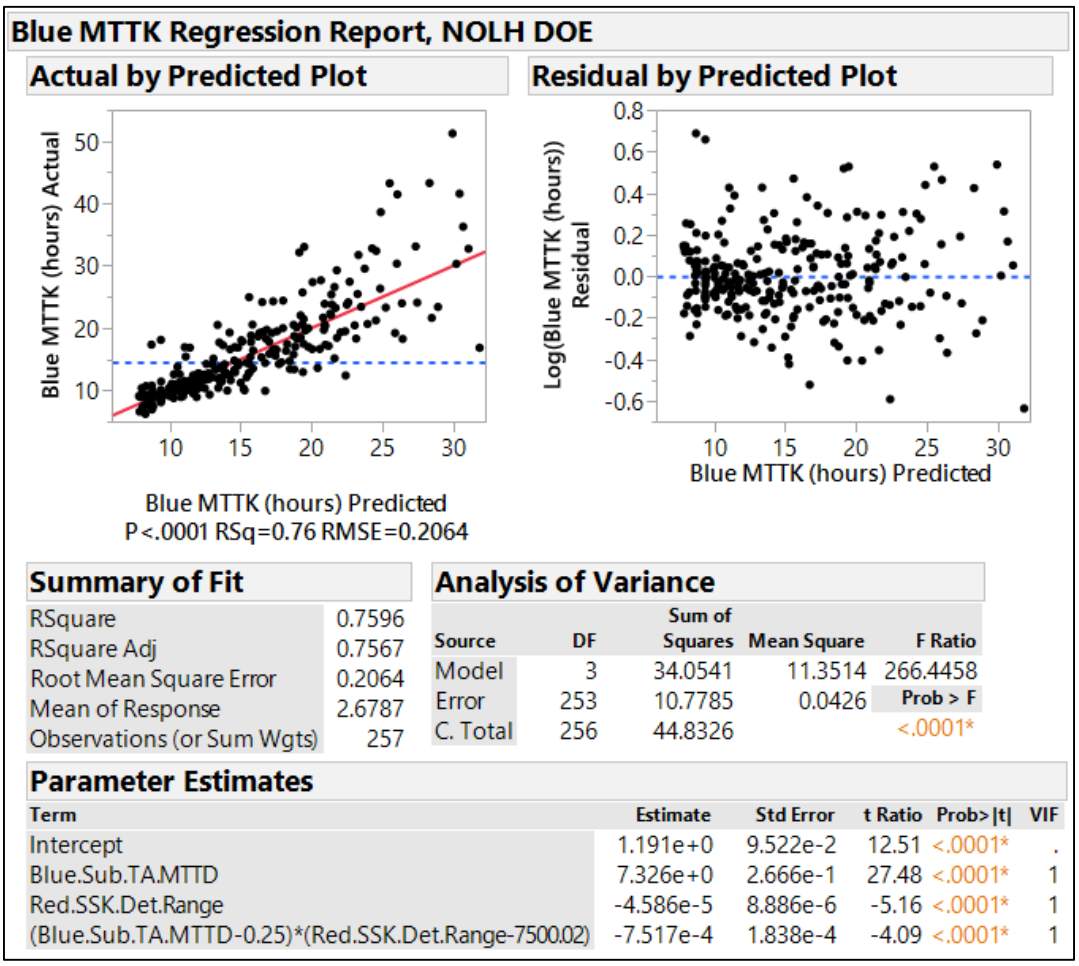
2. Linear Regression Analysis

A linear regression analysis is performed to determine which mission-level measures of performance (MOPs) are most significant in determining the mission level MOEs, MTTK, and P[Blue Wins]. Although the focus of this thesis is not to determine how to engineer a better submarine, the regression analysis helps validate the mission level model and provide some insight into the most important factors affecting unit interactions. If the model is performing well, then there will be a strong relationship between one or more input factors on Blue and Red MTTK.

A linear model is fit to Blue MTTK for each experimental design to determine which factors are most significant in predicting the MTTK. In order to find the best linear model among several variables, the analysis uses JMP's stepwise platform that iteratively tests predictors and produces the model with the minimum Bayesian information criterion (BIC). Although a metric like time is typically not a linear relationship, this is handled by transforming the predicted values with a natural logarithm (Wackerly, Mendenhall III and Schaeffer 2008). The model considers main effects, quadratic effects, and two-factor interactions between variables. Once JMP produces a recommended model, the analysis conducts additional pruning of variables in order to produce a model that meets the assumptions of a linear model (Wackerly, Mendenhall III and Schaeffer 2008).

The regression reports for Blue MTTK in the NOLH and R5FF DOE fitted with a natural logarithm transformation on the predicted values are displayed in Figure 7 and Figure 8. The NOLH model has fewer parameters and lower R^2, 0.76 versus 0.88, but

does a better job at meeting all of the assumptions of a linear model (Wackerly, Mendenhall III and Schaeffer 2008; Carver 2010). Therefore, despite its lower R^2, the NOLH design of experiments produces data that better fits a linear model than the R5FF design. The residual by predicted plot for the R5FF does not illustrate the random scatter pattern that would indicate that the residuals are normally distributed with constant variance, and the large number of terms included in the stepwise regression indicates that the model may be over-fit or may not capture a non-linear relationship. However, removing the interaction terms does not achieve a normally distributed residual versus fitted plot, so the final model includes them.



Summary of Fit

RSquare	0.7596
RSquare Adj	0.7567
Root Mean Square Error	0.2064
Mean of Response	2.6787
Observations (or Sum Wgts)	257

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	34.0541	11.3514	266.4458
Error	253	10.7785	0.0426	Prob > F
C. Total	256	44.8326		<.0001*

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|	VIF
Intercept	1.191e+0	9.522e-2	12.51	<.0001*	.
Blue.Sub.TA.MTTD	7.326e+0	2.666e-1	27.48	<.0001*	1
Red.SSK.Det.Range	-4.586e-5	8.886e-6	-5.16	<.0001*	1
(Blue.Sub.TA.MTTD-0.25)*(Red.SSK.Det.Range-7500.02)	-7.517e-4	1.838e-4	-4.09	<.0001*	1

Figure 7. Blue MTTK regression report, NOLH DOE. The parameter estimates are fitted in a model with a natural logarithm transformation of the estimated Blue MTTK.

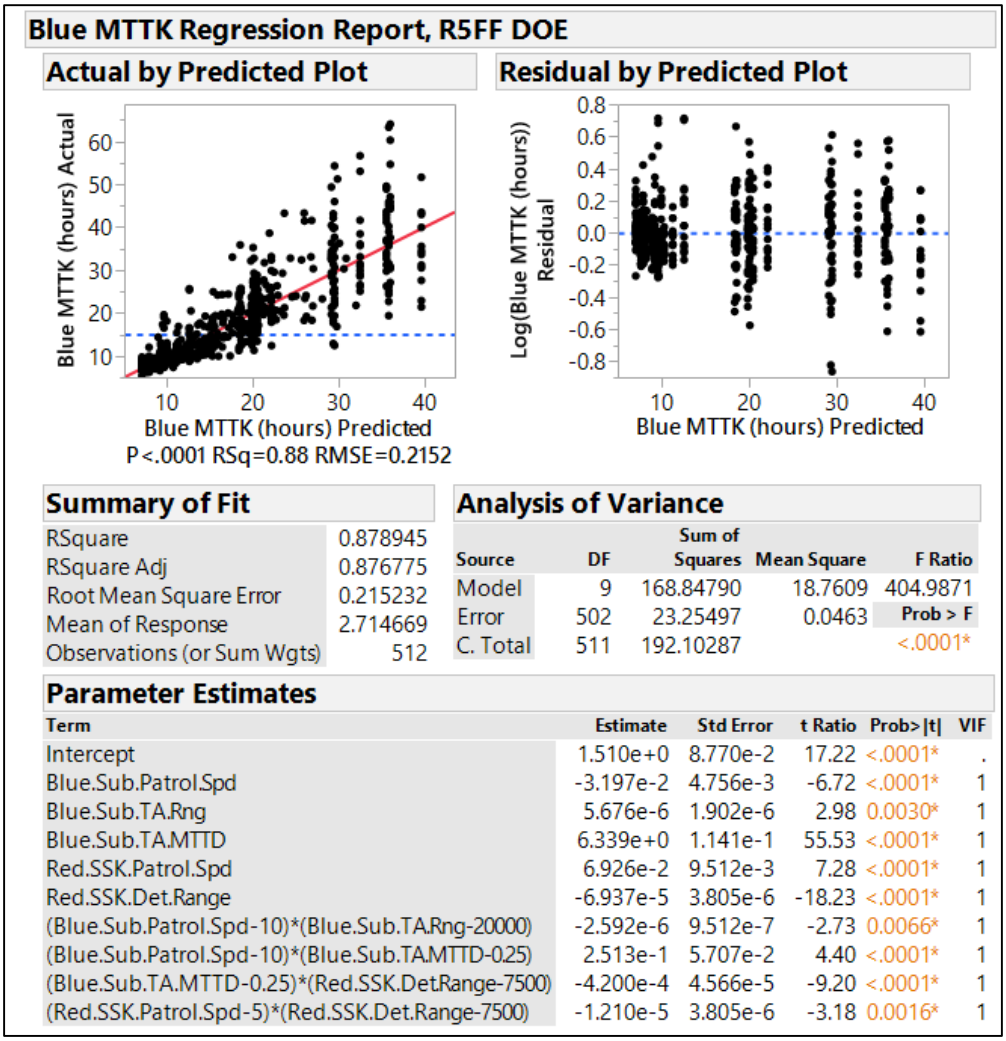


Figure 8. Blue MTTK regression report, R5FF DOE. The parameter estimates are fitted in a model with a natural logarithm transformation of the estimated Blue MTTK.

In addition, both analyses fail the Shapiro-Wilks test for normally distributed residuals, although the departure is small according to the normal quantile plots in Figure 9. The departure from normality is due to the unpredictability of MTTK values higher than 25 hours, which are outliers according to the previous distributions. The lack of normally distributed residuals is not crucial to further analysis since the purpose of this regression is to ensure that the statistically important factors have real-world significance; if it were, it could be eliminated by filtering MTTKs over 25 hours, or using survival analysis.

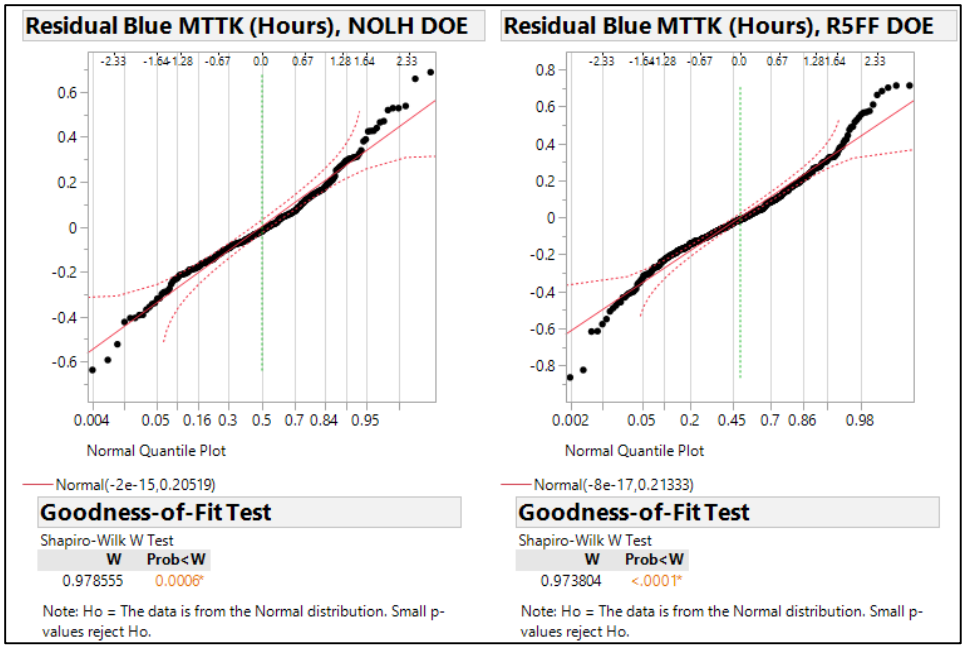


Figure 9. Blue MTTK test for normality of residuals.

Both models indicate that the most significant factor is the Blue submarine's towed array time to detect. The second most significant parameter in both models is the range at which SSKs detect Blue submarines. Therefore, according to the MANA model, the most important factor for detecting a Red submarine is having a capable towed array that can resolve signal from noise and well-trained operators to detect the submarine on the displays. Since a towed array is Blue's primary ASW sensor, and in reality there is a distinct advantage to having the first shot in submarine warfare, the model is performing adequately enough to proceed with further analysis.

B. STOCHASTIC LANCHESTER ANALYSIS

Recall that the MTTK and P_W are used to construct the attrition coefficient for a stochastic Lanchester campaign model, where:

$$b = \frac{P[Blue\,Wins]}{MTTK_{Blue}} \tag{4.1}$$

$$a = \frac{P[Red\,Wins]}{MTTK_{Red}} \tag{4.2}$$

$$P[Red\,Wins] = 1 - P[Blue\,Wins] \tag{4.3}$$

The P[Blue Wins] use the means displayed in Figure 5 and Figure 6 while each force's MTTK will be sampled from the MANA output using a variety of methods. The goal is to compare whether the output of the stochastic Lanchester simulation is significant, both statistically and in terms of real-world numbers. The measures of effectiveness (MOE) for this analysis are μ_W and μ_A in a fight-to-the-finish. In addition, to explore how variance propagates through hierarchal models, the analysis will statistically compare E[σ_W] and E[σ_A] by sampling method. The mean is selected because the summary statistics become approximately normal as the data sets are summed, so they can be tested using student's t-test (Wackerly, Mendenhall III and Schaeffer 2008).

In addition, recall that the first step in output processing is to average the effects by each design point. After this is performed, a subsequent test on variance of the output would be testing differences in V[μ_W] and V[μ_A], the variance in the summarized average means for W and A. This is not the value the analysis is interested in obtaining, and would give a false, narrow estimate of V[W] and V[A]. JMP allows the user to retain the standard deviation of interest as data sets are summarized. Since sums of random variables are approximately normal, a traditional t-test can be used to compare the average variance for W and A.

1. Preliminary Exploration: Sample from Raw versus Summarized Output

a. Overview and Methodology.

The first set of experiments utilizes either the mean MTTK from the MANA output or random sampling from the MANA output, with replacement, repeated ten times with different random seeds. The sampling methods are summarized in Table 4. This gives unique attrition coefficients for Red and Blue that are input into stochastic Lanchester simulations. From herein, each unique combination of attrition coefficients is a "design point," each of the 10 random samples is referred to as a "sampling index," and each method of sampling from the MANA output is an "experiment." Each design point is run in a stochastic Lanchester campaign for 30 replications. Ten sampling indices and 30 repetitions are chosen because it mimics the computing limitations on large-scale

DOD simulations that are computationally expensive. In addition, ten indices serve two analysis purposes: first, to determine if there is a difference in the MOEs based on luck, and second because it allows use of the student's t-test to compare averages.

The overview of the analysis workflow is displayed in Figure 10. The blue region is the mission model analysis, the orange region is the sampled attrition coefficient campaign model analysis, the green region is the mean attrition coefficient campaign model analysis, and the gray region is the final statistical comparison. When sampling from the summary output, the experiment takes a 25% sample across the 512 design points for the R5FF data set and 257 design points for the NOLH data set, each design point has the 40 MANA replications summarized by μ_{MTTK}. When sampling from the raw output data, the random sampling takes 25% of each design point before the MTTK summarized by its mean over 40 replications. In the event where both Red and Blue MTTK values are sampled, the amount is limited to 25% of Red's output because using 25% of Blue's output would result in always sampling 100% of Red's MTTK. This is because there are significantly more cases where Blue wins the one-on-one engagement. Finally, when constructing the experiment for the mean case, the analysis runs a total of 300 repetitions, but divides this output into 10 distinct groups. This is performed to create the same 10 data points obtained in the sampled cases to perform statistical comparisons.

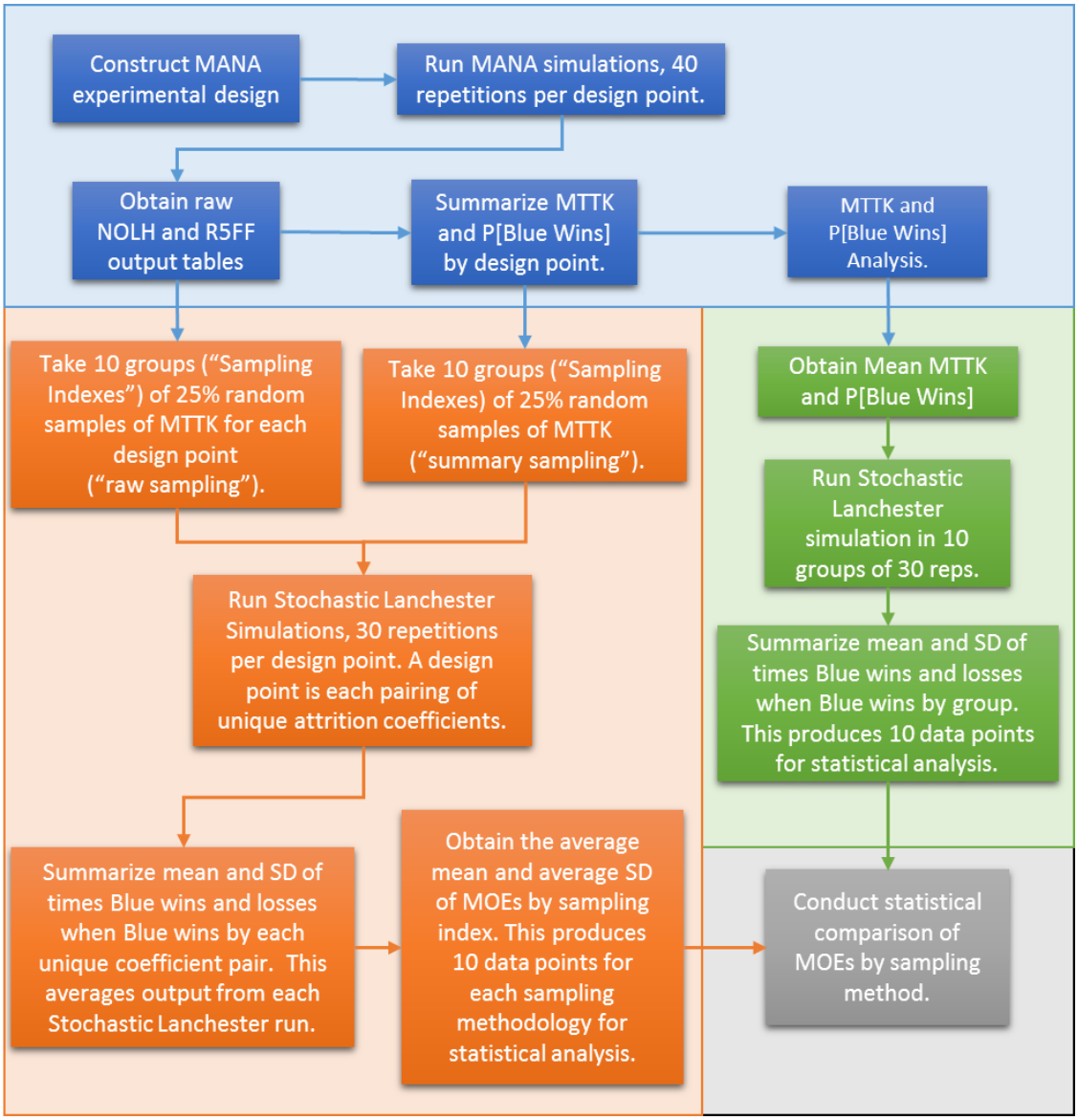


Figure 10. Overview of analytical workflow. The blue region is the mission model analysis, the orange region is the sampled attrition coefficient campaign model analysis, the green region is the mean attrition coefficient campaign model analysis, and the gray region is the final statistical comparison.

Table 4. Summary of sampling methodologies used to construct the stochastic Lanchester experiments. The total number of replications for each experiment is equal to the number of design points × 30 replications per design point × 10 sample indexes.

Sampling Method Name	Description	Data Sampled	Design Points
FF Average Both	Uses average MTTK from the MANA R5FF data set for both Red and Blue Attrition Coefficients.	NA	1
NOLH Average Both	Same as FF Average Both, except MTTK is taken from the MANA NOLH Data Set.	NA	1
FF Raw Sample Blue	Samples only Blue MTTK from the un-summarized MANA R5FF output. Only uses data points when Blue won the 1 versus 1 battle because otherwise the data is blank.	25% of MANA FF Blue MTTK results.	3,721
NOLH Raw Sample Blue	Same as FF Raw Sample Blue, except MTTK is sampled from the MANA NOLH Data Set.	25% of MANA NOLH Blue MTTK results.	720
FF Raw Sample Red	Samples only Red MTTK from the un-summarized MANA output. Only uses data points when Red won the 1 versus 1 battle because otherwise the data is blank.	25% of MANA FF Red MTTK results.	123
NOLH Raw Sample Red	Same as FF Raw Sample Red, except MTTK is sampled from the MANA NOLH Data Set.	25% of MANA NOLH Red MTTK Results.	1,859
FF Raw Sample Both	Samples both Blue and Red MTTK from the un-summarized MANA R5FF output. Uses same amount of samples as the FF Raw Sample Red because using 25% for Blue MTTK results in sampling 100% of Red output.	25% of MANA FF Red MTTK Results.	418
NOLH Raw Sample Both	Samples both Blue and Red MTTK from the un-summarized MANA output. Uses same amount of samples as the NOLH Raw Sample Red because using 25% for Blue MTTK results in sampling 100% of Red output.	25% of MANA NOLH Red MTTK Results.	418
FF Summary Sample Both	Samples both Blue and Red MTTK from the summarized MANA R5FF output.	25% of MANA R5FF results.	1,230
NOLH Summary Sample Both	Samples both Blue and Red MTTK from the summarized MANA R5FF output.	25% of MANA NOLH results.	64

The output of each stochastic Lanchester replication produces the Blue attrition and a binary representation of whether Blue wins. This is then summarized in JMP, first by design point, and then by random sampling index, with the result being a table that has 10 values for every sampling methodology to conduct statistics on. Although 10 values do not seem statistically significant, each value will be the summary of the amount of data points equal to the right hand column in Table 4, and therefore will give good indication of statistical differences. Further exploration with 50 column indices will be performed based upon the results of the preliminary analysis. As with the MANA output, this summary is performed because the analysis is interested in the difference in results across a variety of conditions, and not on the randomness of results within particular conditions.

b. Exploring the Effects of Experimental Design

The first comparison is to determine if there is a difference between the campaign MOEs when using the MTTK generated from the R5FF and the NOLH MANA data sets. The distributions of the MOEs are reproduced in Figure 11 and Figure 12. Figure 11 displays the distribution of the Blue attrition given Blue wins a submarine campaign with 18 Blue SSNs versus 25 Red SSKs, modeled with a stochastic Lanchester simulation. The distributions are separated by the way the MANA model is designed in order to determine if there is an effect of mission-model DOE on campaign model output. Figure 12 displays the distribution of the odds that Blue wins this engagement, separated by the way the MANA model is designed. It's clear that the campaign MOE distributions are different according to the mission model DOE. While the mean and the standard deviations of the outcomes are very close, the shape of the distributions is quite different. An interesting observation regarding the Figure 11 is that Blue never wins when average losses exceed half its force.

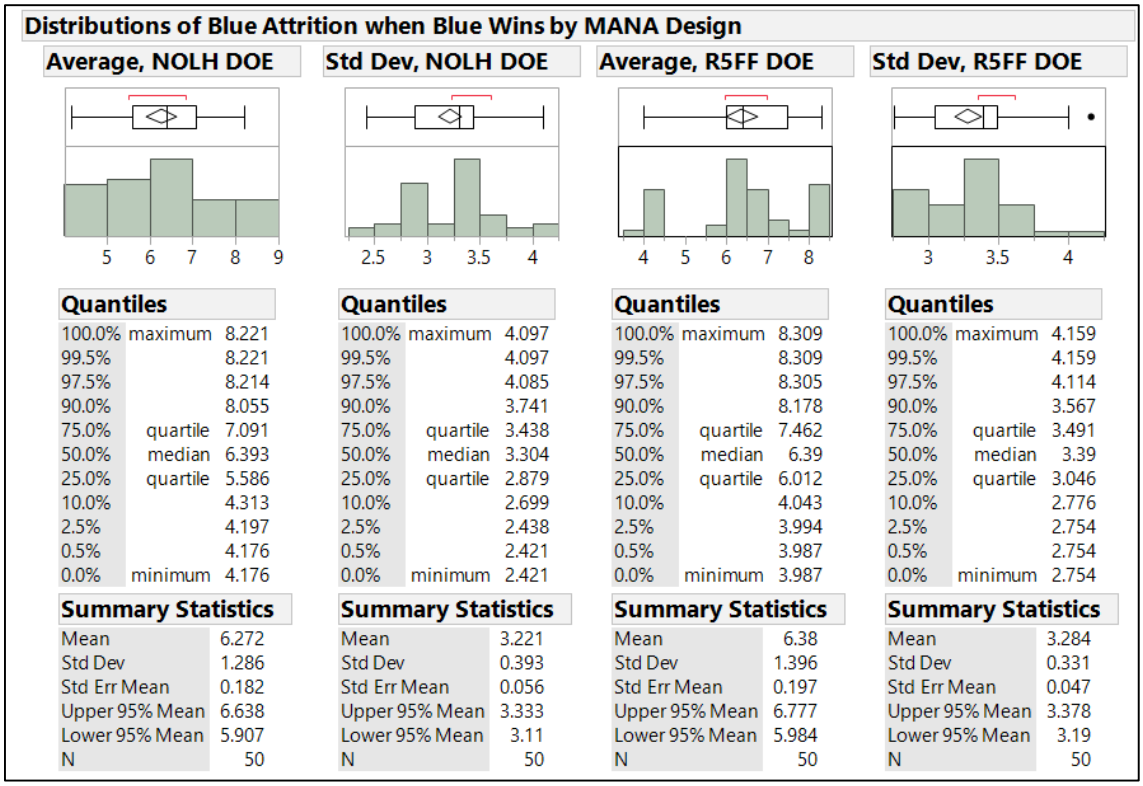


Figure 11. Distributions of Blue attrition when Blue wins according to MANA experimental design.

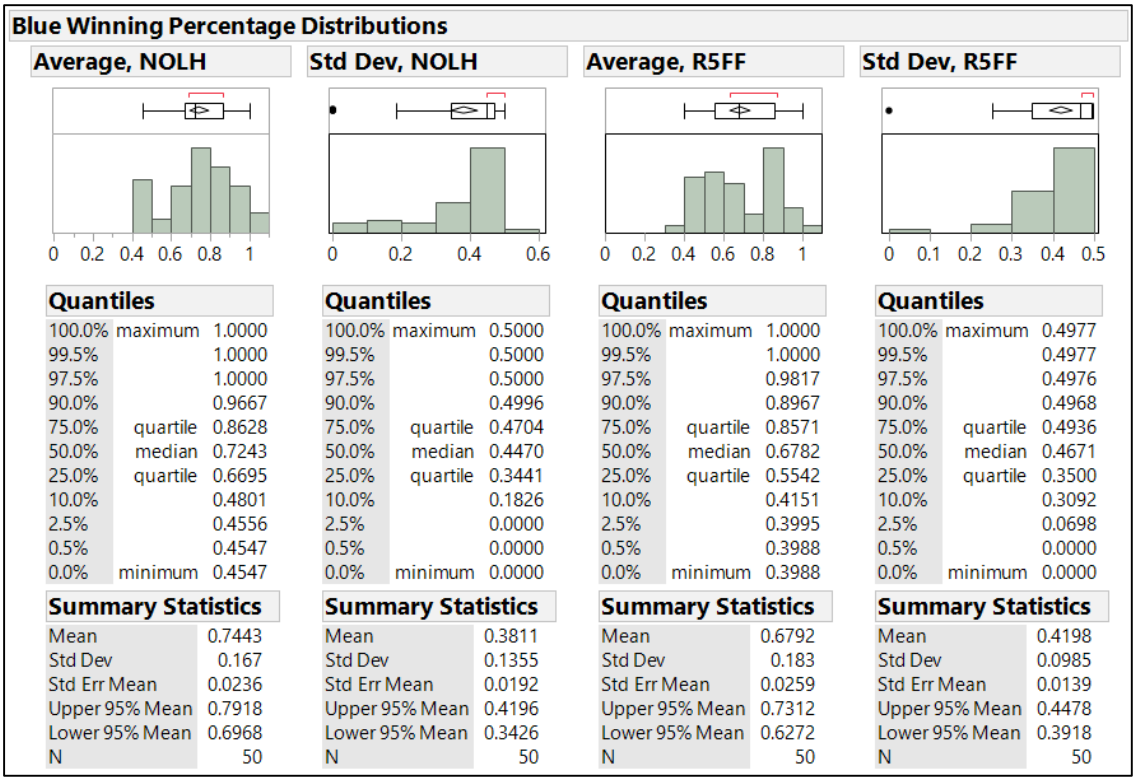


Figure 12. Distribution of Blue winning percentage according to MANA experimental design.

The analysis uses t-tests with α = 0.05 to determine if the MANA experimental design has a statistically significant effect on the μ_A, E[σ_A], μ_W, and E[σ_W]. The reason that a t-test is used to compare E[σ_A] and E[σ_W] summarized by column index rather than using a Levene test to compare variance across each unique attrition coefficient is because the latter test has over 70,000 data points; this virtually guarantees that the test will find a statistically significant difference as long as the values are not identical.

There is a significant difference at the 0.05 level in μ_W for the NOLH (μ = .7443, σ = .167) and R5FF (μ = .6792, σ = .167) data sets in a one-tailed t-test; t_{98} = 1.857, p-value = .0331. There is no statistically significant effect of experimental design on any of the other parameters in a one-tailed test, and there is no statistically significant effect of experimental design in a two-tailed t-test. There is also a practical difference of 13% change in μ_W when changing the experimental design for the MANA simulation. Because

experimental design was significant in at least one campaign MOE, it is kept as a factor in further exploratory analysis.

c. Analyzing Raw versus Summarized Data

The next step of the analysis examines whether sampling from the unprocessed MANA data, or raw output, versus sampling average MTTK by design point, or summarized output, affects the campaign MOEs. When the raw output is sampled, 25% of each MANA design point output is sampled. In the case where both Blue and Red MTTK are sampled, the analysis takes 25% of Red's data because there are fewer cases where Red gets the kill. Refer to Table 4 for sampling methods and descriptions.

The analysis uses JMP's one-way platform to conduct statistical comparisons. This feature is useful because it provides a graphical representation of multiple t-tests that is easy to interpret for analysis. Figure 13 and Figure 14 display the results of these tests for μ_A and E[σ_A], respectively. In the top part of each figure, the green diamonds represent the average values and their 95% confidence interval by sampling methodology. Each dot represents the overall mean of one of the ten sample indices, and may overlap depending on the results. The black line in the center represents the overall mean. The circles to the right of the figure are centered on each mean. If two circles to the right of this figure overlap, then those corresponding means are statistically the same. If not, the values are statistically different. In the bottom half of each figure, the blue banded region represents the overall mean and 95% confidence interval. If the colored dots fall within the banded region, then they are statistically similar to the group mean.

According to Figure 13, the mean Blue attrition values are statistically different under most cases by sampling methodology. Only two values fall within the 95% confidence interval for the overall group mean. The R^2 value for a one-way ANOVA of mean Blue attrition by sampling method is .956, indicating a strong relationship between sampling methodology and μ_A. The difference in variance is also significant, but not as pronounced. Many of the E[σ_A] values are statistically similar, and four values fall within the 95% confidence interval for the overall group mean. The R^2 for one-way ANOVA of E[σ_A] by sampling method is 0.596, indicating a weaker association. It is clear that

sampling methodology has a strong relationship, although this analysis confirms that there is a weak statistical difference between the R5FF and NOLH data sets.

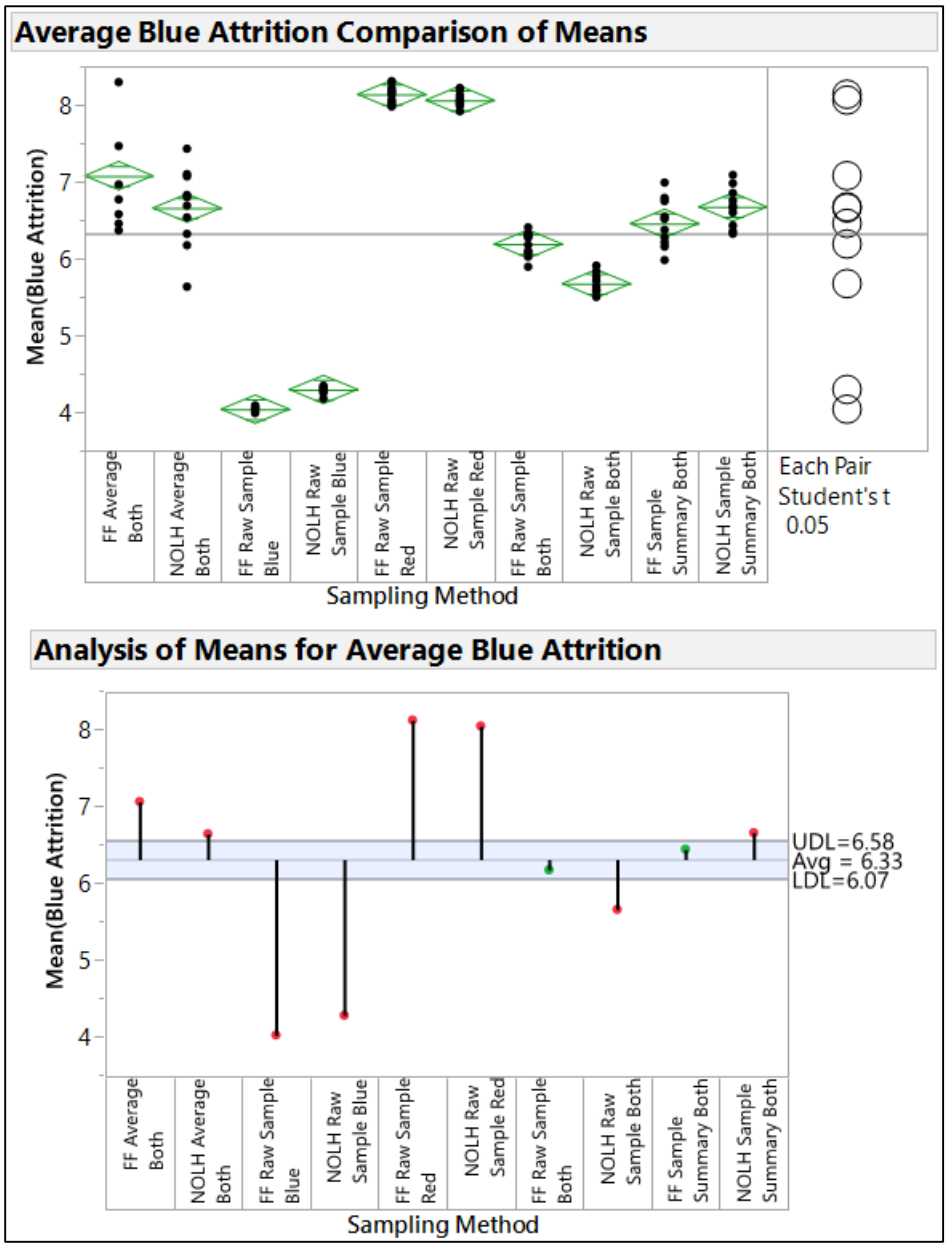


Figure 13. Statistical comparison of average Blue attrition when Blue wins. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

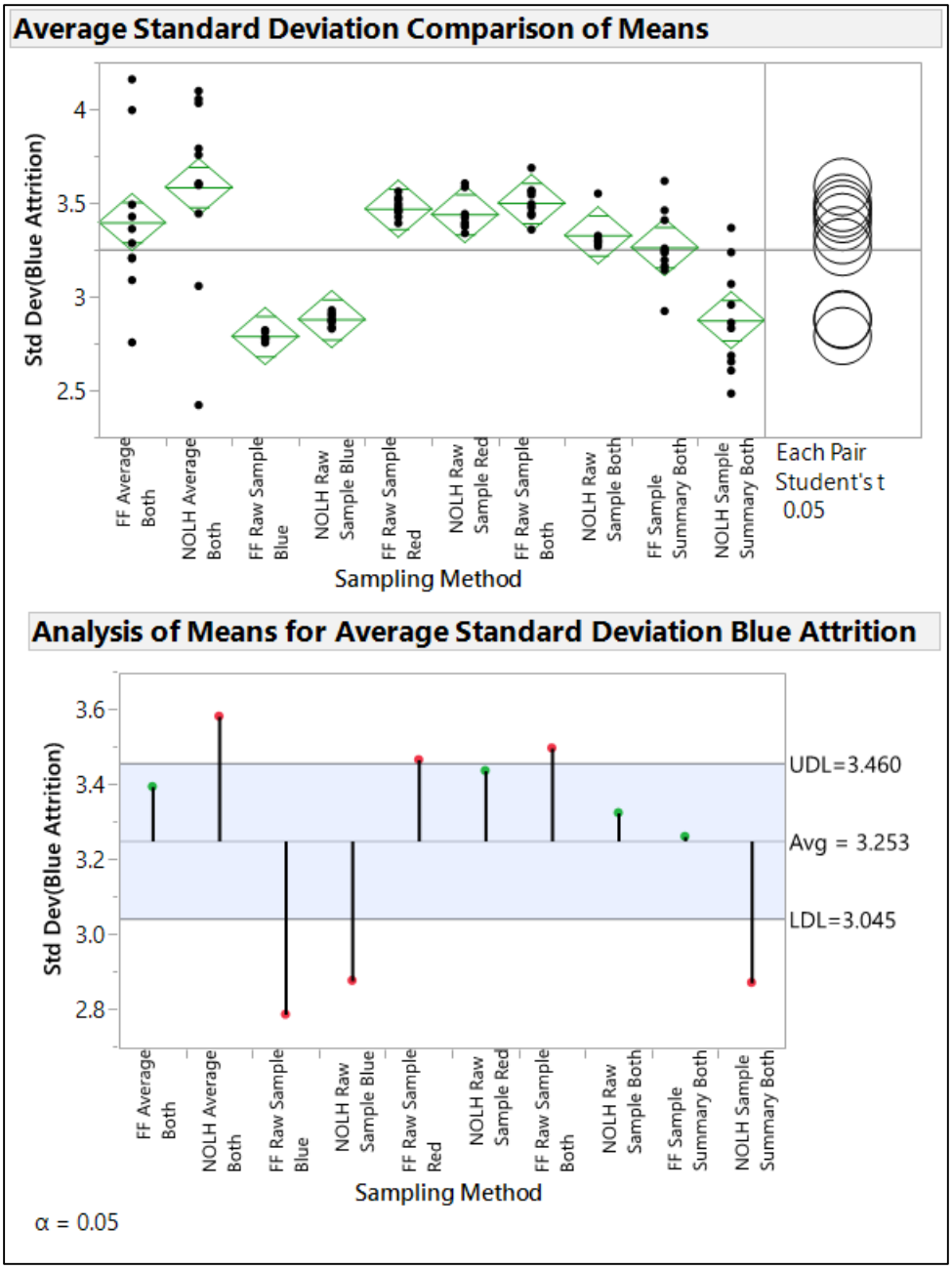


Figure 14. Statistical comparison of the average standard deviation of Blue attrition when Blue wins. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

Next, the analysis explores μ_W and E[σ_W] using the same technique. The graphs are displayed in Figure 15 and Figure 16. The effect of sampling method is more pronounced on μ_W than it was for Blue attrition. For the average, only two sets of data are statistically the same and two values fall within the 95% confidence interval of the group mean, and the R^2 is 0.974, indicating a strong correlation between sampling method and μ_W. The effect on E[σ_W] is slightly weaker, with more values falling within statistical range of each other and an R^2 of 0.833, indicating a moderate correlation. However, no values fall within the 95% confidence interval of the group average.

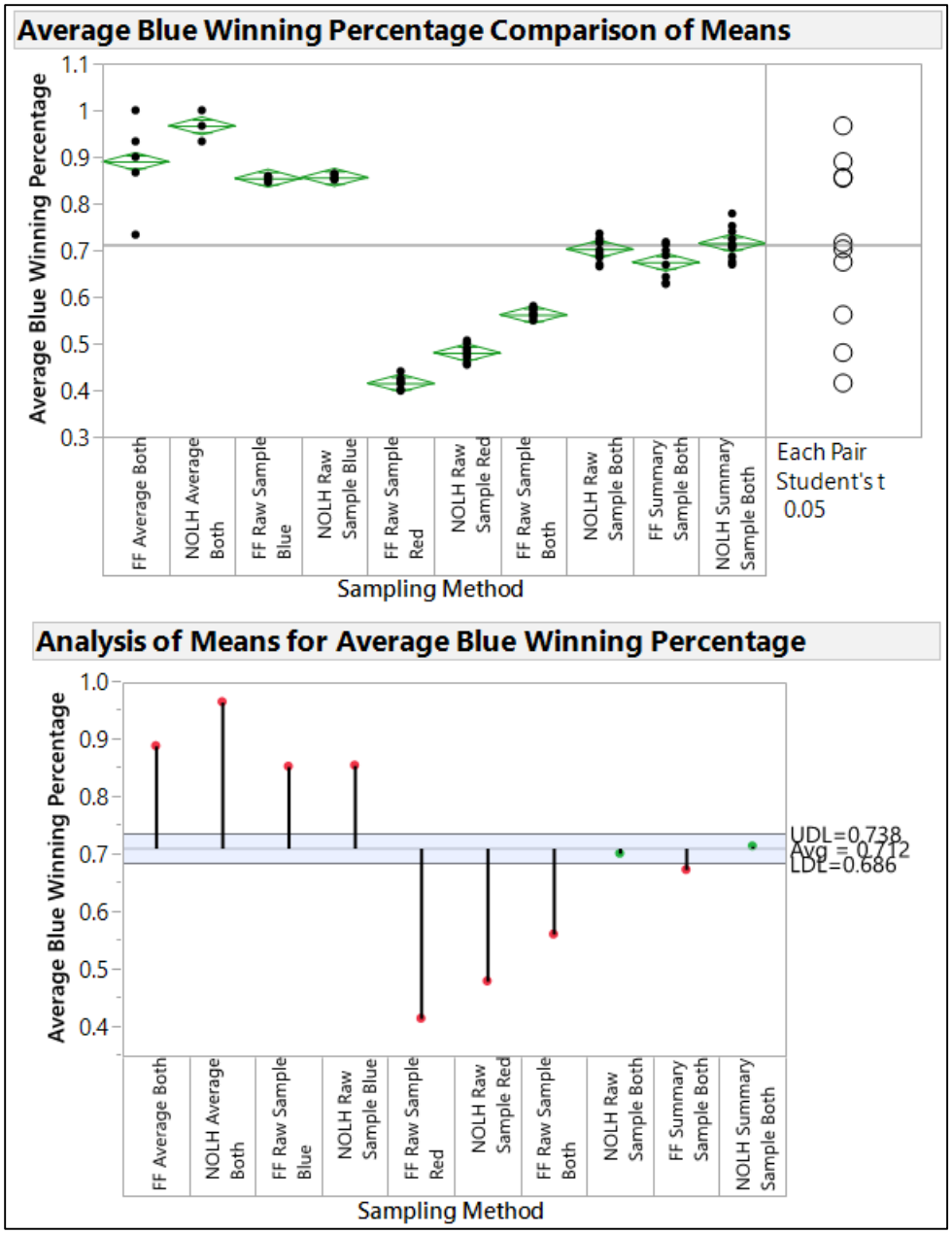


Figure 15. Statistical comparison of average Blue winning percentage. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

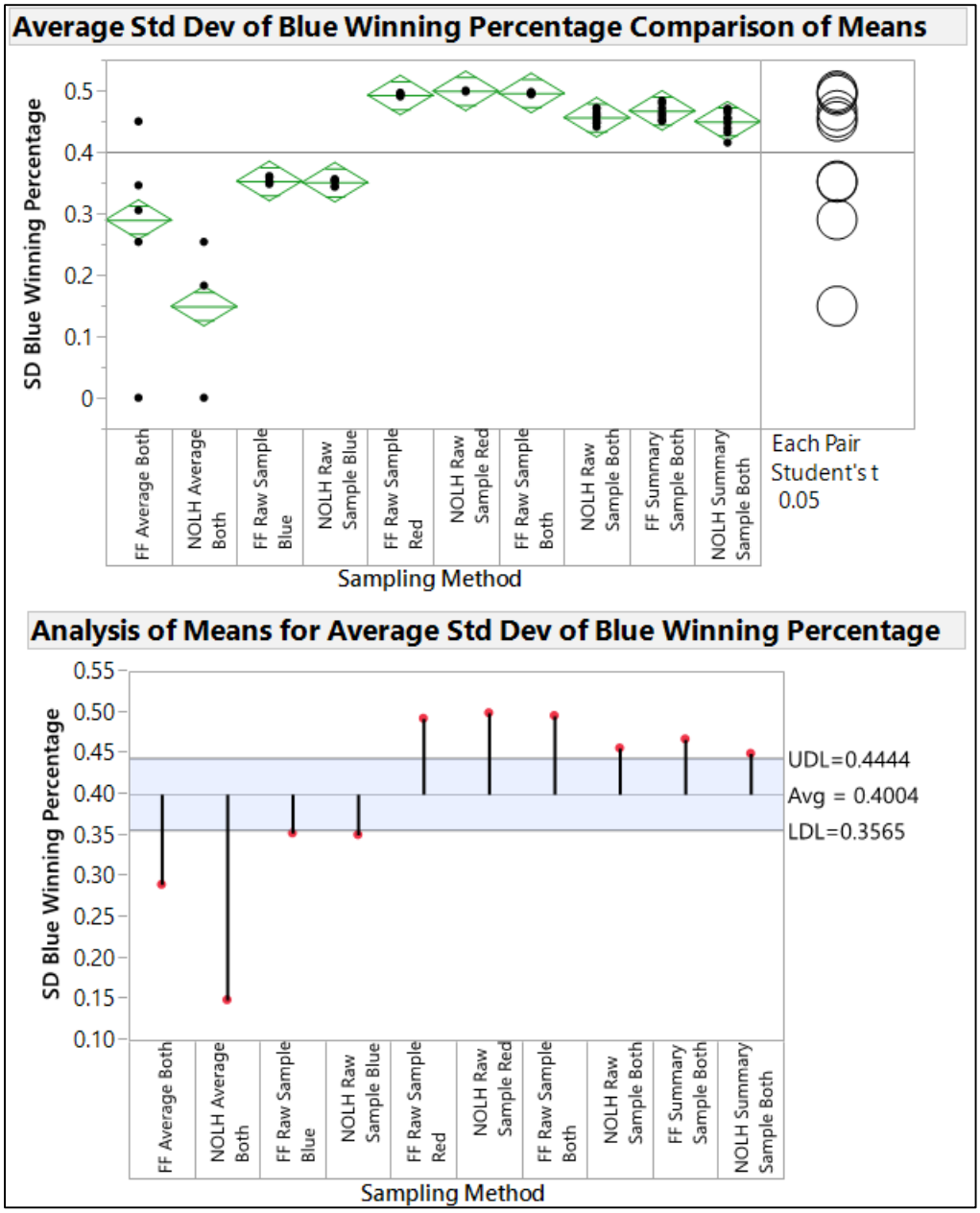


Figure 16. Statistical comparison of average standard deviations of Blue winning percentage. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

It is interesting that the effect of sampling methodology is more pronounced on μ_W than on μ_A. However, E[σ_W] is relatively constant around 0.4, except in the case where both average MTTK values were used, while E[σ_A] fluctuates more. The reason for this may lie in their underlying distributions. Since the stochastic Lanchester model is constructed using exponentially distributed events, the attrition is two-dimensional continuous time Markov Chain. The result of decreased variance means that the results of stochastic simulations will be closer together. On the other hand, the chances Blue wins the campaign, P_W, is derived from a binomial distribution with mean *np* and variance *np(1-p)*, where *p* is the chances Blue wins the one-on-one engagement from the MANA model. In addition there is no filtering for when Blue wins leading to a wider range of conditions.

Another interesting result is that the worst-case μ_A and μ_W for Blue occurs when the sampling method uses the average Blue MTTK and samples Red MTTK, while the best case occurs when the sampling method, aside from using both means, uses average Red MTTK and samples Blue MTTK. Let B and R denote the distribution of MTTK for Blue and Red, respectively. Let *X* and *Y* denote random samples from B and R, respectively:

$$F(X,Y) = P[X > Y] \tag{4.4}$$

$$G(X) = \mathrm{P}[X > \mathrm{E}[Y]] \tag{4.5}$$

$$H(Y) = \mathrm{P}[\mathrm{E}[X] > Y] \tag{4.6}$$

What this experiment reveals is that if E[*B*] > E[*A*] in any distribution, then:

$$G(X) > H(Y) > F(X,Y) \tag{4.7}$$

With sampling the distributions displayed in paragraph I.B.1.b, P[$b > a$] = 0.920, P[b > E[a]] = 0.993, P[E[b] > a] = 0.977. The discrepancy arises because this simulation uses the Lanchester square law, so for Blue to win in a deterministic case:

$$b > a\frac{y^2}{x^2} \tag{4.8}$$

The stochastic simulation allows for some variance in *b* and *a* based on luck, but in general this means that $b > 1.92a$ for Blue to win in this simulation. Therefore, P[b >

$1.92a$] = 0.727, P[b > 1.92(E[a])] = 0.862, and P[E[b] > 1.92a] = 0.652. This is more aligned with the empirical results.

2. Exploratory Analysis of Summarized Output.

a. Overview and Methodology.

The experiments conducted above used 10 column indices. The problem arising from this is that results of the stochastic Lanchester simulations did not begin to approximate the normal distribution. In addition, some may claim that the use of 10 sample indices is too weak statistically, so an additional experiment is conducted using 50 column indices. However, because the amount of runs would become too computationally expensive when sampling 25% of the raw MANA output, and because the previous analysis revealed that there is usually a bigger difference between the R5FF and NOLH data sets than the sample raw both and sample summary both data sets, the follow-on analysis conducts all sampling from MANA output with summarized means by design point.

The analysis will continue to explore the difference between a R5FF and NOLH constructed mission model propagated through a campaign model. In addition, the analysis adds two sampling methodologies used to obtain MTTK values to calculate attrition coefficients for the stochastic Lanchester model: "DOE All" and "DOE Outliers." The former constructs a NOLH set of MTTK based on the min and max of the MANA output, while the latter constructs a NOLH set of MTTK based on the min and max MANA output excluding statistical outliers. Statistical outliers are anything greater than the 3rd Quartile + 1.5 × (Interquartile Range), displayed in Figure 5 and Figure 6. The sampling methods are summarized in Table 5. Because the MANA output is summarized to obtain the mean of each MANA design point prior to sampling, there are no blank points and thus no variance in the amount of design points. All experiments of the stochastic Lanchester model except the DOE cases use 50 sample indices × 50 design points × 30 replications, where the experiments that use the average MTTK just have additional replications at the same value. The DOE sampling methods use 50 column indices × 33 design points × 30 replications; this is simply because of what the Naval

Postgraduate School NOLH algorithm, written by Dr. Tom Lucas, outputs for up to 11 factors.

Table 5. Summary of exploratory sampling methods

Sampling Method Name	Description
FF Average Both	Uses average MTTK from the MANA R5FF data set for both Red and Blue attrition coefficients.
NOLH Average Both	Same as FF Average Both, except MTTK is taken from the MANA NOLH data set.
FF Sample Blue	Samples only Blue MTTK from the summarized MANA R5FF output.
NOLH Sample Blue	Same as FF Sample Blue, except MTTK is sampled from the MANA NOLH data set.
FF Sample Red	Samples only Red MTTK from the summarized MANA output.
NOLH Sample Red	Same as FF Raw Sample Red, except MTTK is sampled from the MANA NOLH data set.
FF Sample Both	Samples both Blue and Red MTTK from the un-summarized MANA R5FF output.
NOLH Sample Both	Samples both Blue and Red MTTK from the un-summarized MANA output.
DOE All	Samples both Blue and Red MTTK from the summarized MANA R5FF output.
DOE No Outliers	Samples both Blue and Red MTTK from the summarized MANA R5FF output.

b. Analysis with More Statistical Power

The same analytical procedure discussed in section IV.B.1.c is performed here. Once again, JMP's one-way analysis platform provides a simple way to graphically present the results across a range of multiple values in Figure 17 and Figure 18, which display the one-way analysis of μ_A and E[σ_A] versus sampling method. This time, there are 50 dots in the top part of the figure that represent the mean of each sampling index. As expected, the greater statistical power gained from conducting 50 column indexes results in more values becoming statistically different. It is interesting that the "DOE All"

sample method produces extremely biased results. This occurs because the NOLH algorithm seeks to provide equal sample points along the entire range provided, and Red's MTTK distribution had a longer tail than Blue's. Therefore, the NOLH algorithm performs poorly with tailed distributions, such as the MTTK data produced from the MANA experiment. However, the bias is removed by removing outliers. Care must be used when constructing NOLH designs to avoid outliers. Alternatively, using a range of means, rather than a minimum and maximum value from the raw distribution, would produce better results. When the DOE All set is not included in the ANOVA analysis, the R^2 changes from 0.864 and 0.863 for the one-way analysis of μ_A and E[σ_A] versus sampling method, respectively, to 0.787 and 0.867.

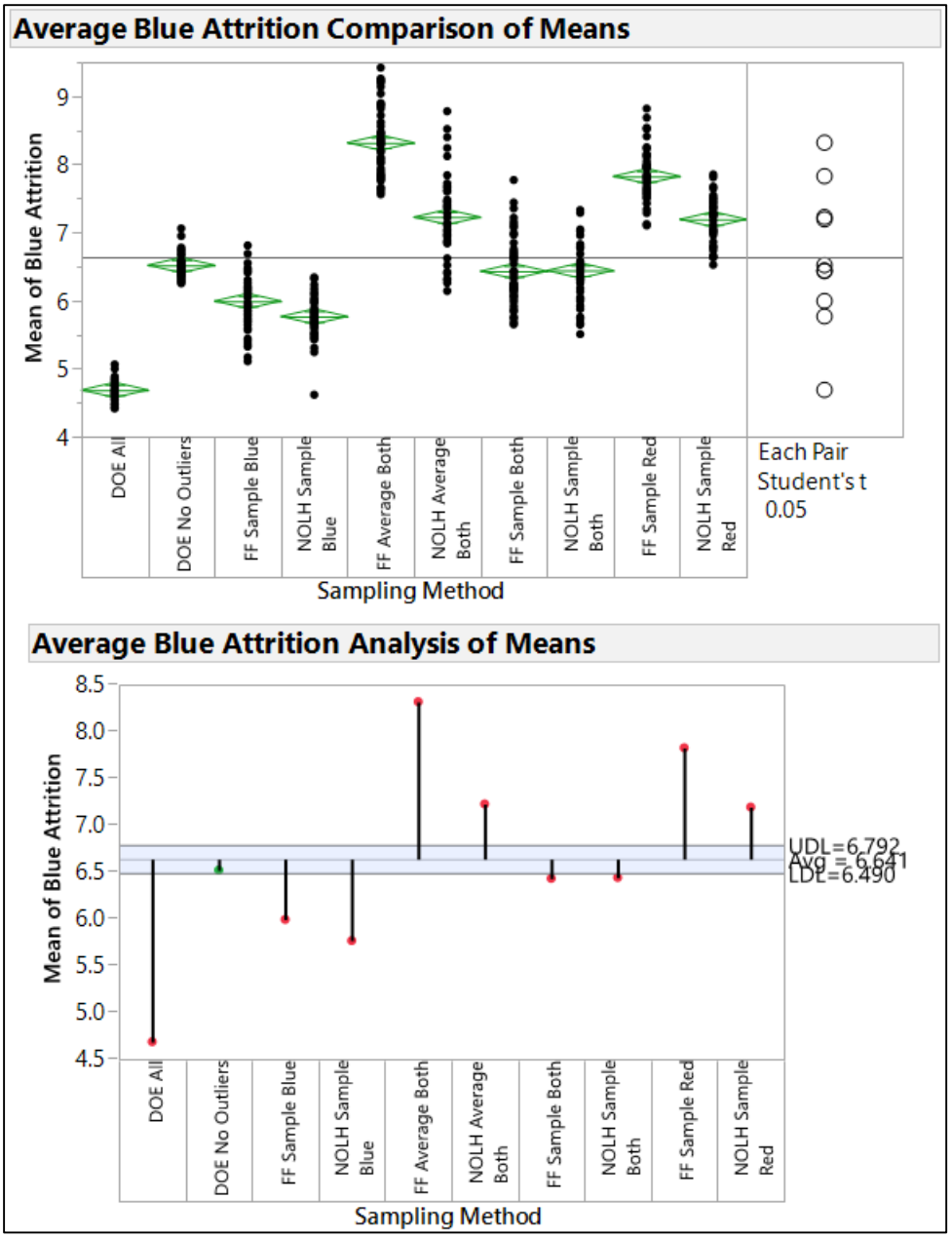


Figure 17. Statistical comparison of average Blue attrition with a bigger sample. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

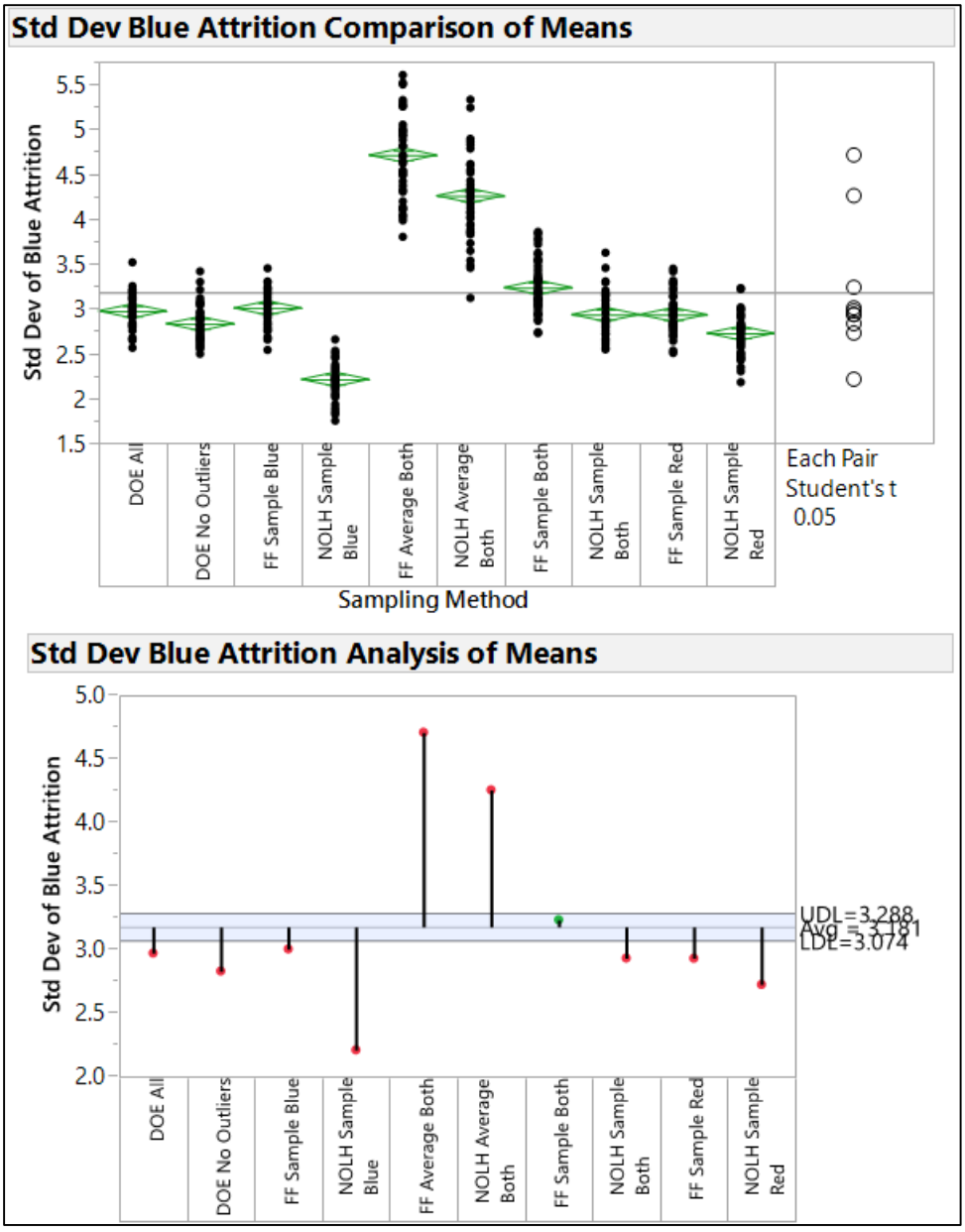


Figure 18. Statistical comparison of the average standard deviations of Blue attrition with a bigger sample. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

The analysis is repeated one more time with μ_W and E[σ_W] versus sampling method. The results are displayed in Figure 19 and Figure 20. Again, the additional statistical power results in almost all values being statistically different. There is a significant relationship between μ_W and E[σ_W] versus sampling method, with R^2 = 0.878 and 0.682, respectively.

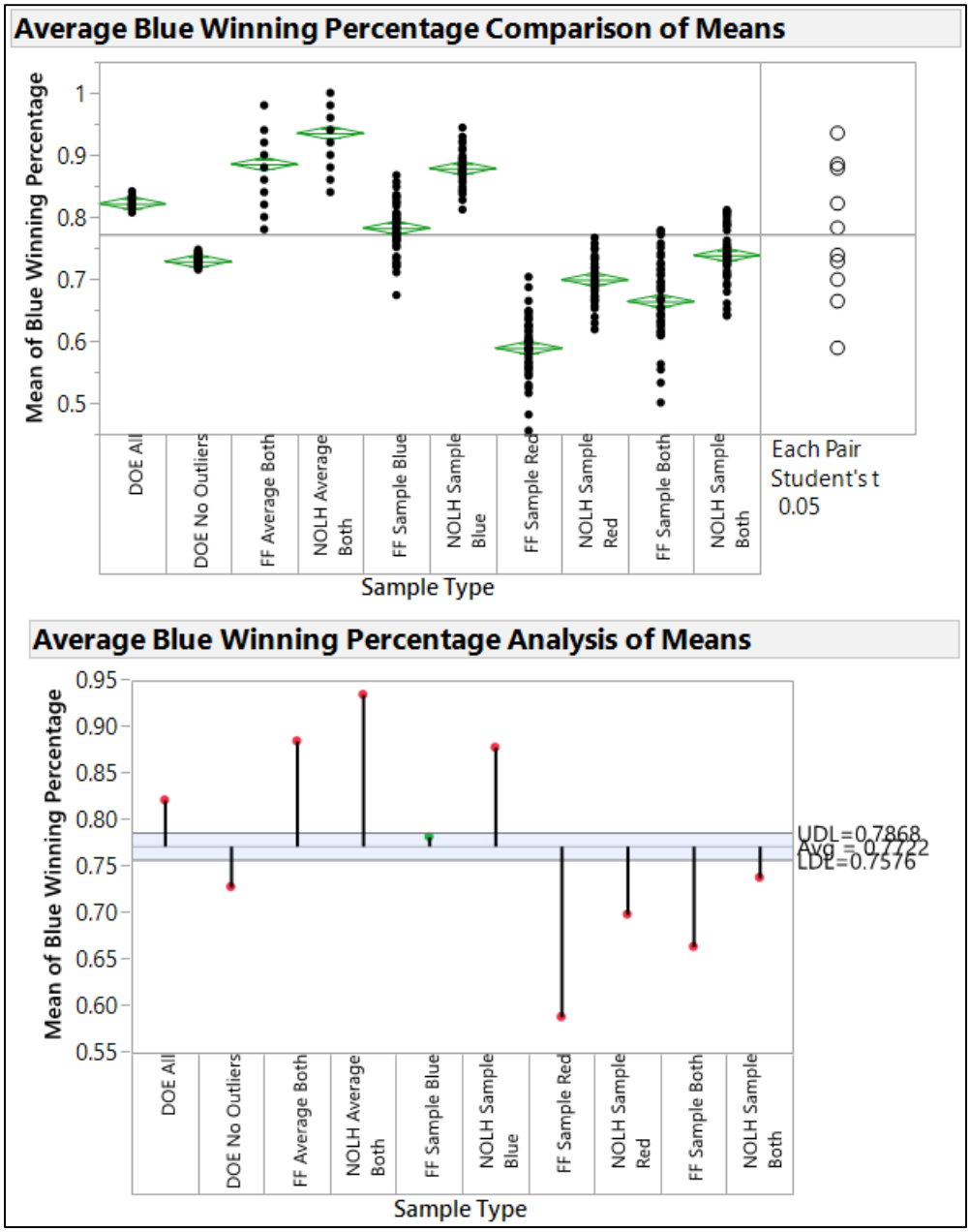


Figure 19. Statistical comparison of average Blue winning percentage with a bigger sample. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

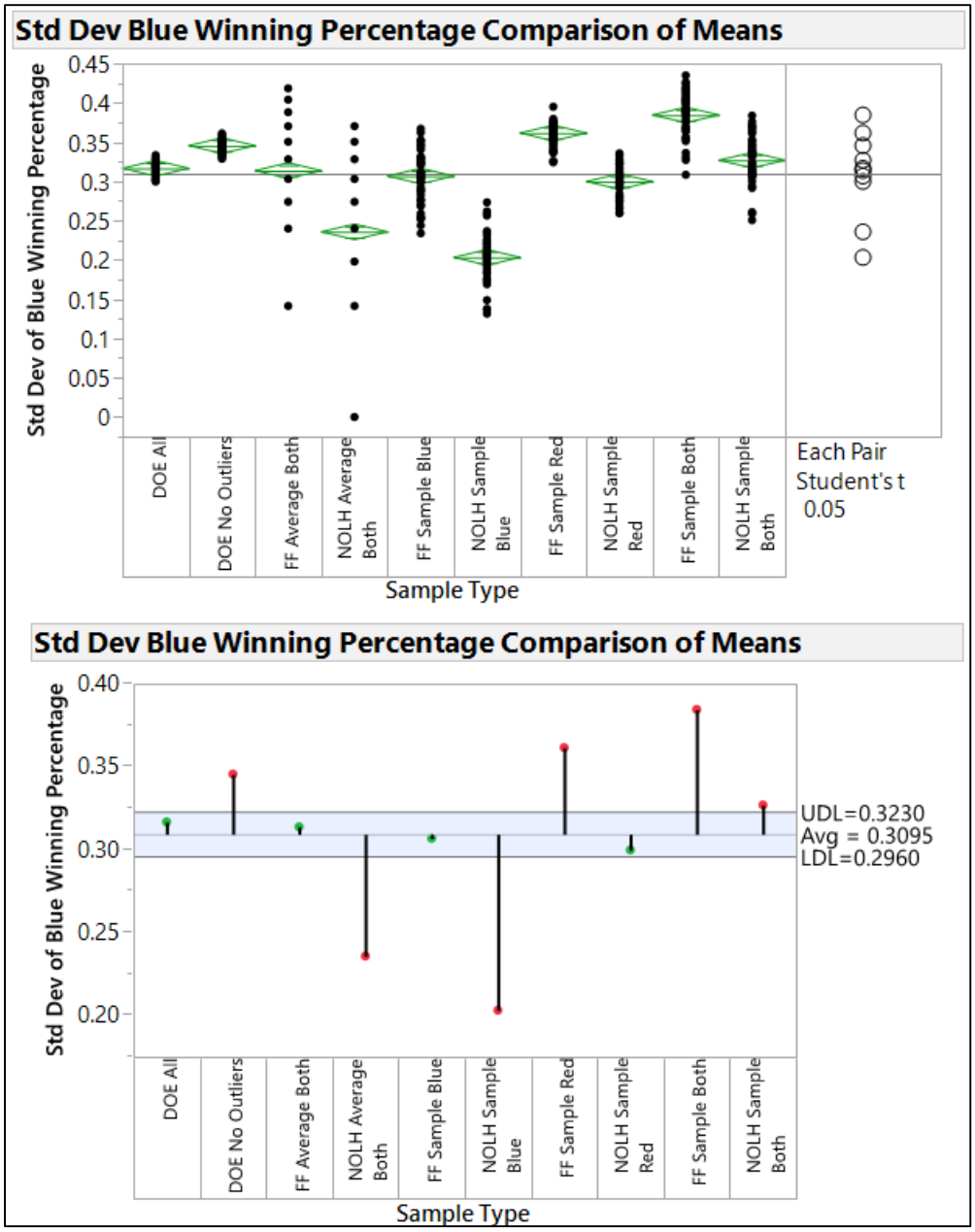


Figure 20. Statistical comparison of average standard deviations of Blue winning percentage with a bigger sample. Circles that don't overlap indicate values that are statistically different, and points outside the blue banded line indicate when values differ statistically from the overall mean among all data points.

Similar to the 10 column index case, using the average MTTK for both Red and Blue or just sampling Blue produces the most optimistic results for μ_W and μ_A. Additionally, while E[σ_A] varies significantly in practical terms by experimental design, from as high as 5 to as low as 2, E[σ_W] is practically constant at 0.30 +/- 0.05. Therefore, the sampling method typically produces biased results in the mean output, but may not always produce a practically biased estimate of the variance. Future work can focus on determining which method produces maximum likelihood of obtaining the minimum variance unbiased estimate of the true mean.

V. DISCUSSION

This thesis explores how error can propagate through hierarchal combat models. Paul K. Davis in his technical report for RAND corporation in 2007 stated that feeding both mean and variance information from the lower-level output to the higher-level input can increase accuracy (Davis and Henninger 2007). This work employed a variety of sampling methods to include not just the variance of the mission model output, but also the distribution of values. The results in Chapter IV demonstrate that random sampling may not eliminate bias in the mean nor misrepresentation of the variance.

A. QUANTIFYING THE RISK

These results have significant practical significance when quantifying risk to military commanders. The results found that the average chance of winning can be as high as 0.94 when the simulation is constructed with average mean time to kill (MTTK) for each force to as low as 0.66 when sampling from both forces. The standard deviation for this estimate is around 0.25 - 0.35 regardless of how the campaign model is constructed. With such a large discrepancy in the mean outcome and such a relatively large standard deviation, the best estimate one can give on these results is "better than half."

The other measure of effectiveness (MOE) for the campaign model is average losses given Blue wins. While this is a useful metric for simulation analysis, it is not useful in communicating risk to a military commander because it is inaccurate (Savage 2012). A battle such as this would only be fought one time. Nevertheless, one can make use of the mean and variance of the campaign output to provide a better estimate of risk by estimating the chances of success and the chances that a certain amount of units are lost. This risk assessment is displayed in the graph in Figure 21. This graph displays the odds that Blue loses a particular amount of submarines by sampling methodology. This graph demonstrates that the risk estimate is heavily dependent upon the way the hierarchal combat model is constructed rather than factors that are input into the model itself.

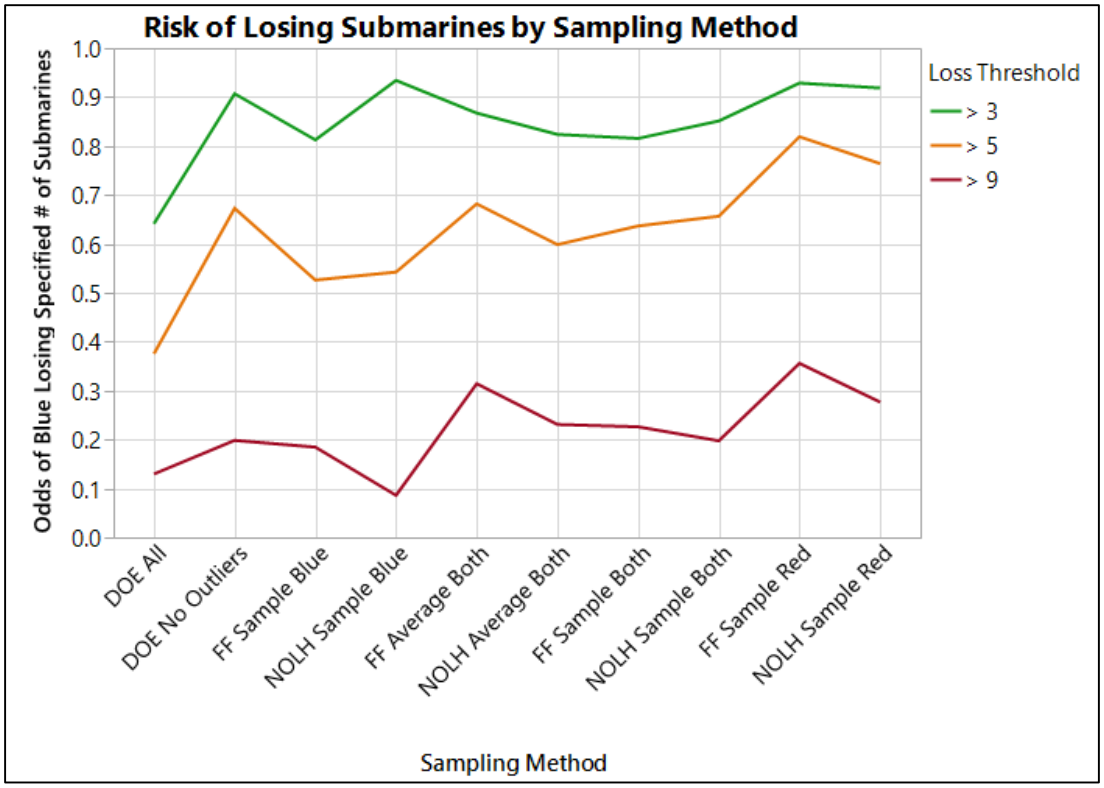


Figure 21. Graph of risk profiles by sampling methodology. Each line indicates the chances of losing at least 3, 5, or 9 submarines. Notice that the risk differs substantially based on how the mission model output is sampled.

B. DESIGN DECISIONS

The most useful element of stochastic combat simulations is not its predictive capability, but its descriptive capability. This is because data for combat simulations is hard to find, often inaccurate, and dependent on environmental conditions. As discussed in paragraph I.A, the basis for this process is to inform decisions on new platforms and technologies.

Unfortunately, the errors discovered in this work also affect this process. Consider a scenario where the Chief of Naval Operations (CNO) must decide between funding improvements between multiple platforms, one of them being fast attack submarines. OPNAV N81 runs a simulation using the current practice of propagating means and provides an optimistic estimate for winning the battle. Moreover, they estimate that investing in new submarine technology will only raise the chances of winning from 0.78

to 0.80, save one submarine in the battle, and cost $10 billion. Clearly, this is not a very large payoff and the Navy would get more value by investing in other technology. On the other hand, if the simulation were constructed by sampling the mission-model output with more pessimistic estimates, now the improvement might be from 0.6 to 0.7 and save five submarines in the battle for $10 billion. This is a substantial improvement that would warrant further investments. The key here is that even if both models reveal the same *relative* importance of design factors, the *practical impact* of those factors can be drastically different.

C. HIERARCHAL META-MODELS

In addition, the current process of using linear regression to construct hierarchal meta-models of stochastic combat simulations cannot accurately estimate how error propagates through the results. When linear models provide a confidence interval, they are estimating the interval for the line of regression and not the confidence interval for the actual values themselves (Wackerly, Mendenhall III and Schaeffer 2008).

The best use of hierarchal meta-models is to provide a coarse estimate of the mean output. They can be useful to provide fast estimates and to determine which combination of factors leads to the best results. However, they cannot accurately quantify, in practical terms, the effect of those factors on the output.

D. FUTURE WORK: IMPLICATIONS FOR DOD COMBAT MODELS

The models employed in this thesis are not accredited by the Department of Defense (DOD). Nevertheless, they can still provide useful insight into the practice of hierarchal combat modeling. This study demonstrates that the way in which variance in the lower-level model output is handled can have significant effects in the higher-level model output. The current practice of propagating averages is almost certainly providing a false estimate of campaign results.

There is potential for several future projects on this subject. First, empirical data can be collected by running a similar experimental design using accredited models. This will confirm that the bias in the mean and inaccuracies in variance also propagate through

these models. Secondly, there is potential work to mathematically derive formulae that can prove which method provides the minimum variance unbiased estimate of the mean. Third, an analysis can use a force-on-force mission-level simulation to better capture spatial effects and use U.S. Arms algorithms to adjust the attrition coefficients (Yildirim 1999). Finally, there is the opportunity for deriving a mathematical formula that can propagate error through hierarchal combat models.

VI. APPENDIX

Below is the code for a stochastic Lanchester campaign model using the squared law. The model takes in a spreadsheet of MTTK values that are randomly sampled from MANA output. It iterates through each pair of coefficient and simulates a fight to the finish between Red and Blue. It calculates the attrition coefficients by dividing the average P[kill] by MTTK for each side, then conducts the Lanchester simulation for 30 reps. It returns the starting and ending conditions to a .csv file as output to be analyzed. The code can be modified to substitute the Blue or Red average values instead of reading them from a spreadsheet.

```
import numpy as np
import scipy as sp
from scipy import stats
import random
import csv
import matplotlib.pyplot as plt
import os

# change working directory
os.chdir('C:\Users\Russell\Documents\OA5000 - Thesis\Campaign Models\ASW\NOLH\Big Samples Different Seed')

# function to import csvs.
def getParameters(filename):
    par = np.loadtxt(filename, delimiter = ',', skiprows = 1)
    return par

### the stochastic lanchester changes X forces to m, Y forces to N. The rate at which x forces are killed (square law) is a*n, the rate at which Y forces are killed is b*m. a = rate at which 1 red kills 1 blue; b = rate at which 1 blue kills 1 red.

def EulerStochLan(m0, n0, a, b, mbp, nbp, tlimit):
    m = m0
    n = n0
    t = 0.0
    endOfBattle = False

    while endOfBattle == False:
        rate = m*b + n*a                # determine time to next kill and increment time. n/m
                                        # because P-3s only attack subs in its sector
        t = t + random.expovariate(rate)
        prob_x_killed = n*a/(n*a + m*b)    # determine probability of a casualty and flip the
                                                #coin
        check = random.uniform(0,1)
        if check < prob_x_killed:
            m -= 1
        else:
            n -= 1
        if m <= mbp or n <= nbp or t >= tlimit:
            endOfBattle=True
```

```
    if m <= mbp:
        winner = 0          # 0 means red wins, used so you can calculate the numerical mean of winning.
    elif n <= nbp:
        winner = 1          # 1 means blue wins.
    elif m0 - m < n0 - n:
        winner = 1
    else:
        winner = 0
    blue_losses = m0 - m
    red_losses = n0 - n
    output = [m, n, blue_losses, red_losses, t, winner]
    return output

def subbattle_sampled_summary():
    blue_mttk = getParameters('Blue NOLH MTTK 50 Samples.csv')
    red_mttk = getParameters('Red NOLH MTTK 50 Samples.csv')   # import the randomly sampled list of
mean time to kill
    blue_mttk = blue_mttk.transpose()
    red_mttk = red_mttk.transpose()
    pkill_blue = 0.761                  # Based on mean time Blue wins from MANA data
    pkill_red = 1 - pkill_blue
    results = []                        # stores table of starting parameters and results for analysis
    out = []                            # stores temporary output
    dp = 0                              # design point for summary statistics
    Type = 'NOLH Sample Both'
    for row in range(0, len(blue_mttk)):
        dp += 1
        for item in range(0, len(blue_mttk[row])):
            a = pkill_red * 1.0 / red_mttk[row][item]       # rate at which red kills blue
            b = pkill_blue * 1.0 / blue_mttk[row][item]      # rate at which blue kills red
            m0 = 18.0                   # set initial blue forces
            n0 = 25.0                   # set initial red forces
            trial = 0
            for i in range(0, 30):          # stochastic simulation runs. 30 is for statistical power.
                trial += 1
                mbp = 0.0 * m0
                nbp = 0.0 * n0
                out = EulerStochLan(m0, n0, a, b, mbp, nbp, 336)
                out.insert(0, b)
                out.insert(0, a)
                out.insert(0, n0)
                out.insert(0, m0)
                out.insert(0, dp)
                out.insert(0, Type)
                results.append(out)

    fileout = open('SubOnSub_results_sampleboth.csv', 'wb')
    writer = csv.writer(fileout, dialect = 'excel')
    writer.writerow(["Sample Type," "Design Point," "Initial Blue," "Initial Red," "a," "b," "Final Blue,"
"Final Red," "Blue Losses," "Red Losses," "time," "winner"])
    writer.writerows(results)
    fileout.close()

                out.insert(0, a)
```

```
            out.insert(0, n0)
            out.insert(0, m0)
            out.insert(0, dp)
            out.insert(0, Type)
            results.append(out)

    fileout = open('SubOnSub_results_filling_nooutliers.csv', 'wb')
    writer = csv.writer(fileout, dialect = 'excel')
    writer.writerow(["Sample Type," "Design Point," "Initial Blue," "Initial Red," "a," "b," "Final Blue,"
"Final Red," "Blue Losses," "Red Losses," "time," "winner"])
    writer.writerows(results)
    fileout.close()
```

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF REFERENCES

Cappellini, Ronald. 2011. *Error Propagation in Hierarchal Combat Models.* Master's thesis, Naval Postgraduate School, Monterey, CA.

Carver, Robert. 2010. *Practical Data Analysis With JMP*. Cary, NC: SAS Institute.

Cioppa, Thomas M, and Thomas W Lucas. 2007. "Efficient Nearly Orthogonal and Space-filling Latin Hypercubes." *Tehnometrics* 49 (1): 45–55.

Davis, Paul K. 2001. *Exploratory Analysis Enabled by Multiresolution, Multiperspective Modeling.*Santa Monica, CA: RAND.

Davis, Paul K., and Amy Henninger. 2007. *Analysis, Analysis Practices, and Implications for Modeling and Simulation* (Technical Report OP-176-OSD). Santa Monica, CA: RAND.

Defense Industry Daily Staff. 2014. *Team Torpedo: U.S. Firms Sell & Support MK48s and MK54s.* August 21. http://www.defenseindustrydaily.com/team-torpedo-raytheon-partners-to-support-mk48-and-mk54-requirements-02533/.

Kelley, C.T., Jr. 1974. *Naval Campaigns in the Mediterranean and the Use of Land-Based Air for Carrier Task Force Defense.* Santa Monica, CA.:RAND.

Lanchester, Frederick W. 1916. *Aircraft in Warfare: The Dawn of the Fourth Arm.* Edited by J. R. Newman and Simon and Schuster. Vol. 4. Charleston, SC: BiblioLife.

Lucas, Thomas W. 2000. "The Stochastic Versus Deterministic Argument for Combat Simulations: Tales of When the Average Won't Do." *Military Operations Research* 5 (3): 9–27.

McIntosh, G.C. 2009. *MANA-V (Map Aware Non-Uniform Automata Vector) Supplementary Manual.* Auckland, N.Z.: Defense Technology Agency.

Osborn, Kris. 2014. *Navy Considers Future After Virginia-Class Subs.* February 12. http://defensetech.org/2014/02/12/navy-considers-future-after-virginia-class-subs/.

Osorio, Carolina, and Linsen Chong. 2015. "A Computationally Efficient Simulation-Based Optimization Algorithm for Large-Scale Urban Transporation Problems." *Transportation Science* 61, no. 6 (2013): 1–15.

Sanchez, Susan M., and Paul J Sanchez. 2005. "Very Large Fractional Factorials and Central Composite Designs." *ACM Transactions on Modeling and Computer Simulation* 15: 362–377.

Sanchez, Susan M., Thomas W. Lucas, Paul J. Sanchez, and Hong Wan Nannini. 2012. "Designs for Large-Scale Simulation Experiments with Applications to Defense and Homeland Security." Edited by K. Hinkelmann. *Design and Analysis of Experiments, Volume 3*. Hoboken, NJ: John Wiley & Sons 413–441.

Savage, Sam L. 2012. *The Flaw of Averages: Why We Underestimate Risk in the Face of Uncertainty.* Hoboken, NJ: Wiley.

U.S. Navy. 2009. U.S. *Navy Fact File: F/A-18 Hornet Strike Fighter.* 26 May. http://www.navy.mil/navydata/fact_display.asp?cid=1100&tid=1200&ct=1.

Wackerly, Dennis D., William Mendenhall III, and Richard L. Schaeffer. 2008. *Mathematical Statistics with Applications, 7th Ed.* Belmont, CA: Brooks/Cole.

Yildirim, U.Z. 1999. "Extending the State-of-the-Art for the COMAN/ATCAL methodology." Master's thesis, Naval Postgraduate School, Monterey, CA.

INITIAL DISTRIBUTION LIST

1. Defense Technical Information Center
   Ft. Belvoir, Virginia

2. Dudley Knox Library
   Naval Postgraduate School
   Monterey, California