



NAVAL POSTGRADUATE SCHOOL

MONTEREY, CALIFORNIA

THESIS

LEVERAGING PREDICTIVE ANALYTICS TO ASSESS 7TH FLEET SUSTAINMENT

by

Edwin J. Stevens

June 2020

Thesis Advisor:

Jeffrey A. Appleget

Co-Advisor:

Ying Zhao

Second Reader:

Douglas J. MacKinnon

Approved for public release. Distribution is unlimited.

THIS PAGE INTENTIONALLY LEFT BLANK

REPORT DOCUMENTATION PAGE			<i>Form Approved OMB No. 0704-0188</i>	
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 June 2020		3. REPORT TYPE AND DATES COVERED Master's thesis
4. TITLE AND SUBTITLE LEVERAGING PREDICTIVE ANALYTICS TO ASSESS 7TH FLEET SUSTAINMENT			5. FUNDING NUMBERS	
6. AUTHOR(S) Edwin J. Stevens				
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) N4			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.				
12a. DISTRIBUTION / AVAILABILITY STATEMENT Approved for public release. Distribution is unlimited.			12b. DISTRIBUTION CODE A	
13. ABSTRACT (maximum 200 words) Increasing operational availability through logistics readiness is critical to ensure maximum safety and capability of forward-deployed ships. The increased operational tempo that forward-deployed ships encounter causes high-impact failures that affect operational availability and the commander's decision-making process. As forward-deployed 7th Fleet ships encounter casualties, unacceptable operational impacts such as out-of-stock or "redlines" occur. Repair rates may be improved by leveraging predictive analytics to understand failure data. Critical systems such as propulsion, communications, or radar are our primary focus. Areas for optimization are identified through statistical analysis of data that we have collected from 7th Fleet organizations such as NAVSUP, NAVSUP WSS, NAVSEA, CTF-73, and SURFGUWESTPAC. We analyzed and sorted the data to identify any trends and risks that may not currently be tracked. We assessed distributions of potential failure rates, developed a peacetime model using Lexical Link Analysis, and then analyzed used JMP. Using our model, potential situations are built to analyze the output and optimize wait times within the constraints of manpower, funding, and availability. Future work using this model may include efforts to evolve to Phase II operations, with increased naval presence resulting in increased sustainability needs, simulated in SIMIO to test casualty and repair rates, as is shown in sample models in this thesis.				
14. SUBJECT TERMS global distribution chain, supply, logistics, data sciences, pull and push logistics, sustainment, demand prediction, route optimization			15. NUMBER OF PAGES 77	
			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	

THIS PAGE INTENTIONALLY LEFT BLANK

Approved for public release. Distribution is unlimited.

**LEVERAGING PREDICTIVE ANALYTICS TO ASSESS 7TH FLEET
SUSTAINMENT**

Edwin J. Stevens
Lieutenant Commander, United States Navy
BA, Linfield College, 2005

Submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE IN OPERATIONS RESEARCH

from the

**NAVAL POSTGRADUATE SCHOOL
June 2020**

Approved by: Jeffrey A. Appleget
Advisor

Ying Zhao
Co-Advisor

Douglas J. MacKinnon
Second Reader

W. Matthew Carlyle
Chair, Department of Operations Research

THIS PAGE INTENTIONALLY LEFT BLANK

ABSTRACT

Increasing operational availability through logistics readiness is critical to ensure maximum safety and capability of forward-deployed ships. The increased operational tempo that forward-deployed ships encounter causes high-impact failures that affect operational availability and the commander's decision-making process. As forward-deployed 7th Fleet ships encounter casualties, unacceptable operational impacts such as out-of-stock or "redlines" occur. Repair rates may be improved by leveraging predictive analytics to understand failure data. Critical systems such as propulsion, communications, or radar are our primary focus. Areas for optimization are identified through statistical analysis of data that we have collected from 7th Fleet organizations such as NAVSUP, NAVSUP WSS, NAVSEA, CTF-73, and SURFGRUWESTPAC. We analyzed and sorted the data to identify any trends and risks that may not currently be tracked. We assessed distributions of potential failure rates, developed a peacetime model using Lexical Link Analysis, and then analyzed used JMP. Using our model, potential situations are built to analyze the output and optimize wait times within the constraints of manpower, funding, and availability. Future work using this model may include efforts to evolve to Phase II operations, with increased naval presence resulting in increased sustainability needs, simulated in SIMIO to test casualty and repair rates, as is shown in sample models in this thesis.

THIS PAGE INTENTIONALLY LEFT BLANK

TABLE OF CONTENTS

I.	INTRODUCTION.....	1
A.	BACKGROUND.....	1
B.	CHALLENGES.....	3
II.	LITERATURE REVIEW	7
A.	PRIOR TOPICS REVIEWED	8
1.	Thesis Review of Social Network Collaborations in Haiti (Reid 2011).....	8
2.	Benefits of Haiti Study for this Thesis	8
3.	Thesis Review of Operational Readiness Impacts with Additive Manufacturing (Lopez 2019).....	9
4.	Benefits of Topic for this Thesis	10
5.	Using a Design of Experiments to Analyze Performance- Based Logistics (Kang, Sanchez, and Doerr 2006).....	11
6.	Benefits of Topic for this Thesis	12
III.	MODEL, METHODOLOGY, AND TOOLS.....	13
A.	SOURCES OF DATA.....	13
B.	TOOLS.....	19
1.	R.....	19
2.	Lexical Link Analysis (LLA).....	20
3.	JMP	22
4.	Additional Tools and Calculations Involved	24
C.	METHODOLOGY	28
IV.	ANALYSIS	33
A.	LLA EXPERIMENTS.....	33
1.	Demand Data.....	33
2.	Failure Data	36
B.	COMPARING LLA TO PRISM AND HISTORICAL DATA	43
V.	CONCLUSIONS AND FUTURE WORK.....	51
A.	CONCLUSIONS	51
B.	FUTURE WORK.....	52
	LIST OF REFERENCES.....	55
	INITIAL DISTRIBUTION LIST	57

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF FIGURES

Figure 1.	Map with overlay showing the numbered U.S. fleet areas of responsibility (AORs). Source: World Map with Nations 2009.....	2
Figure 2.	Chart showing the different phases of an operational plan. Source: Joint Chiefs of Staff 2017.	5
Figure 3.	Output from different additional manufacturing capabilities and their respective benefits to AO. Source: Lopez 2019.....	10
Figure 4.	Flowchart showing the steps to be used to predict potential redline demands.	13
Figure 5.	Plot showing demand rates for IMEC 3, 4, and 5 NSNs.	15
Figure 6.	Screenshot showing initial R based dataset after being exported to a CSV file.	16
Figure 7.	Screenshot showing a sample of CASREP data after removing non-pertinent information.	17
Figure 8.	Sample output of PRiSM analysis of MRDB Data.....	18
Figure 9.	Image showing LLA working to find correlations between dataset entries used in this thesis.....	20
Figure 10.	Sample figure showing correlation with the word “energy” in LLA.....	21
Figure 11.	Correlation of independent variables to the expected number of failures of any given NSN studied.	23
Figure 12.	Sample image of prediction profiler to test the effects of changes in the independent variables in a regression.	24
Figure 13.	PRiSM presentation to NAVWAR after test deployment on BOXER ARG ((Naval Information Warfare Systems Command (NAVWAR) 2020).	25
Figure 14.	Formula for AO.....	26
Figure 15.	Basic sawtooth graph to explain reorder points.....	26
Figure 16.	Phases of part selection and resourcing using PRiSM. Source: Naval Information Warfare Systems Command (NAVWAR) 2020.....	27

Figure 17.	Calculations used in LLA to determine Probability, Lift, Difference, and Correlation. Source: Quantum Intelligence 2015.....	30
Figure 18.	Sample data after being prepped for use in LLA.	31
Figure 19.	LLA GUI used to visualize and manipulate data showing various correlations.....	31
Figure 20.	R Screenshot showing a summary of the first dataset obtained for analysis.....	34
Figure 21.	Screenshot of the first experiments run in LLA.....	35
Figure 22.	Example of high correlation found among orders in failure data.	37
Figure 23.	A graphic illustrating all connections in the merged dataset for this experiment.....	38
Figure 24.	Illustration showing the correlations between the failure of 013926982, 013228417, and various other parts.	39
Figure 25.	A search for correlations with a high lift related to 014639935.	40
Figure 26.	Another finding with high lift between 013722818 and its associated parts.....	41
Figure 27.	NSN 013833385 and its associations, which had the highest lift of the experiment.....	42
Figure 28.	Average downtime caused by lack of onboard supply	45
Figure 29.	Total days lost per NSN by lack of onboard supply	46
Figure 30.	Comparison of causes of downtime by NSN.....	46
Figure 31.	Improvement by NSN across the potential levels of supply available	47
Figure 32.	AO improvement comparing PM recommendation to the analysis outcome.....	48
Figure 33.	Chart showing AO improvement considering worst case MTBM	49
Figure 34.	SIMIO model running and showing sawtooth plot created by spikes in demand causing backlog.....	54

LIST OF TABLES

Table 1.	Number of demands, P values, impact, and likelihood of failure for the 11 matching NSNs between LLA and PRiSM analysis.....	44
Table 2.	Demand rates and processing rates for 7F supply	53

THIS PAGE INTENTIONALLY LEFT BLANK

LIST OF ACRONYMS AND ABBREVIATIONS

5F	Fifth Fleet
7F	Seventh Fleet
AO	Operational Availability
AOR	Areas of Responsibility
OPTEMPO	Operational Tempo
C4I	Command, Control, Computers, Communications, and Intelligence
CG	Guided Missile Cruiser
CIWS	Close-in Weapon System
CLA	Collaborative Learning Agents
CSV	Comma Separated Value
DLA	Defense Logistics Agency
GUI	Graphic User Interface
HADR	Humanitarian Assistance and Disaster Relief
IMEC	Item Mission Essentiality Code
LLA	Lexical Link Analysis
MDT	Mean Down Time
MRDB	Material Readiness Database
MTBF	Mean Time Between Failure
MTBM	Mean Time Between Maintenance
NAVSUP	Naval Supply Systems Command
NAVSUP WSS	Naval Supply Systems Command, Weapon System Support
NAVWAR	Naval Information Warfare Systems Command
N-ERP	Navy Enterprise Resource Planning System
NPS	U.S. Naval Postgraduate School
NSN	National Stock Number
NSWC	Naval Surface Warfare Center
PM	Program Manager
PRiSM	Navy's predictive logistics and predictive risk sparing matrix
PUK	Pack-Up Kits
RBS	Readiness Based Sparing
UICs	Unit Identification Codes

THIS PAGE INTENTIONALLY LEFT BLANK

EXECUTIVE SUMMARY

The U.S. Navy Supply Corps has been performing logistics at sea since 1795 – during times of war and peace. Over the years, new techniques were developed to improve at sea sustainment as technology improved. This included refueling at sea which began for the U.S. in World War 1 and continues to this day to give the U.S. Navy a notable advantage over adversaries with regard to on station time and operational availability.

Challenges occur as the technology race and the global power competition continues to accelerate. Modern techniques such as *just in time* logistics, supply chain optimization, and predictive analytics help logisticians to determine increasingly improved support infrastructure, while trying to maintain 100% operational availability. Weapons systems have improved with technology, but also become much more complex. There are hundreds of thousands of parts in any given combat system, making it very difficult to track failures and assess which parts should be carried in which location in order to minimize downtime due to failures.

This thesis follows the path of many research projects in the past, in the sense that it searches for ways to improve bottom line operational availability, with the potential additional benefit of operating more efficiently. While some research examines methods to improve operational availability by manufacturing parts onboard ships, and others search for methods to increase reliability, our effort searches for methods of predicting the reliability of parts and the probability of failure on a forward-deployed ship using data mining methodology leveraging historical failure and demand data.

Challenges arise in data mining because there are so many systems that the military uses to track various parts on various ships. Sometimes the data is incomplete and sometimes there is redundant and inapplicable data that is obtained. During the research for this thesis, it was found that Lexical Link Analysis (LLA) can be used to search for commonalities in failure data after the dataset is prepared by cleaning and merging information that is available across the spectrum of systems that are tracked. Data can include ship name and hull number, ship type, ship location, part number, system name,

time the part was in service, time to repair the part, criticality of impact when the part fails, where and how many repair parts were available, and various engineering specifications such as expected service time of the part, just to name a few. By mining the data, preparing it, and then using LLA for big data analysis, many combinations of failures can be identified that are apparent once identified, yet in the massive amount of information considered, many issues and connections might not have been noticed before. By identifying trends and predicting failures, parts can be accurately assessed and placed on ships, or in forward locations, for faster repair and less downtime due to failure, therefore increasing operational availability.

To test this, first NAVSUP failure data was analyzed, then CASREP demand data was analyzed, and datasets were merged with multiple other datasets from the maintenance readiness database (MRDB). These were compared and contrasted against other predictive analytics tools currently being developed such as the Predictive Risk Sparing Matrix (PRiSM). When deriving comparison p-values of the datasets against each other, it was found that there were very high similarities between them. This could be expected, but with the differences noted, it helped to validate that LLA was performing as expected. During a test run of PRiSM, analyzed parts on the USS BOXER Amphibious Readiness Group, 34 of 52 high impact C4I parts that were now identified, had actually failed. This would have raised operational availability to 87%, or 2% higher than the requirement.

When comparing PRiSM with Lexical Link Analysis (LLA) output, failures can be validated, and a larger selection of data can be reviewed since LLA uses data mining and can analyze large datasets. As shown in the research and the findings, there have been valuable tests of LLA when working with PRiSM as a validation tool, or as a tool to help identify additional parts that may fail. This thesis project shows the applicability of this big data methodology in researching historical data to predict failures, and also shows that it is possible to use those predictions to test the possible impact of increased stress on systems when the operational tempo is raised. Finally, this method can be used to test the supply systems' ability to keep pace with increased demand so that weaknesses can be assessed and resolved.

ACKNOWLEDGMENTS

First and foremost, I would like to thank Professor Jeffrey Appleget and Professor Ying Zhao for their outstanding support, guidance, and patience during this thesis process. Professor Appleget's help in designing the thesis layout, advice in writing, and recommendations for addition or removal has been instrumental to its success. Professor Zhao's guidance in coding and data mining was invaluable, and her natural understanding of what I was trying to accomplish was often even better than my own when things got difficult. I would also like to thank Professor Doug Mackinnon and Professor Bonnie Johnson for their advice and direction whenever it was needed. I'd like to give special thanks my Operations Research study group; Shuchang Liu, Kuo-Wei Yao, Alex Gramann, Alex Amper, Win Thaw, and Pete Rivera. I could not have gotten through the program without them. Finally, I would not have been able to get through the toughest times without the love and support of my wife, Yujung and my daughter, Ari.

THIS PAGE INTENTIONALLY LEFT BLANK

I. INTRODUCTION

The U.S. Navy Supply Corps has been performing logistics at sea since 1795 – during times of war and peace. Over the years, new techniques were developed to improve at sea sustainment as technology improved. This included refueling at sea which began for the U.S. in World War 1 and continues to this day to give the U.S. Navy a notable advantage over adversaries with regard to on station time and operational availability. There is a famous idiom originating from 1870s news articles in the United States that speaks of “the tail wagging the dog” (“Tail Wagging the Dog” n.d.). This idiom can be applied very directly to the U.S. Navy, stating “don’t let the *logistical* tail wag the *operational* dog.” This is a profound use of the idiom, and it applies directly to the mission of the U.S. Navy’s logistics operations. In naval terminology, the goal is to attain an operational availability (AO) of 100% at the tactical level. When AO is 100%, operational level commanders can task their units without encumbrance due to loss of critical mission capabilities. This thesis aims to explore ways to improve the U.S. Navy’s ways of predicting demand, in order to improve AO.

A. BACKGROUND

The challenge of improving AO is particularly of interest in forward locations such as Seventh Fleet (7F, Western Pacific) and Fifth Fleet (5F, Middle East), shown in Figure 1.

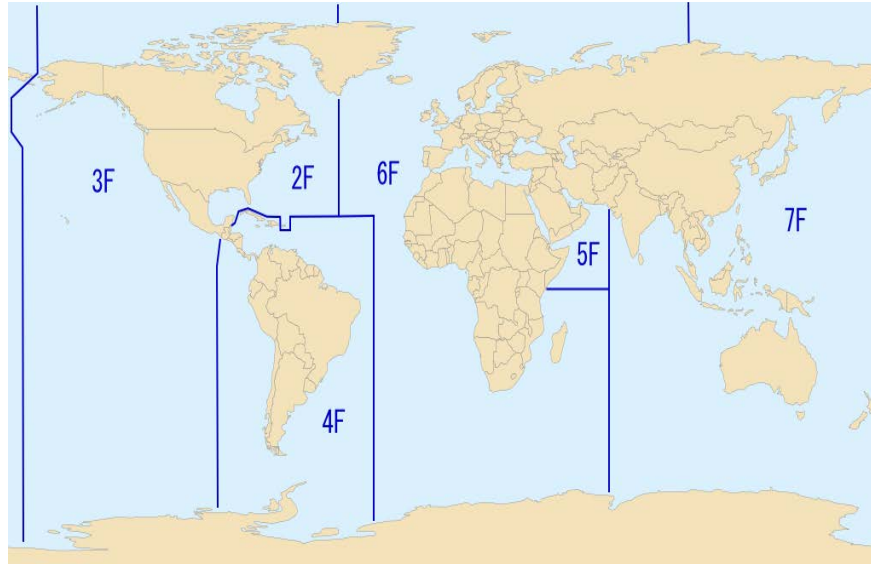


Figure 1. Map with overlay showing the numbered U.S. fleet areas of responsibility (AORs). Source: World Map with Nations 2009.

Due to extended logistical lines, increased operational tempo (OPTEMPO), and different environmental conditions, logisticians have constant challenges in providing logistical support when deployed units have part failures or other supply demands. While this thesis focuses primarily on 7F data, the techniques used can be applied to any AOR.

The highest priority challenges arise when a high impact part fails and a replacement is not immediately available. When this happens, it can cause a “redline,” or a failure that stops the unit affected from being able to complete its mission until the failure is resolved. The challenge is compounded when units are forward deployed, because in potentially adversarial situations an operational unit needs to be 100% mission capable. Even if a unit has been certified to deploy, new failures in certain mission areas overseas will stop the mission, for example if a guided missile cruiser (CG) has a failure in its close-in weapon system (CIWS), it will be unable to continue escorting the carrier in its strike group. If this is in an area like the Taiwan Strait, this redline failure could stop the entire strike group from transiting. In the CIWS system alone there are thousands of parts which could cause a redline failure if they stop functioning and a replacement is not available.

B. CHALLENGES

The constraining factors of manpower, funding, and available resources are all in high demand so a good system needs to be in place to determine the most efficient and effective method of stocking, forward staging, or procuring the repair parts which may be required. Since even one CG has hundreds of thousands of repair parts, thousands of which could cause a redline upon breaking, it is of critical importance that the supply system is as intelligently designed as possible. The AO of deployed ships is tracked all the way up to the flag officer level, and if the AO drops below mandatory requirements, that unit will be “in the spotlight” until it is back to 100%, or an AO of 1.0. To achieve this difficult goal, high priority failures must be repaired quickly and with minimal disruption to mission and schedule.

There are currently government service employees who work as liaisons in forward deployed locations, whose job is to identify and track critical parts. They must work with DOD organizations such as Defense Logistics Agency (DLA) and Naval Supply Systems Command (NAVSUP) to find ways to improve unit reliability. Currently, the supply system assumes that every demand that comes in is legitimate and stocking levels are based on the number of demands of a specific part, tracked by National Stock Number (NSN). Since there are so many parts that could fail, it is difficult to accurately ascertain which spare parts should be stocked in a warehouse in forward location, which ones should be stocked in the United States, and which ones should be carried onboard the individual unit. An apparently easy solution would be to simply carry every part that could cause a redline, but there is neither enough room to carry these parts nor is there enough funding to pay for them all. Under the Navy’s just-in-time logistics plan, some depot level repairable parts are left on the shelf at the repair facility in a non-operational status and are only repaired to be put back into use if a demand is registered. These repair times can cause significant delays. The reason for this system is that in a large scale operation it is more economical to fix or buy the part just-in-time than it is to use limited funds for parts which may just sit on the shelf, when the funding could be used instead for a more critical immediate requirement. In some low demand high impact cases, such as the redline parts that this thesis looks to study, it may be better to risk the potential unnecessary expenditure to have working spares

immediately available when needed. The current stocking model is good at selecting high failure rate parts, but fails to identify low failure rate, high impact parts.

To add to potential challenges with redline parts is the uncertainty that comes with military operations. The Navy has been operating at “Phase 0,” or general theater shaping peacetime operations, since the end of the Cold War, where our military advantage is unchallenged, meaning the defensive posture is setup for relatively low risk compared to war time (Mattis 2018). While there are sporadic real-life incidents that occasionally happen, and units do train for the worst-case scenario, the Navy hasn’t seen all out ship-to-ship combat since the end of World War II.

Particularly when operating in 7F, there is a large threat of asymmetric warfare, peer and near peer naval forces, and a potential need to use Navy ships to defend allies. If the United States Navy was required to shift to Phase 1 (deterrence efforts), Phase 2 (efforts to seize the initiative in the combat zone), or even higher, the logistics system would be strained in a way that it has not been for many years. It is unknown how well the system would be able to support the effort of combat operations. A chart showing the notional phases of an operational plan is shown in Figure 2.

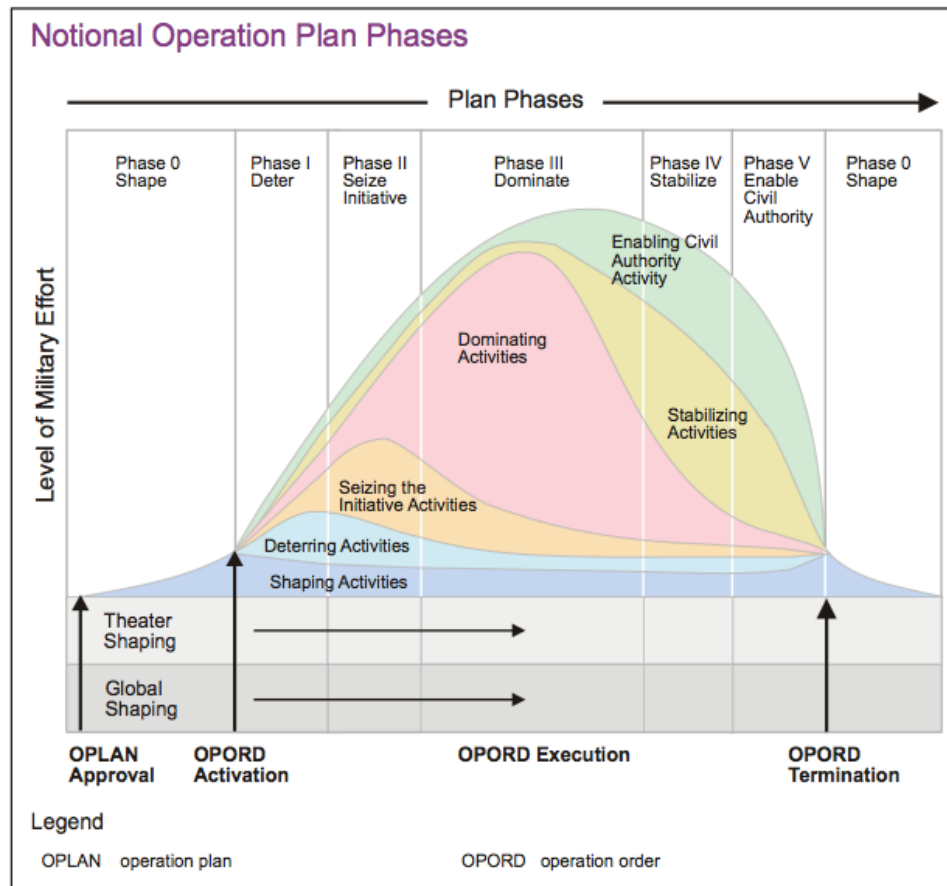


Figure 2. Chart showing the different phases of an operational plan. Source: Joint Chiefs of Staff 2017.

While how well the current system will work under a higher OPTEMPO, as units use their systems more often and in ways that may cause them to fail more, it is known that the current forecasting and stocking methods will not be sufficient to support the fleet's needs and to keep an AO of 1.0. With a more comprehensive model in place to predict current Phase 0 logistics, historical data and engineering specifications can be used to predict increased failure rates with increased usage, and the model could be used to predict needs as OPTEMPO increases. New methods studied in this thesis will use decision aids such as lexical link analysis (LLA) to better estimate potential demands. This topic is in line with issues of importance to the fleet, and although it is specific to 7F, it could be modified to work in other environments as well.

This thesis uses coding languages such as R and Perl with LLA and statistical analysis tools such as JMP and Microsoft Excel to better forecast required low failure rate, high impact parts that can become necessary during deployment to increase AO for the fleet. It can also help to improve AO without a substantial impact on the manpower, funding, and resource availability restrictions that exist.

For the following chapters, Chapter II will explain the literature review performed in preparation for this thesis, Chapter III will describe the tools that will be used, the methodology, and the model used, Chapter IV will be the analysis of the output from the model, and Chapter V will be the conclusion and future work.

II. LITERATURE REVIEW

The importance of optimizing the logistics infrastructure is not a new concept but coming up with solutions can quickly become very complex. In modern times, with many high-tech tools keeping logistics flowing more efficiently than ever before, some warfighters can think that logistics is just about keeping the parts supplied and doing so in a way that saves money when possible without impacting the mission. While that is the visible result at the tactical level, logistics is also important at the operational and strategic levels of war. Many great strategists and warfighters in history have emphasized superior logistics as a path to winning wars. Sun Tzu stated, “The line between disorder and order lies in logistics” (Sun Tzu, n.d.). Carl Von Clausewitz also stated, “The end for which a soldier is recruited, clothed, armed, and trained, the whole objective of his sleeping, eating, drinking, and marching is simply that he should fight at the right place and the right time” (Clausewitz 1989).

The idea that logistics is one of the key factors in warfighting is as important in modern military practice as it ever has been. Within the last 100 years, there are also many phrases from great students of war that can be attributed to the importance of logistics. Admiral Lynde McCormick, first Supreme Allied Commander of NATO forces in the Atlantic stated that “logistics is all of war-making, except shooting the guns, releasing the bombs, and firing the torpedoes” (Furer 1960), Fleet Admiral Ernest King stated “I don’t know what the hell this logistics is that Marshall is always talking about, but I want some of it” (Kuehn 2014), and then the year after WW2 ended he attributed the victory to the U.S. logistics advantage, and General Robert H. Barrow, Commandant of the Marine Corps in 1980, has stated “Amateurs talk about strategy and tactics. Professionals talk about logistics and sustainability in warfare” (Barrow 1979).

The importance of logistics can be studied throughout history, emphasizing that no matter how good the system is, a logistician’s purpose is to try to work himself or herself out of the job, making the support good enough that the end result for the warfighter is seamless functionality, and the end result for the commander is operational availability of 1.0, meaning his or her decisions are not affected by broken equipment.

A. PRIOR TOPICS REVIEWED

Many prior thesis topics at Naval Postgraduate School (NPS) have recognized that there is opportunity to improve AO by implementing various ideas and methods. Many prior studies have also utilized LLA as a method to identify trends in data. This thesis however is the first time that LLA has been used in a quantitative manner to study large data sources, and the first time using the output as verification against current systems being developed such as the Navy's predictive logistics and predictive risk sparing matrix (PRiSM). Some prior NPS thesis topics that were studied are as follows;

1. Thesis Review of Social Network Collaborations in Haiti (Reid 2011)

The thesis topic "Social Network Collaboration for Crisis Response Operations: Developing a Situational Awareness (SA) Tool to Improve Haiti's Interagency Relief Efforts" was the first topic reviewed related to LLA, and the topic that showed LLA's potential use in predictive analytics for logistics. This thesis used LLA to perform data mining on resource allocation in the Haiti disaster, and revealed trends in organization themes. The thesis team was able to perform data mining on social networking websites in order to focus in on a specific time frame within the disaster and develop correlations between the trends that were happening. The end result of the data mining and analysis was that if one can find a trend and visualize the organizational dynamic, better predictions can be made to improve situational awareness during the response to a disaster. With better situational awareness, one can better understand the situation as it is developing and can be better prepared to predict and respond to possible outcomes.

2. Benefits of Haiti Study for this Thesis

While the data that was mined in the Haiti study was strictly text, it was discovered that the same techniques could be used to study demand data gathered directly from NAVSUP. To use LLA, the user writes script in a coding language such as R or Perl to do some basic data manipulation in order to prepare the data to be mined by LLA. In this case, datasets were merged, and matching part numbers or document numbers were collocated. Useful data columns were kept, extraneous ones were removed, and in some cases new columns were created to record calculated information such as downtime caused by a

failure (or a demand, depending on the dataset). The user can determine what types of “baskets,” or groups of data that match specific parameters which LLA should try to correlate. In the case of this thesis, baskets were organized by unit and date. Instead of looking for correlations among groups of words, the data was prepared for LLA to look for collaborations by part number and nomenclature. LLA searched all of the data imported and looked for correlation between parts ordered on a certain ship and on a certain date. LLA then listed the number of times specific parts that were ordered at the same time as other parts. Multiple experiments were run, changing the parameters and the resolution required for the baskets, such as changing the date from a year to a quarter to a week to a day. Next, the output was explored, and it was determined that the correlations could be useful in making predictions. More of this process will be explained in Chapters III and IV.

3. Thesis Review of Operational Readiness Impacts with Additive Manufacturing (Lopez 2019)

Another area of interest while searching through various prior topics was to see what entirely new directions researchers may have gone in an attempt to improve AO. One topic found was “Operational and Mission Readiness Impact of Additive Manufacturing Deployment in the Navy Supply Chain” completed by Jose A. Lopez II in September 2019. The goal of the thesis was to explore the impact on AO, total cost, and mission availability by giving ships and forward located repair facilities added manufacturing capabilities. This could be an item like a 3-D printer on a ship, a CNC machine, or an intermediate repair facility which would be in CONUS in most cases. The thesis discussed meeting the Chief of Naval Operations’ core tenet of improved lethality for the Navy, and searched for ways to improve lethality by improving our ability to operate forward faster and longer without needing a larger logistics infrastructure. It then went into the goal of also being able to do it cheaper or without substantial cost increases. The author experimented with various designs to compare the base case with results from potential model designs as follows;

<i>Metric</i>	<i>Scenario #1 Quantity</i>	<i>Scenario #2 Quantity</i>	<i>Scenario #3 Quantity</i>
Total Operating Cost	\$105,414.92	\$25,299.55	\$66,043.10
Operational Availability	.49	.79	.75
Mission Availability	.75	.93	.92

Figure 3. Output from different additional manufacturing capabilities and their respective benefits to AO. Source: Lopez 2019.

In this study, The author concluded that by absorbing the increased initial costs, additional manufacturing capabilities such as advanced 3-D printers provide a vast improvement over the current system in which ships have a small machine shop where sailors can manufacture a very limited set of new parts. This is especially true for smaller ships like Cruisers and Destroyers, while Patrol Craft and Minesweepers have zero capability to manufacture their own parts. The additive manufacturing thesis showed the benefits of accepting innovation and willingness to risk the up-front cost for the potential increased capability in the future.

4. Benefits of Topic for this Thesis

While this thesis looks to use predictive analytics to improve the logistics infrastructure by stocking parts in a smarter way, the additive manufacturing research introduced a capability to produce new parts onboard a ship. This review helped to transform this topic into what it has become. Originally this topic was going to use a coding language to clean NAVSUP data to determine average shipping times in the current Phase 0 operating environment, and then increase demand and simulate how the supply system could keep up with the increased strain. While initial simulations were created and ran, much difficulty was encountered in how to increase strain on the system. Operational environment, operational tempo, outside influences such as attacks or tests on specific systems, and human error can all affect different parts in different ways. Much trial and error as well as study on individual engineering specifications would be required and

separate simulations would be needed for separate parts under separate conditions. Since the goal of this thesis is to leverage predictive analytics to improve 7F sustainment, Lopez's thesis helped direct the bigger picture direction that this thesis takes. Both topics seem to have realized that using broad techniques, areas for improvement in logistics can be found. More of how the bigger picture direction can be useful and how it can be applied to specific parts is discussed further in Chapters III and IV.

5. Using a Design of Experiments to Analyze Performance-Based Logistics (Kang, Sanchez, and Doerr 2006)

Another topic studied comes from the Operations Research department at NPS. The topic is "A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics," written by Dr. Keebom Kang, Dr. Susan Sanchez, and Dr. Kenneth Doerr in September 2006. This topic was discovered when searching for methods that had previously been studied to improve AO of critical systems on ships. The authors focused specifically on improving the average AO of the weapon systems, and on reducing the probability that the weapon system will fail. In order to do this, the study "designed experiments to evaluate simulation scenarios in an intelligent way" (Abstract, page i).

Both of the goals listed above are very similar to this predictive analytics thesis in that the goal is to reduce AO and identify parts that will cause redline failures. The design of experiments paper mentions developing "a simulation model to aid in identifying and evaluating promising alternatives to achieve improvements in weapon system-level availability when outsourcing logistics services for system components." (Kang, Sanchez, and Doerr 2006) This predictive analytics thesis initially started with the intent to develop a complex simulation but eventually changed direction towards the bigger picture attempt at finding correlation in demand to predict potential future demands. Simulation was used in the design of experiments study to focus on improving performance-based contracts by minimizing risk and cost while maximizing useful output from the contract. Interestingly, when this paper was written in 2006, Performance-Based Logistics was a relatively new term, but today it is commonplace and widely used in DOD contracts.

6. Benefits of Topic for this Thesis

The design of experiments thesis topic brought out the idea of using simulation modelling techniques to predict the effects of increased OPTEMPO on the supply system. Simulation is the best way to test these effects, but it was discovered that prior to simulation, much work must be done to accurately predict individual parts that are important to focus on, how the increased OPTEMPO affects them, and how they will affect the AO of the ship. The goal of this predictive analytics thesis is to setup the model to use basic simulation techniques, which have been designed, but to develop data and a model that can be used to focus on increasing AO in peacetime before testing for wartime.

The topics listed above all helped to shape the direction of this study involving predictive analytics to assess and improve 7F logistics, and they have helped to widen the horizons on what work has already been done. This enables the research in this topic to build upon work that has already done and possibly to develop new techniques to study big data with the result of improving AO in the fleet. The next chapter will focus on the tools used, the methodology developed, and the model built to study the supply system data that has been obtained.

III. MODEL, METHODOLOGY, AND TOOLS

This thesis uses a methodology that involves multiple tools used together to first source demand data, then review, clean, and organize it, then mine the data and select points of interest. Next, those points of interest are separated from the original dataset and various analysis techniques are used to determine where specific parts should be stocked and what improvement there will be if the recommendations are followed. Below is a flowchart that summarizes the methodology used with the various tools and modelling techniques which will be described in more detail in the following paragraphs.

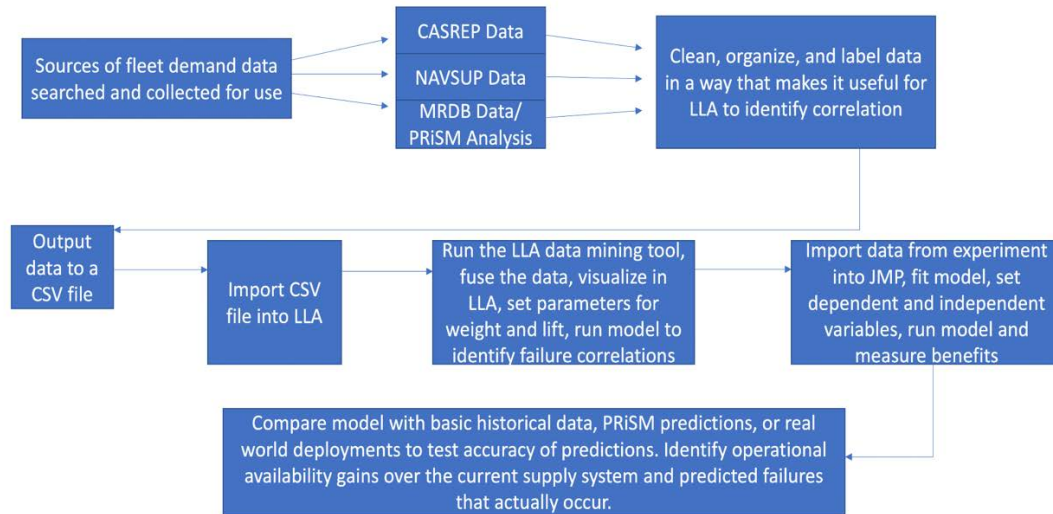


Figure 4. Flowchart showing the steps to be used to predict potential redline demands.

A. SOURCES OF DATA

Over 1,000,000 items of current demand and failure data have been collected from three different datasets. These datasets are in the form of (1) demand data from Naval Supply Systems Command, Weapon System Support (NAVSUP WSS) and (2) Casualty Report (CASREP) data, as well as (3) data output from the Material Readiness Database

(MRDB) which has been analyzed using PRiSM. Each of those three datasets will be described.

Initially, demand data was requested from NAVSUP WSS for analysis. This dataset came in an R format and included parts from 18 Unit Identification Codes (UICs) from different 7F ships from 2011 to 2019. Variables include NSN, Item Mission Essentiality Code (IMEC), line item number, location processed, quarter processed, and number demanded. The dataset included each part number 18 times, once for each UIC, regardless of whether a demand was registered. This meant about 1.2 million rows of data. This data was useful because it was uniform across all UICs by showing one row per quarter per ship, making it useful to verify how many demands a specific NSN had even if it was zero. The issue is that the data does not include UIC per part. Script had to be written to run through each set of parts and add the UICs as an additional column, and assumptions made that each set of 18 NSNs was in order by UIC throughout the dataset. Additionally, while demand by quarter was useful for a low-resolution exploratory approach, as the model was explored, higher resolution data was desired. Additionally, demand data does not necessarily equate to failure or CASREP data. While the ships' crew should be uploading information about failures even if remedied onboard, that isn't always the case and failures that were remedied without registering a demand to NAVSUP were not included in this dataset. This may make some of the demands look more important than they are in the big picture due to them being the only data points available.

The dataset was delineated into the most important parts, or the highest IMECs only. Failure of an IMEC 3 part causes potential loss of redundancy to a critical system, IMEC 4 is a single point of failure for the critical system, and IMEC 5 will cause the loss of the critical system and possible loss of life. Figure 5 shows a plot of the original dataset after analyzing it with R, showing the spread of demands by IMEC 3, 4, and 5.

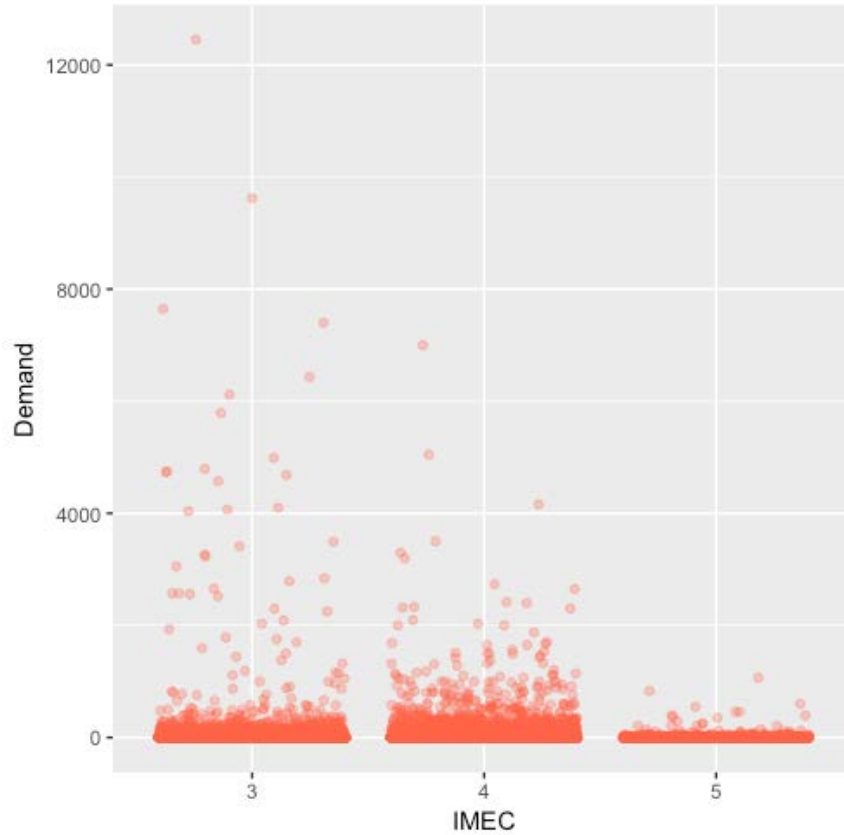


Figure 5. Plot showing demand rates for IMEC 3, 4, and 5 NSNs.

In this dataset there were some outliers, with the max value being 12,353 demands over the nine-year period for an IMEC 3 part. Most demands however only had a few demands over the entire nine-year period covered, with the mean being 2.397 demands over 1,200,000 demands for 18 ships. The majority of parts actually had a demand of zero, with the overall median of the dataset being zero. This causes challenges in determining which parts may fail next, adding to the need for a better analysis of potential future failures. Adding to the challenge is a need to follow proper ordering and stocking procedures. Many repair parts will be acquired through trades or stockpiles that have been accrued over time, and therefore the new failures and demands are under the table and not tracked in the supply system. Figure 6 shows the format and some entries in the dataset obtained.

Edwin_data

Line Item	Material (NSN)	IMEC	AAC	COG	LRC	LRC Tech	Site	Quarter	Demand Qty
2	75	4	J	7H	GWA	DD	M	20192	0
5	75	4	J	7H	GWA	DD	M	20163	0
10	75	4	J	7H	GWA	DD	M	20151	0
13	75	4	J	7H	GWA	DD	M	20174	0
17	75	4	J	7H	GWA	DD	M	20143	0
18	75	4	J	7H	GWA	DD	M	20173	0
19	75	4	J	7H	GWA	DD	M	20183	1
23	75	4	J	7H	GWA	DD	M	20181	0
24	75	4	J	7H	GWA	DD	M	20182	0
25	75	4	J	7H	GWA	DD	M	20171	0
28	75	4	J	7H	GWA	DD	M	20144	0
31	75	4	J	7H	GWA	DD	M	20164	0
35	75	4	J	7H	GWA	DD	M	20153	0
36	75	4	J	7H	GWA	DD	M	20154	0
37	75	4	J	7H	GWA	DD	M	20172	0
40	75	4	J	7H	GWA	DD	M	20191	0
41	75	4	J	7H	GWA	DD	M	20152	0
45	75	4	J	7H	GWA	DD	M	20184	0
46	75	4	J	7H	GWA	DD	M	20161	0
47	75	4	J	7H	GWA	DD	M	20162	0
51	11632	3	C	7H	APB	CC	M	20192	2
53	11632	3	C	7H	APB	CC	M	20153	3
55	11632	3	C	7H	APB	CC	M	20154	2
56	11632	3	C	7H	APB	CC	M	20144	4
57	11632	3	C	7H	APB	CC	M	20172	1
59	11632	3	C	7H	APB	CC	M	20173	3
60	11632	3	C	7H	APB	CC	M	20163	1
62	11632	3	C	7H	APB	CC	M	20184	1

Figure 6. Screenshot showing initial R based dataset after being exported to a CSV file.

For the second dataset, NAVSUP WSS was contacted again, and higher resolution data was requested. This time, all CASREP data was received. The information was still delineated by IMEC 3,4,5, but this time it was also listed by document number for the specific failure. This dataset included approximately 50 columns of information from the Navy Enterprise Resource Planning System (N-ERP) which NAVSUP monitors. Many of the columns were not applicable to this project so much data cleaning was necessary. For the first step columns like part location, CASREP data, material availability date, where the order came from (CONUS or OCONUS), ship name and UIC, document number, order quantity, net price, and supply type available were kept and organized. Also useful was the status of any available parts listed in the system as “A” (operational, waiting to be ordered),

“M” (currently being maintained or repaired), or “F” (on the shelf broken, awaiting repair). This dataset was very useful because it had data points that made it much better to break down into smaller baskets when looking for correlations among the data with LLA. The smaller and more specific the basket size, the more reliable the output will be. The problem with this data set is that while it was broken out as only IMEC 3, 4, and 5, it wasn’t clear which parts correlated to which IMEC, and while each failure had a date it wasn’t organized by quarter and by ship. It only registered a failure if it happened, so the data points needed much cleaning and organizing to be usable in a coherent way. Figure 7 shows a small selection of the CASREP dataset obtained, with the bold rows being parts that correlated showing potential failures in both LLA and PRiSM;

MATNR	Cognizance Code	Repair Indicator	pricing date	Requisitioner_UIC	purchase order type	Description
015863541	7H	Y	07/09/2011	22993	Z101	IMAGE INTENSIFIER,N
015863541	7H	Y	07/09/2011	22993	Z101	IMAGE INTENSIFIER,N
015813701	7H	Y	10/01/2012	21660	Z111	CIRCUIT CARD ASSEMB
015773917	7H	Y	10/01/2013	21660	Z111	CIRCUIT CARD ASSEMB
015740007	7G	Y	11/25/2011	21640	Z101	DISK DRIVE UNIT
015740007	7G	Y	11/25/2011	21640	Z101	DISK DRIVE UNIT
015740007	7G	Y	11/25/2011	21640	Z101	DISK DRIVE UNIT
015740007	7G	Y	11/25/2011	21640	Z101	DISK DRIVE UNIT
015466884	2J	Y	11/24/2011	21387	Z101	DISPLAY UNIT
015442493	7G	Y	05/28/2011	21686	Z101	COMPUTER SUBASSEMBL
015433991	1H	N	06/18/2011	21640	Z111	DISK DRIVE UNIT
015432853	7H	Y	05/07/2011	21560	Z101	CIRCUIT CARD ASSEMB
015369451	7H	Y	11/26/2013	21660	Z111	FAN,CENTRIFUGAL
015369449	7H	Y	10/01/2012	21660	Z111	HEALTH MONITOR UNIT
015369448	7H	Y	10/01/2012	21660	Z111	POWER SUPPLY
015327063	7H	Y	04/30/2011	3012A	Z101	MOTOR,AC 15 SHP
015323562	1H	N	10/04/2018	21943	Z111	JOYSTICK,DATA ENTRY
015310946	7G	Y	10/01/2012	21660	Z111	CIRCUIT CARD ASSEMB
015242629	7H	Y	10/01/2012	21660	Z111	DISK DRIVE UNIT
015206485	7H	Y	10/01/2012	21660	Z111	CIRCUIT CARD ASSEMB
015180145	7G	Y	10/01/2012	21660	Z111	CIRCUIT CARD ASSEMB
015160829	1H	N	03/02/2010	21387	Z111	TRANSMITTER,RADIO
015160829	1H	N	03/02/2010	21531	Z111	TRANSMITTER,RADIO

Figure 7. Screenshot showing a sample of CASREP data after removing non-pertinent information.

The final dataset obtained for use in this thesis comes from the PRiSM analysis of MRDB data. This data source was stumbled upon when networking with sponsors and various agencies that might be interested in the research as part of the literature review process. PRiSM uses extremely high-resolution data and algorithms to calculate useful information, but only has one year of data to analyze due to the program still being in the

experimental phase. The PRiSM program also only received data for command, control, computers, communications, and intelligence (C4I) systems as opposed to all critical systems of IMEC 3, 4, and 5 in the other two datasets. These systems are included in the overall critical systems used in this thesis, but the other data covered nine years instead of just one year. PRiSM selects parts by using a complex system of algorithms with multiple inputs to determine time in system for the studied parts, mean time between failure (MTBF) for the parts, impact of the failure, where the parts are and how many of them are available, as well as other parameters that overlap with the other two datasets. As opposed to being organized by document number or quarter and IMEC, the PRiSM data is organized by specific part with all pertinent information available, so it is at an even higher resolution than the CASREP dataset obtained. To collect data on each part requires engineering specs, historical data, and much information from the program office to predict probability of failure. PRiSM has had high success rates in its' initial uses, specifically on a test run with a USS Theodore Roosevelt Carrier Strike Group deployment to 7F. This gives good real-life test data to compare the prediction results from this thesis with. During a meeting with NSWC Corona and Naval Information Warfare Systems Command (NAVWAR), there was much interest expressed in the ability to merge these datasets and collaborate with LLA to cover a larger time period with a wider range of parts with less comprehensive data collection required to come up with potential results. Figure 8 shows a sample of the PRiSM output collected, with some of the columns showing the valuable information it provides;

System Name	Hull Number	Activity Receiving Part	Block Number	Part Number	Part Nomenclature	Impact to System
AN/SSQ-130: OVERALL	DDG 91	NAVJUP/FDNR	AS-3202 Antenna	AS-3202/SRD-19	ANTENNA HP OF	Reduced Redundancy
AN/SSQ-137: SSQ INC 1 Fleetwide	CG 52	CG 52	9102 Unit Chassis	213715/G-20	CIRCUIT BREAKER	Single Point of Failure
NAVY MULTIBAND TERMINAL: AN/NSC-9(V)	CVN 71	CVN 71	MISC	D38999/26W8985N	CONNECTOR PLUG ELEC	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 100		DIGITAL DATA PROCESSOR CP-1575A/URC-107 [1A1A3]	A567A028-101	CRYSTAL CONTROLLED OSCILLATOR	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 100		RADIO RECEIVER-TRANSMITTER RT-1613/URC-107 [1A1A1]	651-0996-050	RECEIVER SYNTHESIZER	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 60		RADIO RECEIVER-TRANSMITTER RT-1613/URC-107 [1A1A1]	651-0996-050	RECEIVER SYNTHESIZER	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 60		RADIO RECEIVER-TRANSMITTER RT-1613/URC-107 [1A1A1]	651-0996-050	RECEIVER SYNTHESIZER	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	CG 52		HIGH POWER AMPLIFIER AM-7389/URC-107 [1A2A1]	676-4527-030	PROCESSOR	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 91	NAVJUP/FDNR	HIGH POWER AMPLIFIER AM-7389/URC-107 [1A2A1]	676-4529-001	DIGITAL WAVEFORM GENERATOR	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	CG 52		HIGH POWER AMPLIFIER AM-7389/URC-107 [1A2A1]	622-7961-081	RADIO FREQUENCY AMPLIFIER	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	CG 52		HIGH POWER AMPLIFIER AM-7389/URC-107 [1A2A1]	676-4525-030	AMPLIFIER-CONVERTER	Single Point of Failure
AN/URC-107(V) JTIDS: OPERATE ALL	DDG 100		HIGH POWER AMPLIFIER AM-7389/URC-107 [1A2A1]	676-4525-030	AMPLIFIER-CONVERTER	Single Point of Failure

Figure 8. Sample output of PRiSM analysis of MRDB Data

B. TOOLS

The primary tools used to determine potential parts for stocking are R, LLA, and JMP. R is used to prepare the raw data gained from the various sources used. First excess columns are removed to simplify the dataset and the higher priority parts are separated. Next, demand rates are assigned levels such as low, medium, or high and the data is sorted by date and by ship. The data is then output to a comma separated value (CSV) file and imported into LLA. LLA is used to identify parts that have links with other parts, and a high “lift,” meaning they have strong correlation. The link weights are also shown, representing the probability that one part will fail, assuming the tested part fails. In one test, 51 high impact, low demand part numbers were identified as likely to fail by LLA. These parts are then separated from the dataset and ran through JMP. Downtime is calculated by subtracting the demand date from the fill date in order to develop the dependent “Y” variable for testing. Based on cost to repair, parts available, source location, and other data gathered, the effects of each independent variable can be assessed, and the dependent variable of average downtime can be estimated by running a regression and adjusting the input. Improvement across parts with IMEC 3,4,5 and various CASREP data can then be logged and compared with real life data to validate the process. As the first step, each of the three tools used will be described.

1. R

The first tool to describe is the coding language of R, specifically the program RStudio, which assists by providing a user interface for R. R is an open source coding language which is comprised of multiple libraries independently developed to give the user access to certain data analysis tools (R Foundation 2020). R Studio is a program based on R that saves the users’ scripts, makes it easier to view graphics, and has a user interface making it more convenient to load and manipulate datasets. R can be used to write code for almost any task needed, but it is particularly useful for statistical analysis and for manipulating large datasets. For this thesis, R was primarily used to organize the dataset and to pull the useful information from a larger dataset. Other data was also obtained, and if it is a small enough file, less than 60,000 entries, a program like Microsoft Excel can be

used just as effectively and was used in many cases for this thesis. R is more efficient for automated processes required to transform large datasets and in many cases is more accurate for statistical analysis.

2. Lexical Link Analysis (LLA)

The next program that is heavily used in this thesis is part of the Collaborative Learning Agents (CLA) package, the LLA tool which has already been somewhat covered in prior paragraphs. CLA is from Quantum Intelligence, Inc., created by Professor Ying Zhao at Naval Postgraduate School. CLA, specifically LLA has many capabilities with respect to data mining and creating graphical visualizations to help determine trends in datasets (Zhao, MacKinnon, and Gallup 2015). Figure 9 shows an example of a visualization developed using the datasets obtained.

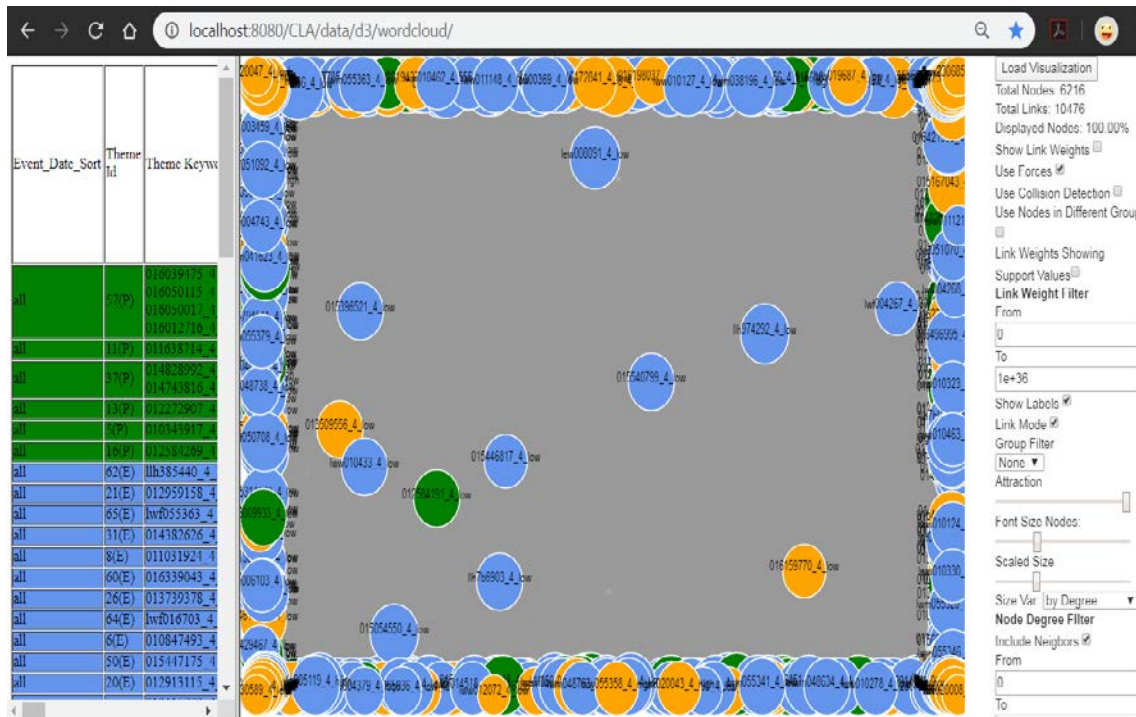


Figure 9. Image showing LLA working to find correlations between dataset entries used in this thesis.

From the article "Using Lexical Link Analysis as a Tool to Improve Sustainment"
(Stevens and Zhao 2019);

LLA is an unsupervised machine learning method and describes the characteristics of a complex system using a list of attributes or features, or specific vocabularies or lexical terms. Because of the potentially vast number of lexical terms from big data, the model can be viewed as a deep model for big data. For example, we can describe a system using word pairs or bigrams as lexical terms extracted from text data. LLA automatically discovers word pairs, and displays them as word pair networks. Figure 10 shows an example of such a word network discovered from data. "Clean energy" and "renewable energy" are two bigram word pairs. For a text document, words are represented as nodes and word pairs as the links between nodes. A word center (e.g., energy, below) is formed around a word node connected with a list of other words to form more word pairs with the center word energy.

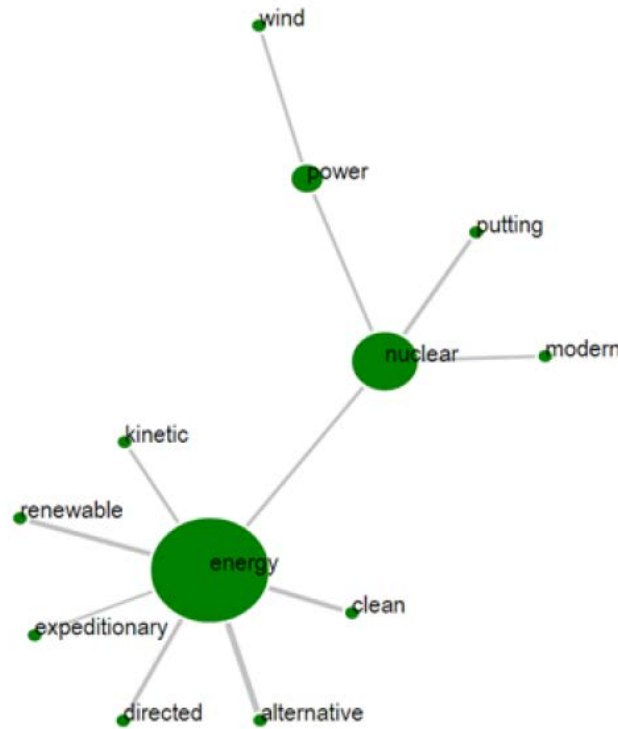


Figure 10. Sample figure showing correlation with the word "energy" in LLA.

LLA is a dynamic tool that can be used to recognize qualitative (as shown above) or quantitative relationships (as in this thesis). Similar to its ability to find word pairs, LLA

can be used to find relationships in part numbers, quantity of demands, time periods, and any other variables that are properly imported. In one dataset used, ship names and part nomenclature were not available, but quarter, material number, demand level, and number of failures were used to find correlation. The wider the range of information that is available, the better a job LLA can do to find relationships, in other words high resolution data sets with less overall data may be more useful, but one of LLA's strengths is that it can work on high resolution data or low resolution data.

3. JMP

Once the LLA model is finished running, the JMP program is utilized to find the most useful correlations among the factors in the data points discovered. JMP is a statistical discovery and analysis tool that has a useful Graphical User Interface for running regressions, estimating values, determining residuals, and creating various types of charts to show the information ("Statistical Software" n.d.). A dependent "Y" variable is selected (such as downtime due to the failure), and independent variables that feed into downtime are selected as the "X" variables. One valuable feature of JMP is a stepwise regression, which allows the user to run a regression in steps, moving through the independent variables from the ones with the highest R squared value to the lowest R squared value. By doing this, the user can easily stop at the most useful number of steps, selected the variables with the highest impact on the dependent variable that is affected. Figure 11 shows a stepwise regression that was run against the expected number of failures.

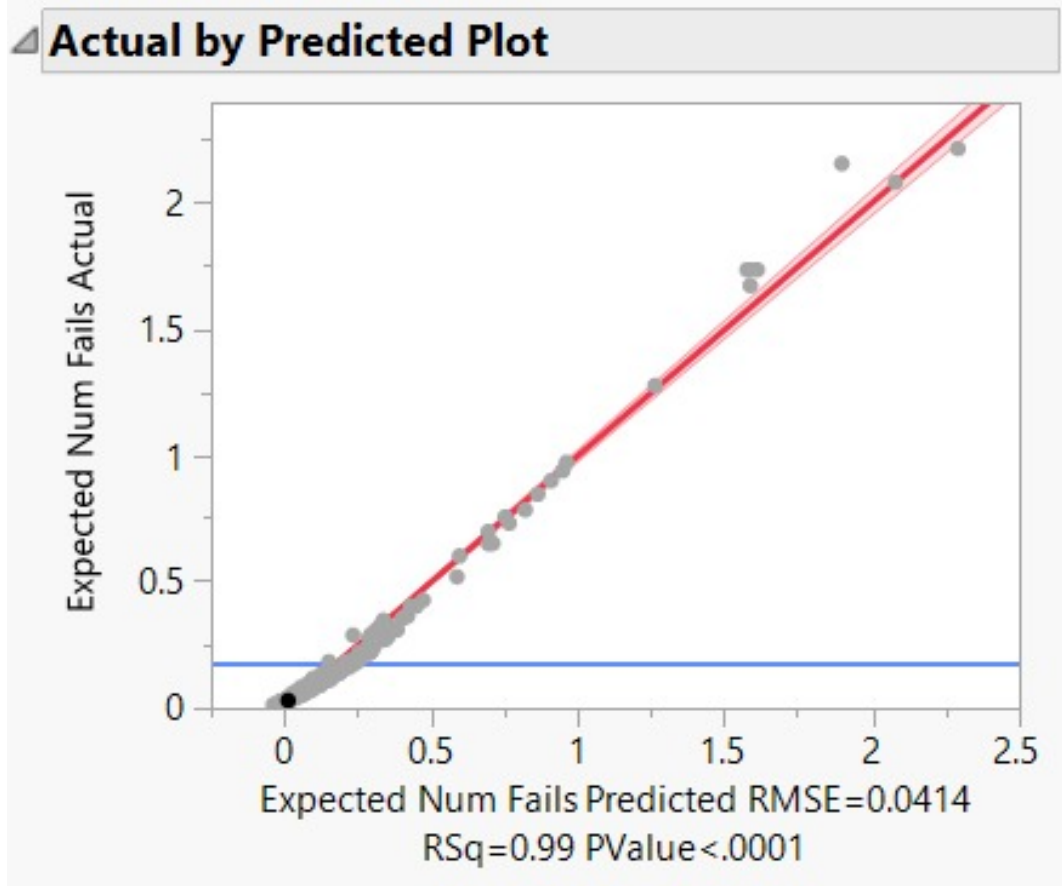


Figure 11. Correlation of independent variables to the expected number of failures of any given NSN studied.

Another valuable feature in JMP is the ability to run a prediction profiler. In Figure 12, various parameters that were selected as being highly correlated to expected number of failures can be adjusted to show the effect on the expected number of failures. In the case below, Class A supply available, Time in Service of the parts, location of the spares, and the upper value of the expected number of failures were all strongly tied to the expected number of failures of a given NSN based on the data used.

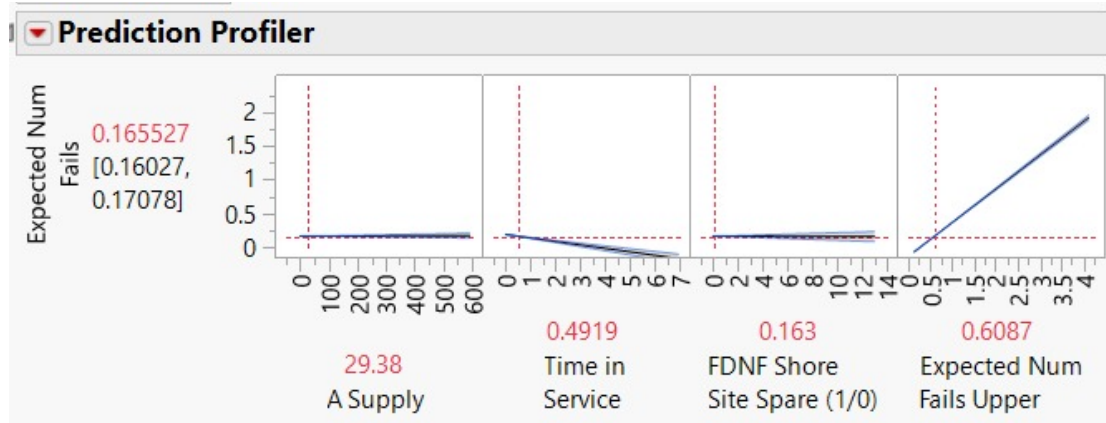


Figure 12. Sample image of prediction profiler to test the effects of changes in the independent variables in a regression.

In the case of this thesis, JMP was used to determine which factors are most likely to impact AO, and then those factors were chosen to visually show the impact of carrying the parts onboard, forward stocking, having them available in the United States, or not having them available, requiring procurement or repair in the case of a failure.

4. Additional Tools and Calculations Involved

a. PRiSM

During the search for data collection and networking for possible methods for this thesis the PRiSM tool that is currently being developed at Naval Information Warfare Systems Command (NAVWAR) and Naval Surface Warfare Center (NSWC) Corona was discovered. While the focus of this thesis is on failures and unfilled demands that can cause redlines, PRiSM has been focused on C4I systems for Amphibious Readiness Groups and Carrier Strike Groups deploying to 7F. The goal of the PRiSM program is to increase AO while decreasing the need for onboard tech assists (and lengthy wait times) which has the benefit of decreasing downtime and cost. There is some overlap between PRiSM and the analysis in this thesis, as many C4I parts will cause a redline if they fail. During a meeting with the PRiSM program lead at NAVWAR it was determined that PRiSM could be used in collaboration with or verified by the methods used in this thesis. PRiSM has seen some success with a test run on USS BOXER and on USS Theodore Roosevelt during their 7F deployments. During these deployments, the formulas and algorithms used in PRiSM were

used to identify 52 C4I components to transfer from other ships or to stock from warehouses, and of those parts 66%, or 34 parts were actually used during the deployments. These parts would have otherwise had to be shipped or procured from other sources, greatly decreasing AO. Many parts actually failed within one month from when the program said they would fail. Figure 13 from the PRiSM program office show which components were identified and added for the USS Boxer deployment.

▼ PRiSM Update

- USS BOXER ARG PRiSM Analysis Review Conducted on 13 December 2018 at SPAWARSCOM 4.0
 - PRiSM Data Analysis & Worksheet used for review
 - Attendees included PEO C4I and SPAWAR FRD Logisticians and SME Support, NAVSUP WSS N961 Program Managers, and NSWC Corona Program Manager and Data Scientists
 - PRiSM identified the following C4I Systems as End of MTBF Life Parts

NAVSSI (17 Components)	JTT (1 Component)
CND (1 Component)	ADSI (1 Components)
JTIDS (9 Components)	CDLMS (9 Components)
ISNS (1 Component)	ADNS (3 Components)
* NMT (7 Components)	SRS (2 Components)
CBSP (1 Component)	
- * Part identified on PRiSM experienced CASREP on 8-Jan-19**

Figure 13. PRiSM presentation to NAVWAR after test deployment on BOXER ARG ((Naval Information Warfare Systems Command (NAVWAR) 2020).

The PRiSM structure starts with some of the techniques used currently in Readiness Based Sparing (RBS) models, such as the basic calculation of AO by dividing system uptime over total system time, or mean time between maintenance (MTBM) divided by MTBM + mean down time (MDT).

$$A_o = \frac{UPTIME}{UPTIME + DOWNTIME} = \frac{MTBM}{MTBM + MDT}$$

Figure 14. Formula for AO.

Additionally, basic inventory models such as sawtooth plots are considered, which attempt to optimize how much to reorder (Q), and when to reorder it (S), to keep costs down and efficiency up.

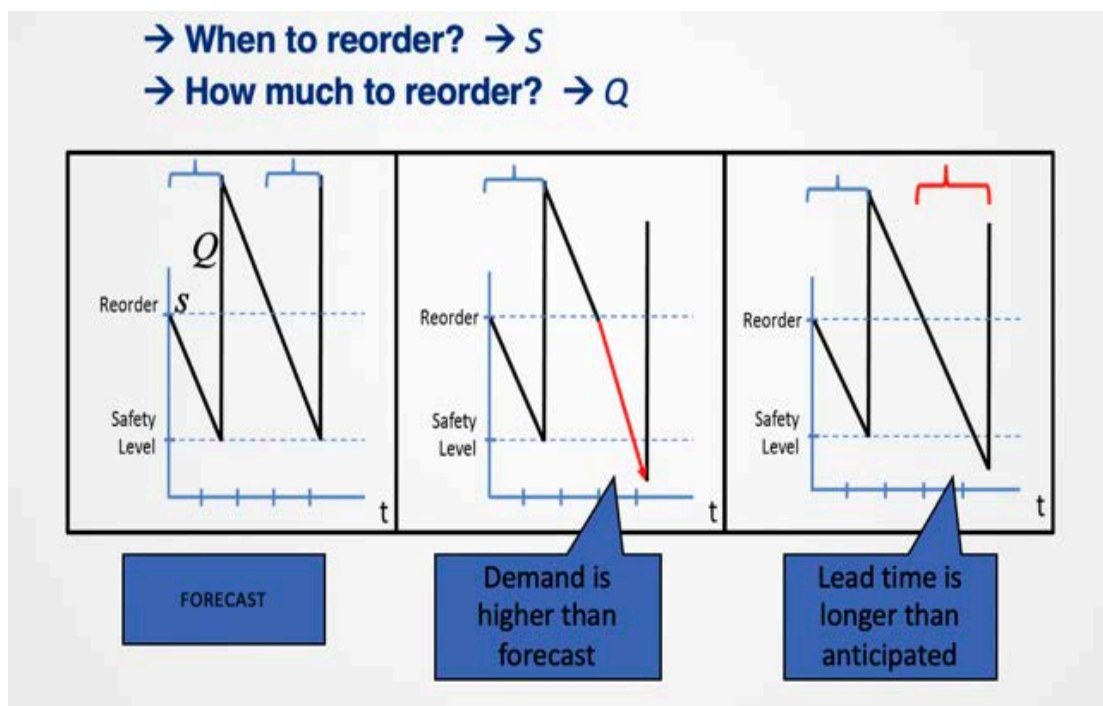


Figure 15. Basic sawtooth graph to explain reorder points.

The PRiSM model goes a step beyond basic reorder points and inventory management, however. NSWCCOR keeps a database of failures based on CASREP information, engineering information, and other pertinent data on a part by part basis. This could be a massive amount of data, and for now only C4I parts on certain test ships have

been assessed. Figure 16 from the PRiSM standard operating procedures shows that there is a three-phase assessment process.

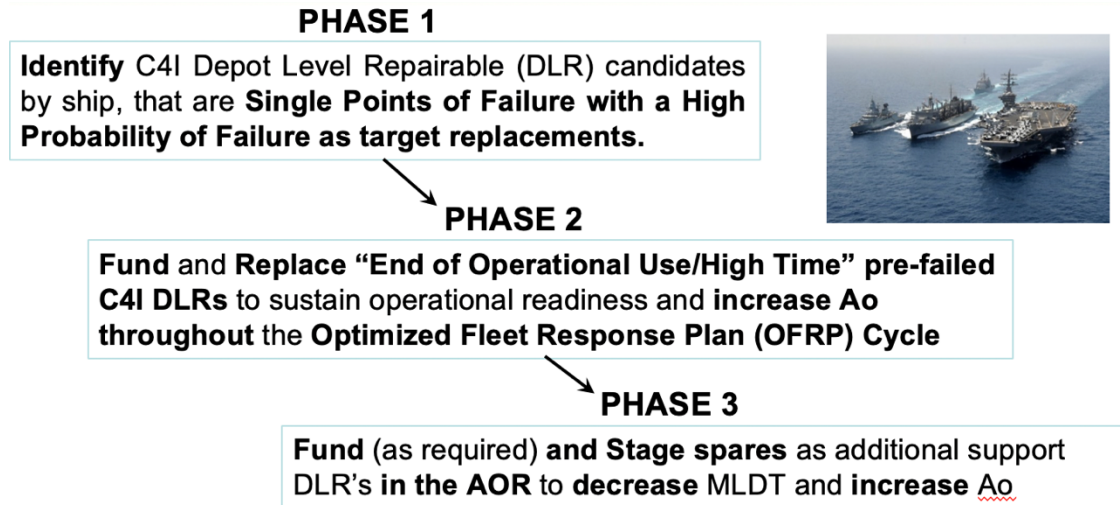


Figure 16. Phases of part selection and resourcing using PRiSM. Source: Naval Information Warfare Systems Command (NAVWAR) 2020.

Interestingly, upon researching PRiSM, many of the ideas for the methodology for selecting parts are very similar to the ideas being researched in this thesis. PRiSM searches for parts that are a single point of failure, or redline, and have a high probability of failure, while this thesis searches for redline parts with a lower probability of failure but a possible relationship to a more commonly demanded part. PRiSM takes an additional step through the extensive data collection and storage capabilities (which are beyond the capabilities for this thesis) in order to keep track of operational time in use by part, compared with the mean time between failure. Using this technique, PRiSM can accurately assess when parts will potentially fail. Finally, exactly the same as the goals of this thesis, PRiSM recommends ideas such as ship pack-up kits (PUK), forward deployed spares at NAVSUP and DLA owned warehouses, and staging on Military Sealift Command ships for replenishment at sea. These tools would be used by liaisons in forward deployed locations to better assess what parts can be forward staged, and additionally there is potential to use the methodology developed in this thesis as a verification or collaborative tool to further improve the development of PRiSM.

b. SIMIO

There is potential for future work on this project using SIMIO, which is a valuable tool that can model complex event graphs with stochastic or deterministic nature. A major benefit of using computer-based simulation is that a stochastic model can be run thousands of times with reliable data gained from training data or assumed data and based on the inputs given results can be generated within a fairly certain confidence interval. From the SIMIO, LLC website, “The company SIMIO LLC was established to deliver leading edge solutions for the design, emulation, and scheduling of complex systems” (Simio LLC n.d.). The name SIMIO comes from Simulation Modeling framework based on Intelligent Objects. SIMIO works with a GUI on a chart overlay, where the user can drag different types of “servers,” or objects that push the process through the system, “entities,” or the objects being pushed through the system, “sinks,” or the point of completion, and different types of “paths” and “connectors,” all of which have their own rules for processing entities. SIMIO can model different statistical distributions such as triangular, normal, exponential, etc., depending on the distribution of the subject being simulated. The tool can also use work schedules, different numbers of servers on different shifts, multiple entities feeding into a system, different service rates, and many other valuable functions to simulate a realistic routine. A sample model has been built in SIMIO to demonstrate its potential effectiveness regarding this thesis and will be discussed in Chapter V.

C. METHODOLOGY

After cleaning and preparing the various datasets in R, multiple experiments were then prepared by merging the datasets and matching parts. The new datasets would be prepared for importation into LLA by selecting the different parameters to focus on predicted parts. These datasets were manipulated using R to create CSV data files that were then imported into LLA to identify link between demands and ship type, time of year, location, and criticality code of the parts. Individual parts that are at higher risk of failure and have not been identified using prior methods can then be studied with more detail.

The next step in the methodology used was to prepare the various experimental datasets for input into LLA. Scripts were built to pull different sets of information from the

data to build categories for LLA to correlate, such as failures on certain ships and on certain dates. A total of six LLA experiments were run on the various datasets. The first two experiments were on the original dataset organized by IMEC, Quarter, and ship. The next two experiments were run on the CASREP data, on a section that utilized the recurring requisitions, and on a subsection that only used the non-recurring requisitions. Next, the three datasets (demand, recurring CASREPs, non-recurring CASREPS) were merged, only using parts that occurred in all three, bringing the total amount down to 2960 unique items and 7197 total lines of data, and this dataset was also experimented within LLA. The final experiment involved the PRiSM analysis as well, and experiments were done to use the data in different ways such as by testing on items with higher than standard processing time (two days or greater), segregating further by comparing predictions from LLA to predictions from PRiSM (11 NSNs matching).

As the experiments were run, more correlations were found, and more confident probabilities of failure were able to be discerned. This was based both on simply having more data, and also on expanding the number of useful parameters to use. From “Using Lexical Link Analysis (LLA) as a Tool to Analyze A Complex System and Improve Sustainment” (Stevens and Zhao 2019), the probability of failure of a part can be affected by many factors including uncertainty, a disruption, or a perturbation in the system. These failures can disrupt the logistics system as a whole, particularly when low failure rate parts suddenly come into high demand. The complexity of predicting the probability of a failure calls for the integration of methods in data fusion, data mining, causal learning, and optimization of all the elements in a logistics system when facing particular uncertainty and perturbation.

In the first experiment, using the historical demand dataset, a selection of demand data over the last nine years was prepared and imported. This dataset contained over 45,000 individual lines of demand with 7197 unique NSNs. This data was sorted by impact code and demand was listed as low, medium, or high. Next, LLA was applied the data to help discover historical associations among the demands, within the same context (i.e., the documentation date and ship ID). As shown in the figure below, each grouping represents

all item demands occurred within the same date and same ship. In this dataset, there were 13193 correlations among the 7197 NSNs was discovered.

In order to discover item associations using LLA, bi-grams, or baskets, are built to contain various numerical categorical data (like ship name, date, system, part type). Within the script to prepare the data for LLA, these bi-grams are discretized to categorize the values into word-like features. These can include part numbers and ships' UICs, but as a string vice a numerical value. The idea is that this data mining can discover initial statistical correlations and associations from big data. Next, LLA associates each item in the dataset with each other item, and calculates the weight and lift using the below formulas, with W_k being the instance, n being the number of instances, and t being the point within the summation from t to T . Figure 17 shows the various calculations used by LLA.

$$\begin{aligned}
 \text{Probability: } P(W_k|n) &= \frac{\text{\# of instances when } W_k \text{ follows (or precedes) } n}{\text{\# of instances of } n} \\
 \text{Lift: } L(W_k, n) &= \frac{P(W_k|n)}{P(W_k)} \\
 \text{Difference: } D(W_k, n) &= P(W_k|n) - P(W_k) \\
 \text{Correlation: } R(W_k, n) &= \frac{\sum_{t=1}^T (\# W_k \text{ in } t) * (\# \text{ of } n \text{ in } t)}{\sqrt{\sum_{t=1}^T (\# \text{ of } W_k \text{ in } t)^2} * \sqrt{\sum_{t=1}^T (\# \text{ of } n \text{ in } t)^2}}
 \end{aligned}$$

Figure 17. Calculations used in LLA to determine Probability, Lift, Difference, and Correlation. Source: Quantum Intelligence 2015.

Figure 18 illustrates data that has been prepared for entry into LLA for mining, broken down by UIC, NSN, IMEC, and demand rate.

```

1 20143 000044583_4_high 000424968_4_high 000532068_4_low 00053
005560091_4_low 005966274_4_low 006011317_4_low 006120730_4_l
009644662_4_high 009873744_4_low 009894589_4_high 009910948_4
010299566_4_low 010299760_4_low 010300069_4_low 010300091_4_l
010533898_4_low 010543561_4_high 010548439_4_low 010561592_4_
010654686_4_low 010655668_4_low 010658740_4_low 010667415_4_l
010817174_4_low 010834065_4_low 010847920_4_low 010848690_4_l
010985692_4_low 011006765_4_low 011007732_4_low 011007733_4_l
011146655_4_low 011182640_4_high 011194879_4_low 011196580_4_
011359896_4_low 011359903_4_low 011359939_4_low 011366872_4_l
011468786_4_low 011468788_4_low 011484022_4_low 011485018_4_l
011632843_4_low 011637699_4_low 011637736_4_low 011637737_4_l
011704221_4_low 011704376_4_low 011706847_4_low 011719254_4_l

```

Figure 18. Sample data after being prepped for use in LLA.

After the data is prepared, it is fused, mined, and the visualization is loaded, resulting in a GUI that can be manipulated in various ways to search for linkage weights, probability of failure, and specific NSN correlations.

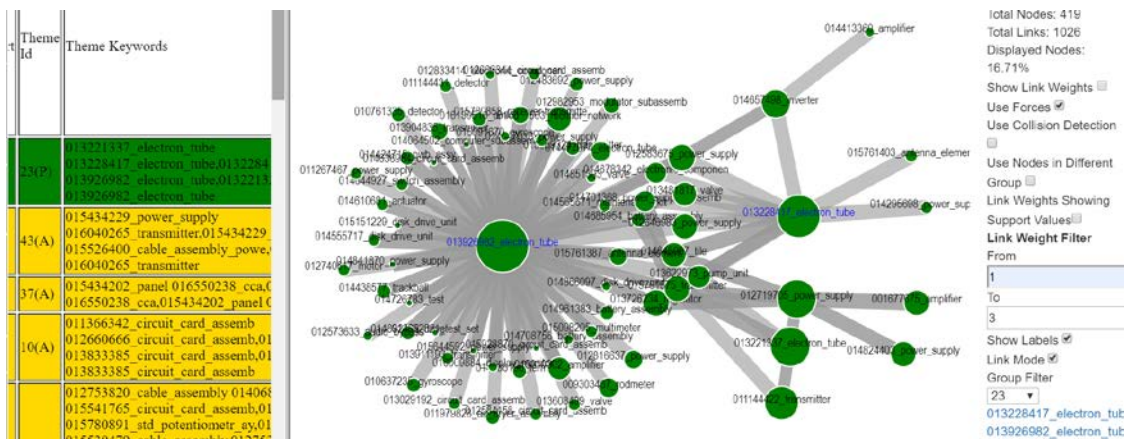


Figure 19. LLA GUI used to visualize and manipulate data showing various correlations.

The information to the left of the GUI shows the correlations identified, in order from the highest lift down, with green being the popular correlations that have the highest number of mutually connected pairs. Emerging correlations are colored blue and contain the next highest set of mutually connected pairs, and anomalous correlations are colored

yellow, meaning there are mutually connected data points but perhaps not enough to warrant further exploration. On the actual diagram in the GUI, the size of the node is based on the number of data point for that item, and the thickness of the connector represents the weight of the link, which can also determine the probability of failure of one part based on the failure of the other part. In the figure above, if electron tube NSN 013926982 fails, there are various probabilities that many other parts can also fail as well based on the data used.

By performing this method, many commonsense correlations are found. It was noticed that pistols, rifles, machine guns, and shotguns were often ordered at the same time on each ship. While this doesn't mean that they failed because of each other, regardless they were ordered on the same day so it could be stated that if one of those NSNs is not carried with the others, perhaps it should be. In the case of potential parts to avoid redline failures, it can be said that if one part is not carried onboard, perhaps it should be to be prepared for the event of a critical system going down.

Once an LLA experiment has been run, the output can be searched on the dataset and the other parameters can be inspected. For example, if LLA focused on ship, date, and criticality code as above, those selections are separated from the original dataset, and then a regression is run with variables such as downtime, mean time between failure, net price, A supply (ready for issue), M supply (failed, in repair), or F supply (failed, awaiting repair or disposition). The preferred regression to use was a stepwise regression, with downtime as the dependent x variable, and the other parameters as the independent y variables.

The following chapter will discuss the output from the experiments run on this model and will describe an analysis of the output.

IV. ANALYSIS

The first step of the analysis portion of this thesis will be to describe the six LLA experiments that were run. Next, JMP analysis will be introduced to show the value of carrying the parts selected. Finally, calculations that were performed to show overall goodness of the extra parts stocked as well as plots comparing the improvement to the basic model will be shown.

A. LLA EXPERIMENTS

A total of six LLA experiments were run on the various datasets. The first two experiments were on the original dataset organized by IMEC, Quarter, and ship. The next two experiments were run on the CASREP data, on a section that utilized the recurring requisitions, and on a subsection that only used the non-recurring requisitions. Next, the three datasets (demand, recurring CASREPs, non-recurring CASREPS) were merged, only using parts that occurred in all three, bringing the total amount down from over 100,000 items and nearly 1,000,000 total lines of data to 2960 unique items and 7197 total lines of data, and this dataset was also experimented within LLA. The final experiment involved the PRiSM analysis as well, and experiments were done to use the data in different ways such as by testing on items with higher than standard processing time (two days or greater), segregating further by comparing predictions from LLA to predictions from PRiSM.

1. Demand Data

The first two experiments performed in LLA were on the demand data originally obtained from NAVSUP WSS. This data included data points for line item number, NSN, IMEC, site, quarter, and demand among others as shown in Figure 20.

```

> tail(Edwin_data)
  Material IMEC AAC COG LRC LRC Tech SiteCd Quarter Demand
1210263 LWX230965 4 D 6X FBB 91 M 20173 0
1210264 LWX230965 4 D 6X FBB 91 M 20161 0
1210265 LWX230965 4 D 6X FBB 91 M 20151 0
1210266 LWX230965 4 D 6X FBB 91 M 20152 0
1210269 LWX230965 4 D 6X FBB 91 M 20184 0
1210272 LWX230965 4 D 6X FBB 91 M 20163 0
> summary(Edwin_data)
  Material          IMEC          AAC          COG          LRC          LRC Tech
Length:504280 Length:504280 Length:504280 Length:504280 Length:504280 Length:504280
Class :character Class :character Class :character Class :character Class :character Class :character
Mode :character  Mode :character  Mode :character  Mode :character  Mode :character  Mode :character

  SiteCd          Quarter          Demand
Length:504280 Min. :20143 Min. : 0.000
Class :character 1st Qu.:20154 1st Qu.: 0.000
Mode :character  Median :20168 Median : 0.000
                Mean :20168 Mean : 2.397
                3rd Qu.:20181 3rd Qu.: 0.000
                Max. :20192 Max. :12453.000

```

Figure 20. R Screenshot showing a summary of the first dataset obtained for analysis.

First, the data was arranged into the categories of low, medium, and high demand rates and separated by NSN and UIC. Next, the data was exported to a CSV file and imported into LLA for experimentation and analysis. The first of the two experiments used a 60,000 row subset of the original 504,280 line items, using IMEC 4 and IMEC 5 data only as a sample to experiment with the ability to find correlations among the quantitative data. While this dataset covered the same period from 2011 to 2019, this experiment did not specifically divide the demands by quarter, but instead had individual dates of failure. Using this dataset, multiple correlations were also. The apparent strong correlations were promising, but it was determined that to get more accuracy more input parameters were still needed. In the second experiment more bi-grams were added, further separating the data by the quarter in which the demand happened, as well as the NSN and the UIC. This experiment showed promising results, with parts from similar systems showing strong correlations, high lift, and reasonable probabilities of failure. Figure 21 shows a sample of the visualization of the data in the first experiments, with logical correlations, but too many to slim down the data to a useful number of highly probable demands to occur.

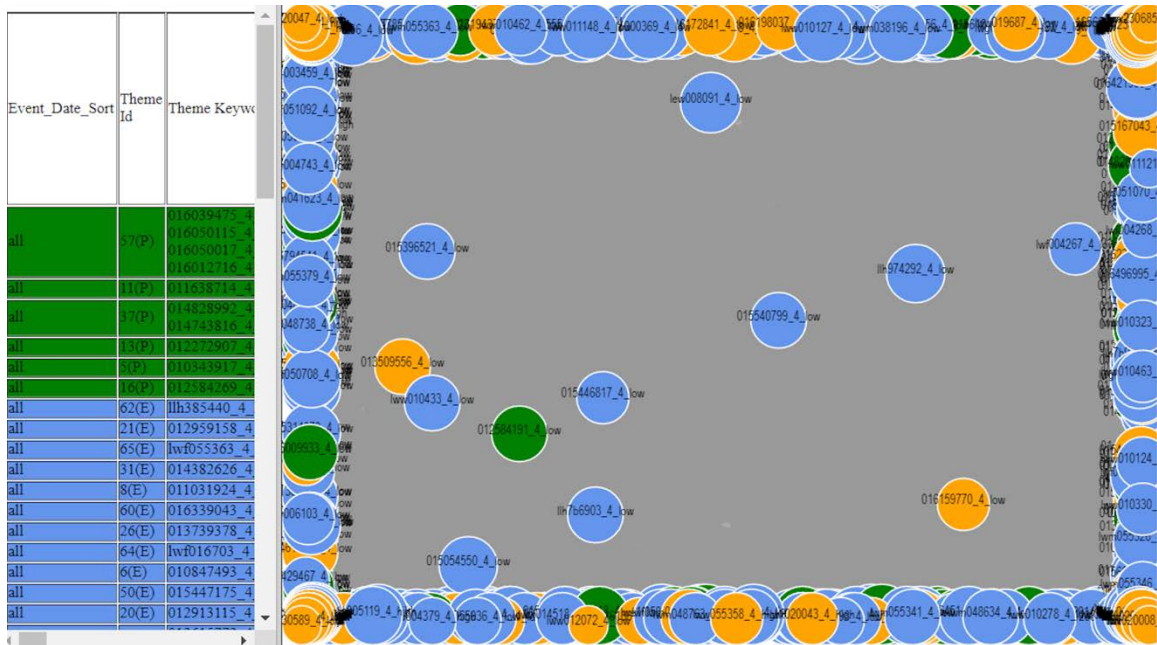


Figure 21. Screenshot of the first experiments run in LLA.

As can be observed in the left margin of the figure above, while there were many parts that did show correlation, there were six groupings that were popular, meaning they showed the highest lift of the group. Those groupings involved NSNs 016039475, 016050115, 016050017, 016012716, 011638714, 014828992, 014743816, 012272907, 010343917, 010343917, and 012584269, which were all from the same C4I system on the ship. This shows that if any one of these parts fails, there is a high likelihood that one of the others will also fail, and it could be determined that if the ship carries one of those systems but not these parts, they should actually be carried. All of these parts had between low demand levels, meaning they would most likely not be identified using the current stocking system which only adds parts to the COSAL if there are two or more failures in one year.

While the initial two experiments were a good proof of concept regarding the usefulness of LLA in predicting future failures or demands, they also helped to determine that more information was needed in order to make a coherent analysis of what, where, and why the parts should be stocked. For the next step, failure data in the form of CASREP tracking was requested from NAVSUP WSS.

2. Failure Data

The failure data received from NAVSUP WSS was in the form of CASREP reporting from the N-ERP system. The N-ERP system consolidates data from 26 other Navy systems, and there were 158 columns of information in the CASREP listings received. This data included categories such as NSN, part description, document number, data that the failure occurred, UIC, whether the ship was CONUS or OCONUS, location where the part is coming from, and the number of replacement parts available in A, M, or F condition, price to repair, and price for a new part. The data also contained columns of repeat information, as there can be overlap in the various N-ERP systems, for example there was pricing data, billing date, and document date, all of which were generally the same or within one day of each other. There were also many categories containing information specific to the supply system which were not of use for this research, such as acquisition codes and units of issue. The first step to make the data usable was to remove the excess columns, bringing the total number down to just 18. Next, since data was received in two separate files, recurring and non-recurring failures, the two files were made uniform in format and merged into one file. The recurring data contained 45,584 individual document numbers, and the non-recurring data contained 8,662 individual document numbers, giving a total of 54,246 rows of individual IMEC 3, 4, and 5 failure data from 2011 until 2019 among the 7F ships studied.

Similar to the original demand data obtained, the first step was to sort the failure data into bi-grams. The first experiment involved the individual datasets and the second experiment involved both together, and both experiments yielded similar results as the non-recurring data was a small subsection of the recurring data. It was also noted that even though the smaller dataset was labelled non-recurring, there was actually multiple cases where the demands did in fact repeat, although that was over multiple years and on multiple ships. Due to the similar nature of the datasets, the second experiment with the merged data will be discussed. Bi-grams were sorted by specific date of failure (as opposed to quarter), UIC, NSN, and description. The ability to sort into much smaller baskets made the results much more precise than the quarterly demand data. There were cases that showed up in

some experiments that told the story of how parts are ordered, for example on multiple ships pistols, machine guns, rifles, and shotguns were all ordered at the same time.

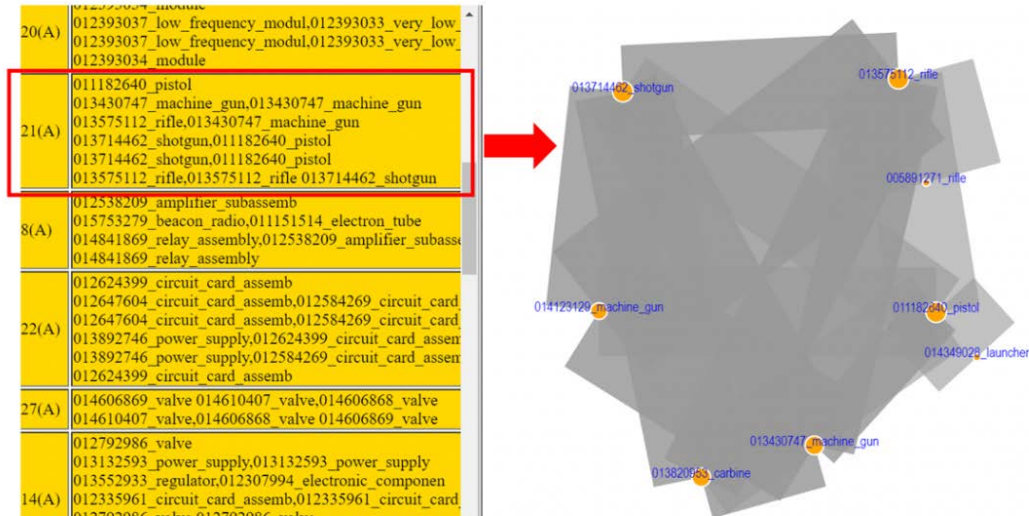


Figure 22. Example of high correlation found among orders in failure data.

This correlation is interesting because not only does it have a very high lift, as visualized by the thick grey bars connecting the orange nodes in the graphic on the right, but it also shows that while the failures obviously were not connected there is an important correlation to the ordering practice. As theorized in the beginning of this project, demand can be caused by weather, change in OPTEMPO, business practice, or human error. In this case, it seems that as ships prepare to deploy, they make orders of multiple armaments at the same time. While this may seem obvious due to the easy to read nature of the parts ordered based on the descriptions, the same practice could be said for less obvious systems such as the amplifier subassemblies, circuit card assemblies, valves, and other parts shown in the figure above. Whether the parts above were ordered at the same time due to any of the perturbations in the supply system is not of consequence, what is of consequence is if they are ordered and supply is not available causing a disruption in mission for the ship and a decrease in AO.

For the next step, link weight filters were adjusted, meaning that the requirement for parts to be considered correlated was higher, and different groupings of correlations among the parts were individually viewed. Among the 54,246 individual rows of data there were multiple correlations, as the system went through each part listed and found links with each other part other part in the dataset. This process was ran for every row of demand. The below image shows the initial run of this experiment had 1026 total links. Also, as can be seen by the number of grey connectors covering the screen in Figure 23, and the amount of nodes, this setting has too much information to use.

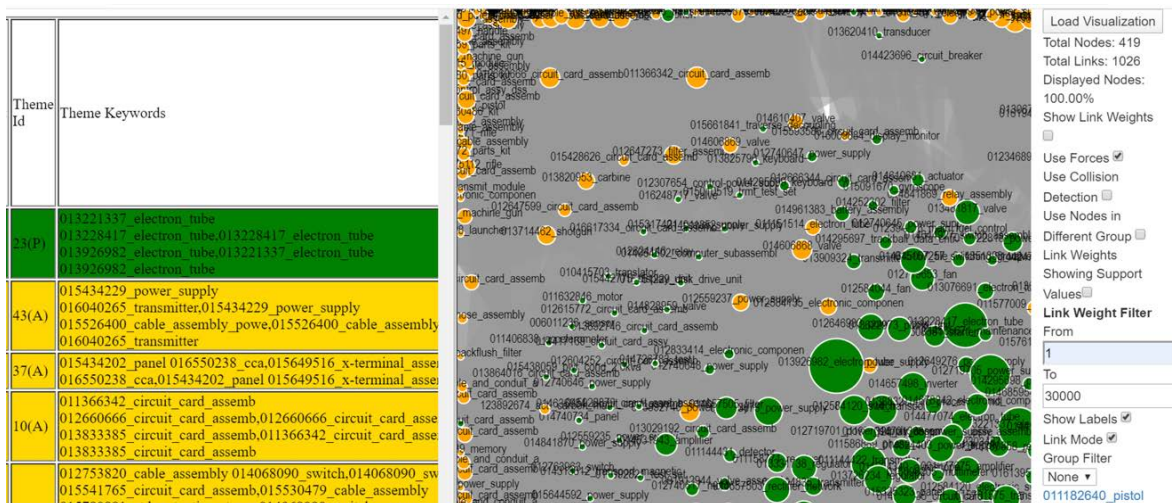


Figure 23. A graphic illustrating all connections in the merged dataset for this experiment.

In Figure 24, by adjusting the link weight parameters in the same experiment even further to ignore lower weights, and to focus in on individual groupings of interest, more specific and interesting correlations were discovered.

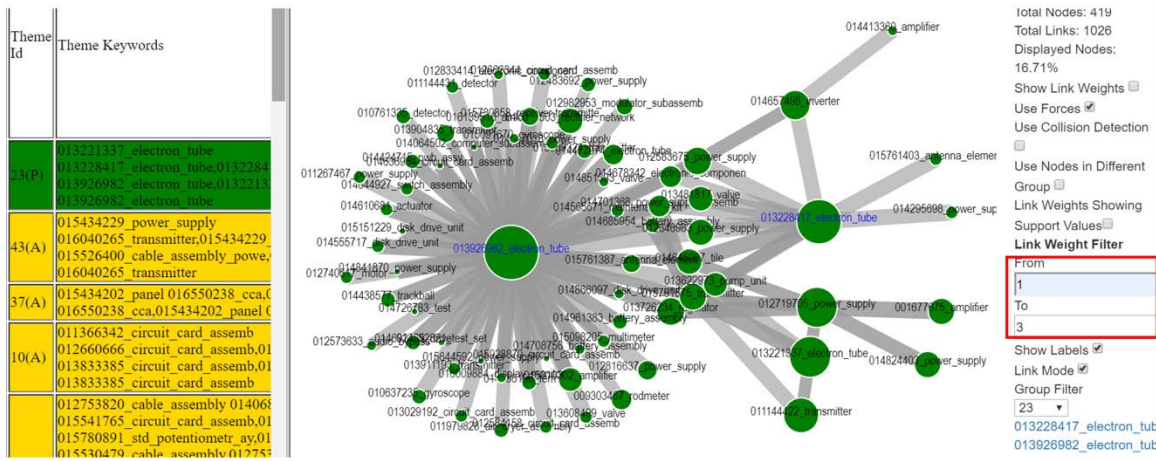


Figure 24. Illustration showing the correlations between the failure of 013926982, 013228417, and various other parts.

As can be seen in the figure above, the two electron tubes selected have related multiple related failures in systems that are also related to the C4I system that they are part of. Searching a bit further, some electrical components had more specific related failures, which can be determined by setting a higher parameter for lift in the data mining.

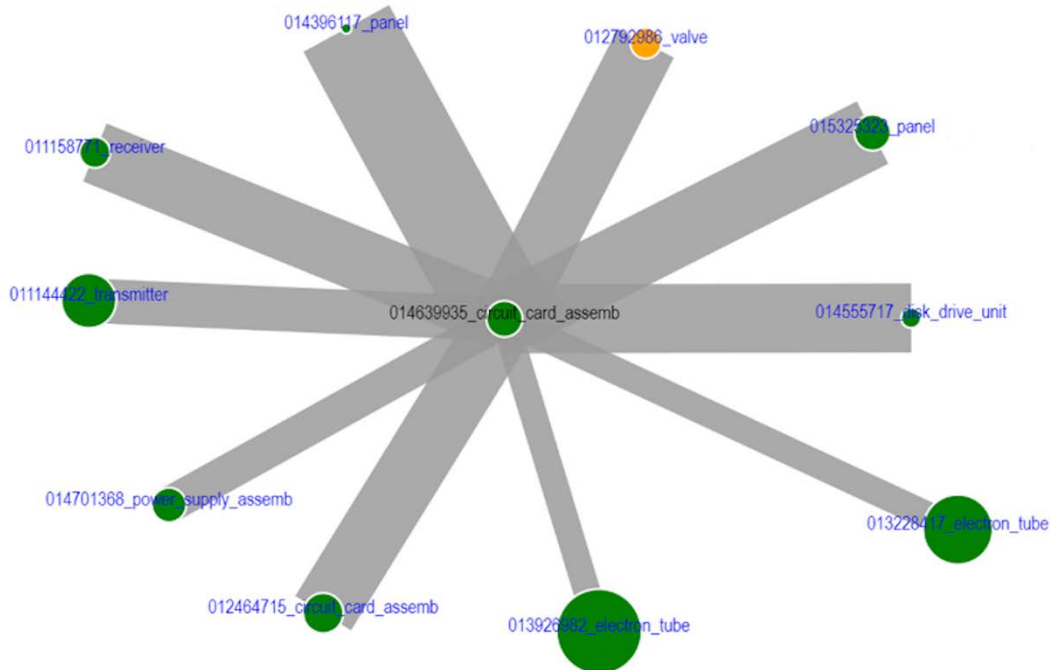


Figure 25. A search for correlations with a high lift related to 014639935.

In Figure 25, NSN 014639935 showed 69 demands over the nine-year period that the data covers. Of these demands, many did have availability in the supply system, so the only wait time was based on shipping the part to the ship, however there were nine examples showing no availability in the supply system, with the longest wait being 148 days plus shipping time. Further analysis showed that the supply system carries 13 of these parts INCONUS in A condition, and keeps 228 of them in F condition, meaning they are in a warehouse awaiting repair. The COSAL for this part is zero, in other words ships do not normally carry it onboard. By finding the correlations between this failure and the others it could be determined that these parts should be forward staged or carried as part of the ship's COSAL. Interestingly, the program office in charge of this NSN recommends it to be forward staged as found in the notes column in the data. Similarly, a search focusing on the grouping around another NSN, 01372218 found results that point in the same direction.

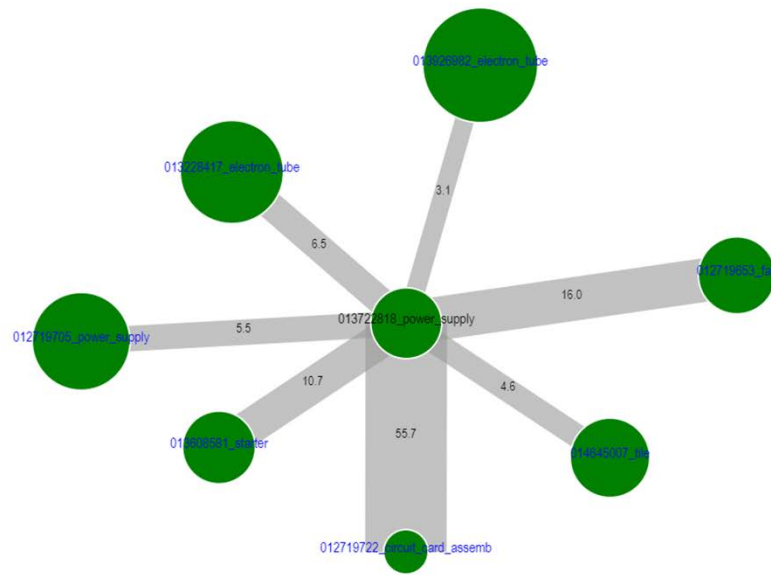


Figure 26. Another finding with high lift between 013722818 and its associated parts.

A final finding from this experiment that is noteworthy was the search for the grouping around NSN 013833385, which showed very high lift.

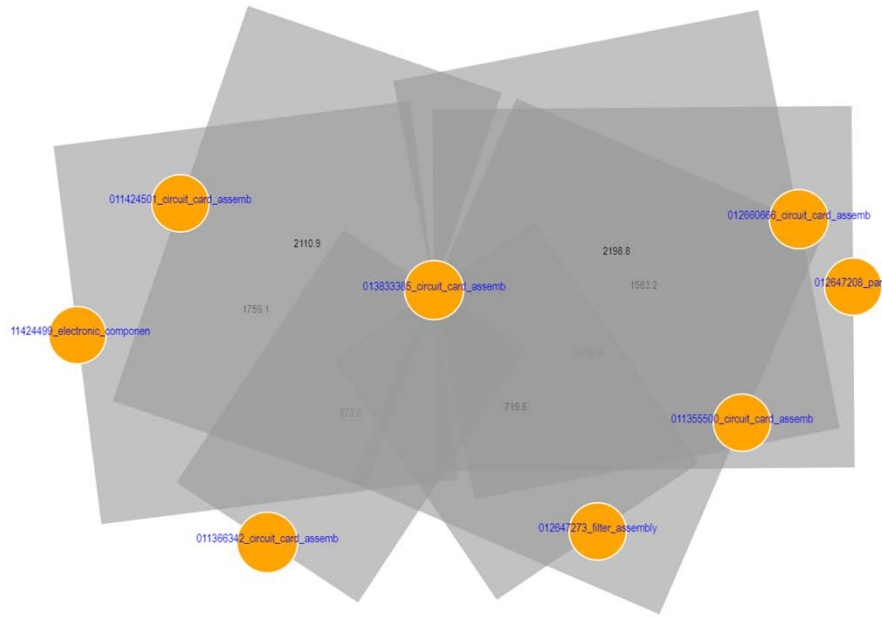


Figure 27. NSN 013833385 and its associations, which had the highest lift of the experiment.

While there were only five failures associated with 013833385, each time it failed there was over 500 days wait time for the replacement, as there are zero held in A condition and zero held in F condition in the supply system. This could be representative of a replacement part with a new NSN being implemented, but there is still current demand data being shown for this part. Also, each time this part failed, there were orders for each of the seven other related parts by the same ship on the same day, as shown by the thick grey bars and high lift values. Overall, when using the 54,246 rows of demand in the merged dataset, there were 242 demands among 95 individual NSNs that had greater than two days of delay time based on the supply system not having supply readily available. This is whether the supply was CONUS or OCONUS and does not include possible additional delays for shipping if the part is not forward staged. Of the 242 lines of demand, 187 of them had supply related delays of two months or more.

While there were many promising findings among the correlations found in the CASREP data obtained, there was still a desire for higher resolution in the data and a solid way to show goodness in the form of improvement in AO being directly affected by carrying the parts selected. In the search for data, a new project in development called

PRiSM, ran by a team in NSWC Corona in Southern California, and led by joint efforts between NAVSEA and NAVWAR was discovered. The next section will cover an attempt to use LLA data to verify the selections made by PRiSM, and vice versa, and to show improvement in AO made possible by comparing the datasets with the status quo.

B. COMPARING LLA TO PRISM AND HISTORICAL DATA

From Chapter III of this thesis and based on the PRiSM guide, the PRiSM project focuses on three key areas (Naval Information Warfare Systems Command (NAVWAR) 2020);

- To utilize the NSWC Corona Predictive Risk Sparing Matrix (PRiSM) Analysis to apply predictive logistics to increase AO and decrease on-site tech assists, MLDT and system Mean Down Time.
- Success is measured by the validation of PRiSM analytics to accurately identify high failure parts against actual CASREP/Non-CASREP incidents and the ability to support those parts onboard the ship or in FDNF locations that occur for CSG/ARG deployers.
- Identify validated Fleet C4I parts requirements to determine the appropriate combination of actions to implement.

Upon discovering this project in development and working together over the phone and in person during a meeting at NAVWAR, it was agreed upon that the direction of this thesis is much aligned with the direction that NSWC Corona is taking. It was determined that LLA could be used to verify the results from PRiSM and that the extra data collected from MRDB and analyzed using PRiSM could be used to supplement the method used in LLA.

To start, the parts selected from the merged dataset with the highest lift values were compared to the parts selected by the PRiSM team. Keeping in mind that nine years of data and a large amount of demands across multiple IMEC 3, 4, and 5 parts was being compared to one year of demand data for C4I systems only, but with very high levels of information available on each part. When comparing the datasets, 11 recommended NSNs matched, covering 587 demands over nine years. The p values for the chance of failure using the combined data were all below .05, meaning we reject the null hypothesis that the values are not related. Additionally, every part selected except for one was listed as either a single point of failure or a loss of redundancy for a critical C4I system and many had a low

likelihood of failure and would not be caught in the normal COSAL process, which is what this thesis set out to find. The values are listed in Table 1.

Table 1. Number of demands, P values, impact, and likelihood of failure for the 11 matching NSNs between LLA and PRiSM analysis.

NSN	#of demands	P Value	Impact	Likelihood of Failure
014295698	65	.003184	Loss of Redundancy	Most Likely
014636984	42	.004548	Single Point of Failure	Least Likely
014636990	3	.001819	Single Point of Failure	Least Likely
014639935	70	.003714	Single Point of Failure	Least Likely
014639937	20	.001289	Single Point of Failure	Least Likely
014685954	179	.00902	Non-Mission Essential	Less Likely
015415748	43	.002198	Single Point of Failure	Most Likely
015428626	39	.00235	Single Point of Failure	Least Likely
015434470	13	.00091	Single Point of Failure	Most Likely
015438059	54	.001971	Single Point of Failure	Less Likely
015553364	69	.00144	Single Point of Failure	Least Likely

Once the 11 parts that showed high lift and matched strongly between the two analyses were selected, further analysis was done in an attempt to show the improvement by carrying them onboard the ship, forward stocking in a close location to the ship, maintaining A condition parts, and not maintaining any A condition parts in the case of a failure. To do this, an assumption for shipping time had to be made. Shipping time is variable, but has a mean of 42 days to 7F, and using two standard deviations the range of 30 to 64 days was used was a random number generator for each row of demand. Also included in the calculation was average downtime for each NSN, calculated by using the

downtime across the range of demands. It's worth considering that some of the NSNs had many zero-day downtimes, but than a large jump once the A condition supply ran out, but for the purpose of this analysis and calculating average AO, the average was still used. Also calculated was the total days of downtime caused by lack of the specific NSN across the 7F ships, total days downtime caused by lack of shipping, and total days lost by repair time if the part was onboard. Some of the impacts were graphed as shown in Figures 28, 29, and 30.

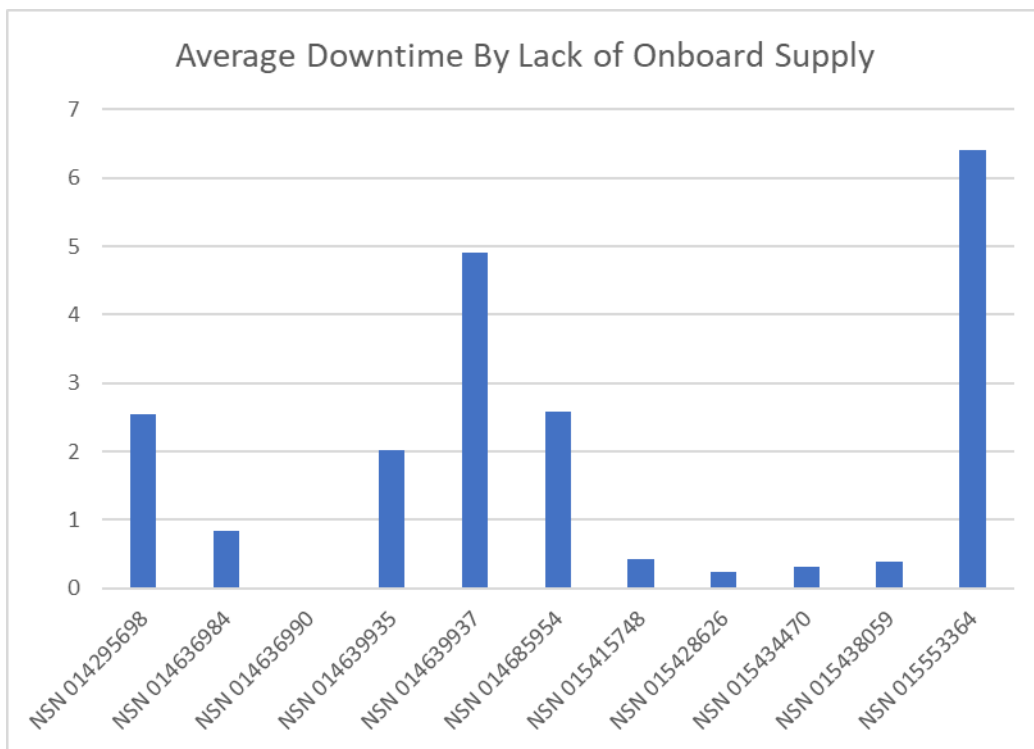


Figure 28. Average downtime caused by lack of onboard supply

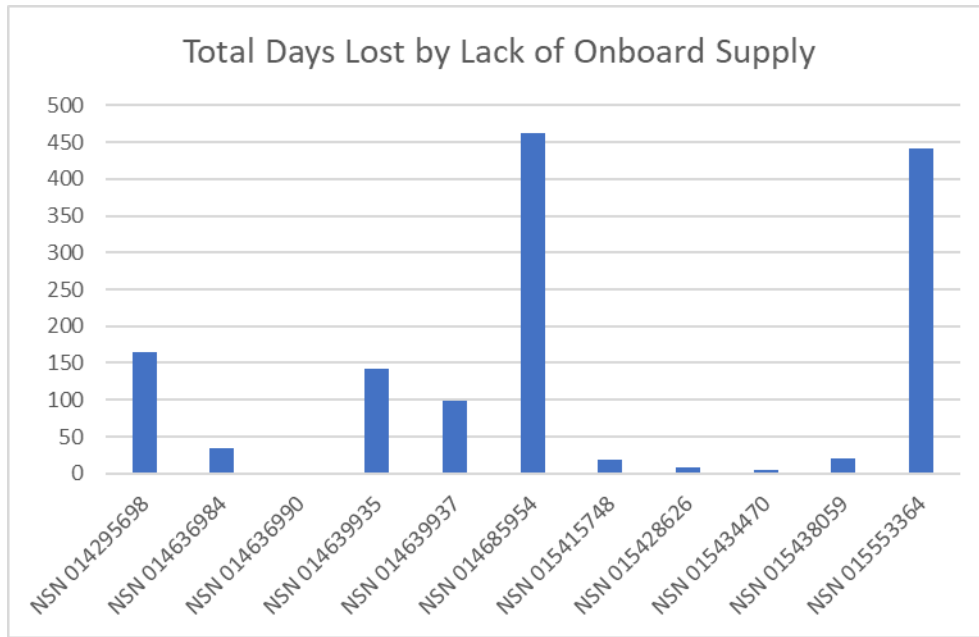


Figure 29. Total days lost per NSN by lack of onboard supply

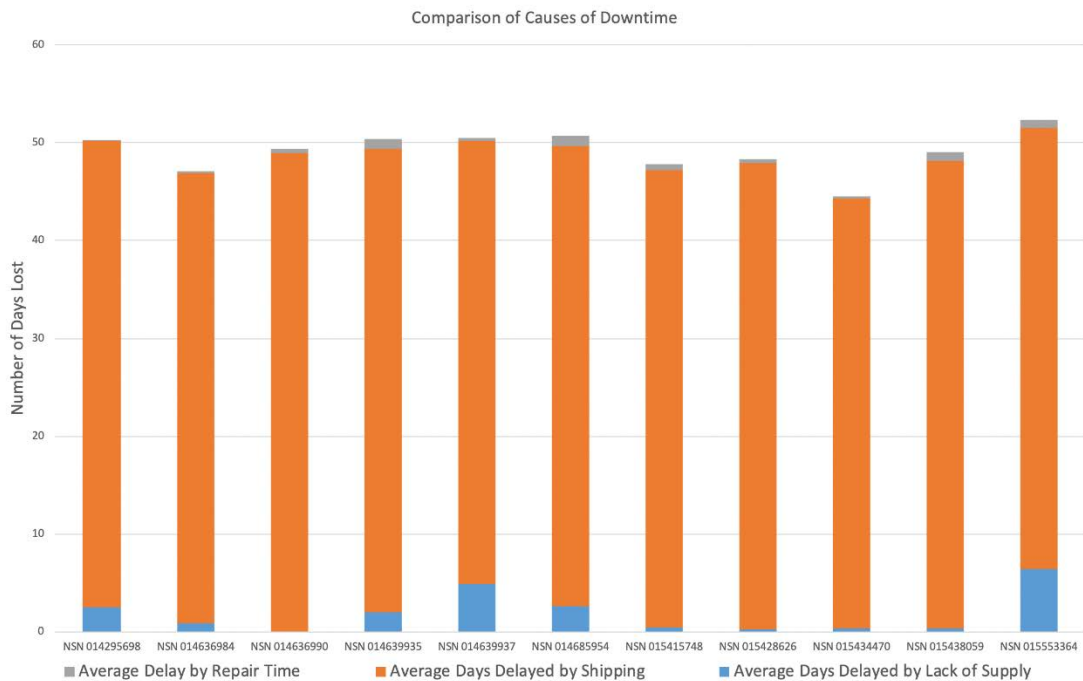


Figure 30. Comparison of causes of downtime by NSN

As can be seen in the above figure, and as expected, the largest cause of downtime on average is shipping if the part is not available in a location close to the ship. In this calculation the lack of supply is significant, but also may be underrepresented as the large outliers in some cases are averaged with the demand that was immediately filled. As stated earlier in Chapter IV, some demands jumped to over one year of delay once the A condition supply ran out. Also as expected, onboard repair time was the lowest cause of downtime, but it's worth noting that some parts (not in this list) require onboard technical assistance to repair and can have a long delay waiting for the representative to make it to the ship, especially when it is forward deployed and in a remote location. To compare the causes better in each scenario where the supply can be available, the improvement from no A condition supply, to A condition supply CONUS, to supply onboard was calculated next.

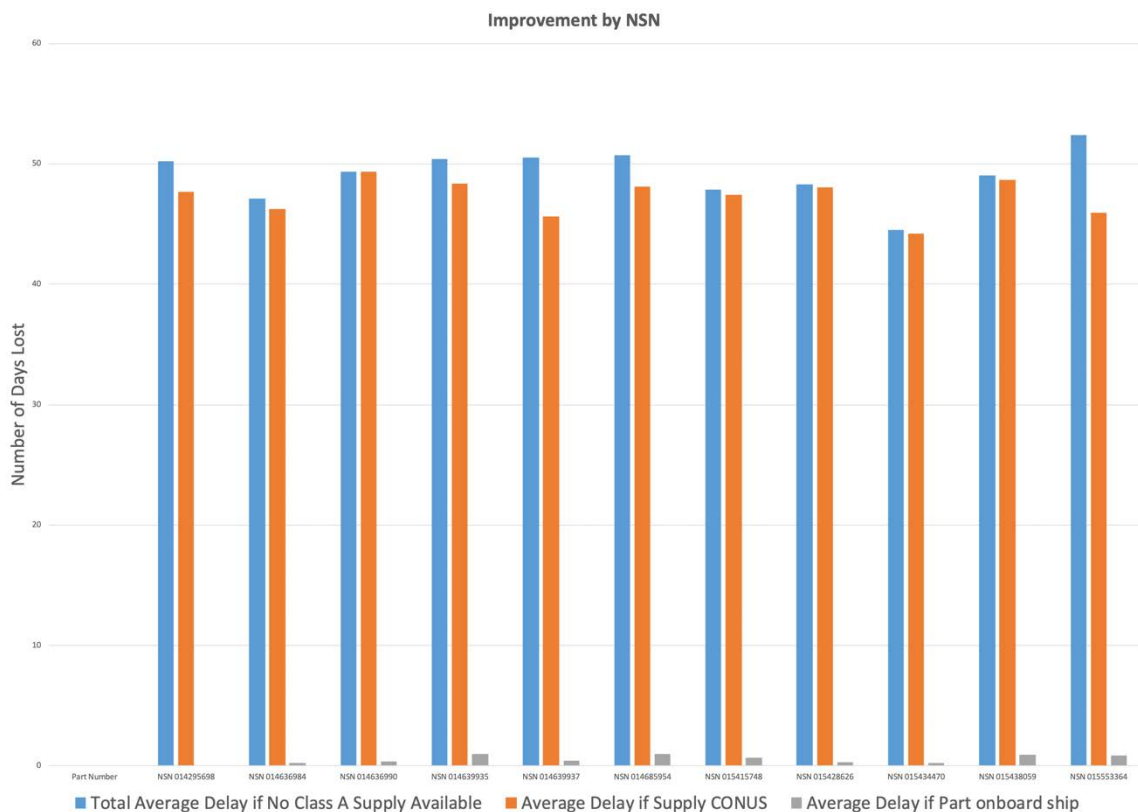


Figure 31. Improvement by NSN across the potential levels of supply available

The next step to show improvement, and to obtain the goal of this thesis was to calculate AO for each of the NSNs selected. As stated in Chapter III, AO can be calculated once mean downtime and mean time between maintenance (MTBM) is calculated. The below chart shows the difference in AO using the program manager (PM) recommendations for carrying the selected NSNs versus the improvement that would be gained if carrying the parts that are selected from this analysis, calculated by both the average observed MTBM and the worst case MTBM calculated.

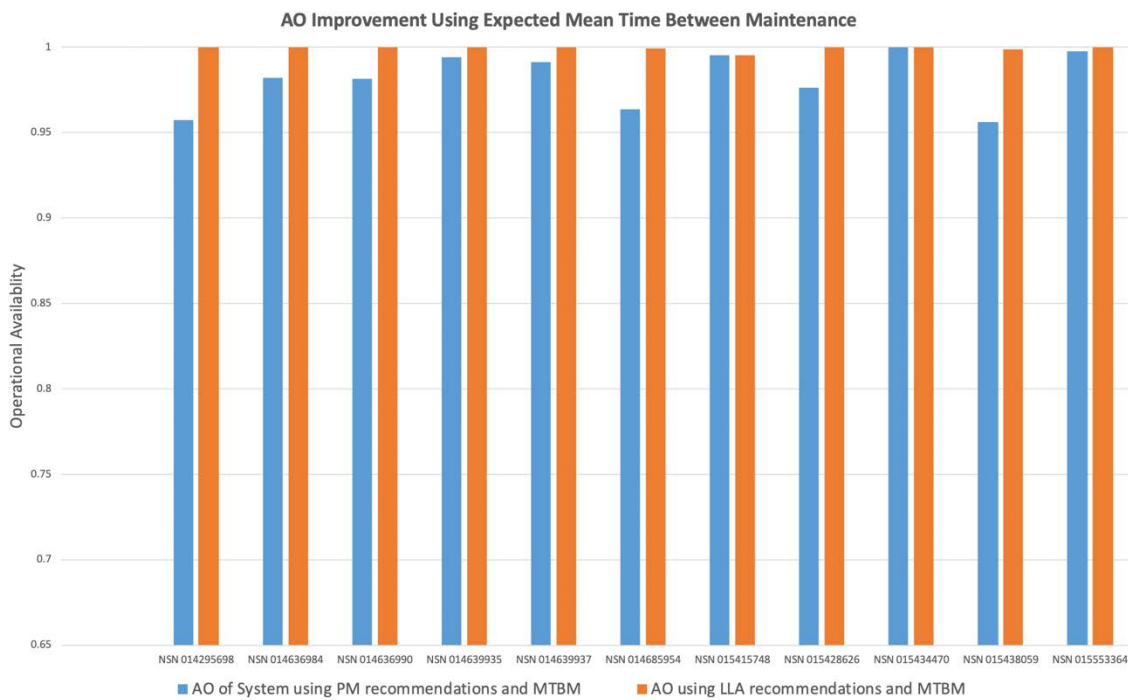


Figure 32. AO improvement comparing PM recommendation to the analysis outcome

It is worth noting above, in some cases the PM did *recommend* carrying the part onboard, but if supply was out, and if the part was not part of the COSAL AO would be lower. Even though it shows the same AO due to this, it is listed to show that all levels agree, and the parts should be carried. As stated in Chapter III, these parts have been carried during test deployments based on PRiSM recommendations and AO was improved by 67%. Figure 33 shows the worst case MTBM AO improvement calculated.

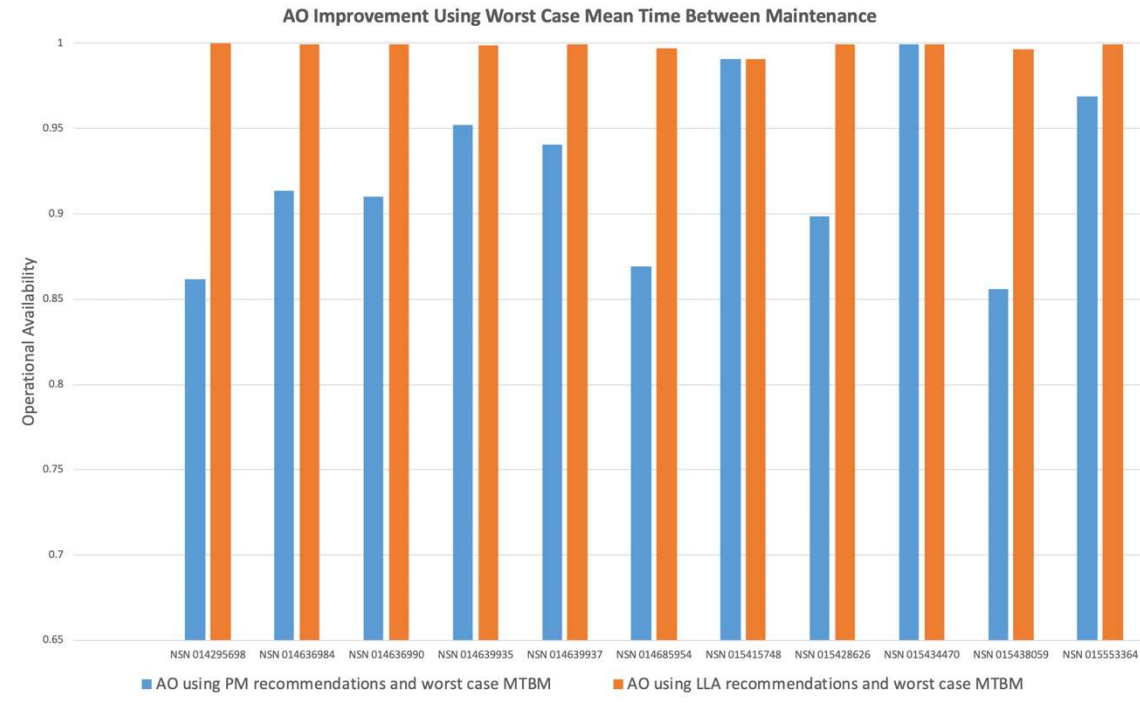


Figure 33. Chart showing AO improvement considering worst case MTBM

Considering that MTBM is volatile based on operating conditions, personnel operating the equipment, or human error, worst case MTBM may be a better comparison to reflect the volatility. When attempting to build a peacetime model that can be later used to simulate higher phases of military operations, the worst case MTBM should be the starting point. While it can be said that AO will improve based on almost any part carried, these C4I parts are all small enough to be carried onboard, all except for one listed as potential redline parts, and are all currently not carried as part of the COSAL due to low demand. If selecting additional parts to carry or forward stage to gain the maximum improvement in AO, these would be a good start based on the historical data and the impact calculated.

THIS PAGE INTENTIONALLY LEFT BLANK

V. CONCLUSIONS AND FUTURE WORK

A. CONCLUSIONS

As mentioned in Chapter I, the goal of any logistician should be to work himself/herself out of the job. This means making the process as seamless, efficient, and error-proof as possible. In searching through prior research and in researching this topic, there is constant room for improvement in the field of logistics, particularly when it comes to improving AO.

With limited resources and funding available, and critical time constraints for forward-deployed ships, it may never be possible to reach 100% AO for any given set of circumstances. NAVSUP has listed AO levels to meet, 85% for critical systems and 65% for non-critical systems, but at those levels occasional redline failures still happen. This thesis has shown that within the same constraint levels it is possible to improve AO for forward-deployed ships by using predictive analytics algorithms to better determine which parts that may fail.

This thesis considered time in service for the parts, MTBF, downtime, and AO among other parameters. When looking at what related failures happen along with these variables, trends have been identified and parts have been identified that were shown to have real-world failures on deployments, which would have been a redline or at least a loss of redundancy if they weren't onboard as part of the COSAL.

Now that demand and failure data is fed into systems such as MRDB and N-ERP, there are many years' worth of data of hundreds of thousands of individual demands that can be analyzed in many different ways. As was determined during the research for this thesis, depending how the bi-grams were arranged multiple outcomes could be discerned. The benefit of doing the analysis this way is that the large datasets can be downloaded in nearly real-time terms and mined relatively quickly once the parameters are determined and scripted to prep the data for use in LLA. The next challenge is to determine the level of correlation to search for between demands, and to discern the noise from valuable data.

This can be done by comparing the output with real world results or by cross validating with other analysis like PRiSM.

This project is ongoing with the teams at NAVWAR and NSWC Corona to continue to use LLA to validate PRiSM as it is developed, and vice versa. There is still much room for growth and development in cleaning data to use the best parameters, to determine what aspects of the demands are most highly correlated to future failures, and to build the database of engineering specifications and other points that PRiSM is working to collect for analysis. As the analysis process is refined, other systems can also be analyzed, as PRiSM focuses primarily on C4I, and the initial LLA analysis focused on IMEC 3, 4, and 5 parts. Luckily, forward warehouse storage, supply ships, and even space onboard forward-deployed ships is enough that small parts and large parts can be placed in a more optimal location than where they currently are.

B. FUTURE WORK

To add another level of productive use for this process is in determining supply levels needed for a potential increase in military operations. As the Navy has operated in “Phase 0” or peacetime operations for the majority of the currently active personnels’ careers, it is currently unknown how the supply system would react to a real world naval engagement beyond Humanitarian Assistance and Disaster Relief (HADR), the occasional adversarial encounter, and training accidents. By selecting the most critical parts through the analysis done in this thesis, estimating the cause of failure, and determining how individual parts would handle different circumstances such as more use, harsher use, and so on, the increases in OPTEMPO can be simulated to show the increase in demand and the delays that result from it.

Initial models have been run using SIMIO, and a simplified version with Excel, using the current warehouse and manpower capabilities in place at DLA Distribution in Yokosuka, which is the main Navy supply hub for Seventh Fleet. The goal of the simulation is to minimize spikes in backlog due to increased demand. The less extreme these spikes in the sawtooth effect of the workflow are, the better. Some areas that could logically be optimized were not considered, such as working on Sunday, or adding excessive overtime

when there is not a special event (like an unexpected CSG deployment). While there are areas that can be expanded to add complexity, i.e., creating 7 individual warehouses vice averaging the rates into a warehouse, for this model a certain level of simplicity was used to start the process. The starting completion rates in SIMIO are shown in Table 2.

Table 2. Demand rates and processing rates for 7F supply

Initial Settings	Weight (%)	Arrival Rate	Processing Rate (in minutes)
Low-Pri Demand Source	60	1/30 minutes	N/A
High-Pri Demand Source	40	1/36 minutes	N/A
DDYJ Completion	100	First in First Out	Triangular(1,3,10)
Military Delivery Completion	50 (50% local deliveries)	First in First Out	Triangular(.1,.2,.3)
FedEx DHL Truck (for foreign port)	25 (with a random distribution for at sea or not at sea)	First in First Out	Triangular(.01,.02,.03) (Basically immediate service)
MSC COD (if at sea)	25 (with a random distribution for at sea or not at sea)	First in First Out	Triangular(.1,.2,.3)
Local Completions	50	Dependent on Worker	N /A
Non-Local Completions	50	Dependent on Worker	N/A

A benefit of SIMIO is that in addition to arrival and completion rates being programmed with a choice of various distributions, work schedules can also be programmed. In this case, from 1–19 workers were programmed in, varying by the day and the hour including lunch breaks and turnover times. Arrival rates were also varied by day and by hour, but averaged over the entire demand vice by part or ship. By doing this, small spikes can be viewed during turnover and on weekends, and the backlog slowly flattens back out during the workday. The final SIMIO example programmed resulted in a very

similar result to what is actually happening day to day in 7F during peacetime. The SIMIO model created daily sawtooth plots that can be seen in Figure 34.

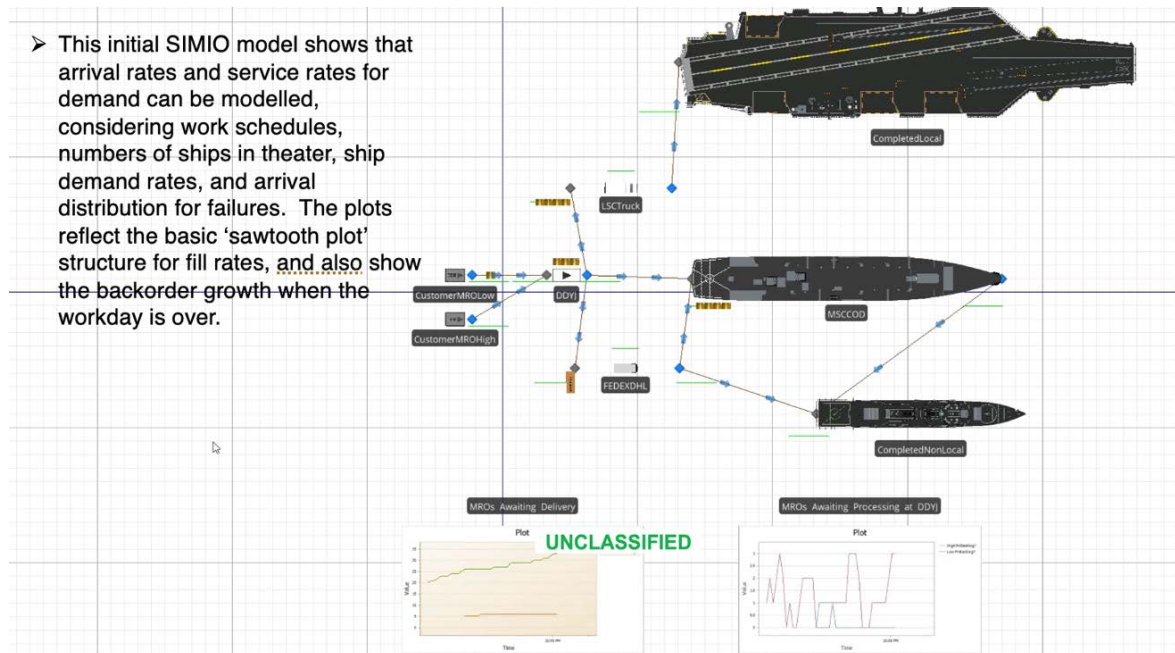


Figure 34. SIMIO model running and showing sawtooth plot created by spikes in demand causing backlog.

The next steps would be to simulate increased demand, such as in 20%, 40%, and 60% increases, but even at the 20% level the system quickly became unstable, meaning that more demand was coming in than being completed. To better predict the outcomes of increased demand, individual parts that are selected should be researched more thoroughly to determine individual reaction to increased OPTEMPO, and expected increase in workers should also be determined. After doing that, the system can be simulated and optimized for Phase 0, Phase 1, Phase 2, and other events that may cause different demand data than is currently experienced. If this can be modelled and validated, that would provide good grounds for expected logistical effectiveness during these various future events during wargames and operational planning events.

LIST OF REFERENCES

- Barrow, Robert H. "Q&A: Marines' Barrow Backs SALT—and Conventional Rearming." *San Diego Union*, November 11, 1979 , C4.
- Clausewitz, Carl von. 1989. *On War*. Translated by Michael Howard and Peter Paret. Princeton, NJ: Princeton Uni. Press.
- Furer, Julius Augustus. 1960. *Administration of the Navy Department in World War II*. Washington, DC: U.S. Government Printing Office.
- Grammarist: Idiom Definitions. "Tail Wagging the Dog." n.d. Accessed May 23, 2020. <https://grammarist.com/idiom/tail-wagging-the-dog/>.
- JMP Statistical Discovery. "Statistical Software." n.d. Accessed May 23, 2020. https://www.jmp.com/en_us/home.html.
- Joint Chiefs of Staff. 2017. *Joint Planning*. Joint Pub 5–0. Washington, DC: Joint Chiefs of Staff. https://www.jcs.mil/Portals/36/Documents/Doctrine/pubs/jp5_0_20171606.pdf.
- Kang, Keebom, Susan M. Sanchez, and Kenneth H. Doerr. 2006. *A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics*. NPS-LM-06-037. Monterey, CA: Naval Postgraduate School, Graduate School of Business and Public Policy. <http://hdl.handle.net/10945/33801>.
- Kuehn, Kurt. 2014. "The Role of Logistics in an Age of Disruption." PressReleasePoint. March 28, 2014. <http://www.pressreleasepoint.com/role-logistics-age-disruption>.
- Lopez, Al, Bruce Cormany, and Nick Hymer. 2019. "SPAWAR FRD 100 Maritime Sustainment Summit." Naval Information Warfare Systems Command (NAVWAR) San Diego, CA, Final Draft.
- Lopez, II, Jose A. 2019. "Operational and Mission Readiness Impact of Additive Manufacturing Deployment in the Navy Supply Chain." Master's thesis, Naval Postgraduate School. <http://hdl.handle.net/10945/63477>.
- Mattis, Jim. 2018. *Summary of the 2018 National Defense Strategy of the United States of America*. Washington, DC: Department of Defense. <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>.
- Naval Information Warfare Systems Command (NAVWAR) 2020. *NAVWAR Predictive Logistics and Predictive Risk Sparing Matrix (PRiSM) Standard Operating Procedures*. San Diego, CA: NAVWARSYSCOM Fleet Readiness Directorate.

- Quantum Intelligence. 2015. *Collaborative Learning Agent (CLA): Users' Guide. Version 2.0*. Santa Clara, CA: Quantum Intelligence, Inc.
- R Foundation. 2020. "R: What Is R?" 2020. <https://www.r-project.org/about.html>.
- Reid, Elaine D. 2011. "Social Network Collaboration for Crisis Response Operations Developing a Situational Awareness (SA) Tool to Improve Haiti's Interagency Relief Efforts." Master's thesis, Naval Postgraduate School. <http://hdl.handle.net/10945/5670>.
- Simio n.d. "The Story of Simio." Accessed May 23, 2020. <https://www.simio.com/about-simio/>.
- Stevens, Edwin J., and Ying Zhao. 2019. "Using Lexical Link Analysis as a Tool to Improve Sustainment." White Paper, Naval Postgraduate School, Monterey, CA.
- Sun Tzu. n.d. *36 Strategems of War*. Asiapac 2003
- World Map with Nations. 2009. "US Navy Fleets Areas of Responsibility." Wikimedia Commons. [https://commons.wikimedia.org/wiki/File:USN_Fleets_\(2009\).png](https://commons.wikimedia.org/wiki/File:USN_Fleets_(2009).png).
- Zhao, Ying, Douglas J. MacKinnin, and Shelly Gallup. 2015. "Collaborative Learning Agents (CLA)." Users Guide, 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