AFIT-AM-14-009

ACQUISITION RESEARCH PROGRAM
SPONSORED REPORT SERIES

Enterprise Requirements and Acquisition Model (ERAM) Analysis and Extension

20 February 2014

Dr. John M. Colombi, Assistant Professor, and
Dr. J. Robert Wirthlin, Assistant Professor

Department of Systems Engineering and Management

Air Force Institute of Technology

Dr. Teresa Wu, Associate Professor

School of Computing, Informatics, and Decision Systems Engineering

Arizona State University

Approved for public release; distribution is unlimited.

Prepared for the Naval Postgraduate School, Monterey, CA 93943.

The research presented in this report was supported by the Acquisition Research Program of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request defense acquisition research, to become a research sponsor, or to print additional copies of reports, please contact any of the staff listed on the Acquisition Research Program website (www.acquisitionresearch.net).

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL

Abstract

The Department of Defense (DoD) acquisition system is challenged by schedule and cost overruns that can be attributed to a complex acquisition process. This process drives great research interests in exploring intervention strategies that would help reduce program delays. However, quantitatively evaluating the impact of new policy has been limited due to the lack of system models with appropriate fidelity. The application of a simulation model to address this challenge thus becomes a promising approach. In this research, we explore the application of the Enterprise Requirements and Acquisition Model (ERAM), a discrete event simulation of the DoD acquisition system, to quantitatively examine several interventions. Recent studies indicate that policies that address (1) scope growth, (2) acquisition process variability, and (3) program technology maturity should be investigated because they may have a significant impact on reducing program completion time. Thus, the effect of scope growth frequency and size, technology maturity, and changing variability and mean process times in several government and contractor pre–Milestone C activities are investigated. Additional research includes an engineering bottleneck analysis, the effects of requiring that all program types conduct an analysis of alternatives (AoA), and modeling extensions for understanding post–Milestone C space launch delays. The insights gained from the simulation experiments can potentially help formulate new policies to improve DoD acquisition.

Keywords: defense acquisition program, process modeling, simulation, sensitivity analysis, bottleneck analysis, systems engineering processes

THIS PAGE INTENTIONALLY LEFT BLANK

Acknowledgments

We want to thank our students for their hard work on these research efforts.

Danielle Worger is a student in Barrett, the Honors College at Arizona State University. She is concurrently pursing her BS and MS degrees in industrial engineering at the School of Computing, Informatics, and Decision Systems Engineering of Arizona State University.

Eugene Rex Jalao is a PhD candidate of industrial engineering from Arizona State University. He is also an assistant professor on study leave from the University of the Philippines, Industrial Engineering and Operations Research Department, where he obtained both his BS and MS degrees in industrial engineering. His research interests are multi-criteria decision-making, business intelligence and analytics, and supply chain management.

Christopher Auger, Lars Baldus, and Brian Yoshimoto are students at the Air Force Institute of Technology. Majors Auger and Yoshimoto are Air Force officers, and Lieutenant Colonel Baldus is from the German Air Force. All are pursuing MS degrees in systems engineering.

THIS PAGE INTENTIONALLY LEFT BLANK

About the Authors

Dr. John Colombi is an assistant professor of systems engineering at the Air Force Institute of Technology (AFIT). He received his PhD in electrical engineering from AFIT. Dr. Colombi served 21 years in the U.S. Air Force, with experiences in C4ISR systems integration and engineering activities, program management, and information systems security and communication networking research. Currently, he leads research in human systems integration, enterprise and system architecture evaluation, complex adaptive systems, and acquisition policy analysis.

John M. Colombi, PhD
Department of Systems Engineering and Management
Air Force Institute of Technology
2950 Hobson Way
Wright-Patterson AFB OH 45433
Tel: 937-255-3636 x3347
E-mail: john.colombi@afit.edu

Dr. J. Robert Wirthlin, LtCol, USAF, is an assistant professor of engineering systems at the Air Force Institute of Technology. He holds a PhD and MS from the Massachusetts Institute of Technology, and a BS from the United States Air Force Academy. He is a member of the faculty of both the AFIT Graduate Engineering Management and Systems Engineering programs. His research interests include new product development, risk, and application of lean-to product development.

Dr. J. Robert Wirthlin
Department of Systems Engineering and Management
Air Force Institute of Technology
2950 Hobson Way
Wright-Patterson AFB OH 45433
Tel: 937-255-3636

Teresa Wu is an associate professor of industrial engineering at the School of Computing, Informatics, and Decision Systems Engineering of Arizona State University. She is the NSF CAREER award winner (2003) and the ASU tenure and promotion exemplar winner (2006). She received her PhD from the University of Iowa. Her main areas of interests are in distributed decision support, systems simulation, distributed information systems, and supply chain management.

Dr. Teresa Wu
School of Computing, Informatics, and Decision Systems Engineering
Arizona State University
1151 S. Forest Ave.
Tempe, AZ 85281
Tel: (480) 965-4157
E-mail: teresa.wu@asu.edu

THIS PAGE INTENTIONALLY LEFT BLANK

AFIT-AM-14-009

ACQUISITION RESEARCH PROGRAM SPONSORED REPORT SERIES

Enterprise Requirements and Acquisition Model (ERAM) Analysis and Extension

20 February 2014

Dr. John M. Colombi, Assistant Professor, and
Dr. J. Robert Wirthlin, Assistant Professor

Department of Systems Engineering and Management

Air Force Institute of Technology

Dr. Teresa Wu, Associate Professor

School of Computing, Informatics, and Decision Systems Engineering

Arizona State University

THIS PAGE INTENTIONALLY LEFT BLANK

Table of Contents

Executive Summary

The Department of Defense (DoD) acquisition system is challenged by schedule and cost overruns that can be attributed to a complex acquisition process. This process drives great research interests in exploring intervention strategies that would help reduce program schedule. Using the Enterprise Requirements and Acquisition Model (ERAM), a discrete event simulation of the DoD acquisition system, this research seeks to determine the acquisition activities that contribute significantly to the time a program takes to reach Milestone C. ERAM was created and validated by Dr. Robb Wirthlin, LtCol, USAF, in 2009 during his MIT doctoral dissertation. The Air Force Institute of Technology (AFIT), Department of Systems Engineering and Management, has continued to apply and extend this simulation over the last few years. Such work has examined pre-acquisition effects, rapid acquisition (quick reaction capability and urgent operational needs), and unique space acquisition extensions.

This research used Monte Carlo analysis and *t*-tests, along with sensitivity analysis and critical path analysis to determine the impact of excursions in both the pre–Milestone B and C processes. This research examines the role that an acquisition model can provide to analyze process changes to improve program completion time. ERAM suggests that while some process enhancements do improve and the end-to-end acquisition process would be worthwhile, others have little or no effect.

This research first examines the effect of systems engineering processes by varying the mean completion times and having them occur earlier in the acquisition process. In particular, the impact of requiring programs, regardless of Acquisition Category (ACAT) to complete a formal analysis of alternative was investigated. Recent studies also indicate that policies that address (1) scope growth, (2) acquisition process variability, and (3) program technology maturity should be investigated because they may have a significant impact on reducing program completion time. However, quantitatively evaluating the impact of the new policy has been limited due to the lack of system models with appropriate fidelity. The application of a simulation model to address this challenge thus becomes a promising approach. We explore the application of ERAM to examine scope growth, technology maturity, and various process completion times. Lastly, as an extension to ERAM, we examine the ability to add specialized process models, such as space launch acquisition processes. This extension demonstrates how empirical and pedigreed data can be incorporated and supplemented by subject matter expert input to extend ERAM for a complex iterative model for acquisition areas of interest.

The insights gained from all these simulation experiments can potentially help formulate new policies to improve DoD acquisition.

Specific observations were as follows:

- Previously, the analysis of alternatives (AoA) was only required for all ACAT 1 programs and ACAT 2 and 3 programs with sufficient funding. If significantly more ACAT 2 and 3 programs require AoA, their overall schedule increases significantly, at the 95% confidence.
- Often taking a year (360 days) for the AoA, making the AoA more consistent (less variable), has no effort on ACAT 1 programs reaching Milestone C (MS-C).
- However, reduction of the AoA mean down to 202 days has a significant effect on all programs (ACAT 1–3) reaching MS-C.
- A critical path analysis can capture which branches of activities cause significant delays, prior to Milestone B (MS-B). Simulation indicated that the systems engineering (SE) activities and their communication with requirements are the bottleneck of the pre–MS-B portion of the acquisition system.
- Simulation also indicated that focusing on reforms that address this bottleneck has the potential to decrease the total time spent on MS-B activities by approximately 7%; this corresponds to a process time reduction of approximately six months.
- Findings suggest that from a purely statistical simulation standpoint, decreasing the amount of scope growth that occurs during acquisition programs will not have a statistically significant impact on end-to-end program time. In addition, it appears that programs are sensitive only to very large increases in scope growth, with the pre–MS-C activities having a greater sensitivity than pre–MS-B activities.
- Results indicated improvements that reduce the mean process time for the entire fabrication; assembly and testing portion of the acquisition process could have significant effects.
- Pressure to push a program through the milestone with optimistic Technical Readiness Level (TRL) assessments may be preventing the acquisition system from reaching reduced cycle times. From literature and our simulation research, it can be seen that from a statistical standpoint, the potential schedule benefits of being critical of the technology maturity level at MS-B outweigh the schedule penalties of implementation

- The Space Launch process includes six types of delays experienced by space acquisition programs, and demonstrates a bimodal distribution beyond MS-C.

The organization of this technical report includes four investigations. The first (found in Section II) examines the effect of analysis of alternative (AoA) process time interventions. A second article contained in Section III was submitted to the Naval Postgraduate School (NPS) Acquisition Research Symposium for 2013 and is entitled “Bottleneck Analysis on the DoD Pre-Milestone B Acquisition Processes.” This work examines the pre–MS-C engineering activities and their sensitivity. Section IV is a paper entitled “Intervention Strategies for the DoD Acquisition Process.” This work quantitatively captures the effects of (1) scope growth, (2) acquisition process variability, and (3) program technology maturity on completion time. Lastly, Section V is a paper entitled “Modeling Space Launch Process Delays to Improve Space Vehicle Acquisition Planning.”

THIS PAGE INTENTIONALLY LEFT BLANK

Enterprise Requirements and Acquisition Model Analysis and Extension

Effect of the Analysis of Alternatives on the DoD Acquisition System

The Enterprise Requirements and Acquisition Model (ERAM) is a discrete event simulation that models the major tasks and decisions within the Department of Defense (DoD) acquisition system. A majority of DoD acquisition projects are being completed behind schedule and over budget. ERAM suggests process improvements can have salutary effects. Hence, enhancements in improving the end-to-end acquisition process would be worthwhile. Until 2008, the analysis of alternatives (AoA) process was a mandatory task for acquisition category (ACAT) level 1 projects. As such, expected program completion time for ACAT 2 and ACAT 3 categories is shorter. Since 2008, the AoA became a required procedure for all programs. However, to the best of our knowledge, the impact of requiring all programs to complete an AoA has not yet been studied in literature. This research addresses this gap with two main contributions. First, this research seeks to quantify the amount of delay on total completion time when the AoA is required for all ACAT programs. Second, the sensitivity of the processing time and variability of the AoA process is simulated, and its effect is studied on total program completion time. Viable policies and intervention strategies are then inferred from these contributions to further improve acquisition program completion time.

Introduction

It is a known fact that a large number of DoD projects are being completed behind schedule and over-budget (Schwartz, 2010). A Government Accountability Office (GAO) report released in 2009 states that for the DoD's 2008 portfolio, on average a program faced a 22-month delay and exceeded the original budget (Sullivan et al., 2009). Generally, total cost growth has been consistent over the past few decades with a recent assessment by Arena et al. (2006) of 1.44 or 44% growth. The current DoD acquisition system, which is composed of three separate and distinct processes, including the Joint Capabilities Integration Development System (JCIDS); the Planning, Programming, Budgeting & Execution (PPBE) process; and the formal acquisition development system outlined by the DoD 5000 series of instructions, does not exist in a static environment. The system is constantly being adjusted, either through policy changes or statute (Chairman of the Joint Chiefs of Staff [CJCS], 2012; Under Secretary of Defense for Acquisitions, Technology, and Logistics [USD(AT&L)], 2008; Weapon Systems Acquisitions Reform Act, 2009).

Because the acquisition process is a large, complex, socio-technological system, it is difficult to determine which processes or factors affect performance metrics, such as time, cost, and resource utilization. Hence, alternative modeling tools to improve the DoD acquisition process are the subject of current research.

In 2009, a discrete event simulation (DES) model called the Enterprise Requirements and Acquisition Model (ERAM) was developed by Wirthlin (2009). This model simulates the actual acquisition processes of the DoD using the Air Force implementation of acquisition processes for ACAT levels as the basis of the model. This is done in order to provide further insight and understanding of the complex system's behavior. However, this research did not include new policies set forth by the DoD, specifically the DoD Instruction 5000.02 that requires all programs to go through an AoA process (USD[AT&L], 2008) before reaching Milestone A. The AoA is a requirement for all military acquisition programs. By definition, the AoA is an analytical comparison of a multiple alternatives process that needs to be performed prior to committing resources to a given acquisition program (Georgiadis, Mazzuchi, & Sarkani, 2012). According to the DoD 5000.02 Instruction,

> The AoA shall focus on identification and analysis of alternatives, measures of effectiveness, cost, schedule, concepts of operations, and overall risk. The AoA shall assess the critical technology elements (CTEs) associated with each proposed materiel solution, including technology maturity, integration risk, manufacturing feasibility, and, where necessary, technology maturation and demonstration needs. (USD[AT&L], 2008)

Through this requirement, an implicit assumption is being made that this step will actually shorten the overall life-cycle development time for a given acquisition program and increase the quality of the final form of the materiel solution. However to the best of our knowledge, this policy effect has not yet been quantitatively studied in existing literature. Nevertheless, the AoA theoretically contributes to longer program time completion of all DoD acquisition projects because it is an additional task that must be performed during the process. Hence potential policies could be developed to counter the effect of requiring the AoAs (e.g., acknowledging better quality solutions earlier in development, which could be easily translated to viable policies to further improve not only the duration but the entire end-to-end DoD acquisition system).

Against this background, this research addresses these limitations by performing additional simulation and statistical analysis on the ERAM model. The primary goal of this research is to determine the effect on the total acquisition program completion time by requiring all ACAT programs to go through the AoA process. Furthermore, the effect of reducing the variability and the time spent on the AoA process is studied and potential intervention strategies are developed.

Additionally, this research provides viable policies and discussion points from the intervention strategies that could further reduce program completion time.

Review of Literature

DoD Enterprise Requirements Acquisition Model

ERAM was originally an Arena simulation model that provided the foundations for research on applying discrete event system simulation to the DoD acquisition process. Extensive validation of the model is done by comparing the performance results of the model to the actual DoD acquisition data and expert reviews. Initially, 20 intervention strategies are explored using the ERAM model to assess the potential of the total completion time reduction of DoD acquisition projects. If all 20 interventions are implemented, a 20% reduction in the total program time could be realized. It has been found that the most effective interventions to improve the system are those that reduce the variability of the processes. Since its publication in 2009, the Arena model has been translated to ExtendSim and extended by the Aerospace Corporation's Developmental Planning and Architectures Division for use in the Concept Development Center of the Space and Missile Systems Center at Los Angeles AFB, CA (Leach & Searle, 2011). Moreover, Montgomery (2012) provided the research for Aerospace to extend the model to further include ACAT 2/3 programs along with modeling the rapid acquisition process for space programs. Table 1 summarizes the different versions of the ERAM model. Please note the earlier version of ERAM (1.0) does *not* implement AoA on all ACAT projects.

Table 1. ERAM Versions
(Adapted from Houston, 2012)

Author	**Version Number**	**Changes**
Wirthlin (2009)	ERAM 1.0	Baseline translation from Arena to ExtendSim
Leach and Searle (2011)	ERAM 1.1	Updates by the Aerospace Design Team and served as new baseline model
	ERAM 1.2	Implemented new DoD 5000.02 policies
	ERAM 2.0	Incorporated the global variables that modify acquisition capabilities
	ERAM 2.1	Incorporated the JCIDS review process
Montgomery (2012)	ERAM 2.2	Added more capabilities for ACAT 2/3 and rapid acquisition process

Verification and validation of the baseline distributions included hand modeling, iterations of correction from feedback of experts in all three branches of acquisition, and comparison of schedule and budget information from the DAMIR and SMART databases to distributions of the schedule time of model-generated data (Wirthlin, 2009).

Implementation of the Analysis of Alternatives

There are also several articles that illustrate the implementation of the AoA in the DoD acquisition process. Cervantes, Enderton, and Power (2012) applied a rapid AoA implementation to find a NATO Special Operations Headquarters (NSHQ) Air Wing. Georgiadis et al. (2012) proposed an analytical multiple criteria decision-making methodology to handle the AoA. The AoA also plays a part in the replacement of the aging presidential helicopter fleet (Sullivan, 2012).

Research also explores policies to improve the implementation of the AoA in the end-to-end DoD acquisition system. For example, Schank (2012) identified several important factors for the success of the AoA, specifically, (1) the AoA must have a study plan that considers a wide range of alternatives and must be flexible in the analysis methodology; (2) oversight committees must manage effective relationships; (3) trade-off analysis should be conducted on all alternatives; and (4) good estimation and recognition of technical, design, and production risks is a must. A white paper from the Training and Doctrine Command (TRADOC) Analysis Center (TRAC) addresses two problems of the implementation of the AoA specifically: (1) by clearly describing and differentiating the purpose and scope of each AoA for each milestone decision, and (2) by clearly defining and describing “materiel solution” as it can be used interchangeably with the “alternatives” term (Training and Doctrine Command Analysis Center, 2011). Stadterman (2012) proposed improvements on the AoA in the Weapon Systems Acquisition Reform Act of 2009. Among these recommendations are that (1) all concerned parties within the AoA should build a working relationship; (2) the AoAs should focus on the decision choices and the decision space, and have an achievable, affordable, and operationally relevant set of criteria; and (3) the Army should follow a formal analytical process that supports the AoA throughout the acquisition process. Additionally, Ford,Housel, and Dillard (2010) improved the implementation of the AoA by incorporating benefits in the decision methodology through the use of a system dynamics model of a military operation and integrating it with a knowledge-value-added methodology. Furthermore, in a GAO report (Sullivan et al., 2011), the GAO recommended that the Joint Requirements Oversight Council establish a review mechanism for AoA earlier in the acquisition process. According to Roper (2010), all alternatives being considered within the AoA must go through consistent analysis methodologies and assumptions in order to ensure comparability. Hence, adopting these quality improvement suggestions also imply a reduction in the length and/or the variability in the AoA process. Doing so will allow AoAs to obtain consistent AoA process times.

Based on this review, the following gaps in literature can be gleaned:

- Currently, a quantitative study to assess the impact of requiring AoA implementation in all ACAT programs is lacking.

- Secondly, the effect of reducing the variability and length of the AoA process has not yet been addressed as a potential intervention to further improve the program acquisition completion time.

Simulation Analysis

This research is composed of three distinct phases. The first phase is a simulation study to determine the effect of the AoA for ACAT 2 and ACAT 3 programs on the total completion time. The second phase consists of two simulation experiments in which the variability and the mean of the AoA process time are reduced and its effect on total completion time is determined. Lastly, the third phase is the translation of results into recommendations and further research.

Requiring an AoA on a Percentage of ACAT 2 and 3 Programs

The original ERAM 1.0 model is utilized because it is the baseline model that did not require an AoA on all programs. By varying the ACAT 2 or ACAT 3 funding process, specifically the percentage of programs with funding already available for an AoA, the effect of requiring a certain percentage of programs to undergo an AoA on the total program completion time, or the time the program completes Milestone C (MS-C), is determined. Figure 1 presents the screenshot of the module. This module decides whether the program has enough funding to perform an AoA. If it is found that there is enough funding, the program will have to undergo an AoA.

ACAT 1 programs are not included in the analysis because AoAs were required previously. The baseline scenario is set as such the probability that the program would be required to undergo an AoA is 1%. This setting was set before the DoD policy change in 2008 on AoA implementation. Please note in Wirthlin (2009), only ACAT 2 and 3 programs with sufficient funding are needed to undergo an AoA. Simulation trials each with 3000 iterations are run separately for ACAT 2 and ACAT 3 programs. During each trial, the probability that a program would need to perform an AoA is increased incrementally, and its corresponding completion time is tabulated. Each increase is compared to the baseline trial. A *t*-test is then performed in order to determine whether there is a statistically significant difference between the trial results and the baseline trial in terms of the time until MS-C. Hence, the probabilities are incrementally increased until a statistical significant difference from the baseline is obtained at a 95% confidence level.

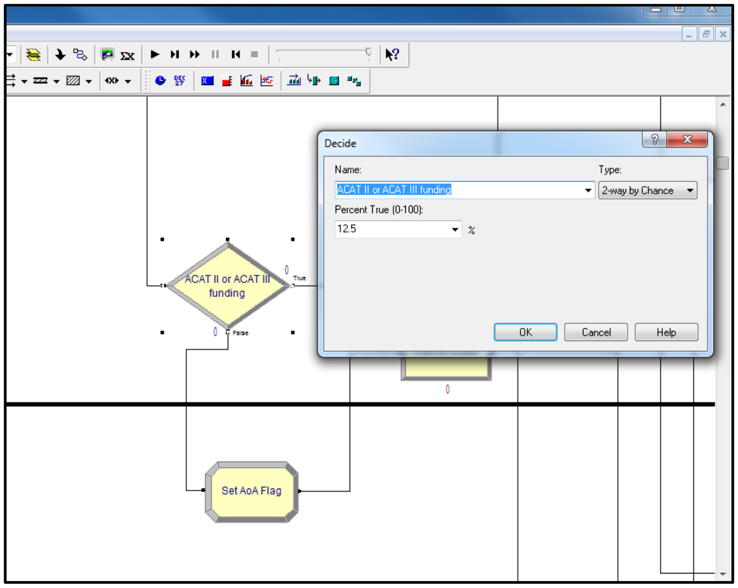

Figure 1. Arena Screenshot of "ACAT 2 or ACAT 3 Funding" Process

Sensitivity Analysis on the AoA Process Parameters

The objective of this phase is to identify the impact of improving the AoA process, specifically its variability and its program completion time. The ExtendSim model is utilized over the Arena model because the model reflects the most recent policy changes. The current AoA process is distributed according to a triangular distribution with a lower limit of 180, most likely value of 360, and an upper limit of 720 days (Tria(180, 360, 720)), and all programs are assumed to undergo an AoA. These settings constitute the baseline trial for this phase. The ACAT 1 level projects are utilized in this phase. The ExtendSim simulation module being modified in the ERAM model is an activity block from the items library called "Analysis of Alternatives." Figure 2 presents a screenshot of the ExtendSim activity.

This phase is composed of two simulation experiments. The first experiment studies the effect of reducing the variability of the AoA process time, while the second section deals with reducing the mean of the AoA process time. For the first experiment, 3,000 iterations are performed; and during each trial, the variability of the time to perform the AoA is adjusted. The variability is adjusted by reducing the difference between the mode and the maximum/minimum by a constant. Hence, the triangular distribution would then have less variance. The variance of the triangular distribution is reduced incrementally until a statistical significant difference is obtained from the baseline trial. A *t*-test is then performed in order to compare the time to MS-C in the trial to the baseline trial. Furthermore, the second simulation experiment deals with reducing the AoA process time to determine its effect on the time ACAT 1 programs reach MS-C. Three thousand iterations are again performed and for each trial, the mean of the AoA program length distribution, which is distributed according to a triangular distribution (Tria(180, 360, 720)) is reduced. The ratio between the minimum, most likely, and maximum of 1:2:4 is maintained as the parameters are changed. The mean completion time of the triangular distribution of the AoA is again reduced until a statistically significant difference from the baseline trial is obtained. A *t*-test is again performed in order to compare the time in the trials to the baseline trial.

Interpretation

The final phase in this paper compiles the results from the tests and translates them into recommendations for viable policy changes. This phase includes identifying which parameter adjustments improved the performance metrics and makes recommendations on further research, like interactions among parameters.

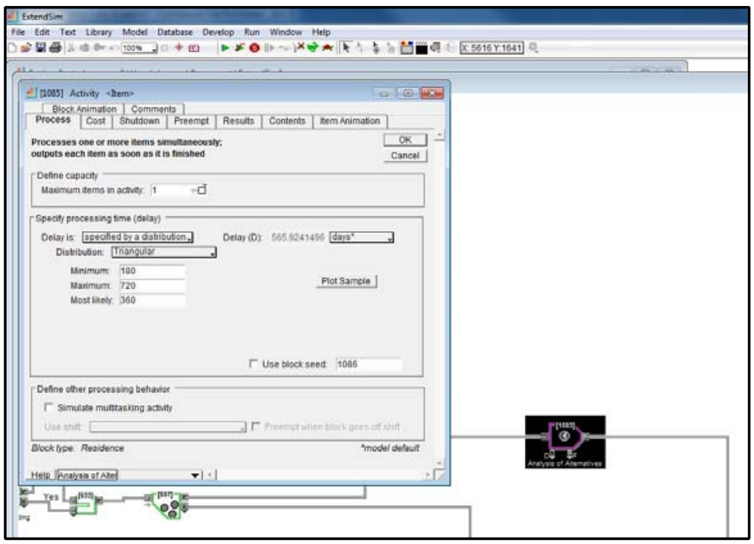

Figure 2. ExtendSim Screenshot of "Analysis of Alternatives" Activity

Results and Discussions

Table 2 summarizes the results of the *t*-tests performed for ACAT 2 programs. The table shows a subset of trials corresponding to 50%, 85%, 87.5%, 90%, 95%, or 99% of programs are required to undergo an AoA. These settings are selected to show the sensitivity of changing the percentage of programs required to undergo an AoA. These are compared to the baseline AoA setting in which only the funded programs (1%) are required to undergo an AoA. The null hypothesis for the *t*-tests is $H_0: \mu_{base} = \mu_{i^{th}\%}$, which corresponds to an insignificant difference between the baseline and the i^{th} percentage, if not rejected, and alternative hypothesis $H_1: \mu_{base} \neq \mu_{i^{th}\%}$, if there is significant difference.

Table 2. Summary of *t*-Test Results of ACAT 2 Programs

	% of Programs Required to Undergo an AoA						
	1% (Baseline)	50%	85%	87.5%	90%	95%	99%
Average Time to MS-C (Days)	3898.099	4001.104	4080.293	4093.409	4108.58	4115.981	4141.088
Standard Deviation (Days)	1411.37	1550.684	1625.738	1634.814	1650.243	1655.492	1673.537
t-Value		-1.34133	-1.78231	-1.90576	-2.04267	-2.04267	-2.34049
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0	Reject H_0	Reject H_0	Reject H_0

It is evident that when requiring an AoA on at most 87.5% of the ACAT 2, programs will result in a failure to reject that the i^{th} percentage is similar to the baseline scenario. Any percentage less than 87.5% will have no effect on the average time a program arrives at MS-C. This means that a majority of the ACAT 2 programs can be subject to the AoA, and no significant increases in the total completion time can be obtained.

Table 3 summarizes the results of the *t*-tests performed for ACAT 3 programs. The table shows a subset of trials corresponding to 50%, 55%, 57.5%, 60%, 65% or 75% of programs are required to undergo an AoA. Furthermore, these settings are selected to show the sensitivity of changing the percentage of programs required to undergo an AoA. These are compared to the baseline in which only the funded programs (1%) are required to undergo an AoA. The null hypothesis for the *t*-tests is $H_0\colon \mu_{base} = \mu_{i^{th}\%}$, which corresponds to no significant difference between the baseline and the i^{th} percentage, if not rejected, and alternative hypothesis $H_1\colon \mu_{base} \neq \mu_{i^{th}\%}$, if there is significant difference.

Table 3. Summary of *t*-Test Results of ACAT 3 Programs

	% of Programs Required to Undergo an AoA						
	1% (Baseline)	50%	55%	57.5%	60%	65%	75%
Average Time to MS-C (Days)	3334.926	3470.918	3493.278	3512.661	3524.566	3551.2	3604.434
Standard Deviation (Days)	1143.19	1314.598	1350.852	1365.653	1372.66	1392.929	1419.016
t-Value		-1.63342	-1.86606	-2.08392	-2.2198	-2.51299	-3.1095
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Reject H_0	Reject H_0	Reject H_0	Reject H_0

It is evident that when requiring an AoA on at most 55% of the ACAT 3, programs will result in a failure to reject that the i^{th} percentage is similar to the baseline scenario. On the other hand, the AoA contributes to a significant change in the total program completion time when at least 57.5% of the programs require AoA. In general, it is evident from Tables 2 and 3 that requiring an AoA on all programs significantly increase the total completion time of ACAT 2 and 3 programs when programs are required to undergo an AoA.

Table 4 summarizes the results of the *t*-tests performed in terms of reducing the variability of the AoA process time. Four settings are tested specifically, Tria(270, 360, 540), Tria(315, 360, 450), Tria(337.5, 360, 405), and Tria(360, 360, 360). These settings are selected to illustrate the sensitivity of the variability of the AoA processing time from the baseline scenario to a deterministic scenario (time is fixed at 360 days) as is done in (Wirthlin, 2009). These settings are compared to the baseline AoA process time, which is distributed Tria(180, 360, 720). The null hypothesis for the *t*-tests is $\mathrm{H_0}: \mu_{\mathrm{base}} = \mu_{\mathrm{i^{th}}_T}$, which corresponds to no significant difference between the baseline and the $\mathrm{i^{th}}$ AoA triangular distribution variance setting, if not rejected, and alternative hypothesis $\mathrm{H_1}: \mu_{\mathrm{base}} \neq \mu_{\mathrm{i^{th}}_T}$, if there is significant difference.

Table 4. *t*-Test Results for the Variance Reduction of AoA Process Time

	AoA Triangular Distribution Settings (Low, Medium, High)				
	180, 360, 720 (Baseline)	270, 360, 540	315, 360, 450	337, 360, 405	360, 360, 360
Average Time to MS-C (Days)	6903.567	6883.159	6867.149	6865.280	6848.717
Standard Deviation (Days)	1584.258	1591.61	1591.442	1590.301	1576.526
t-Value		0.2528	0.4512	0.4745	0.6828
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0

It is evident that reducing the variability of the AoA process time does not have an effect on the total time ACAT 1 programs reach MS-C. Hence, any improvement on the AoA process to make it more consistent and standard would not have a significant effect on the total program completion time.

On the other hand, Table 5 presents the results of reducing the mean process time of the AoA. Four settings are tested specifically: Tria(135, 270, 540), Tria(112.5, 225, 450), Tria(101.25, 202.5, 405), and Tria(90, 180, 360). These settings are selected to illustrate the sensitivity of the length of the AoA processing time from the baseline scenario to the fastest scenario (Tria(90, 180, 360)). These settings are compared to the baseline AoA process time, which is distributed Tria(180, 360, 720). The null hypothesis for the *t*-tests is $\mathrm{H_0}: \mu_{\mathrm{base}} = \mu_{\mathrm{i^{th}}_T}$, which corresponds to no significant difference between the baseline and the $\mathrm{i^{th}}$ AoA triangular mean time distribution setting, if not rejected, and alternative hypothesis $\mathrm{H_1}: \mu_{\mathrm{base}} \neq \mu_{\mathrm{i^{th}}_T}$, if there is significant difference.

Table 5. *t*-Test Results for the Mean Reduction of AoA Process Time

	AoA Triangular Distribution Settings (Low, Medium, High)				
	180, 360, 720 (Baseline)	135, 270, 540	112, 225, 450	101, 202, 405	90, 180, 360
Average Time to MS-C (Days)	6903.567	6827.941	6752.989	6732.038	6681.659
Standard Deviation (Days)	1584.258	1597.607	1578.558	1567.143	1574.53
t-Value		0.9351	1.8732	2.1415	2.7640
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Reject H_0	Reject H_0

It is evident that reducing the AoA process time to at the least Tria(112.5, 225, 450) will result in a failure to reject that the i^{th} percentage is similar to the baseline scenario. This implies that reducing the AoA process time to at most Tria(101.25, 202.5, 405) contributes to a significant change in the total program completion time.

Conclusions

The analysis of alternatives (AoA) process is a new requirement that all acquisition category (ACAT) level projects to be completed by the U.S. government are required to undergo. Previously, this process is required only on all ACAT 1 programs, and ACAT 2 and 3 programs with sufficient funding. Hence, program completion is expected to increase further, given that a majority of the programs finish over budget and with a delayed schedule when AoA is required on all programs. This concept is validated through simulation documented in this paper. By requiring all programs to undergo an AoA, significant completion delays are observed. Furthermore, this paper provides quantitative insights on the amount of time that the AoA contributes to the overall DoD end-to-end process time through Wirthlin's (2009) discrete event system simulation model.

Through additional simulation analysis, it is inferred that a majority of the ACAT 2 programs can be subject to the AoA without significantly affecting completion time. However, requiring an AoA on ACAT 3 programs has a significant effect on program completion time. A frequently cited reason for not conducting an AoA is that an AoA causes a significant delay in the program. The simulation results in this paper should provide strength to the argument towards providing funds for full AoAs. However, as there is a percentage of programs where the AoA can cause significant delays in program completion (87.5% and 55% for ACAT II and ACAT III, respectively), policy improvements with regards to selection criteria for programs that must undergo an AoA may be beneficial.

Support for this conclusion can be found in the system engineering case study for the A-10 Thunderbolt II performed by Jacques and Strouble (2010). From this case study, it can be seen that the correct level of focus is needed for an AoA to be beneficial. During the development of the A-10, three prototyping studies were performed. These prototyping efforts are analogous to an AoA. One prototype is a system level test in the form of a fly-off. The other two are subsystem prototyping of the guns and ammunition for the aircraft. Although the system level prototyping provided little meaningful information, the two subsystem prototyping efforts provide benefits to the DoD in terms of both overall cost reduction and design improvements. The findings from the case study in conjunction with the simulation results in this paper indicate that policy applying tailored criteria for ACAT 2 and ACAT 3 AoAs for programs that must undergo an AoA could be beneficial.

Furthermore, findings from this paper show that any variance reduction improvements or standardizations on the AoA process would not affect the total program completion time of ACAT 1 programs. However, any improvements that reduce the processing time of the AoA, or, in other words, making it more efficient and lean, would significantly reduce program completion time. With the DoD's requirement to conduct an AoA, an implicit assumption has been made that the step will shorten the overall life-cycle development. The simulation's finding suggests that, from a statistical perspective, the DoD must be expeditious during the AoA process, or the value sought from requiring the AoA decreases. Hence, policy improvements with regards to an AoA should focus on reducing processing and execution time of the AoA without sacrificing quality in its output.

The use of simulation in this study places some limitations on this research, as well as provides opportunities for further research. One limitation is that the simulation takes a program rather than a portfolio perspective. Previously, fewer AoAs were performed because of lack of funding. The need to allocate funding to AoAs for ACAT 2 and ACAT 3 programs may result in more delays across the DoD portfolio because funding may be spread too thin. This effect cannot be illustrated by the model if it is present. Further research may be needed to investigate this possibility. In addition, AoAs were implemented with the assumption that overall program quality would increase with the inclusion of an AoA. The simulation does not incorporate the interactions between the AoA and quality. Further study can be conducted to investigate these interactions or add them to the ERAM simulation.

References

.

Arena, M. V., Obaid, Y., Galway L. A., Fox, B., Graser, J. C., Sollinger, J. M., Wu, F., & Wong, C. (2006). *Impossible certainty: Cost risk analysis for Air Force systems* (MG-415-AF). Santa Monica, CA: RAND.

Cervantes, M. A., Enderton, C., & Powers, J. S. (2012). *Rapid execution of an analysis of alternatives for NATO special operations HQ: A smart defence approach.* Monterey, CA: Naval Postgraduate School.

Chairman of the Joint Chiefs of Staff (CJCS). (2012). *Joint Capabilities Integration and Development System* (CJCS Instruction 3170.01H). Washington, DC: Author.

Ford, D. N., Housel, T., & Dillard, J. T. (2010). *Integrating system dynamics modeling and knowledge value added for improved analysis of alternatives: A proof of concept study* [DTIC document].

Georgiadis, D. R., Mazzuchi, T. A., & Sarkani, S. (2012). Using multi criteria decision making in analysis of alternatives for selection of enabling technology. *Systems Engineering, 16*(3). doi:10.1002/sys.21233

Houston, D. (2012). *Enterprise requirements and acquisition model.* Unpublished manuscript.

Jacques, D. R., & Strouble, D. D. (2010). *A-10 Thunderbolt II (Warthog) systems engineering case study.* Wright-Patterson AFB, OH: Air Force Center for Systems Engineering (AFIT/SY).

Leach, D., & Searle, C. (2011). *Department of Defense enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Montgomery, J. (2012). *Department of Defense update to the enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Roper, M. (2010). *Risk considerations in pre-Milestone-A cost analysis*. Washington, DC: Office of the Deputy Assistant Secretary of the Army for Cost and Economics.

Schank, J. F. (2012). *Analysis of alternatives: Keys to success* (ADA563514). Santa Monica, CA: RAND Corp.

Schwartz, M. (2010). *Defense acquisitions: How DoD acquires weapon systems and recent efforts to reform the process.* Darby, PA: Diane Publishing.

Stadterman, T. J. (2012). *Improving U.S. Army analysis of alternatives to better address the Weapon Systems Acquisition Reform Act of 2009* (SSCF research report). Retrieved from http://www.dau.mil/SSCF/SSCFdocs/Stadterman.pdf

Sullivan, M. J. (2012). *Presidential helicopter acquisition: Effort delayed as DOD adopts new approach to balance requirements, costs, and schedule* (GAO-12-381R). Washington, DC: Government Accountability Office.

Sullivan, M. J., Andrew, C., Bowman, R., Fairbairn, B., Neill, S., Oppenheim, J., & Schwenn, R. (2009). *Defense acquisitions: Measuring the value of DOD's* *weapon programs requires starting with realistic baselines* (GAO-09-543T) Washington, DC: Government Accountability Office

Sullivan, M. J., Schwenn, R. E., Bleicher, N. B., Marchesani, S. V., Patton, K. E., Richey, K. A., Russell, A. K., & Tranquilli, N. A. (2011). *DoD weapon systems: Missed trade-off opportunities during requirements reviews*. Washington, DC: Government Accountability Office.

Training and Doctrine Command Analysis Center. (2011). *TRADOC assessment of AoA guidance in DoD 5000 Series* [Unpublished manuscript].. Fort Eustis, Virginia: Author.

Under Secretary of Defense for Acquisition, Technology, & Logistics (USD[AT&L]). (2008). *Operation of the Defense Acquisition System* (Department of Defense Instruction 5000.02). Washington, DC: Author.

Weapon Systems Acquisition Reform Act of 2009, 111-23 Cong. (2009).

Wirthlin, J. R. (2009). *Identifying enterprise leverage points in Defense Acquisition Program performance* (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology.

THIS PAGE INTENTIONALLY LEFT BLANK

Bottleneck Analysis on the Department of Defense Pre–Milestone B Acquisition Processes

The current Enterprise Requirements and Acquisition Model (ERAM), a discrete event simulation of the major tasks and decisions within the Department of Defense (DoD) acquisition system, identifies several what-if intervention strategies to improve program completion time. However, processes that contribute to the program acquisition completion time were not explicitly identified. This research seeks to determine the acquisition processes that contribute significantly to the time a program reaches Milestone B and provide interventions to improve program completion time. In order to solve this problem, this research uses critical path analysis to determine the bottleneck activities in the pre–Milestone B (MS-B) processes using additional simulation analysis. Results show that the systems engineering processes are the bottleneck activities in pre–MS-B acquisition stage. Furthermore, this research then examines the effect of these processes by varying the mean completion times and having them occur earlier in the acquisition process. Potential policies are formulated from the results to further reduce program acquisition completion time.

Introduction

A large number of DoD projects are often being completed behind schedule and over-budget (Schwartz, 2010). For example,a Government Accountability Office (GAO) report released in 2009 stated that for the DoD's 2008 portfolio, on average, a program faced a 22-month delay and exceeded the original budget (Sullivan et al., 2009). Generally, total cost growth of 44% has been consistent over the past few decades with a recent assessment by RAND Corporation (Arena et al., 2006). Hence, potential intervention strategies and policies to improve the acquisition processes would be worthwhile. On the other hand, since the end-to-end DoD acquisition process is a large, complex, socio-technological system, it is difficult to analyze and determine which processes or factors affect performance metrics like time, cost, and resource utilization. The current DoD acquisition system, which is composed of three separate and distinct processes (the Joint Capabilities Integration Development System [JCIDS]; the Planning, Programming, Budgeting, & Execution [PPBE] process; and the formal acquisition development system outlined by the DoD 5000 series of instructions), does not exist in a static environment. The system is constantly being adjusted, either through policy changes or statute (Chairman of the Joint Chiefs of Staff [CJCS], 2012; Weapon Systems Acquisition Reform Act of 2009; Under Secretary of Defense for Acquisition, Technology, and Logistics [USD(AT&L)], 2008). Hence, other viable analysis methodologies must be utilized to fully comprehend this complex system.

In 2009, Wirthlin (2009) created a discrete event simulation (DES) model called the Enterprise Requirements and Acquisition Model (ERAM). This model was created to simulate the actual acquisition processes of the DoD, using the Air Force implementation of acquisition processes as the basis of the model, in order to provide further insight and understanding of the complex system's behavior. Furthermore, ERAM has benefited from additional research since the original 2009 Wirthlin version (Leach & Searle, 2011; Montgomery, 2012). These new versions have added additional functionality and options for model users to manipulate (Wirthlin, Houston, & Madachy, 2011). According to the ERAM model, during the acquisition process, approximately 80% of the time, a program was undergoing parallel processes when it is in the acquisition system. It was also observed that one of the main portions of the model during which these parallel processes take place are within the pre–MS-B) stage. However, Wirthlin's (2009) research did not identify the significant processes that affect the total program time for a project to reach MS-B.

Against this background, this research addressed these limitations and issues by additional simulation and statistical analysis on the ERAM Arena version of the model. The end goal of this research was to determine the bottleneck of the pre–MS-B processes, investigate interventions to alleviate the bottleneck, and translate them into implementable policy changes. The rest of this paper is organized as follows: an overview of the current literature on bottleneck analysis and the ERAM model, the simulation analysis methodology, then the results of the analysis. Finally, the conclusions of this research are presented, as well as viable intervention policies for reducing the time a program takes to reach MS-B.

Review of Literature

The Enterprise Requirements and Acquisition Model

The ERAM simulation model extends from the generation of capability requirements in the JCIDS process to MS-C, the review before the production stage begins. Additionally, the ERAM is abstracted at a very high level (Wirthlin, 2009). This high-level of abstraction allows overall system performance to be more easily studied. For each replication, ERAM produces schedule time for programs that reach MS-C. Although cost is not measured, it was found that cost over-runs were closely related to schedule over-runs (Wirthlin, 2009). The validation and verification of ERAM included hand modeling, iterations of correction from feedback of experts in all three systems that comprise the entire acquisition system, and comparison of schedule and budget information from the DAMIR and SMART databases to distributions of the schedule time of model-generated data (Wirthlin, 2009).

The original version of ERAM was created in Arena Simulation software; however, it was translated into an ExtendSim version (ERAM 1.0) to serve as a schedule and success estimation tool of space programs for the Concept Design Center of Aerospace Corporation (Leach & Searle, 2011). Leach and Searle (2011) further modified the model introducing ERAM 1.1 to 2.1 by correcting discrepancies between the Arena and ExtendSim model, adding user-controlled variables, incorporating space-acquisition specific elements, and updating the model to include policy in the newly released DoDI 5000.02 document. Montgomery (2012) continued developing the model in order to add the rapid acquisition process and include ACAT 2/3 programs. A summary of the versions of the ERAM is presented in

Table 1. ERAM Versions
(Adapted from Houston, 2012)

Author	**Version Number**	**Changes**
Wirthlin (2009)	ERAM 1.0	Baseline translation from Arena to ExtendSim
Leach and Searle (2011)	ERAM 1.1	Updates by the Aerospace Design Team and served as new baseline model
	ERAM 1.2	Implemented new DoD 5000.02 policies
	ERAM 2.0	Incorporated the global variables that modify acquisition capabilities
	ERAM 2.1	Incorporated the JCIDS review process
Montgomery (2012)	ERAM 2.2	Added more capabilities for ACAT 2/3 and rapid acquisition process

Because the ExtendSim version of ERAM was designed with the purpose of allowing Aerospace Corporation to create estimates of the schedule and success of a particular project, it has a distinctly different scope and utility from the Arena model of ERAM. The Arena model allows the user to view the behavior of the overall portfolio, while the ExtendSim version allows the user to investigate a specific program. For example, while the ExtendSim requires the user to select a specific ACAT level for the program being tested, the Arena version assigns ACAT levels based on the distribution of programs observed in the actual acquisition system. Although the ExtendSim version of ERAM was designed with the intention of allowing the user to perform what-if scenarios, as far as the researcher is concerned, no literature of the evaluation of possible intervention strategies using the ExtendSim version of ERAM has been published. In his dissertation, Wirthlin (2009) investigated the effect of 20 interventions on the effect of end-to-end acquisition time in the Arena version. When all 20 interventions were implemented, a 20% reduction in end-to-end acquisition time was achieved. However, more interventions can be developed to further study and improve the DoD end-to-end acquisition process.

Critical Path Analysis

To the best of our knowledge, no literature has attempted to identify the critical path of the acquisition process (Monaco & White, 2005). Although long cycle times continue to plague DoD acquisition programs, relatively few studies have focused on identifying significant processes that dictate program cycle time. Despite the Packard Commission's assertion that schedule drives costs, most studies and policy changes have focused on cost reduction rather than reducing cycle time (Al-Harbi, 2001; McNutt, 1999). Drenzer and Smith (1990) performed a statistical analysis of 10 programs in order to hypothesize factors that affect the original plan and/or program deviation. Tyson, Nelson, Om, and Palmer (1989) examined schedule variance and its causes. They found that prototyping, sole-source procurement, fixed-priced contracts, and multiyear procurement reduced schedule variance. They also found that programs awarded through full and open competition experienced more schedule growth than those programs that did not. Another possible schedule driver is presented by Brown, Flowe, & Hammel (2007). They compared the schedule quality of joint and single-system programs. Brown et al. (2007) found that joint system programs has significantly more schedule breaches; however, the research did not identify the root cause of this difference.

In summary, to the best of our knowledge, there exists no research conducted that isolates and identifies bottleneck activities and its effect on the program completion time throughout the DoD acquisition process. Hence, intervention strategies to be developed must be focused on addressing bottleneck issues to obtain maximum improvement of the end-to-end DoD acquisition process.

Simulation Analysis Methodology

This section describes the analysis performed to identify bottleneck operations within the pre–MS-B stage. After identifying bottleneck operations, intervention strategies were also formulated to reduce total program completion time. Hence, this research was performed in two phases. A brief description of these phases is presented as follows:

- The first phase performed a critical path analysis on the pre–MS-B activities to identify a bottleneck
- The second phase focused on investigating the effect of reducing the process times of the identified bottleneck activities from phase 1 and determining the effect of allowing them to be executed earlier in the process

Identification of Bottleneck Activities

In order to perform critical path analysis, the pre–MS-B phase was mapped by hand to assist in visualization of the complex network of separation and batches in the acquisition system. The processes between each separate and batch method were left out for simplicity and ease of interpretation. The line segment between any two nodes was labeled. Figure 1 shows the mapped version of the pre–MS-B activities, and Table 2 shows the activities associated with each section.

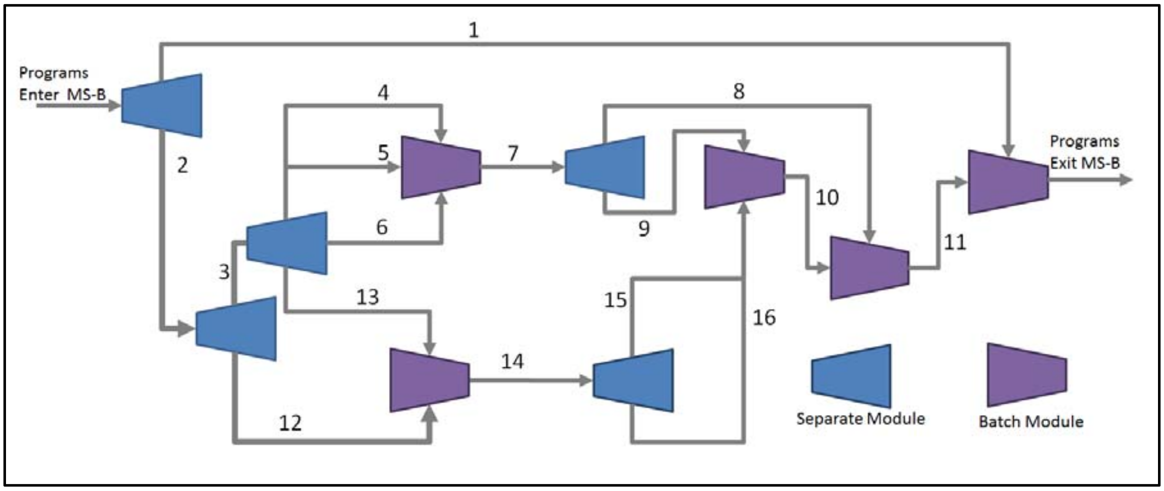

Figure 1. Pre–MS-B Flowchart

Table 2. List of Activities in the Pre–MS-B Flowchart

Section	Description of Activities
1	Requirements generation: KPP Development, high performance team work, etc.
2	RFP release, contract awarding
3	Waiting period for start of contract
4-6	Cost estimates (contractor, program office, and independent)
7	Affordability assessment
8	Set acquisition program baseline
9-10	No processes
11	Prepare and conduct acquisition panels
12	Early Systems Engineering (SE) activities: EOA, developmental testing, SRR, etc.
13	Acquisition planning activities
14	Draft RFP
15	RFP coordination
16	Source selection plans

Several Assign and Record modules were added to the Arena model in order to determine the time to complete each segment. Next, a trial of 3,000 runs was performed in Arena and the times for each segment were collected. A spreadsheet was then used to analyze results and determine the time of every possible path from the beginning of the pre–MS-B activities to the MS-B decision. The path that took the longest amount of time was deemed as the critical path. By comparing segments in the longest paths to the sections found in shorter paths, the bottleneck activities for the pre–MS-B processes were identified.

Design of Pre–Milestone B Bottleneck Interventions

In order to improve the performance and alleviate the delay caused by this bottleneck, two intervention strategies were developed and tested in ERAM. The first intervention performed was to test the effect of decreasing the process time for all bottleneck activities. In order to test the effect of reducing total process time, the minimum, maximum, and mode for these activities was reduced by a fixed percentage. A paired *t*-test was then performed to compare each trial to the baseline at 95% confidence level. The reduction by using a fixed percentage was performed until a statistically significant change was obtained. Furthermore, the second intervention was a sensitivity analysis to determine the effect of allowing the bottleneck to be performed earlier in the pre–MS-B process to determine its effect on the total process time. The results of these interventions are illustrated in the next section.

Results and Discussions

This section presents the results of both simulation analysis phases performed on the ERAM Arena model. Specifically, we present the results of the identification of the critical path and bottleneck activities. Additionally, the results of the interventions performed on the bottleneck analysis are shown to improve program completion time.

Pre–Milestone B Critical Path Analysis Results

During the critical path analysis, times for all 11 paths through the system were calculated. The paths were labeled by letters. Each path was composed of segments. A subset of the paths and their corresponding activities is shown in Table 3.

Table 3. List of Paths and Segments for Pre–MS-B

Path Name	**Corresponding Segments From** Figure 1
A	1
B	2, 12, 14, 16, 10, 11
C	2, 12, 14, 15, 10, 11
D	2, 3, 6, 7, 8, 11
E	2, 3, 5, 7, 8, 11
F	2, 3, 4, 7, 8, 11
G	2, 3, 6, 7, 9, 10, 11
H	2, 3, 5, 7, 9, 10, 11
I	2, 3, 4, 7, 9, 10, 11
J	2, 3, 13, 14, 15, 10, 11
L	2, 3, 13, 14, 16, 10, 11

As seen in Table 3, Paths B and C heavily overlap while Path A has no overlap with any other path. From the total time for each path, the longest was deemed the critical path. The second longest and third longest paths were also determined. A subset of this data can be seen in Table 4.

Table 4. Length of Longest Paths to MS-B

Run	Percentage of Runs as Longest Path	Percentage of Runs as Second Longest Path	Percentage of Runs as Third Longest Path	Percentage of Runs in the Top Three Longest Path
A	45	8.75	42.5	96.25
B	43.75	38.75	7.5	90.00
C	6.25	38.75	31.25	76.25
D	3.75	8.75	8.75	21.25
F	0	2.5	5	7.5
J	0	1.25	1.25	2.50
K	1.25	1.25	3.75	6.25
E, G, H, I	0	0	0	0

As can be observed in Table 4, the critical path was most often A, B, and C. In approximately 95% of the trials, either A, B, or C composed the critical path. Specifically, 50% of the time Path B or C was the critical path, and 45% of the time Path A was the critical path. We note that Paths B and C have significant overlap; therefore, they are considered a single path, Path B/C. Because the critical path was very evenly split between Path A and Path B/C, it can be deduced that a pre–Milestone-B process common to both of paths would be the bottleneck of the process.

In examining the ERAM, it can be gleaned that there was some interaction between Path A and Path B/C. One of the last modules of Path A was a hold module called “Wait for Evaluation of Analysis (EOA) completion.” A screenshot of this module can be seen in Figure 3.

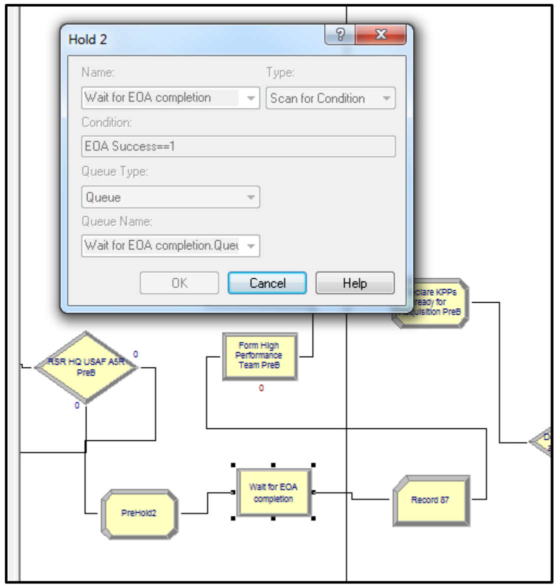

Figure 3. Wait for EOA Completion Screenshot

As seen in the figure above, Path A must wait for the EOA to be complete before the path can finish. A second communication occurs between the two paths. In order for the System Engineering (SE) activities, like the EOA, to occur, the key performance parameters (KPPs) must be complete. The hold model called "Wait for Test and Evaluation (T&E) start" facilitates this communication and can be seen in Figure 4.

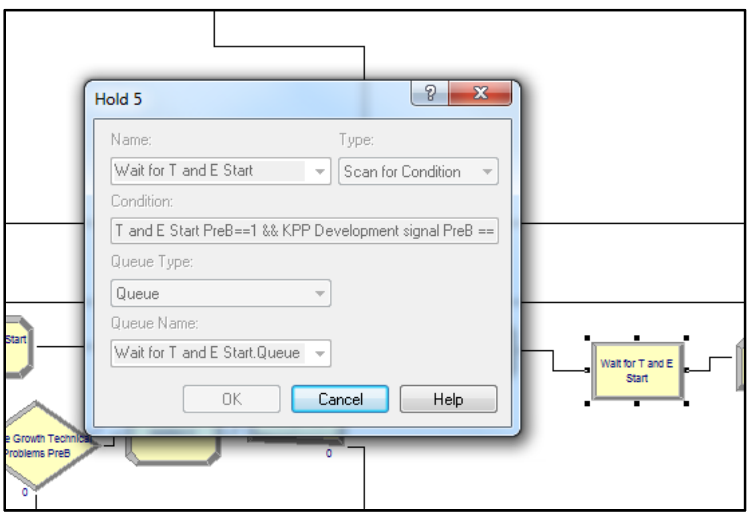

Figure 4. Wait for T&E Start Screenshot

However, we note that this hold module also waits for 75% of the contract length to elapse. At the default settings, the KPPs will always be completed in less than 75% of the contract length. Therefore, at the default settings, this hold does not serve as communication between the paths.

Because the completion of the EOA was the only communication between the two critical paths, the SE activities that begin before the EOA completion was determined to be the bottleneck of the pre–MS-B activities. If this bottleneck activity were removed, the time to MS-B would be reduced by an average of 6.8%.

Additional Pre–Milestone B Bottleneck Interventions

Table 5 summarizes the results of the *t*-tests performed for when the process time for MS-B system engineering activities was reduced. The tables show a subset of trials corresponding to a reduction in process times by 0%, 20%, 35%, or 50%. These settings were selected to show the sensitivity of the model to various degrees of process time reduction. From these simulation analyses, the mean ($\mu_{i^{th}\%}$) and standard deviation of the total completion time for each trial were calculated. These calculated means were compared to the mean of the baseline setting (μ_{base}) in the default settings, or 0% process time reduction. The null hypothesis for the *t*-tests is $H_0: \mu_{base} = \mu_{i^{th}\%}$, which corresponds to a failure to reject the claim that the baseline and the i^{th} percentage are similar and alternative hypothesis $H_1: \mu_{base} \neq \mu_{i^{th}\%}$ if there is significant difference.

Table 5. Summary of *t*-Test Results of Process Time Reduction for System Engineering (SE) Activities

	% Reduction of Process Time			
	0% (Baseline)	20%	35%	50%
Average Time to MS-B (Days)	3418.01	3274.90	3211.564	3164.25
Standard Deviation (Days)	1701.08	1636.108	1557.816	1515.48
p-Value		0.281	0.109	0.046
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Reject H_0

From Table 5, it is evident that when the process time for SE activities was decreased by less than 50%, there will not be a statistically significant decrease in the time to MS-B. However, when the process times for SE activities are reduced by more than approximately 50%, the model exhibits a statistically significant decrease in time to MS-B.

Based on the identified bottleneck, which were the SE activities, a second intervention was developed. Specifically a sensitivity analysis was done to test the effect of allowing the bottleneck activities to occur earlier in the contract. This was implemented by adjusting the module called “Begin Testing PreB,” which can be seen in the Figure 5.

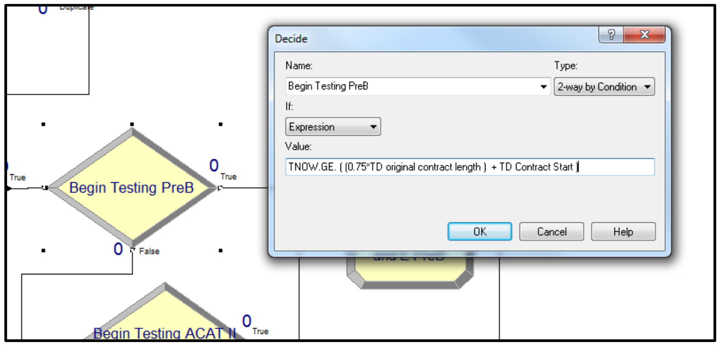

Figure 5. Begin Testing PreB Screenshot

The “Begin Testing PreB” module is a Decide module that when set to true, triggers the beginning of the SE activities. The original criterion for the decide module, as verified and validated by Wirthlin (2009) when creating the ERAM model, was that 75% of the contract length must pass before these activities can occur. During this research, this percent was decreased to simulate the SE activities occurring sooner and more resources being applied at the beginning of the contract.

In addition, to allow SE tasks to begin sooner, the KPPs must be completed sooner in the process because their completion is also needed to trigger the start of the SE tasks. A more complete discussion of this interaction can be found in critical path analysis section. The process time of the KPP development was reduced in order for the KPPs to be completed in a manner that does not delay the SE activities. A paired *t*-test was then performed to compare each trial to the baseline at 95% confidence level.

Table 6 and Table 7 summarize the results of the *t*-tests performed by allowing pre–MS-B contractor activities to occur earlier in the contract. Specifically, Table 6 shows the effect of allowing the SE activities to occur earlier in the contract when the KPPs’ generation process time was not decreased; and Table 7 shows the effect of allowing the SE activities to occur earlier in the contract in conjunction with the KPPs’ generation process performing faster. Table 6 shows a subset of trials corresponding to the SE activity starting when 75%, 50%, 33%, or 25% of the contract has elapsed. Table 7 shows a subset of trials corresponding to the SE activity starting when 75%, 65%, 60%, or 55% of the contract has elapsed.

These settings were selected to show the sensitivity of the model to various start times of SE activities. From these simulations, the mean ($\mu_{i^{th}\%}$) and standard deviation of the total MS-B completion time for each trial were calculated. These calculated means were compared to the mean of the baseline setting (μ_{base}) in the default settings, or starting after 75% of the contract has elapsed. The null hypothesis for the *t*-tests is $H_0: \mu_{base} = \mu_{i^{th}\%}$, which corresponds to a failure to reject the claim that the baseline and the i^{th} percentage of contract elapsing before start is similar in terms of program completion time and alternative hypothesis $H_1: \mu_{base} \neq \mu_{i^{th}\%}$, if there is significant difference.

Table 6. Summary of *t*-Test Results of SE Activity Start Time Adjustments With Original KPP's Process Time

	% of Contract Elapsed Before Start			
	75% (Baseline)	50%	33%	25%
Average Time to MS-B (Days)	3418.01	3379.09	3379.09	3379.09
Standard Deviation (Days)	1701.08	1670.31	1670.31	1670.31
p-Value		0.770	0.770	0.770
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0

Table 7. Summary of *t*-Test Results of SE Activity Start Time Adjustments With Reduced KPP's Process Time

	% of Contract Elapsed Before Start			
	75% (Baseline)	65%	60%	55%
Average Time to MS-B (Days)	3418.01	3305.44	3200.75	3139.95
Standard Deviation (Days)	1701.08	1628.08	1599.04	1553.38
p-Value		0.392	0.099	0.032
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Reject H_0

In Table 6, it is evident that the time to MS-B is not sensitive to an earlier start time for SE activities when the KPP process time is set to the default distribution. In fact, when the start time is at 50%, 33%, and 25% of the contract time, the time to MS-B, standard deviation of time to MS-B, and *p*-value are identical. This is due to the hold module in the SE path described earlier. As previously discussed, in order for the SE activities to begin, a percent of the contract must elapse, and the KPPs must be complete. Once the SE start time occurs earlier than 50% of the contract length, the KPP's completion is the determining factor of the SE activity start time.

Table 7 takes this into account by reducing the KPP's process time to a point where it does not dictate the start of the SE activities. From Table 7, it is evident that when SE activities begin at 60% of the contract length or later, there will not be a statistically significant decrease in the time to MS B. However, when SE activities begin at 55% of the contract length or sooner and the KPP's generation process are shortened the same degree, the model exhibits a statistically significant decrease in time to MS-B.

Conclusions

The critical path analysis performed in this research indicated that the SE activities and their communication with the requirements branch are the bottleneck of the pre–MS-B portion of the acquisition system. In addition, the research indicated that focusing on reforms that address this bottleneck has the potential to decrease the total time spent on MS-B activities by approximately 7%; this corresponds to a process time reduction of approximately six months.

This research also tested two strategies to address this bottleneck. The first was reducing the process time of all SE activities. The second was to allow the SE activities to have an earlier start time. This research showed that the latter policy has the potential to be the most beneficial. This research showed that the process times for all SE activities must be decreased by approximately 50% in order for a statistically significant decrease in time to MS-B to occur. This degree of process time reduction may be infeasible. On the other hand, allowing the SE activities to occur after 55% of the contract time has elapsed rather than the current 75%, produces a statistically significant decrease in time to MS-B.

The increased sensitivity of program time to start time, rather than process length, suggests that schedule benefits may be achieved if some resources, both financial and human, are transferred from the SE activities to the activities prior to test and development. However, this re-allocation of resources must be accompanied by responsiveness from the JCIDS branch, which is the branch that generates the KPPs. This research indicates that there was a large amount of co-dependence between the JCIDS and SE activities and that communication and coordination between these branches is needed in order to address the bottleneck.

References

Al-Harbi, K. M. (2001). Application of the AHP in project management. *International Journal of Project Management, 19*(1), 19–27.

Arena, M. V., Obaid, Y., Galway L. A., Fox, B., Graser, J. C., Sollinger, J. M., Wu, F., & Wong, C. (2006). *Impossible certainty: Cost risk analysis for air force systems* (MG-415-AF). Santa Monica, CA: RAND Corp.

Brown, M. M., Flowe, R. M., & Hamel, S. P. (2007). *The acquisition of joint programs: The implications of interdependencies*. Washington, DC: Office of the Secretary of Defense.

Chairman of the Joint Chiefs of Staff (CJCS). (2012). *Joint Capabilities Integration and Development System* (CJCS instruction 3170.01H). Washington, DC: Author.

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (RAND/R-3937-ACQ). Santa Monica, CA: RAND.

Houston, D. (2012). *Enterprise requirements and acquisition model*. Unpublished manuscript. Los Angelos, CA: Space and Missile Center (SMC).

Leach, D., & Searle, C. (2011). *Department of Defense enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

McNutt, R. T. (1999). *Reducing DoD product development time: The role of the schedule development proces* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Monaco, J. V., & White, E. D., III. (2005). Investigating schedule slippage. *Defense Acquisition Review Journal*, *12*(2), 176–193.

Montgomery, J. (2012). *Department of Defense update to the enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Schwartz, M. (2010). *Defense acquisitions: How DoD acquires weapon systems and recent efforts to reform the process*. Darby, PA: Diane Publishing.

Sullivan, M. J., Andrew, C., Bowman, R., Fairbairn, B., Neill, S., Oppenheim, J., & Schwenn, R. (2009). *Defense acquisitions measuring the value of DOD's weapon programs requires starting with realistic baselines* (GAO-12-381R). Washington, DC: Government Accountability Office.

Tyson, K. W., Nelson, J. R., Om, N. I., & Palmer, P. R. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (No. IDA-P-2201). Alexandria, VA: Institute For Defense Analyses.

Weapon Systems Acquisition Reform Act of 2009, 111-23 Cong. (2009).

Under Secretary of Defense for Acquisition, Technology, & Logistics (USD[AT&L]). (2008). *Operation of the Defense Acquisition System* (Department of Defense Instruction 5000.02). Washington, DC: Author.

Wirthlin, J. R. (2009). *Identifying enterprise leverage points in Defense Acquisition Program performance* (Doctoral dissertation). Cambridge, MA: Massachussetts Institute of Technology.

Wirthlin, J. R., Houston, D. X., & Madachy, R. J. (2011). Defense acquisition system simulation studies. In *Proceedings of the 2011 International Conference on Software and Systems Process.* Honolulu, HI (pp. 189–193).

THIS PAGE INTENTIONALLY LEFT BLANK

Intervention Strategies for the DoD Acquisition Process

The Department of Defense (DoD) acquisition system is marred by schedule and cost overruns that can be attributed to a complex acquisition process. This process drives great research interests in exploring intervention strategies that would help reduce program delays. Recent studies indicate that policies that address (1) scope growth, (2) acquisition process variability, and (3) program technology maturity should be investigated because they may have a significant impact on reducing program completion time. However, quantitatively evaluating the impact of the new policy has been limited due to the lack of system models with appropriate fidelity. The application of a simulation model to address this challenge thus becomes a promising approach. In this research, we explore the application of the Enterprise Requirements and Acquisition Model (ERAM), a discrete event simulation of the DoD acquisition system to quantitatively examine the effect of scope growth, technology maturity, and decreased variation and means process times in post–design readiness review contractor activities. The insights gained from the simulation experiments can potentially help formulate new policies to improve DoD acquisition.

Introduction

The DoD acquisition system is composed of three separate and distinct processes: the Joint Capabilities Integration Development System (JCIDS), the Planning, Programming, Budgeting & Execution (PPBE) process, and the formal acquisition development system. The system is inherently dynamic and complex and is constantly being adjusted through policy changes or statute (Chairman of the Joint Chiefs of Staff [CJCS], 2012; Weapon Systems Acquisition Reform Act of 2009; Under Secretary of Defense for Acquisition, Technology, & Logistics [USD(AT&L)], 2008). As a result, a high percentage of DoD projects are being completed behind schedule and over-budget (Schwartz, 2010). A 2009 Government Accountability Office (GAO) report indicated that for the DoD's 2008 portfolio, on average, a program faced a 22-month delay and exceeded the original budget (Sullivan et al., 2009).

Due to the problematic performance of the DoD acquisition, extensive research has explored ways to identify intervention strategies in hope of improving the acquisition process. Three intervention strategies are of special interests from existing literature, including (1) scope growth, (2) the issue of process variability, and (3) technology maturity (McNutt, 1999; P. Montgomery, Carlson, & Quartuccio, 2012; Moorman, 2005; Sullivan et al., 2009). Scope growth, usually reflected as schedule slippage according to the Packard Commission (1986), is the root cause of the delays. However, to the best of our knowledge, studies have focused mainly on evaluating the existing system based on historical data. Process variability, and the closely related technique of Lean Six Sigma, has been adopted to improve the operational processes within the end-to-end acquisition practice (Moorman, 2005). The application of the lean concept into the earlier phases of acquisition (e.g., before the Milestone C [MS-C] review [Labedz, Initiative, & Harvey, 2006; Moorman, 2005]) has been limited. The third outstanding issue is related to technology maturity. According to the DoD Instruction 5000.02 (DoDI 5000.02), the objective of the Technology Development Phase is to determine the mature technologies that will be integrated into the full system (USD[AT&L], 2008). Surprisingly, recent GAO reports indicate that of 26 programs reviewed, only 11 entered the post–Milestone B (MS-B) phases with mature technology. It is apparent that the Technical Readiness Level (TRL) 7 requirement of MS-B has not been consistently met. We conclude that these intervention strategies have been well recognized in DoD acquisition process to great extent. However, the research outcome to date is less than satisfactory. This is probably due to the lack of a detailed system model that enables the quantitative assessment of these what-if scenarios.

In 2009, a large scale Discrete Event Simulation (DES) model called the Enterprise Requirements and Acquisition Model (ERAM) was developed (Wirthlin, 2009) to comprehensively depict the actual acquisition processes of the DoD using the Air Force implementation as the base for validation and verification. The ERAM opens opportunities to not only evaluate acquisition process, but also re-evaluate the performance after intervention strategies have been implemented. Hence, the purpose of this research is to explore the capability of ERAM in implementing interventions on scope growth, process variability and technology maturity. We enhance the ERAM model by introducing different scenarios that represent these interventions. The program completion time is then evaluated to gain the insights on the efficacy of the interventions.

Literature Review

Simulation Research in the DoD Acquisition System

The DoD’s interest in modeling and simulation of its internal systems has been constantly growing at a rapid pace (Kölsch, 2011). The DoD Modeling and Simulation coordination office recognizes this need and provides the needed acquisition education to the acquisition workforce using simulation (Olwell, Johnson, Didoszak, & Few, 2012). In addition to the ERAM model (Wirthlin, 2009), there are notable efforts in exploring the applications of simulation to the end-to-end DoD acquisition system. For example, McQuay (1997) summarized the growing trend of applying simulation to the DoD acquisition and provides the implementation guidelines for future research. Keane, Lutz, Myers, and Coolahan (2000) proposed an architecture for simulation-based acquisition that covers the operational system and technical views of the acquisition system, to name just a few. Simulation has also been used as a teaching tool to accelerate the experience of new DoD acquisition employees (Bodner et al., 2012). In addition, simulation is used in conjunction with risk analysis tools to study an F-35 Joint Strike Fighter program acquisition (Bodner, 2012), and portfolio management for project optimization (Mun & Housel, 2010). Given the importance of validation and verification in simulation research, Pace (2004) and Harmon and Youngblood (2005) proposed a methodology for the validation of DoD simulation acquisition models. Elele and Smith (2010), as well as Balci (1997) proposed a risk-based verification, validation, and accreditation process for simulation models in the DoD setting. The contribution of this research, using ERAM as a foundation, is to quantitatively evaluate intervention strategies that could improve the DoD acquisition process.

Intervention Strategies for DoD Acquisition Improvement

Scope Growth in Acquisition

According to the Packard Commission report (1986), the key issue from which most acquisitions problems stem is the long cycle time of programs. Some of the problems that find root in long acquisition time are high costs of development, fielding of obsolete technology, and slow response time to military threats (McNutt, 1999; Packard Commission, 1986). A study in 1983 found that the development time has increased significantly since the 1950s. The 2009 report from the GAO indicated that on average a program experience s 22-month delay. In 1989, Tyson, Nelson, Om, and Palmer examined schedule variance and its causes, and concluded that prototyping, sole-source procurement, fixed-priced contracts, and multiyear procurement may reduce schedule variance. The same study also found that competitive programs experience more schedule growth than non-competitive programs. Drenzer and Smith (1990) performed a statistical analysis of 10 programs aiming to identify the factors that affect the original plan and/or program deviation. A more recent research by Brown, Flowe, and Hamel (2007) compared the schedule quality of joint and single-system programs and concluded that joint system programs have significantly more schedule breaches. Reig (1995) studied 24 programs that had completed MS-C behind schedule and proposed the use of an indicator for schedule slippage, which is defined as the percentage of Low Rate Initial Production (LRIP) test quantity divided by the total planned production quantity. However, using a larger sample size, Gailey (2002) repeated the study performed by Reig and found no correlation between LRIP quantities and schedule slippage. In conclusion, all the literature reviewed above addressed schedule slippage from a historical perspective and used a relatively small sample size, which, as Reig (1995) and Gailey (2002) show, may lead to inaccurate conclusions.

Reducing Process Variance Through Lean Six Sigma

Lean focuses on adding maximum value for the customer while reducing waste, and the goal of six sigma is to improve quality by identifying root causes of defects and minimizing variability. Despite the success of Lean Six Sigma projects in industry, it is not utilized as widely in the DoD end-to-end process. As Apte and Kang (2006) summarized, four primary obstacles facing the application of Lean Six Sigma in the DoD setting are as follows: (1) Lean depends on empowerment of the employee. The traditional and strict hierarchy in the DoD opposes this. (2) Frequent rotation in officers causes frequent change of culture. (3) Few incentive programs for contributions to improvement efforts. (4) Uncertainty in supply and demand due to the nature of the industry. Still, there are notable efforts in DoD community in using lean in acquisition (Labedz et al., 2006). For example, the Letterkenny Army depot has a successful lean implementation (Labedz et al., 2006). Red River Army Depot Repair Facility implemented a lean project leading cost of repair decreased from $89,000 to $48,000. The Global Hawk team from the Air Force broke down the value stream of the contracting process and eliminated wasteful steps. These changes reduced the process time of the three processes by 37%, 43%, and 73% (Moorman, 2005). Although exciting, the implementation has been on operational level practice, and the improvement on the overall acquisition process to MS-C is currently lacking, which may be due to the complexity and dynamic nature of the system.

Technology Maturity

Mature technology is a critical facet in decreasing cycle time. Monaco and White (2005) reviewed the literature on improving schedule time and concluded that the most often cited driver for schedule time is technology readiness. Sherman and Rhoades (2010) listed four pre-conditions necessary for reducing cycle time, among which, three are related to technology maturity.

During the DoD acquisition process, the primary phase where technology maturity is developed and assessed is during the pre–MS-B processes known as the Technology Development Phase (USD[AT&L], 2008). According to DoDI 5000.02, “The purpose of [the Technology Development Phase] is to reduce technology risk, determine and mature the appropriate set of technologies to be integrated into a full system and to demonstrate Critical Technology Elements (CTEs) on prototypes” (USD[AT&L], 2008). During this assessment, a review known as a technology readiness assessment is performed. The Technology Readiness Level (TRL) is assigned to each element of technology that is critical to the function for the program or that poses a large risk. The TRL system was proposed by NASA in the mid-1970s and adopted by the Air Force and DoD in the 1990s and 2001, respectively (Mankins, 2009). The definitions of the nine TRLs are presented in Table 1.

Table 1. Technology Readiness Levels
(Adapted from Mankins, 2009)

TRL Level	Description
1	Basic principles observed and reported
2	Technology concept and/or application formulated
3	Analytical and experimental critical function and/or characteristic proof-of-concept
4	Component and/or breadboard validation in laboratory environment
5	Component and/or breadboard validation in relevant environment
6	System/subsystem model or prototype demonstration in relevant environment (ground or space)
7	System prototype demonstration in the planned operational environment
8	Actual system completed and “qualified” through test and demonstration (in the operational environment)
9	Actual system “proven” through successful system and/or mission operations

Although the TRL of all CTEs is required to be 6 in order to pass MS-B, the GAO asserted that a majority of the programs’ current problems of being behind schedule and over budget are caused by technology immaturity (Gallegos et al., 2011). Specifically, 45% of the programs that had reached the Critical Design Review had immature technologies. As stated by Meier (2008), it is the immature technology that causes delayed schedule and cost growth. In addition, overzealous advocacy is a major cause for the large amount of technology growth seen in the DoD. This may explain the problematic performance of DoD acquisition to some extents.

The very earlier attempt to correlate technology maturity with schedule slippage may be Dubos, Saleh, and Brown (2007). From a space acquisition standpoint, Dubos et al. (2007) developed TRL-Schedule Risk curves and a model that characterizes the random variable, schedule slippage, as a function of TRL. From observations of 28 space programs, the relation between TRL and schedule slippage determined by Dubos et al. (2007) can be seen in Table 2. With the increasing maturity level, the slippage decreases dramatically. However, even at the TRL 8, there is still probability of schedule slippage due to a number of unforeseeable reasons.

Table 2. Mean Relative Schedule Slippage and TRL
(Adapted from Dubos et al., 2007)

TRL	Observed mean relative slippage (data)	Modeled mean relative schedule slippage
4	78%	88%
5	57%	50%
6	20%	29%
7	19%	16%
8	7%	9%

Another method for quantifying the effect of immature technology is presented in Valerdi's (2005) schedule estimation tool called CoSySMo. The model contains a driver that represents the level of technology risk in the program (Valerdi, 2005). The definition and effect on total schedule time of technology risk described in CoSySMo is summarized in Table 3.

We note that both Dubos et al. (2007) and Valerdi (2005) correlated the schedule with cost benefits without considering the costs associated with obtaining higher levels of technology maturity, which limits its application to invention strategy evaluations. This research will fill in the gap by quantifying the effect on technology requirements' rigor in order to compare the value and cost of requiring high levels of technology maturity.

Table 3. CoSySMo Technology Risk Driver Values
(Adapted from Valerdi, 2005)

	Very Low	Low	Nominal	High	Very High
Lack of Maturity	Technology proven and widely used throughout industry	Proven through actual use and ready for widespread adoption	Proven on pilot projects and ready to roll out for production jobs	Ready for pilot use	Still in the laboratory
Lack of Readiness	Mission proven (TRL 9)	Concept qualified (TRL 8)	Concept has been demonstrated (TRL 7)	Proof of concept validated (TRL 5& 6)	Concept defined (TRL 3 & 4)
Rating Scale Value	0.67	0.82	1.00	1.36	1.85

In summary, we reviewed the simulation applications in DoD acquisition and three important intervention strategies explored in DoD research and practices. In the next section, we discuss the application of ERAM to quantitatively evaluate different intervention scenarios, which may provide insights to identify potential policies for DoD acquisition improvement.

ERAM–Discrete Event Simulation Model

The ERAM is a discrete event simulation that simulates the behavior of the end-to-end DoD acquisition system (Wirthlin, 2009). It was initially developed using Arena simulation software and later migrated to ExtendSim software. The scope of the ERAM extends from the generation of capability requirements in the JCIDS process to Milestone C (MS-C), where the review before the production stage begins. Because ERAM is abstracted at a very high level, the overall system performance can be studied. Figure 1 illustrates the architecture of ERAM.

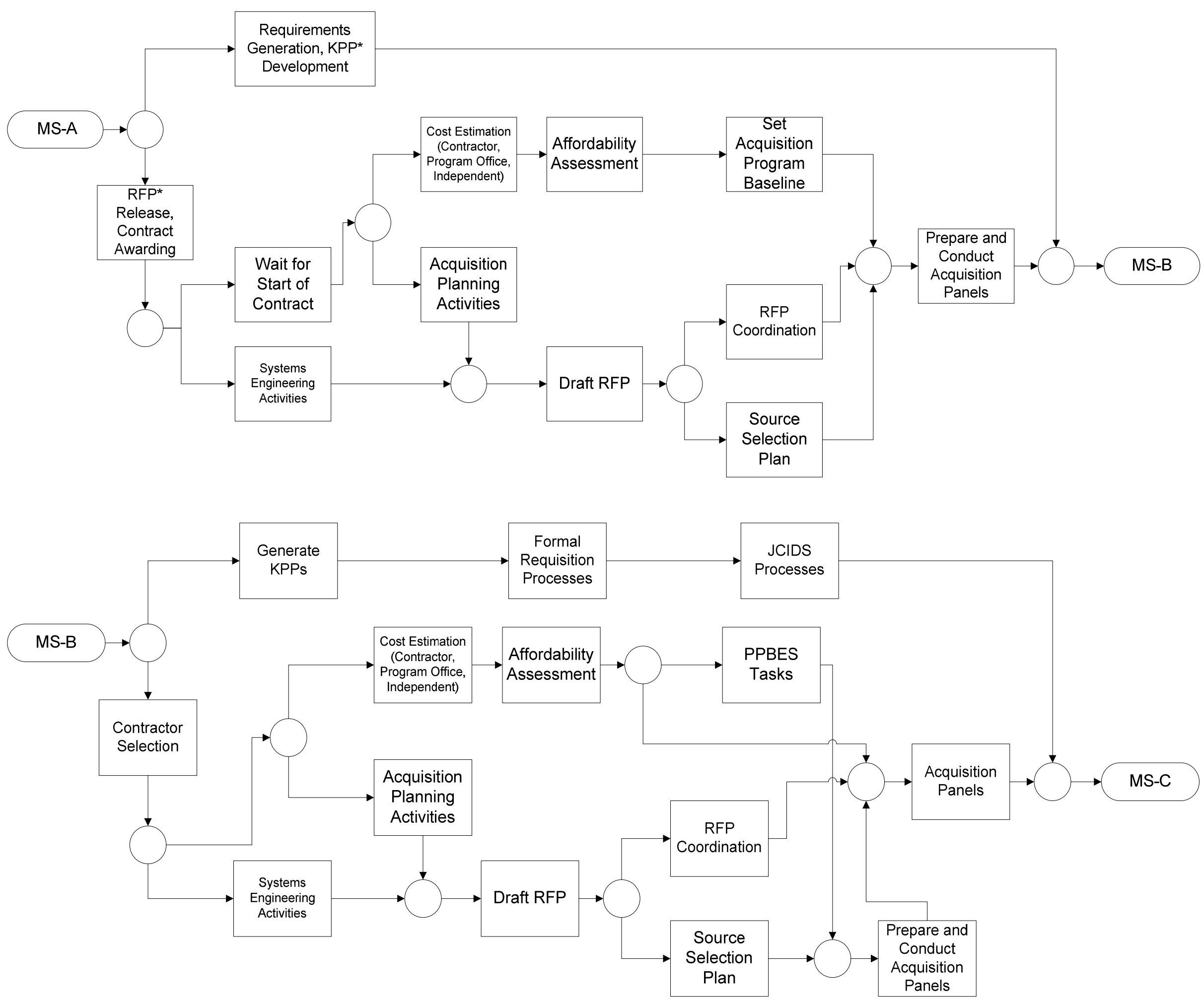

Figure 1. Block Diagram of the ERAM Model

Note. *AoA = Analysis of Alternatives, KPP = Key Performance Parameters, RFP = Request for Proposal

To appreciate the scale of ERAM, Table 4 provides an approximate size reference of the main processes within each milestone in ERAM. Each milestone is composed of several similar minor swimlane processes. Specifically these main processes are

JCIDS where requirements are developed, the Planning, Programming, Budgeting and Execution System (PPBES) process where actual allocation of resources are done, the Acquisition process where acquisition processes are simulated, and lastly the Contractor processes that illustrate the steps a contractor undergo during acquisition.

Table 4. Approximate Size of Model Swimlanes

Swimlane	Major Processes	Number of Modules
Pre A	JCIDS	88
	PPBES	10
	Acquisition	35
Pre-B	JCIDS	32
	PPBES	4
	Acquisition	73
	Contractor	21
Pre-C	JCIDS	30
	PPBES	6
	Acquisition	106
	Contractor	31

Verification and validation of the ERAM included hand modeling, iterations of correction from feedback of experts in all three branches of acquisition, and comparison of schedule and budget information from the Defense Acquisition Management Retrieval (DAMIR) and System Metric and Reporting Tool (SMART) data accesses, which contain schedule and cost data for past DoD projects, to distributions of the schedule time of model-generated data. The validated ERAM model thus provides the baseline for the following experiments.

Experiments and Analysis

Determining Number of Replications

To determine the number of replications needed to achieve the desired level of precision, an initial trial is performed with 3,000 replications. Of those 3,000 replications, 446 reached MS-C. Using an alpha level (α) of 0.05, a relative precision of 0.05 (ε), and the sample standard deviation (S0) from an initial run, Equation 1 is used to determine the number of trials (R) needed to obtain the desired level of precision.

$$R \geq \frac{z\alpha/_2 S_0}{\varepsilon} \quad (1)$$

From the calculations, it is found that 1,008 trials are needed. Since approximately 15% of total replications reach MS-C, the total number of replications needed to obtain 1,008 complete data points is 7,200. Therefore, for all experiments in this research, 7,500 trials are performed in order to ensure that the desired relative precision is achieved.

Experiment I: Scope Growth

The first experiment is to evaluate the impact of scope growth on the end-to-end acquisition time. We first modify the ERAM model to enable this scenario analysis. Specifically, we modify two decide modules entitled "Scope Growth Technical Problems PreB" and "Scope Growth Technical Problems PreC." These modules are located in the contractor swimlane of the model in the pre–MS-B and pre–MS-C portion of the model. Within the simulation model, entities called "Events Happen" are created by a create module in the contractor swimlane. These entities represent an event or issue occurring during the time of the contract. After being created, the entities enter the "Scope Growth Technical Problems" module. If "true," the entity then triggers a delay in the program. If "false," the entity triggers the next program review if an appropriate percentage of the contract has elapsed.

A baseline trial using the setting of 20% for both modules is conducted with 7,500 replications. This setting translates to the pre–MS-B contract length increasing by 36% and the pre–MS-C contract length increasing by 30%. The mean and sample standard deviation of the time for programs to reach MS-C is calculated from the baseline trial. Next, the probability of scope growth in pre–MS-B and pre–MS-C is adjusted independently in order to determine the probability of scope growth that would cause a statistically significant difference in end-to-end process time. The scope growth in pre–MS-B and pre–MS-C are then increased and decreased simultaneously. A paired *t*-test is then performed to compare each trial to the baseline at 95% confidence level.

Table 5, Table 6, and Table 7 summarize the results of the *t*-tests performed for decreased scope growth. Here, the null hypothesis for the *t*-tests is $\mathrm{H}_0: \mu_{\mathrm{base}} = \mu_{\mathrm{i^{th}}_{\%}}$, which corresponds to an insignificant difference between the baseline and the $\mathrm{i^{th}}$ percentage, if not rejected, and alternative hypothesis $\mathrm{H}_1: \mu_{\mathrm{base}} \neq \mu_{\mathrm{i^{th}}_{\%}}$, if there is significant difference. Specifically, Table 5 shows the results of the *t*-tests performed when only contractor activities in the pre–MS-B portion of the model experience changing scope growth. The tables show a subset of trials corresponding to 0%, 20%, 70%, or 100% of events that corresponded to scope growth. Table 6 shows the results of the *t*-tests performed when only contractor activities in the pre–MS-C portion of the model experience scope growth. The tables show a subset of trials corresponding to 0%, 20%, 50%, 55%, and 60% (the rejecting H0 hypothesis) of events that corresponded to scope growth. Table 7 shows the *t*-test results when both pre–MS-B and -C contractor activities experience scope growth to the same degree. The table shows a subset of trials corresponding to 0%, 45%, 47.5%, or 50% (the rejecting H0 hypothesis) of events that corresponded to scope growth.

These settings are chosen to show the sensitivity of the model to various degrees of scope growth. From these simulations, the mean ($\mu_{i^{th}\%}$) and standard deviation of the total completion time for each trial are calculated. These calculated means are compared to the mean of the baseline setting (μ_{base}) in which only 20% of events occurring are scope growth for both pre–MS-B and pre–MS-C.

Table 5. Summary of *t*-Test Results of Pre–MS-B Scope Growth

	% of Events Resulting in Scope Growth			
	20% (Baseline)	0%	70%	100%
Average Time to MS-C (Days)	3755.23	3707.07	3823.57	3842.08
Standard Deviation (Days)	1502.19	1484.137	1612.62	1641.52
p-Value		0.458	0.312	.203
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0

Table 6. Summary of *t*-Test Results of Pre–MS-C Scope Growth

	% of Events Resulting in Scope Growth				
	20% (Baseline)	0%	50%	55%	60%
Average Time to MS-C (Days)	3755.23	3704.53	3849.39	3876.43	3926.10
Standard Deviation (Days)	1503.19	1519.33	1531.38	1557.34	1556.01
p-Value		0.780	0.153	0.069	0.011
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0	Reject H_0

Table 7. Summary of *t*-Test Results of Pre–MS-B and Pre–MS-C Scope Growth

	% of Events Resulting in Scope Growth				
	1% (Baseline)	0%	45%	47.5%	50%
Average Time to MS-C (Days)	3755.23	3646.79	3821.35	3904.69	3911.48
Standard Deviation (Days)	1503.19	1467.97	1567.21	1627.41	1624.54
p-Value		0.093	0.326	0.028	0.022
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Fail to Reject H_0	Reject H_0

From Table 5, it is evident that the end-to-end acquisition time is not sensitive to changing levels of scope growth during the MS-B contractor activities because all tests yielded results that are statistically insignificant as compared to the baseline scenario. Even when the probability of scope growth is set to 100%, corresponding to a 145% increase in contract length, there is not a significant effect on time to MS-C. One explanation of this result could be that in the Acquisition processes, not all programs are required to go through every phase. More than half of all programs begin after the MS-B (Wirthlin, 2009). Therefore, efforts focused on improving processes prior to MS-B will not have a significant effect. In Table 6, it is evident that the end-to-end acquisition time is not sensitive to a decrease of the degree in scope growth during the MS-C contractor activities. The test when scope growth is set to 0% yielded results that are statistically insignificant as compared to the baseline scenario. However, when the degree of scope growth is approximately 60%, the time to MS-C has a statistically significant increase. This 60% setting translates to the contract increasing by 87%. This translation is determined by comparing the average final contract time to the average original contract time for all trials with the 60% setting. In Table 7, it is evident that the model is most sensitive to changes in both the MS-B and MS-C portion of the model. When overall degree of scope growth is approximately 50%, the time to MS-C has a statistically significant increase. This 50% scope growth translates to an approximate 70–75% growth in total contract length due to scope growth and technical issues. However, like the independent testing of MS-B and -C scope growth, reducing the degree of scope did not have a statistically significant effect on the end-to-end acquisition time.

Experiment II: Reduction of Variance and Process Mean in Post–Design Readiness Review Contractor Activities

Similar to the first experiment, to test the effect of reduced variance and process mean in Post–Design Readiness Review contractor activities on the end-to-end acquisition time, five process modules are modified (see Table 8). As seen in Table 8, the triangular (TRIA) distributions of these processes are expressed as a function of the System Development and Demonstration (SDD) contract length and the Acquisition Category (ACAT) level of the program.

The *t*-test results of individual process mean reduction is conducted with 7,500 trials. The mean and sample standard deviation of time to MS-C for this baseline trial is calculated. The variability is adjusted by reducing the difference between the mode and the maximum/minimum by a constant in order to determine the setting that would cause a statistically significant difference in end-to-end process time. Next, the entire distribution is multiplied by a coefficient in order to effectively increase or decrease the process mean time in order to determine the setting that would cause a statistically significant difference in end-to-end process time. This is performed for each of the five modules independently and as a collective unit. A paired *t*-test is then performed to compare each trial to the baseline at 95% confidence level.

Table 8. Baseline Setting of Post–Design Readiness Review Contractor Process Modules

Process Module Name	Baseline Setting
Fabrication	TRIA(.06*SDD original contract length, .1*SDD original contract length, .11*SDD original contract length)
Assembly	TRIA(.06*SDD original contract length, .1*SDD original contract length, .11*SDD original contract length)
Integrated testing	(ACAT Level==1*0.15*SDD original contract length+ACAT Level==2*0.07*SDD original contract length+ACAT Level==3*0.07*SDD original contract length,ACAT Level==1*0.25*SDD original contract length+ACAT Level==2*0.1*SDD original contract length+ACAT Level==3*0.1*SDD original contract length,ACAT Level==1*0.26*SDD original contract length+ACAT Level==2*0.11*SDD original contract length+ACAT Level==3*0.11*SDD original contract length)
Developmental system testing and Live Fire test and Operational Assessment testing	TRIA(ACAT Level==1*0.18*SDD original contract length+ACAT Level==2*0.1*SDD original contract length+ACAT Level==3*0.1*SDD original contract length,ACAT Level==1*0.25*SDD original contract length+ACAT Level==2*0.15*SDD original contract length+ACAT Level==3*0.15*SDD original contract length,ACAT Level==1*0.27*SDD original contract length+ACAT Level==2*0.17*SDD original contract length+ACAT Level==3*0.17*SDD original contract length)
Combined testing	TRIA(.07*SDD original contract length, 0.1*SDD original contract length, 0.11*SDD original contract length)

Table 9, Table 10, and Table 11 summarize the results of the *t*-test for decreasing the variation and mean process in post–design readiness review contractor processes. Here, the null hypothesis for the *t*-tests is $H_0: \mu_{base} = \mu_i$, which corresponds to no significant difference between the baseline and the i^{th} mean or variance reduction scenario, if not rejected, and alternative hypothesis $H_1: \mu_{base} \neq \mu_i$, if there is a significant difference. Specifically, Table 9 shows the effect of decreasing the triangular distributions of each process separately by 75%. Table 10 illustrates the effect of decreasing the process mean for all five processes by a percentage. The percentages tested are 0%, 20%, 25%, and 50%. These setting are chosen to show the sensitivity of the model to changes in post–design readiness review contractor activities as a whole. Table 11 summarizes the results of altering the variance of all five processes. In addition to the baseline, two scenarios are tested. The first scenario equates the minimum and maximum of the triangular distribution to the mode for each process representing all programs performing at the expected level with no variance. The second scenario reduces the mode and maximum so that they equal the minimum. This represents all programs illustrating the best-case-scenario process times with no variance.

Table 9. Summary of *t*-Test Results of Individual Process Mean Reduction

	Process with 75% Mean Reduction					
	None (Baseline)	Fabrication	Assembly	Integrated Testing	Developmental Testing	Combined Testing
Average Time to MS-C (Days)	3755.23	3691.53	3691.53	3755.23	3755.23	3694.95
Standard Deviation (Days)	1503.19	1512.43	1512.43	1503.19	1503.19	1513.81
p-Value Conclusion		0.220 Fail to Reject H_0	0.220 Fail to Reject H_0	1.00 Fail to Reject H_0	1.000 Fail to Reject H_0	0.240 Fail to Reject H_0

Table 10. Summary of *t*-Test Results of Overall Mean Production

	% Decrease in Process Distribution Times			
	0% (Baseline)	20%	25%	50%
Average Time to MS-C (Days)	3755.23	3697.86	3656.43	3607.12
Standard Deviation (Days)	1503.19	1500.596	1494.67	1492.41
p-Value		0.254	0.076	0.012
Conclusion		Fail to Reject H_0	Fail to Reject H_0	Reject H_0

Table 11. Summary of *t*-Test Results of Overall Variation Reduction

	Baseline	No Variance at Mean Performance	No Variance at High Performance
Average Time to MS-C (Days)	3755.23	3796.41	3659.16
Standard Deviation (Days)	1503.19	1540.97	1512.34
p-Value		0.718	0.08
Conclusion		Fail to Reject H_0	Fail to Reject H_0

In Table 9, the reduction of 75% is chosen to represent a scenario that would be difficult to achieve by simply leaning out the process. It is evident from the table that decreasing the process mean of a single process, even to an extreme amount, does not result in a reduction of mean. One particularly interesting result from this section is that decreasing the process mean of integrated testing and developmental testing has no effect on the time to MS-C; both trials show an identical mean and standard deviation to the baseline trial. One possible explanation for this outcome is that the two events are not on the critical path. However, findings in Table 10 indicate that reducing the mean process time in all post–design readiness review contractor abilities has the potential to significantly affect the overall time-to-MS-C. However, the amount of improvement must be significant, approximately 50%.

The findings in Table 11 indicate that performing at the mean performance level with no variance will not have a significant effect on end-to-end process time. In fact, the *t*- statistic from this test is a negative value, indicating that the baseline trial actually had a smaller (although not statistically significantly smaller) time to MS-C than the trial with no variance. A possible explanation for this anomaly is that the triangular distribution for these processes is skewed to the right. On the other hand, performing at a high performance level with no variance yields *p*-value, although not quite reaching the .05 cut-off. In Table 11, it is evident that simply removing variance will not have a significant impact on time to MS-C; however, reducing variance and helping all programs perform at the level that the best programs are currently performing at has the potential to significantly decrease the time to MS-C.

Experiment III: Increased Rigor in Technology Maturity

In experiment III, we plan to test the effect of technology maturity on the end-to-end acquisition time; only the decide module entitled “MDA Milestone Approval PreB” is modified. This module is located in the pre–MS-B portion of the model. The “MDA Milestone Approval PreB” module represents the percentage of programs that pass the MS-B review. If the entity does not pass, the entity goes through a delay before returning to the “MDA Milestone Approval PreB” module. Failing this review twice leads to the program being terminated and the entity is disposed. Passing the review allows the program to continue to pre–MS-C processes and represents a program achieving TRL 6 or higher for all Critical Technology Elements (CTEs).

A baseline setting of 99% was conducted with 7,500 trials. The mean and sample standard deviation of the time to MS-C for this baseline trial was calculated. Next, the probability of passing the program review was increased in order to determine the setting that would cause a statistically significant difference in end-to-end process time. A paired *t*-test was then used to compare each trial to the baseline at 95% confidence level.

Table 12 summarizes the results of the *t*-tests performed for mature technology. The null hypothesis for the *t*-tests is $\mathrm{H_0}: \mu_{\mathrm{base}} = \mu_{\mathrm{i^{th}}\%}$, which corresponds to an insignificant difference between the baseline and the $\mathrm{i^{th}}$ percentage, if not rejected, and alternative hypothesis $\mathrm{H_1}: \mu_{\mathrm{base}} \neq \mu_{\mathrm{i^{th}}\%}$, if there is significant difference. Furthermore, Table 12 shows the *t*-test results when the probability that a program MS-B review is decreased. It presents a subset of trials corresponding to 40%, 45%, 50%, and 99% of events passing the MS-B review. These settings are selected to show the sensitivity of the model to changes in the percentage of programs that pass MS-B, or the rigor level of MS-B reviews. From these simulations, the mean and standard deviation of each trial are calculated. These calculated means are compared to the mean of the baseline setting in which a program has a 99% of passing MS-B.

Table 12. Summary of *t*-Test Results of the Rigor of MS-B Review

	Probability of a Program Passing Milestone			
	99% (Baseline)	50%	45%	40%
Average Time to MS-C (Days)	3772.38	3715.72	3635.21	3578.05
Standard Deviation (Days)	1520.92	1505.43	1491.06	1507.45
p-Value		0.192	0.047	0.005
Conclusion		Fail to Reject H_0	Reject H_0	Reject H_0

In Table 12, it is evident that when the probability of programs passing the MS-B review is more than 50%, there will not be a statistically significant increase in the end-to-end acquisition time. However, when the percentage of programs that pass MS-B is reduced to approximately 47.5%, the model exhibits a statistically significant increase in program time. These results indicate that from a statistical view, more rigorous guidelines can be put in place during MS-B without negative effect.

Conclusions and Future Research

This research investigated three possible intervention strategies: scope growth, reducing the mean and variance of processes, and technology maturity for improving the end-to-end DoD acquisition process using simulation model.

Through simulating the effect of contractor-driven scope growth, findings suggest that from a purely statistical standpoint, decreasing the amount of scope growth that occurs during acquisition programs will not have a statistically significant impact on end-to-end program time. In addition, the program is only sensitive to very large increases in scope growth with the pre–Milestone C (MS-C) activities having a greater sensitivity than pre–Milestone B (MS-B) activities. From an empirical standpoint, these results suggest that, although the program managers should prevent excessive creep in scope, imposing more stringent methods of scope growth management would not be beneficial.

The second intervention investigated is the effect of decreasing the mean and variation in post–design readiness review contractor activities. The results from simulation indicated that while leaning out individual process may not have an effect on the time until MS-C, focusing on improvements that reduce the mean process time for the entire fabrication, assembly, and testing portion of the acquisition process could have significant effects.

However, more importantly, the results from these two interventions have implications about the overall constraints in the DoD acquisition process. These experiments simulated having zero scope growth and performing with no variance at the optimum level for five major contracting processes. These conditions would be considered ideal by most program managers. Despite these ideal circumstances, the model indicated no significant difference in end-to-end program completion time.

These results indicate that according to the model, the contractor processes are not on the critical path of the acquisition process. During the acquisition process, there are many parallel processes occurring. Many of these processes occur in the PBBES and JCIDS branch, as well as the acquisition branch. A delay in any of these parallel paths will cause the time to MS-C to increase. Both interventions indicate that according to the modeled system, until the processes external to the program manager office are improved, improvements in the contractor's branch will have minimal effect on the overall program. In addition, trials where scope growth is increased indicate that unless the contract experiences schedule delays that cause the contract length to increase by more than 70%, it will continue to be off of the critical path for the acquisition project.

The last strategy investigated is increasing rigor in the MS-B requirements. As discussed in the literature review, the GAO has found that many times, programs reach post MS-B activities when one or more CTE is not to TRL 6 or higher (Gallegos et al., 2011). In addition, Dubos et al. (2007) found that when programs are at TRL 4 and 5, the schedule slippage is usually 78% and 57%, respectively. This schedule slippage is significant compared to 20% at TRL 6. Results from the simulation experiments performed indicated that the value gained from have more rigorous requirements for passing MS-B outweighs the schedule costs. Currently, a program has a 99% chance of passing the MS-B review. In this regard, MS-B has become a "rubber stamp" review. From interviews with individuals from various branches of the acquisition service, Wirthlin (2009) found that "there is tremendous institutional pressure to push the activity as far forward in the acquisition system as possible." This pressure comes from the desire to deliver capabilities in a timely manner. The results from this research indicate that this pressure to push a program through the system may be preventing the acquisition system from reaching reduced cycle times. From literature and this research, it can be seen that from a purely statistical standpoint, the potential schedule benefits of being critical of the technology maturity level at MS-B outweigh the schedule penalties of implementation.

The results from all three interventions suggest a pervasive issue facing the DoD acquisition system. From the results of testing the effect of scope growth and reduction of process mean in post DRR contractor activities, there is statistical evidence to support the claim that the myriad of support process designed to oversee the contractor and assist the program manager have overpowered the capacity of the system and are in part responsible for the long cycle times of the acquisition process. The research on technology maturity suggests that the need to push the program through such a large number of activities may potentially turn many into rubber stamps that do not add true value to the system. Overall, this research suggests that DoD decision-makers need to analyze these support and oversight activities to assess the validity of their purpose and investigate whether their implementation is achieving that purpose.

Although this research provides evidence to support that the contractor activities are not on the critical path of acquisition system, it does not identify the critical path. Future research is needed to examine the ERAM and determine the critical path. In addition, research is needed to further validate the model's findings, beyond the original subject matter expert input and face validity analysis of the 2009 ERAM version. One limitation of this research is that the problem is looked at from a purely statistical standpoint. For the interventions tested, while the expected value may have decreased during certain interventions, it is important to note that the sample size is approximately 1,000, while the number of projects in the DoD portfolio at a given time is usually about 96 major acquisition programs (Sullivan et al., 2009). Although the evidence from the experiments suggest that the interventions may decrease the average end-to-end process time, the inherently large standard deviation of the process, the small sample size, and the long-term duration of the programs could prevent this effect from being realized. Further practical, in-field studies could be performed to validate the results of this empirical study. The use of simulation in this study places some limitations on this research, but also provides opportunities for further research. For example, when looking at increasing the TRL for passing MS-B, this study assumed that this increased rigor would lead to higher TRL's and less schedule slippage. However, the simulation does not yet incorporate the complex interactions between early quality (passing MS-B) and future project quality, say at MS-C. Further studies can investigate these interactions and add them into the ERAM simulation.

References

Apte, U., & Kang, K. (2006). *Lean Six Sigma for reduced cycle costs and improved readiness*. Retrieved from http://www.dtic.mil

Azizian, N., Sarkani, S., & Mazzuchi, T. (2009). A comprehensive review and analysis of maturity assessment approaches for improved decision support to achieve efficient defense acquisition. In *Proceedings of the World Congress on Engineering and Computer Science 2* (pp. 20–22).

Balci, O. (1997). Verification validation and accreditation of simulation models. In *Proceedings of the 29th Conference on Winter Simulation* (pp. 135–141).

Bodner, D. A., Wade, J. P., Squires, A. F., Reilly, R. R., Dominick, P. G., Kamberov, G., & Watson, W. R. (2012). Simulation-based decision support for systems engineering experience acceleration. In *2012 IEEE International Systems Conference (SysCon)* (pp. 1–6).

Bodner, D. (2012). *Addressing risk in the acquisition lifecycle with enterprise simulation* (ADA591729). Atlanta GA: Georgia Institute of Technology.

Brown, M. M., Flowe, R. M., & Hamel, S. P. (2007). *The acquisition of joint programs: The implications of interdependencies.* Washington, DC: Office of the Secretary of Defense.

Chairman of the Joint Chiefs of Staff (CJCS). (2012). *Joint Capabilities Integration and Development System* (CJCS Instruction 3170.01H). Washington, DC: Author.

Drezner, J. A., & Smith, G. K. (1990). *An analysis of weapon system acquisition schedules* (RAND/R-3937-ACQ). Santa Monica, CA: RAND.

Dubos, G. F., Saleh, J. H., & Braun, R. (2007). *Technology readiness level: Schedule risk and slippage in spacecraft design: Data analysis and modeling* (AIAA-2007-6020). Long Beach, CA.

Elele, J. N., & Smith, J. (2010). Risk-based verification, validation, and accreditation process (77050E-77050E-11). In *Proceedings of SPIE 7705*, *Modeling and Simulation for Defense Systems and Applications V.* Orlando, FL.

Gailey, C. K., III. (2002). *Predictive power for program success from engineering and manufacturing development performance trends* (TR 1-02). Fort Belvoir, VA: DAU Press.

Gallegos, A., Graveline, W. R., Schwenn, R. E., Fairbairn, B. D., Chitikila, R. C., Smythe, W. P., . . . Cherveny, M. (2011). Defense acquisitions: Assessments of selected weapon programs (GAO-11-233SP). Washington, DC: Government Accountability Office.

Harmon, S., & Youngblood, S. M. (2005). A proposed model for simulation validation process maturity. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 2*(4), 179–190.

Houston, D. (2012). *Enterprise requirements and acquisition model.* Unpublished manuscript. Los Angelos, CA: Space and Missile Center (SMC).

Keane, J. F., Lutz, R. R., Myers, S. E., & Coolahan, J. E. (2000). An architecture for simulation-based acquisition. *Johns Hopkins APL Technical Digest, 21*(3), 348–358.

Kölsch, M. (2011). A snapshot of the modeling and simulation community and education. *Journal of Defense Modeling and Simulation: Applications, Methodology, Technology, 8*(4), 217–224.

Labedz, C. S., Jr., Initiative, L. A., & Harvey, R. K. (2006). *Letterkenny Army Depot: Finance innovations support Lean Six Sigma success*. Cambridge, MA: Lean Aerospace Initiative.

Leach, D., & Searle, C. (2011). *Department of Defense enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Mankins, J. C. (2009). Technology readiness assessments: A retrospective. *Acta Astronautica, 65*(9), 1216–1223.

McNutt, R. T. (1999). *Reducing DoD product development time: The role of the schedule development process*. Wright-Patterson AFB, OH: Air Force Institute of Technology.

McQuay, W. K. (1997). The new vision for modeling and simulation in the DoD acquisition process. In *16th AIAA/IEEE Digital Avionics Systems Conference (DASC), 1 0.2-1-7* (Vol. 1). Irvine, CA.

Meier, S. R. (2008). Best project management and systems engineering practices in the preacquisition phase for federal intelligence and defense agencies. *Project Management Journal, 39*(1), 59–71.

Monaco, J. V., & White, E. D., III. (2005). Investigating schedule slippage. *Defense Acquisition Review Journal*, *12*(2), 176–193.

Montgomery, J. (2012). *Department of Defense update to the enterprise requirements and acquisition model* (Master's thesis). Wright-Patterson AFB, OH: Air Force Institute of Technology.

Montgomery, P., Carlson, R., & Quartuccio, J. (2012). *System definition-enabled acquisition (SDEA)—A concept for defining requirements for applying model-based systems engineering (MBSE) to the acquisition of DoD complex systems* (NPS-AM-12-C9P02Z01-100). Retrieved from http://oai.dtic.mil/oai/oai?verb=getRecord&metadataPrefix=html&identifier=ADA563266

Moorman, R. (2005). Implementing Lean. *Overhaul and Maintenance, 11*(6), 22.

Mun, J., & Housel, T. (2010). *A primer on applying Monte Carlo Simulation, real options analysis, knowledge value added, forecasting, and portfolio optimization.* Monterey, CA: Naval Postgraduate School.

Olwell, D. H., Johnson, J. M., Didoszak, J. M., & Few, S. M. (2012). *Modeling and simulation education for the acquisition and T&E workforce: Final report*. Retrieved from https://calhoun.nps.edu/public/handle/10945/6988

Pace, D. K. (2004). Modeling and simulation verification and validation challenges. *Johns Hopkins APL Technical Digest, 25*(2), 163–172.

Packard Commission. (1986). *A formula for action: A report to the president on defense acquisition*. Washington, DC: Author.

Reig, R. W. (1995). A decade of success and failure in the DOD acquisition system. *Program Manager, 24*, 27–29.

Sauser, B., Ramirez-Marquez, J. E., Magnaye, R., & Tan, W. (2009). *A systems approach to expanding the technology readiness level within defense acquisition* [DTIC document].

Schwartz, M. (2010). *Defense acquisitions: How DoD acquires weapon systems and recent efforts to reform the process.* Darby, PA: Diane Publishing.

Sherman, J. D., & Rhoades, R. G. (2010). Cycle time reduction in defense acquisition. *Research-Technology Management, 53*(5), 46–54.

Sullivan, M. J., Andrew, C., Bowman, R., Fairbairn, B., Neill, S., Oppenheim, J., & Schwenn, R. (2009). *Defense acquisitions: Measuring the value of DOD's weapon programs requires starting with realistic baselines* (GAO-09-543T) Washington, DC: Government Accountability Office.

Tyson, K. W., Nelson, J. R., Om, N. I., & Palmer, P. R. (1989). *Acquiring major systems: Cost and schedule trends and acquisition initiative effectiveness* (No. IDA-P-2201). Alexandria, VA: Institute for Defense Analyses.

Weapon Systems Acquisition Reform Act of 2009, 111-23 Cong. (2009).

Under Secretary of Defense for Acquisition, Technology, & Logistics (USD[AT&L]). (2008). *Operation of the Defense Acquisition System* (Department of Defense Instruction 5000.02). Washington, DC: Author.

Valerdi, R. (2005). *The constructive systems engineering cost model (COSYSMO)*. University of Southern California.

Valerdi, R., & Kohl, R. J. (2004). An approach to technology risk management. In *Engineering Systems Division Symposium 3* (pp. 29 –31).

Wirthlin, J. R. (2009). *Identifying enterprise leverage points in defense acquisition program performance* (Doctoral dissertation). Cambridge, MA: Massachussetts Institute of Technology.

Modeling Space Launch Process Delays to Improve Space Vehicle Acquisition Planning

Department of Defense (DoD) space acquisition programs almost always experience significant schedule growth, which drives cost growth. Even with the advent of more reliable launch vehicles, schedule delays often exceed three years and have the implication of reduced military or national security capabilities, significant increases in costs, and occasionally program cancellations. This paper provides acquisition professionals insight into the DoD's space launch process through modeling and simulation. Seven causal delay types are identified, and these factors are analyzed to draw conclusions about schedule growth considerations. We discuss the implications of these launch delay factors and make recommendations for those involved with space policy, acquisitions, and launch.

Introduction

DoD space acquisition programs are continually plagued with significant unplanned schedule growth. Cost and schedule for required launch capabilities are determined early in the acquisition process based on requirements. An increase in space launch delay affects not only the individual program, but also other space acquisition programs' funding and schedule. The Government Accountability Office (GAO) conducted a study of multiple space programs in 2006 to determine the cause of increased costs to programs. Not surprisingly, the study revealed original schedule estimates were particularly unrealistic and unachievable (GAO, 2006; GAO, 2012).

In the early 1990s, the United States suffered a series of failures of several satellite launch systems that were intended to support DoD and National Intelligence Community (IC) missions. In total, three Titan IV mishaps resulted in nearly $3 billion in fiscal losses. Subsequently, the Delta III launch vehicle experienced two failures during commercial launches. In addition to the launch failures, several in-flight anomalies occurred, casting a shadow of doubt on the nation's ability to guarantee assured access to space (U.S. Air Force, 1999).

As a result of unacceptable failure rates, the DoD introduced major changes in the risk posture associated with space launch. In August 2008, then–Major General Pawlikowski, deputy director of the National Reconnaissance Office (NRO), stated, "In the almost decade since the costly failures of the late 1990s, the Air Force Space and Missile Systems Center (SMC) and the NRO have adopted a back-to-basics approach to mission assurance." These changes, while intended to promote mission success and assure access to space, also had the unintended consequence of increasing costs and extending launch schedules.

In order to drive down cost and schedule impacts, the DoD generated a new approach to space launch. Boeing's Delta IV and Lockheed Martin's Atlas V launch vehicles became the crucible of our Nation's space launch capability for large DoD and intelligence community satellites. This approach became known as the Evolved Expendable Launch Vehicle (EELV) and was intended to solve cost and schedule issues via competitive pricing and assured access to space on either of the two launch vehicles (Saxer, Knauf, Drake, & Portanova, 2002).

In 2005, Boeing and Lockheed Martin formed the United Launch Alliance (ULA), the organization the DoD currently uses to contract Delta IV and Atlas V launch support (Saxer et al., 2013). Moreover, the DoD is actively pursuing the possibility of commercial space launch from vendors such as Space X, the company that has recently supported multiple resupply missions for the International Space Station.

We focused on data from EELV missions because it is the primary launch capability for the United States DoD and intelligence community. The data was collected from the Launch Information Support Network (LISN) and includes missions from June 2006 through March 2013. This effort included the modification and extension of an existing acquisition process simulation, the Enterprise Requirements and Acquisition Model (ERAM). As part of a broader study of simulating DoD acquisition process improvement, this research analyzed the implications of the space launch process on satellite acquisition schedule. With validated simulation of DoD acquisition, research can be conducted on process improvement through critical path and bottleneck analysis, sensitivity analysis of delays, or Lean and Six Sigma task time modifications.

Background

Acquisition Process Modeling

In 2008, ERAM was developed by Lieutenant Colonel J. Robert Wirthlin in an effort to understand key interactions between the requirements generation (i.e., Joint Capabilities Integration Development System), funding (i.e., planning, programming, budgeting, and execution), and acquisition program portions of the DoD's acquisition process from post–Milestone A through Milestone C. Wirthlin's (2009) model focused on the schedule implications of the process and interactions associated with each arm of the acquisition process in an effort to identify critical interactions that regularly led to significant schedule delays and proposed process or policy modifications that could be implemented to reduce schedules for acquisition programs. Specifically, Wirthlin (2009) focused on such questions as "Why does the system behave the way that it does?" and "Are there changes that can significantly improve the schedule?" Majors Leach and Searle (2011) extended ERAM, focusing on space system acquisition. Major Montgomery (2012) later extended ERAM in 2012 by modeling the rapid acquisition process often used by organizations executing Joint Urgent Operational Needs (JUONS). More recent research has used Monte Carlo analysis to examine pre–Milestone B and C bottleneck analysis and alternatives (interventions) from the baseline simulation (Worger, Jalao, Wirthlin, Colombi, & Wu, 2013; Worger, Jalao, Augur, et al., 2013). Extensions to this discrete event simulation for space launch drove this research.

Space Launch Policy and Process

The Air Force Space Command's (AFSPC's) launch scheduling process guidance and lower echelon documentation was reviewed to ensure comprehensive understanding of the manifesting processes. The primary document that provided the information necessary to understand these processes was Air Force Instruction (AFI) 10-1211, Space Launch Operations. AFI 10-1211 outlines the roles and responsibilities of the Air Force as the DoD Executive Agent for Space. Furthermore, it places the SMC commander as the sole focal point for certification of all DoD and NRO launch vehicles. Additionally, this document specifies that "launch schedule execution will be based on national priorities" and designates AFSPC as the responsible agent for establishing the manifest for all DoD, civil, and commercial missions (Chandler, 2006).

The launch manifest process is outlined in AFSPCI10-1213, *Launch Scheduling and Forecasting Procedures* (Weinstein, 2012) and *AFSPC Long Range Launch Scheduling Process* (LeMaitre, 2005). These documents discuss the Current Launch Schedule Review Board (CLSRB) process from the initial launch support request through launch for space systems. Specifically, the Long Range Launch Scheduling Process outlines the National Launch Forecast (NLF) compilation in the 4-to-11 year future and how it flows into the Space Launch Manifest (SLM), which is a near-term, three-year schedule for launches. The CLSRB is a body of stakeholders convened biannually to certify the next 18 months of the SLM (LeMaitre, 2005). The Air Force Space Command Instruction guidance memorandum AFSPCI 10-1213 implements minor changes to the process by creating a series of launch commit reviews (LCRs) to assess risk related to launch vehicle (LV) readiness, space vehicle (SV) readiness, ground/control system readiness, and operations readiness. It further delineates organizational responsibilities for each of these risk assessments, and assesses missions scheduled for the next 18 months (Weinstein, 2012). These documents define launch scheduling and the capability of the U.S. launch industry.

Space Launch Assessments

The single most significant document related to the evolution of the space launch process over the past 15 years is the Space Launch Vehicle Broad Area Review (SLV BAR). The SLV BAR, led by Gen. Larry D. Welch, highlighted several problems with the space launch process, which occurred in the 1990s, specifically, the increase in launch failure rates from one per year over a 12-year period to five failures within 10 months. Mission assurance and quality incidents also raised from 18 incidents in 200 launches to 9 in 51 launches, a 100% increase (U.S. Air Force, 1999). The SLV BAR began a period of intense scrutiny related to launch vehicle mission assurance, but the added attention to detail and slower pace yielded strong success rates (LeMaitre, 2005).

RAND Corporation, a nonprofit research and analysis organization intended to improve policy and decision-making, highlighted additional issues with the space launch segment, discussing the ramifications of a reduced commercial launch requirement on the cost and schedule of government launches. These issues ultimately led to the combination of the Delta IV and Atlas V teams forming the United Launch Alliance to preserve the EELV heavy lift capability (McCartney, et al., 2006).

The GAO’s annual assessments repeatedly highlighted issues with technology, design, and production maturity for the spacecraft. Additional issues included synchronization of space and ground segment activation, changes in prescribed program production rates, software-related delays, and fiscal and manning constraints (GAO, 2006; GAO, 2012).

Methodology and Analysis

A model of the space launch process was developed using grounded theory to gain insight into process times, delay causes, and key integration and decision points (Corbin & Strauss, 2008; Montgomery, 2012). Insight was gathered via discussions with subject matter experts (SMEs) from space-community locations throughout the United States. A total of 14 SMEs were utilized, including members from the Air Force Headquarters staff, Air Force Space Command staff, various space vehicle program offices, and the space launch community. Furthermore, SME volunteers ranged from government civilians, military, technical support contractors from Aerospace Corporation, and industry contract partners from the ULA. Most SMEs had 15 or more years in the industry, and some had as many as 30 years experience with the space launch process, including several active and retired senior military leaders. The SME discussions covered the full spectrum of the space launch process, including space launch requirements, budgeting, space vehicle integration, and launch operations.

Delay Types

In addition to qualitative SME input, historical data was collected from the LISN database maintained by the Launch, Ranges and Networks Division of Headquarters Air Force Space Command. Available data was collected on missions from June 2006 through March 2013. This dataset included 33 missions and a record of 389 launch date changes and causes for these changes. Each mission history yielded many Launch Change Request (LCR) data inputs, ranging from as few as four LCRs to as many as 32 LCRs. The team then binned the individual delays found on LISN into the most appropriate category, and separated them by launch vehicle type (Atlas vs. Delta) to allow for statistical analysis.

Using the coding techniques of grounded theory (Evans, 2013), the team discerned seven primary categories of delay plaguing the space launch process post–Milestone C (Corbin & Strauss, 2008). The delay categories are described in Table 1. Delays either occurred for similar reasons, at similar times within the planned timeline, or were due to common external factors. These categories aided in simplifying the model and provided a venue for later analysis. Most significantly, data coding ensured individual categories fit common and manageable distributions for inclusion into the resulting model.

Atlas and Delta Comparison

SME discussions indicated a potential difference between the space launch timelines associated with Atlas and Delta missions. Specifically, it was believed the launch vehicle long term, launch vehicle short term, re-queuing, and possibly the priority delay categories were dependent upon the launch vehicle type and associated reliability and launch rates. Delays related to the space vehicle, both early and late, as well as weather and miscellaneous delays were expected to be launch vehicle agnostic.

Upon examination of the data shown in Table 2, the delay categories did not appear significantly different based on the launch vehicle type; a direct contradiction to SME expectations. A *t*-test was completed against the null hypothesis that the Atlas and Delta sample means were equal for each factor; all *p*-values shown do not reject the null hypothesis at a 0.05 level of significance.

Table 1. Taxonomy of Space Launch Delay Categories

Space Launch Delay Categories	
Delay Type	**Description**
Space Vehicle—Early (SV Early) (>18 months)	Delay initiated by the SV program office 18 months or more prior the predicted launch date. These delays typically have little impact on the ability to manifest a specific desired launch date at either Vandenberg AFB (VAFB) or Cape Canaveral Air Force Station (CCAFS).
Space Vehicle—Late (SV Late) (<18 months)	Delay initiated by the SV program office within 18 months of the current predicted launch date. These delays often impact the ability to manifest a desired launch date at either Vandenberg AFB or Cape Canaveral AFS, depending on manifest density. The SV late delays were often of shorter duration than SV early delays, leading to a separate distribution.
Launch Vehicle—Long Term (LV Long) (>18 months)	Delay initiated by the launch vehicle or associated leadership due to known manufacturing issues, launch separation requirements, or updates to an Initial Launch Capability (ILC) for the specific mission. These delays typically have little impact on the ability to manifest a specific desired launch date at either Vandenberg AFB or Cape Canaveral AFS.
Launch Vehicle—Short Term (LV Short) (<18 months)	Delay initiated by the launch vehicle or associated leadership due to unforeseen issues with the launch vehicle, near-term launch date change requests by the mission integrator, or a launch vehicle anomaly on a previous mission that has a ripple effect on the mission of interest. These delays may impact the ability to manifest a desired launch date at either Vandenberg AFB or Cape Canaveral AFS, depending on manifest density.
Re-queue	Delay or, in seldom cases, acceleration encountered when a program attempts to re-enter the launch manifest after it was removed due to another delay such as SV Early. This occurs more often as the re-entry attempt is closer to the planned launch date, generally within 18 months.
Priority	Delay or acceleration of the launch date due to mission priorities. This occurs when the CLSRB process or senior leadership determines a launch date must slip or in seldom cases move earlier to accommodate mission requirements.
Weather/Miscellaneous (Wx/Misc.)	Delay of relatively short duration caused by weather, launch window refinement, or launch range support issues.

Table 2. Initial Statistics of Launch Delay Categories

Atlas Delay Category Statistics (Months)							
	SV-Early	SV-Late	LV-Long	LV-Short	Re-queue	Priority	Wx/Misc
Sample Size	25	83	32	28	31	15	17
Mean	3.77	1.18	1.22	0.50	1.75	1.87	0.14
Std Dev	3.99	2.84	2.62	0.63	2.20	4.73	0.20

Delta Delay Category Statistics (Months)							
	SV-Early	SV-Late	LV-Long	LV-Short	Re-queue	Priority	Wx/Misc
Sample Size	20	24	41	28	24	11	10
Mean	4.40	1.09	1.85	0.62	2.30	2.35	0.24
Std Dev	7.68	1.33	2.35	1.64	2.85	3.26	0.52

	SV-Early	SV-Late	LV-Long	LV-Short	Re-queue	Priority	Wx/Misc
T Statistic	-0.3305	0.2197	-1.0600	-0.3645	-0.7885	-0.3094	-0.5674
p-Value	0.7436	0.8266	0.2932	0.7177	0.4349	0.7597	0.5818

Delay Analysis

Based on these results, the data was consolidated into a single EELV data set for all further analysis, as shown in Table 3. This data set was used to build a process model.

It was also insightful to examine the launch delays from a program perspective, with the contribution of overall delay and occurrence by all seven delay categories shown in Table 4. This information can be useful by systems engineers and program managers to baseline their schedules. The delay category appearing to have the most significant impact on a program's schedule is the space vehicle–early delay. The space vehicle–early delay is experienced by the SV program office 18-months or more prior to the predicted launch date. At this point in an acquisition program, significant fixes or changes may be incurred, usually extending schedules due to satellite disassembly, reassembly, test, and analysis involved in the specific resolution. Although this delay occurs on average only 1.36 times per program, its overall time is the largest at 4.05 months per delay.

Table 3. Overall Launch Delay Statistics

Overall Launch Delay Statistics for 33 Programs (Months)							
	SV-Early	SV-Late	LV-Long	LV-Short	Requeue	Priority	Wx/Misc
Sample Size	45	107	73	56	55	26	27
Probability of Delay Within Category	0.51	0.73	0.67	0.61	0.67	0.58	0.42
Mean Delay	4.05	1.16	1.57	0.56	1.99	2.07	0.18
Median	3.03	0.49	0.72	0.21	0.82	1.00	0.03
Std Dev	5.85	2.58	2.47	1.23	2.49	4.10	0.35
Min Delay	-3.09	-5.46	-4.05	0.03	-0.30	-4.38	-0.03
Max Delay	35.00	15.33	14.05	8.75	11.18	12.99	1.71

Table 4. Overall Delay Category Statistics

Overall Delay Category Statistics							
	SV-Early	SV-Late	LV-Long	LV-Short	Re-queue	Priority	Wx/Misc
Avg # of Occurrences per Program	1.36	3.24	2.21	1.70	1.67	0.79	0.82
% Based on Occurrence	11.6%	27.5%	18.8%	14.4%	14.1%	6.7%	6.9%
Avg Time Delay per Occurrence (Months)	4.05	1.16	1.57	0.56	1.99	2.07	0.18
% Based on Average Time per Occurrence	35.0%	10.0%	13.6%	4.9%	17.2%	17.9%	1.5%
Avg Total Time Delayed per Program (Months)	5.52	3.77	3.48	0.95	3.32	1.63	0.14
% Time Delayed per Program	29.3%	20.0%	18.5%	5.1%	17.6%	8.7%	0.8%

Alternatively, the team hypothesizes that satellite assembly, integration, and test issues occurring late in a program have a significant time impact due to their frequency of occurrence and ripple they induce in the overall launch process. Space vehicle–late delays occur on average 3.24 times per program but only incur 1.16 months per delay. More importantly, any delay within 18 months of a planned launch has the potential to induce delays in another delay category, such as priority or re-queue. This ripple effect may be largely eliminated if the initial SV delay is eliminated. Additionally, from a systems engineering perspective, unforeseen issues late in a program tend to have much greater impact on cost and schedule than early delays.

As shown in Table 5, space vehicle–early delays contribute the most to the overall delay a program can expect to encounter, at 29.3%. In fact, roughly half of all program delay (49.3%) is encountered due to the space vehicle program itself, while the other half is contributed by the launch vehicle or process. Additionally, it could be argued the re-queue delay is in some cases due to a late space vehicle program delay, hence increasing delay contributed by the space vehicle program.

Table 5. Consolidated Delay Statistics

Consolidated Delay Statistics							
	SV-Early	SV-Late	LV-Long	LV-Short	Re-queue	Priority	Wx/Misc

	Satellite Vehicle	Launch Vehicle or Process
% Time Delayed per Program	49.4%	50.6%

	Satellite Vehicle	Launch Vehicle	Launch Process or Range
% Time Delayed per Program	49.4%	23.5%	27.1%

	Satellite Vehicle	Launch Vehicle	Launch Process	Wx/Misc
% Time Delayed per Program	49.4%	23.5%	26.3%	0.8%

Model Development

The decision was made to model each of the seven delay categories in series using a double loop for each delay type. The logic used is illustrated in Figure 1. The seven individual delays were programmed in series using ExtendSim® 8 (Imagine That, 2013) as an entire simulation segment, as shown in Figure 2.

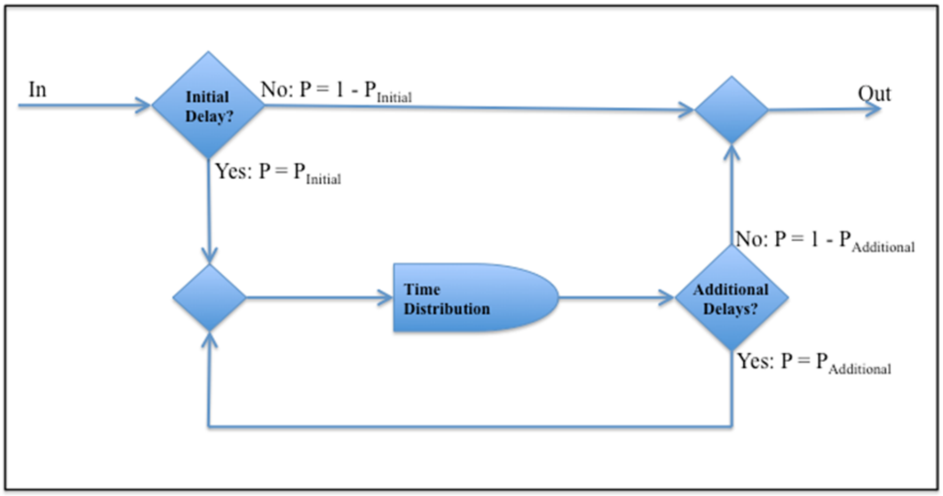

Figure 2. Example of Individual Delay Category Loop

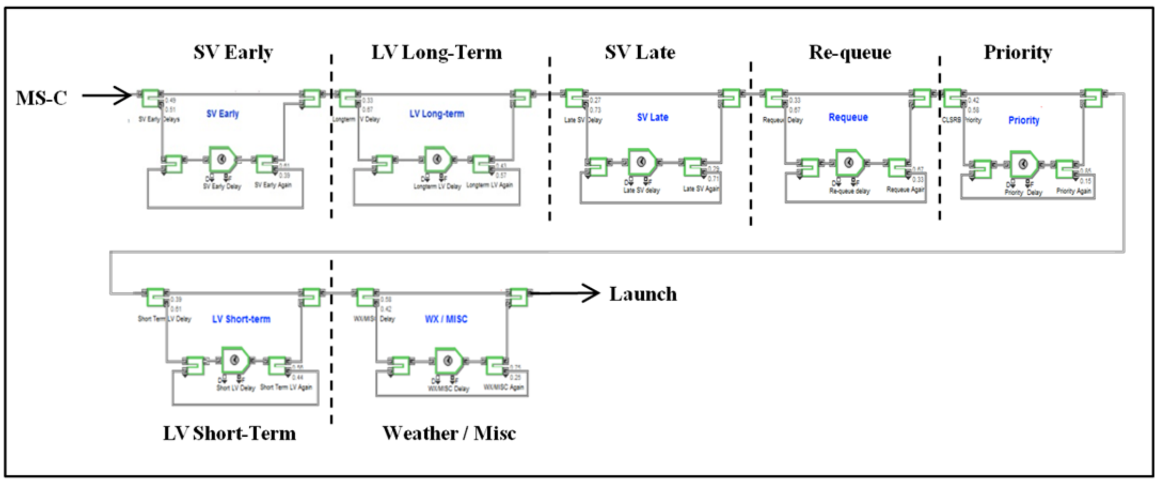

Figure 3. Model as Depicted in ExtendSim

The initial probability block setting for each category was determined by dividing the number of programs that experienced the particular delay by the total number of programs. This determined the probability a particular program would experience a delay within a particular category. This calculation was derived using Table 3 data.

$$P_{initial} = \text{\# of programs delayed in category / \# of total programs} \quad (1)$$

If a program did encounter an initial delay, the second delay block determined the probability of additional delays within the same category. This probability was estimated by taking the reciprocal of the average number of additional delays. This approximation was accurate when the average number of additional delays was greater than one. In the two cases when it was not greater than one, Priority and Wx/Misc, the team used experience to estimate the probability on an additional delay:

$P_{additional}$ = 1 – {1 / [(avg # of additional delays per program per category)]} (2)

The actual time delays themselves were simulated in ExtendSim using activity blocks. The activity blocks simulate a time delay via a randomly-seeded sampling of a pre-assigned distribution. Each pass through the activity block simulates an individual delay occurrence and then flows back into the decision block to determine whether another iteration of the delay will occur. If not, it will flow on to the next delay category, as depicted in Figure 2.

We chose the proper distributions for each activity block after analysis of the actual program delay data. Statistical analysis was conducted on each delay category's actual data, shown in Table 6.

Table 6. Overall Delay Occurrence Statistics

	SV-Early	SV-Late	LV-Long	LV-Short	Requeue	Priority	Wx/Misc
# of "No Delay"	16	9	11	13	11	14	19
Total Programs	33	33	33	33	33	33	33
Probability of First Delay ($P_{initial}$)	0.51	0.73	0.67	0.61	0.67	0.58	0.42
1 - $P_{initial}$	0.49	0.27	0.33	0.39	0.33	0.42	0.58
(Average of "Non-0" Delays) - 1	1.64	3.45	2.31	1.80	1.50	0.36	0.92
Probability of Additional Delays ($P_{additional}$)	0.61	0.29	0.43	0.56	0.67	0.85	0.75

An analytical probability density function (PDF) was chosen to best approximate the empirical histogram data in each category. In most cases, the inverse Gaussian function closely approximated the delay data. Microsoft Excel Solver plugin (Frontline Systems, 2011), was used to minimize the cumulative squared error between the histogram data and the cumulative distribution by optimizing the inverse Gaussian parameters, α and β. The following equation describes this technique:

$$\min_{\alpha,\beta} \sum_{i=1}^{N} \left(F_{\alpha,\beta}(i) - H(i) \right)^2 \quad (3)$$

where:

α and β are parameters to the inverse Gaussian function,

$F_{\alpha,\beta}(i)$ is the cumulative distribution function evaluated at i,

$H(i)$ is the normalized histogram evaluated at point i, and

N is the total number of histogram points.

Two particular delay category histograms, space vehicle–early and priority, did not closely fit an inverse Gaussian distribution or any other distribution. In these cases, simple triangular distributions were used, with minimum, maximum, and most likely values set by observation, excluding outliers (Forbes et al., 2000). This use of triangular distributions was the technique predominantly used within the entire ERAM simulation (Wirthlin, 2009). The distribution parameters used in the model are shown in Table 7. The histograms for each delay category along with the selected overlying distributions, F, are shown in Figure 3.

Table 7. Delay Category Distribution Parameters

	Inverse Gaussian		Triangular		
Delay Category	α	β	**Minimum**	**Maximum**	**Most Likely**
SV - Early			-3	12	4
SV - Late	1.2	3			
LV – Long Term	1.1	1.8			
LV - Short Term	0.01	0.1			
Re-queue	0.59	4.97			
Priority			-4	10	1.3
Wx / Misc	0.152	0.159			

Running and Verifying the Simulation

Our team verified the model by comparing two criteria between actual and modeled data. The first element compared was the average number of times a program or simulation experienced the individual delay categories. This assessment was used to verify the accuracy of the probability blocks used to simulate delay occurrences. Matching the empirical standard deviation, the student's *t*-distribution can be used to determine the minimum number of Monte Carlo replications to achieve a relative precision and significance level. For a 0.05 precision and significance, 473 replications were required; however, a total of 1,000 replications were used.

In each delay category, the model simulations experienced fewer occurrences on average than the actual launch programs. This difference ranged from 16% to

41% among the delay categories. The team considered this difference substantial but accepted it as within reasonable error bounds given the variance of the sample data and relative accuracy of the overall simulation results.

The chosen model most likely underestimated the number of delay occurrences because multiple delays were simulated using only two probabilities, $P_{initial}$ and $P_{additional.}$ This technique was used to simulate the potential for three or more delays of each type, while minimizing overall complexity and maximizing flexibility within the model. The error had a "delay shortening" effect on the overall model results. Additionally, inverse Gaussian distributions used to simulate most of the delay categories have an infinitely long "tail." The team assessed the combination of the above discussed "delay shortening" occurrence estimation error and the "delay lengthening" distribution error actually combined to form an accurate end result. Results of the basic statistics are shown in Table 8. The difference in mean program delay between actual and modeled data is within 1%, with a difference in standard deviation of 4%.

The histograms of the model and actual data, shown in Figure 4, provide a comparison of the respective distributions. Both distributions appear to display a bi-modal nature. Actual data appears to have modes at approximately two and 20–30 months, while the model outputs modes at two and 10–20 months, albeit with a larger tail. Possible reasons for this bi-modal nature are discussed in the Summary and Recommendations section.

Finally, a two-sample Kolmogorov-Smirnov (K-S) test was performed to test for like distributions between the model and actual data. The K-S test is a nonparametric test used to compare a sample's empirical distribution function (EDF) to a reference cumulative distribution function (CDF; Forbes et al., 2000). In this case, the simulated data was treated as the reference CDF; the actual overall program delay data was the sample EDF. The test statistic is calculated under the null hypothesis that both samples are drawn from the same distribution. At a significance level of 0.05, the test statistic was 0.1677, with a *p*-value of 0.3019. The result failed to reject the null hypothesis that both samples are from the same distribution.

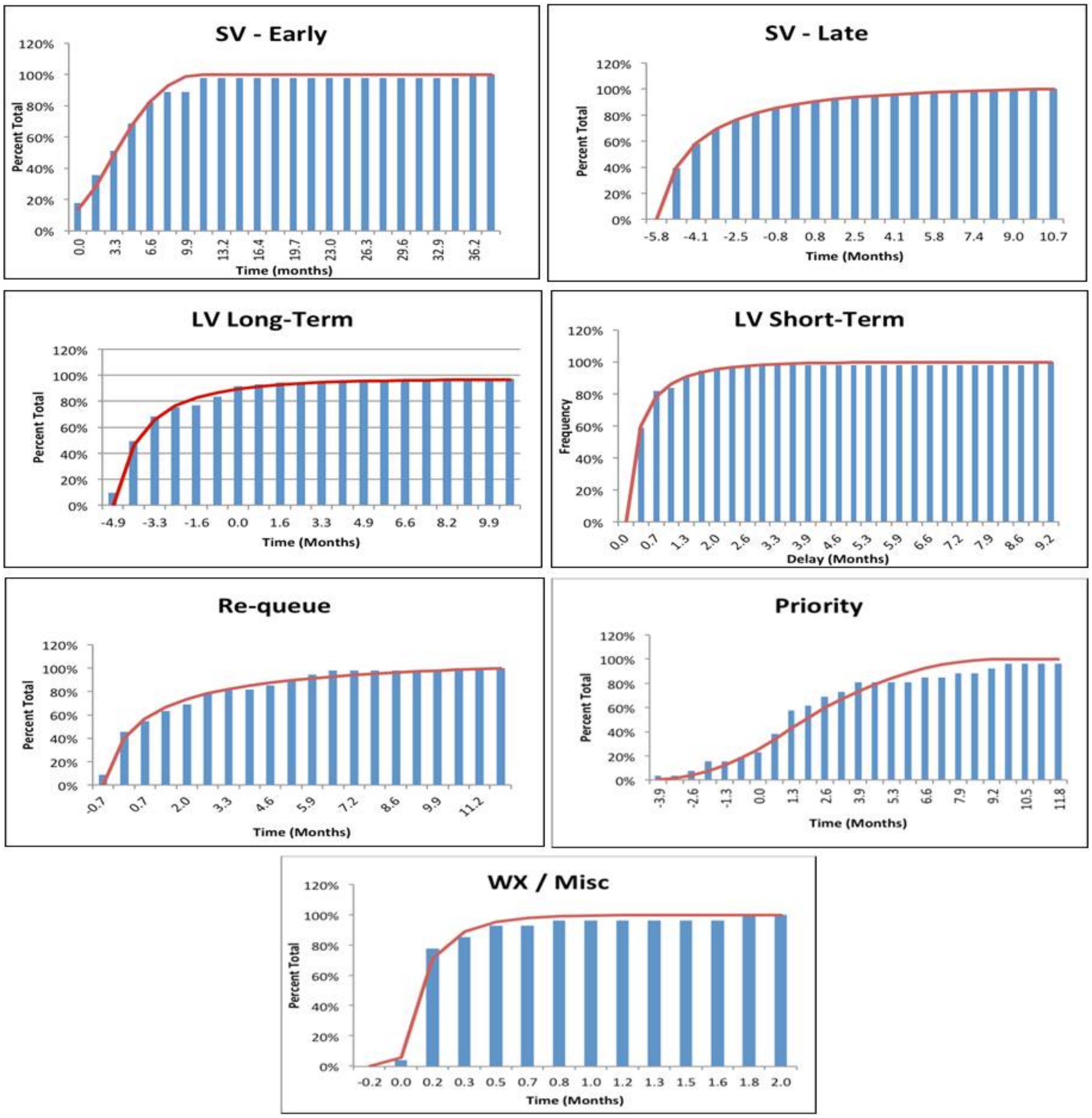

Figure 4. Delay Category Modeled Distributions (Line) and Empirical (Bar)

Table 8. Model (1000 Replications) vs. Actual (33 Program) Delay Statistics

	Mean	Std Dev
Model Results	18.62	11.58
Actual Results	18.82	12.08
Difference	1%	4%

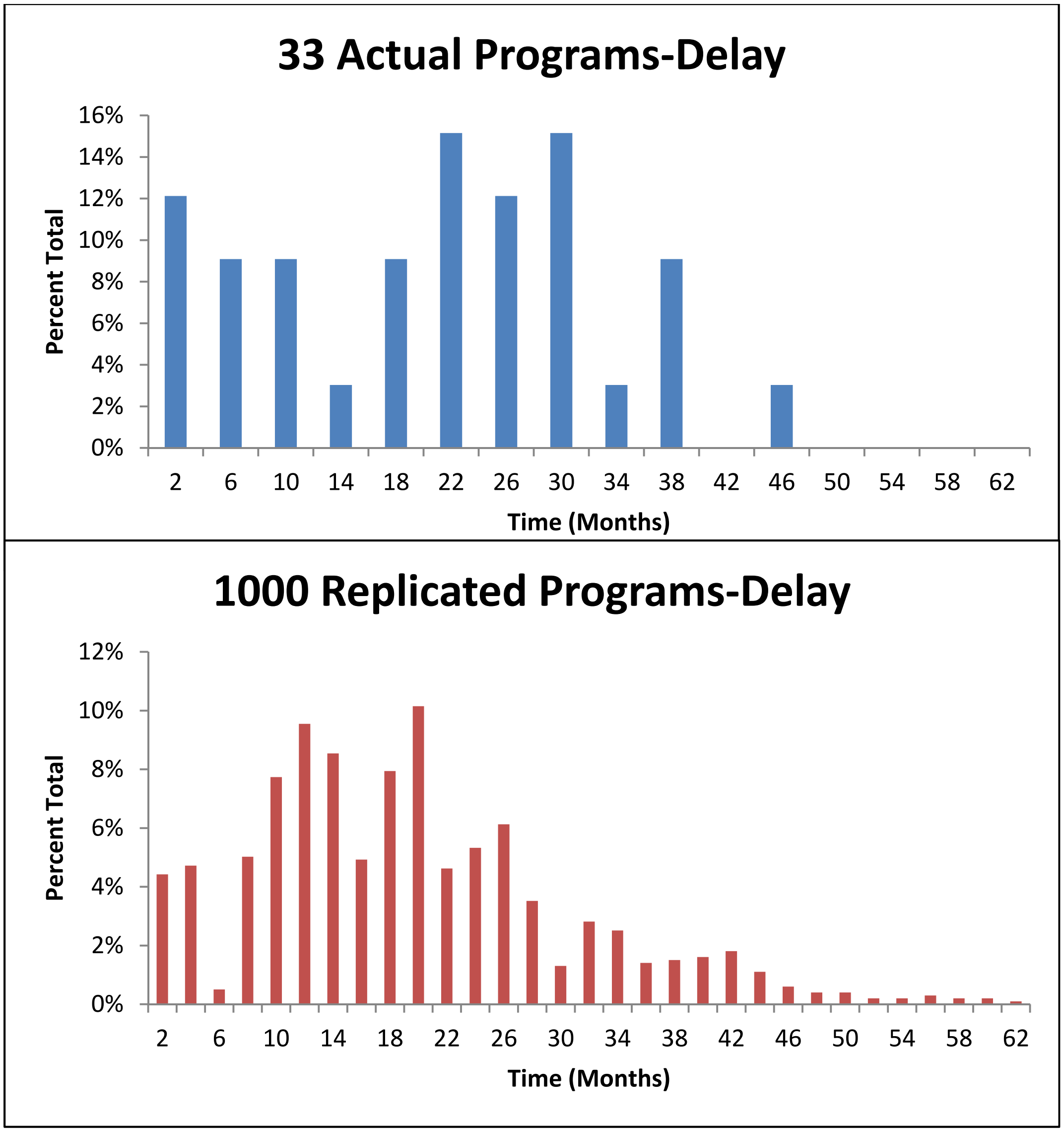

Figure 5. Model and Actual Histograms of Total Program Delay

Summary and Recommendations

Using SMEs, together with LISN launch change request data, we created a serial and iterative model of EELV space launch delays. The average program experienced a delay of 18.82 months, but it appears to experience variation on the characteristics of the satellite program. Technologically mature programs face fewer threats to a schedule and therefore tend to have fewer slips in the integration, testing, and launch phases. In comparison, those space systems with significant complexity, either technological or integrative in nature, tend to experience significantly greater delays.

First, our team identified seven delay categories for use in describing space acquisition schedule delays. These categories were classified as follows: space vehicle–Early (>18 months from launch), space vehicle–late (<18 months), launch vehicle–long term (>18 months from launch), launch vehicle–short term (<18 months), re-queue, priority, and weather/miscellaneous. The statistics on the frequency and length of each category delay was documented.

The second finding was that no statistical difference exists between the Atlas V and Delta IV launch vehicles with respect to schedule delays. This conclusion was based on data collected from 33 historic launch schedules with almost 400 launch change requests. The data showed that delays within each of the seven delay categories can vary significantly for both launch vehicles, but an average delay in each group cannot be statistically separated.

Third, historical delay data exhibited some bimodality, with a mode at a few months and a mode approximately at 20 months. The team hypothesizes that this bimodality may be associated with specific aspects of the space vehicle program. The characteristics most noted were multi-satellite procurements, varying technology risks, varying satellite complexity, contractor risk, and the confidence associated with original programmatic schedule estimates. Often, individual delays are unforeseeable and are caused by manufacturing issues or issues associated with another mission (AFSPC, 2013). A relatively small number of programs caused the short-delay mode at approximately two months; the SMEs reported that programs have the ability to accelerate if there is significant impetus and close coordination between all components of the acquisition process. Lastly, the team found that an urgency of need has shown the ability to drive a program closer to an estimated schedule.

Recommendations and Future Work

Based on SME discussions, current scheduling tools are fairly accurate; however, the perception is that a realistic program schedule often dooms the program in terms of support. This leads to creation of "green-light" program schedules in an effort to compete with other programs for scarce funding. These "green-light" optimistic schedules will only be achieved if every aspect of a long complex space vehicle development goes flawlessly. Based on historical data, the probability of meeting a "green-light" schedule is very low. The acquisition community must overcome this cultural artifact. Some recent efforts have looked at macro-stochastic estimating taking into account empirical changes to baselines (Imagine That, 2013).

In concert with increasing schedule margins to account for expected schedule delays, space programs should continue to assess a "green-light" schedule.

However, similar to the will-cost and should-cost estimation management implemented across DoD acquisition (Carter, 2011), satellite acquisition offices should consider implementing two schedules, a “green-light” and “most-likely” schedule. Acquirers should vigorously pursue the “green-light” schedule with satellite contractors; however, leadership at all levels should be aware that these schedules are optimistic and that the “most-likely” schedule will best suit planning purposes for budgetary and requirements discussions.

Lastly, the space community should implement better practices for tracking historical program timelines and associated causes of delay. This data should be used to ensure lessons learned are properly vetted and passed between programs to alleviate schedule growth issues. Furthermore, future analysis similar to that conducted in this study can target specific areas for schedule improvement.

Research is continuing to analyze the acquisition process using a discrete event simulation, such as ERAM. Such extensions are adding fidelity to the Test and Evaluation subprocesses, and revalidation of the full end-to-end model with more empirical data. Proposals include an Agent-Based Modeling extension to capture complex inter-organizational behaviors, incentives, and rules. ERAM could be extended to encompass other launch vehicles such as the Delta II Medium Launch Vehicle, future launch capability estimates for Space X, Orbital Sciences, and other potential commercial launch vehicles. There exists an opportunity for significant sensitivity analysis of the Space Launch model and its interaction across the DoD acquisition process captured in ERAM.

References

Carter, A. (2011, April 22). *Implementation of will-cost and should-cost management* [Memorandum]. Washington, DC: Under Secretary of Defense.

Chandler, C. H. (2006, July 17). *Air Force Instruction 10-1211: Space launch operations.* Retrieved from http://static.e-publishing.af.mil/production/1/af_a3_5/publication/afi10-1211/afi10-1211.pdf

Corbin, J., & Strauss, A. L. (2008). *Basics of qualitative research: Grounded theory procedures and techniques* (3rd ed.). Thousand Oaks, CA: Sage.

Evans, G. L. (2013). A novice researcher’s first walk through the maze of grounded theory: Rationalization for classical grounded theory. *Grounded Theory Review, 12*(1), 37–55.

Forbes, C., Evans, M., Hastings, N., & Peacock, B. (2000). *Statistical distributions* (3rd ed.). Hoboken, NJ: Wiley.

Frontline Systems. (2011). V11.0 plug-in solver engines user guide. Retrieved from http://www.solver.com

Government Accountability Office (GAO). (2006, March). *Defense acquisition assessment of selected weapons programs* (GAO-06-391). Washington, DC: Author.

Government Accountability Office (GAO). (2012, March). *Defense acquisition assessment of selected weapons programs* (GAO-12-400SP). Washington, DC: Author.

Headquarters, Air Force Space Command (AFSPC). (2013, May 6). Launch Information Support Network. Retrieved from https://lisn.peterson.af.mil/Web/Pages/home.aspx

Imagine That. (2013). Extendsim 8 user guide. Retrieved from http://www.extendsim.com/index.html

LeMaitre, P. (2005). *Air Force Space Command long range launch scheduling process.* Shreiver Air Force Base, CO.

McCartney, F., Wilson, P., Bien, L., Hogan, T., Lewis, L., Whitehair, C.,Freeman, D., ..., Morrell, J. (2006). *National Security space launch report.* Santa Monica, CA: RAND Corp.

Montgomery, J. A. (2012, June). *Expansion of Department of Defense enterprise requirements and acquisition model* (Master's thesis, AFIT/ISE/ENV/12-J03). Wright-Patterson AFB, OH: Air Force Institute of Technology, School of Engineering and Management.

Pawlikowski, E. (2008, August). Mission assurance—A key part of space vehicle launch mission success. *High Frontier, 4*(4), 6–9.

Saxer, R. K., Knauf, J. M., Drake, L. R., & Portanova, P. L. (2002, March–April). Evolved expendable launch vehicle system: The next step in affordable space transportation. *PM*, 2–15.

Saxer, R. K., Knauf, J. M., Drake, L. R., & Portanova, P. L. (2013). Company history: Over 100 years of combined launch history*. United Launch Alliance*. Retrieved from http://www.ulalaunch.com/site/pages/About_ULAHistory.shtml

Searle, C. T., & Leach, D. A. (2011, June). *Department of Defense enterprise requirements and acquisition model* (Master's thesis, AFIT/ISE/ENV/11-J01). Retrieved from http://www.dtic.mil/dtic/tr/fulltext/u2/a545755.pdf

U.S. Air Force. (1999). *Space launch vehicles broad area review report.* Washington, DC: U.S. Air Force.

Weinstein, J. (2012, November 29). *Guidance memorandum (GM) to AFSPCI 10-1213: Launch scheduling and forecasting procedures* (AFSPCI10-1213_AFSPCGM1)*.* Retrieved from http://static.e-publishing.af.mil/production/1/afspc/publication/afspci10-1213/afspci10-1213.pdf

Wirthlin, J. R. (2009, September). *Identifying enterprise leverage points in defense acquisition program performance* (Doctoral dissertation). Cambridge, MA: Massachusetts Institute of Technology.

Worger , D., Jalao, E. R., Auger, C., Baldus, L., Yoshimoto, B., Wirthlin, J. R., Colombi, J., & Wu, T. (2013). Bottleneck analysis for the DoD end-to-end acquisition system. In *Proceedings of the 10th Annual Acquisition Research Symposium*. Retrieved from http://acquisitionresearch.net

Worger, D., Jalao, E. R., Wirthlin, J. R., Colombi, J., & Wu, T. (2013). Effect of the analysis of alternatives on the DoD acquisition system. In A. Krishnamurthy & W. K. V. Chan (Eds.), *Proceedings of the 2013 Industrial and Systems Engineering Research Conference*. San Juan, Puerto Rico.

THIS PAGE INTENTIONALLY LEFT BLANK

ACQUISITION RESEARCH PROGRAM
GRADUATE SCHOOL OF BUSINESS & PUBLIC POLICY
NAVAL POSTGRADUATE SCHOOL
555 DYER ROAD, INGERSOLL HALL
MONTEREY, CA 93943

www.acquisitionresearch.net