

UNCLASSIFIED



SYSTEMS ENGINEERING
Research Center

Valuing Flexibility Phase II

A013 – Final Technical Report SERC-2012-TR-10-2

October 29, 2012

Principal Investigator: Abhijit Deshmukh – Purdue University

Team Members

Barry Boehm, USC

Tom Housel, NPS

Dave Jacques, AFIT

Supannika Koolmanojwong, USC

Jo Ann Lane, USC

Alan Levin, USC

Brandon Pope, Purdue University

Major Erin Ryan, AFIT

Martin Wortman, Texas A&M University

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE 29 OCT 2012	2. REPORT TYPE	3. DATES COVERED 00-00-2012 to 00-00-2012			
4. TITLE AND SUBTITLE Valuing Flexibility Phase II		5a. CONTRACT NUMBER			
		5b. GRANT NUMBER			
		5c. PROGRAM ELEMENT NUMBER			
6. AUTHOR(S)		5d. PROJECT NUMBER			
		5e. TASK NUMBER			
		5f. WORK UNIT NUMBER			
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Purdue University, Systems Engineering Research Center, West Lafayette, IN, 47907		8. PERFORMING ORGANIZATION REPORT NUMBER			
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)		10. SPONSOR/MONITOR'S ACRONYM(S)			
		11. SPONSOR/MONITOR'S REPORT NUMBER(S)			
12. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT <p>Despite its ubiquity in the systems engineering literature, flexibility remains an ambiguous concept. There exist a multitude of definitions, which vary not only by domain, but within domains as well. Despite the confusion, flexibility is an oft purported means for dealing with the well-chronicled cost and time overruns that plague the DoD systems engineering projects. This report provides findings from research conducted under the RT-18a: Valuing Flexibility project. The primary goal of this research project is to identify, develop, and validate sound quantitative methods, processes, and tools (MPTs) that will enable DoD leadership and program managers to make a convincing case for investments in system flexibility when acquisition decisions are made. The research conducted during the first phase of this project (summarized in the Mid-Term Report) focused on identifying current quantitative MPTs for valuing flexibility in DoD contexts, critically evaluated the theoretical underpinnings of these MPTs, and delivered initial capabilities to value investments in flexibility to handle unforeseen sources of change. The current phase of the project focused in three areas: developing a taxonomy for evaluating MPTs for valuing flexibility in DoD contexts (including an overview of a software implementation), extending existing methods by developing new tools for valuing flexibility through life cycle costs, and using real and illustrative scenarios as examples for applying methods to value flexibility, including a detailed case study of flexibility in Ship Maintenance.</p>					
15. SUBJECT TERMS					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT Same as Report (SAR)	18. NUMBER OF PAGES 161	19a. NAME OF RESPONSIBLE PERSON
a. REPORT unclassified	b. ABSTRACT unclassified	c. THIS PAGE unclassified			

Copyright © 2012 Stevens Institute of Technology, Systems Engineering Research Center

This material is based upon work supported, in whole or in part, by the U.S. Department of Defense through the Systems Engineering Research Center (SERC) under Contract H98230-08-D-0171. SERC is a federally funded University Affiliated Research Center managed by Stevens Institute of Technology

Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the United States Department of Defense.

NO WARRANTY

THIS STEVENS INSTITUTE OF TECHNOLOGY AND SYSTEMS ENGINEERING RESEARCH CENTER MATERIAL IS FURNISHED ON AN “AS-IS” BASIS. STEVENS INSTITUTE OF TECHNOLOGY MAKES NO WARRANTIES OF ANY KIND, EITHER EXPRESSED OR IMPLIED, AS TO ANY MATTER INCLUDING, BUT NOT LIMITED TO, WARRANTY OF FITNESS FOR PURPOSE OR MERCHANTABILITY, EXCLUSIVITY, OR RESULTS OBTAINED FROM USE OF THE MATERIAL. STEVENS INSTITUTE OF TECHNOLOGY DOES NOT MAKE ANY WARRANTY OF ANY KIND WITH RESPECT TO FREEDOM FROM PATENT, TRADEMARK, OR COPYRIGHT INFRINGEMENT.

This material has been approved for public release and unlimited distribution except as restricted below.

Internal use:* Permission to reproduce this material and to prepare derivative works from this material for internal use is granted, provided the copyright and “No Warranty” statements are included with all reproductions and derivative works.

External use:* This material may be reproduced in its entirety, without modification, and freely distributed in written or electronic form without requesting formal permission. Permission is required for any other external and/or commercial use. Requests for permission should be directed to the Systems Engineering Research Center at dschultz@stevens.edu

* These restrictions do not apply to U.S. government entities.

ABSTRACT

Despite its ubiquity in the systems engineering literature, flexibility remains an ambiguous concept. There exist a multitude of definitions, which vary not only by domain, but within domains as well. Despite the confusion, flexibility is an oft purported means for dealing with the well-chronicled cost and time overruns that plague the DoD systems engineering projects.

This report provides findings from research conducted under the RT-18a: Valuing Flexibility project. The primary goal of this research project is to identify, develop, and validate sound quantitative methods, processes, and tools (MPTs) that will enable DoD leadership and program managers to make a convincing case for investments in system flexibility when acquisition decisions are made.

The research conducted during the first phase of this project (summarized in the Mid-Term Report) focused on identifying current quantitative MPTs for valuing flexibility in DoD contexts, critically evaluated the theoretical underpinnings of these MPTs, and delivered initial capabilities to value investments in flexibility to handle unforeseen sources of change.

The current phase of the project focused in three areas: developing a taxonomy for evaluating MPTs for valuing flexibility in DoD contexts (including an overview of a software implementation), extending existing methods by developing new tools for valuing flexibility through life cycle costs, and using real and illustrative scenarios as examples for applying methods to value flexibility, including a detailed case study of flexibility in Ship Maintenance.

UNCLASSIFIED

This page intentionally left blank

TABLE OF CONTENTS

Abstract	3
Table of Contents	5
Figures and Tables	6
1 Overview	9
2 Introduction.....	11
3 A Taxonomy of Methods for Valuing Flexibility	16
3.1 Literature Review – Valuing Flexibility	17
3.2 Constructing a Taxonomy	18
3.2.1 Distinguishing System Characteristics.....	19
3.2.2 Model Formulations and Solution Techniques for Valuing Flexibility	21
3.3 Software Implementation: Flexibility Valuation Method Selection Tool (FVMST)	27
3.4 Valuing Flexibility Application: Satellite Design and Operation.....	34
4 Extending Methodologies For Valuing Flexibility	42
4.1 Current Expected Value Life Cycle Cost.....	42
4.2 Operation and Support (O&S) Cost Estimation	50
4.2.1 O&S Reporting	52
4.3 O&S Costs Methodology	53
4.3.1 Enhanced TOC-PL Modeling with Monte Carlo Analysis.....	56
4.3.2 Mixed Model Approach	58
4.3.3 Theoretical Macro-Stochastic Models	65
4.3.4 A Prognostic Macro-Stochastic Model	71
5 Application: Valuing Flexibility In Ship Maintenance	85
5.1 Problem Description	85
5.2 Background and Methodology	86
5.3 Data Collection Results - Royal Dutch Navy (RDN) Fleet Maintenance.....	96
5.4 Integrated Risk Management Modeling and Results	109
6 Summary	124
Appendices	125
Appendix A: Ship Maintenance Resources.....	125
A.1 Informants.....	125
Appendix B: References.....	126
Appendix C: Affordable Systems: Balancing the Capability, Schedule, Flexibility and Technical Debt Tradespace	137
Appendix D: Biorelational Modeling and Adaptability to Unforeseen changes	148

FIGURES AND TABLES

Figure 1: Relationship between phases and life cycle costs [Caro, 1990]	14
Figure 2: Taxonomy Structure - Techniques for Valuing Flexibility.....	24
Figure 3: User Interface	28
Figure 4: Input Panel	29
Figure 5: Solution Panel.....	30
Figure 6: CLIPS rules	31
Figure 7: CLIPS Facts Definition	31
Figure 8: CLIPS Solution Rules and Definition.....	31
Figure 9: File Structure Overview	32
Figure 10: File Structure Details	33
Figure 11: Orbital Transfer Decision	35
Figure 12: Optimal Threshold Values ($\pi_1 * (t)$) to Execute Orbital Shift ($\pi_2, \mu, \sigma, r, T, \delta$) = (10,0,2,1,10,.9)	37
Figure 13: Expected Values with and without Orbital Shift Flexibility (F) Using Optimal Strategies.....	38
Figure 14: Simulated Expected Values for Satellite with Flexibility for Optimal and Suboptimal Policies	39
Figure 15: Value of Case 2 Flexibility by Volatility and Initial Value π_{10}	40
Figure 16: Value of Initial and Upgradable Sensor Flexibility	41
Figure 17: Notational PDFs of Missile Defense Scenario KPPs.....	46
Figure 18: DoD Application Domain and Monte Carlo TOC-PL Results	58
Figure 19: Error in LCC Estimate as a Function of Time (Empirical Data).....	66
Figure 20: Error in LCC Estimate as a Function of Time (Theoretical Macro-Stochastic Model)	66
Figure 21: Error in AUC Estimate as a Function of Time (Empirical Data)	67
Figure 22: Error in AUC Estimate as a Function of Time (Theoretical Macro-Stochastic Model)	68
Figure 23: Error in LCC Estimate as a Function of Time (Prognostic Macro-Stochastic Model)	75

Figure 24: Error in AUC Estimate as a Function of Time (Prognostic Macro-Stochastic Model) 75

Figure 25: Error in LCC Estimate as a Function of Time (Validated Prognostic Macro-Stochastic Model) 77

Figure 26: Error in AUC Estimate as a Function of Time (Validated Prognostic Macro-Stochastic Model) 78

Figure 27: Model Performance as Measured by Mean Magnitude Error per SAR..... 79

Figure 28: Model Performance as Measured by Number of Programs with Lower Overall Error 79

Figure 29: Measuring Output 87

Figure 30: Comparison of Traditional Accounting versus Process-Based Costing 87

Figure 31: Diagram of Royal Dutch Navy Fleet maintenance process98

Figure 32: Royal Dutch Navy Ship Maintenance: Stocks and Flows of the System Dynamics Model 100

Figure 33: U.S. Air Force Cost Analysis Agency Handbook's Probability Risk Distribution Spreads..... 111

Figure 34: U.S. AFCAA Handbook's Probability Risk Distribution Spreads112

Figure 35: Risk Simulation Probability Distribution Parameters113

Figure 36: Risk Simulated Volatility.....114

Figure 37: Strategic Real Options Implementation Pathways and Options 117

Figure 38: Sample Real Options Input Assumptions118

Figure 39: Sample Real Options Values118

Figure 40: Risk-free Rate119

Figure 41: Risk Simulation Confidence and Percentiles..... 120

Figure 42: Risk Simulation Statistics and Percentiles.....121

Figure 43: IRM Analysis Results..... 122

Table 1: Taxonomy Formulation-Solution Pair Examples 24

Table 2: KPPs for Missile Defense Scenario 45

Table 3: Marginal Probability Costs for Directed Energy Architectures.....48

Table 4: Summary of existing O&S studies on defense systems 52

Table 5: Hardware Life Cycle Cost Ratios..... 57

Table 6: Software Life Cycle Cost Ratios 57

Table 7: MDAP Data Used For Model Development..... 62

Table 8: Listing of Independent Variables Evaluated in Error-Correction Models 64

Table 9: Summary of LCC Macro-Stochastic Cost Model Program Categories (Pcats)... 70

Table 10: Summary of AUC Macro-Stochastic Cost Model Program Categories (PCats)... 71

Table 11: LCC Macro-Stochastic Model Variables and Fixed Effects Parameter Estimates 72

Table 12: LCC Macro-Stochastic Model Random Effects Parameter Estimates by Program Category (PCats) 73

Table 13: AUC Macro-Stochastic Model Variables and Fixed Effects Parameter Estimates..... 73

Table 14: AUC Macro-Stochastic Model Random Effects Parameter Estimates by Program Category (PCat)..... 74

Table 15: KVA Metrics 88

Table 16: Royal Dutch Navy Ship Types and Numbers 102

Table 17: Knowledge Value Added Model Results..... 104

Table 18: Variance Analysis of KVA Model Results..... 105

Table 19: Preparation for Maintenance Processes - As-is and Radical ROI Differences [Komoroski, 2005]..... 107

Table 20: Maintenance and Implementation Processes - As-is and To-be ROI Comparison 107

Table 21: Return on Investment: Baseline and Technology Adoption Services 108

1 OVERVIEW

The DOD routinely demonstrates its capability to develop complex systems; however, these accomplishments are often tarnished by substantial cost and schedule over-runs. While defense policies are continually being revised to address these problems, many believe that a more fundamental source of these overruns is the lack of flexibility in the systems being developed. But providing justification to invest in flexibility is a tough sell, as stakeholders struggle to quantitatively demonstrate the potential return on investment, including the return from military capabilities. The RT-18a project took as its mission to identify, critique, and improve on methods for valuing flexibility through rigorous quantitative techniques. A summary of the first phase of research conducted can be found in the Mid-Term Report.

This document highlights the work done in the current phase of the project, and is organized as follows:

1. Section 2 provides a brief introduction to the nature of the problem facing DoD as well as other service, defense, and manufacturing systems in designing systems with appropriate and value-adding flexibility. Since identification and definitional work related to flexibility was a focus of the project's mid-term report, limited literature review is provided here, leaving specific details to the following sections.
2. Section 3 constructs a taxonomy of methods for modeling and valuing flexible systems based on the salient system characteristics that are common to all flexible systems. The taxonomy seeks to identify the most appropriate tool for valuing flexibility based on the number of decision epochs, the number of decision alternatives, and the complexity of uncertainty in the flexible system. Beyond this principle features, we provide a number of secondary characteristics which inform a decision-makers technique selection. For practical implementation, this taxonomy has been coded as a web-based flexibility valuation method selection tool. The overview and documentation of this tool are included in this section of the report.
3. The research efforts highlighted in Section 4 develop a tool for justifying investments in flexibility and valuing the inherent ability of different systems or designs to respond to uncertainty. The tool presented here is essentially a modification of the current life cycle cost (LCC) metric to incorporate uncertainty. This report presents a prognostic cost model that is shown to provide significantly more accurate estimates of life cycle costs for DoD programs. This model adopts a stochastic approach, seeking to identify and incorporate top-level (i.e., "macro") drivers of estimating error to produce a cost estimate that is likely to be more accurate in the real world of shifting program baselines. In this report it is demonstrated that the resulting improved cost estimate accuracy could reduce life cycle costs and/or allow defense acquisition

officials the ability to make better decisions on the basis of more accurate assessments of value and affordability.

4. The research in Section 5 extended prior research that was sponsored by the Office of the Secretary of Defense (OSD) through the Systems Engineering Research Consortium that focused on the potential cost benefits of the value of flexibility that select technology options would provide in core ship maintenance processes. This study compared the Dutch naval and Dutch ship builder experiences with the projections for the use of collaborative product life cycle management (CPLM) and 3D Laser Scanning Technology (3D LST) in the prior US Navy study.

The research team collected data on Dutch ship maintenance operations and used it to build three types of computer simulation models of ship maintenance and technology adoption. The approach included use of the knowledge value added (KVA) models of return on technology investments in those operations, system dynamics models (based on the KVA preliminary ROI results) of ship maintenance operations, and integrated risk management (IRM) models of implementation plans for the technology adoption. The results were then analyzed and compared with the previously developed modeling results of US Navy ship maintenance and technology adoption.

5. Section 6 concludes by providing a summary of the research performed in the current phase of the project.

2 INTRODUCTION

While the U.S. Department of Defense (DOD) routinely fields world-class weapons systems, there is tremendous opportunity for improving the acquisition of these systems, at least with respect to cost and schedule performance. In 2009, the Government Accountability Office found that of the DOD's major ongoing acquisition programs that provided relevant cost data, 69 percent reported an increase in total acquisition costs, with over 40 percent of those pro-grams reporting an increase in acquisition unit costs of at least 25 percent. Moreover, on average, total research and development costs were 42 percent higher than originally estimated and systems were 22 months behind schedule. Moreover, the older the program, the more pronounced the cost overruns and schedule delays. Major defense programs that have been in development more than 15 years have seen an average 138 percent increase in acquisition costs, and over three years of schedule delays [GAO, 2009].

These systemic failings are widely known to those familiar with defense acquisitions, and there is nothing particularly surprising in the latest numbers. Nor is there anything surprising in how the DOD is likely to respond to the problem. If the past is any indication of the future, then we will soon see another acquisition reform effort spawned and promulgated with the expressed intent of reducing monetary waste and/or improving overall mission responsive-ness. This observation is not meant to disparage the various well-intentioned reform efforts and the dedicated professionals that create and implement them; the point is, rather, that the desired improvements are seldom, if ever, realized [Drezner, Jarvaise, et al., 1993; Younossi, Arena, et al., 2007; Christensen, Searle, Vickery, 1999].

One possible explanation for the lack of effectiveness of these acquisition policies is that they are essentially aimed at the cause rather than the symptoms. For most engineering problems, this would be exactly the right approach. One time it is not is when the root cause is ineluctable. When this is the case, resources may actually be squandered by focusing on the cause, and instead should be aimed at how best to mitigate the effects. As an analogy, it is more sensible to design a building to be earthquake-resistant rather than to try to develop a technique for preventing earthquakes entirely.

With respect to acquisition programs, the metaphorical role of the inevitable earthquake is filled by *uncertainty*. Every major program must contend with myriad sources of uncertainty, to include the emergence of new threats, technological setbacks/breakthroughs, requirement creep, test failures, budget fluctuations, market volatility, workforce turnover, and, of course, new acquisition policies. Regarding the last item, the steady barrages of acquisition reform efforts that attempt to overcome uncertainty are arguably futile since *uncertainty cannot be overcome*. Worse, it may be that some of these strategies (e.g., requirement-driven acquisition) contribute to the development of point-solution designs that are ironically less capable of responding to these various sources of uncertainty when they do arise, thereby inevitably wreaking havoc with program budgets and schedules. And at the risk of extending the earth-quake

metaphor to a breaking point, the mounting complexity and rate of change is increasing the frequency and magnitude of the earthquakes. So instead of tilting (or, at least, instead of *only* tilting) at the windmill of uncertainty, a better approach may be to accept uncertainty as a fact of life, and explore how we can design systems to better respond to it.

While the definitional landscape related to a system's ability to respond to uncertainty is large and ambiguous, the term most often associated with this concept is *flexibility* [Ryan, Jacques, and Colombi, 2011]. If systems can be designed in such a way that they are able to more readily respond to various sources of change, then it stands to reason that when uncertainties become realities, the impact to the program will be lessened. Designing flexibility into a system, which paradoxically focuses on the predictable effect, rather than the unavoidable cause, may be vital to achieving the persistently elusive goal of improved cost and schedule performance.

Flexibility is frequently, and almost universally, hailed as a desirable system characteristic, as it is widely perceived as the most effective antidote to the systems engineering scourge of uncertainty [Thomke, 1997; Krishnan and Bhattacharya, 2002; Gershenson, Prasad, and Zhang, 2003; Hastings and McManus, 2004; Nilchiani and Hastings, 2006; Gebauer and Lee, 2007; Shah et al., 2008; Baykasoglu, 2009; Saleh, Mark, and Jordan, 2009; Brown and Eremenko, 2009]. Yet despite its wide usage and high regard, flexibility remains a remarkably ambiguous concept in the systems engineering community. In many cases, the problem extends to—and is exacerbated by—the casual usage of many of the other nontraditional system design parameters. For instance, the terms “flexibility” and “adaptability” are often used interchangeably, or conflated with descriptors like robustness, agility, changeability, scalability, modifiability, and versatility. Consider the latest version of the *INCOSE Systems Engineering Handbook* where *flexibility* and *adaptability* are referred to on numerous occasions without ever being defined, while *robustness* is defined as simply the “ability to adapt” [INCOSE,2011].

The fact that systems engineers do not have a clear, consistent definition for flexibility is lamented by numerous authors [Bordoloi, Cooper, and Matsuo, 1999; Saleh, Hastings, and Newman, 2003; Nachtwey, Riedel, and Mueller, 2009; Fitzgerald et al., 2009]. Moreover, of the “-ilities,” there is reason to believe that the term “flexibility” is the most carelessly employed. In one study [Saleh, Mark, and Jordan, 2009], the authors present evidence showing that the term “flexibility” (and its variants) is used in a colloquial sense far more often than other design terms, concluding that the concept of flexibility lacks “scholarly maturity.” However, these authors also support the notion that the “concept of flexibility is today where the concept of *quality* was some 20 years ago,” (p. 309) suggesting that its definition is destined to mature. The principal aim here is to advance that maturity process. As one of the source authors in this study has keenly observed, “A truly useful set of definitions should be rigorous, empirically grounded, and free from contextual biases” [Ross, Rhodes, and Hastings, 2008]. We seek to achieve this goal not through the typical method of proffering rational arguments that advocate a particular set of definitions. Rather, we take a more *democratic* approach, attempting to consolidate the views of various authoritative sources. Following a thorough review of scholarly definitions of flexibility and flexibility-related terminology, the authors have created a novel ontological framework to deconstruct the extant

definitions into five fundamental components. This framework enables us to identify and discern the most salient characteristics of the subject terminology, and thereby collate the ontological components into a single consolidated set of “majority-rule” definitions.

Resolving academic disagreement via consensus is generally not the preferred approach. After all, simply because a view is advocated by a majority of experts does not necessarily make that view correct; nevertheless, it does tend to shift the burden of cogency on to the minority. And while dissension has long played a crucial role in advancing the state of the art in both science and academics, that role is clearly diminished here. The lack of consensus for flexibility definitions does not relate to “state of the art,” but simply to terminology, and terminology is an abstract creation whose value is intrinsically anchored to consensus. Another benefit of this approach is that the exercise of analyzing how the definitions of flexibility and flexibility-related terms compare and contrast can yield useful insights into the nature and extent of the dissension, as well as where existing terminology remains the most contentious or inadequate.

The Department of Defense (DoD) has long stressed the importance of monitoring and reducing the total life cycle cost (also sometimes referred to as “total ownership cost”) of its systems. Consider the following excerpts from the venerable DOD Directive 5000.01: Defense Acquisition System [DoD, 2007] —

- Programs shall be managed through the application of a systems engineering approach that optimizes total system performance and minimizes total ownership costs. Planning for ... the estimation of total ownership costs shall begin as early as possible.
- To the greatest extent possible, the MDAs (Milestone Decision Authorities) shall identify the total costs of ownership, and at a minimum, the major drivers of total ownership costs.

Similar guidance can be found in a number of other long-standing authoritative sources (e.g., DoDI 5000.02, and DoDM 5000.04 [DoD, 2008; DoD, 1992]), as well as the recently enacted Weapon Systems Acquisition Reform Act (WSARA) of 2009 [DoD, 2009], the importance of which will be discussed later.

The DoD’s definition of Life Cycle Cost (LCC) is the total cost to the government spanning all phases of the program’s life: development, procurement, operation, sustainment, and disposal [DoD, 1992]. Note that this definition includes some costs accrued before a system formally enters the acquisition phase (e.g., requirements definition and concept development) as well as certain costs accrued as the system transitions out of sustainment (e.g., demilitarization and disposal). These initial and final costs—though sometimes sizeable from an absolute perspective—are almost always negligible when compared to the costs incurred during the program’s acquisition phase and its Operations and Support (O&S) phase. Consequently, one can state, to a high degree of accuracy, that a system’s LCC is simply the sum of its total acquisition costs and its total O&S costs.

Of these two cost components, the DoD has historically placed significantly greater emphasis on the acquisition side of the equation. Over the years, a plethora of control and oversight accountability mechanisms—from milestones and congressional reporting to baselines and breaches—have been implemented with the expressed purpose of improving the execution and/or management of the acquisition phase of defense

programs. Meanwhile, sustainability considerations have been perennially neglected or subordinated to acquisition requirements or program survival [DoD, 2009; Choi, Alper, Gessner, et al., 2009].

At first, this disproportionate emphasis on the acquisition phase might seem odd given the well-known fact that the majority of DoD system costs tend to be incurred during the O&S phase (recent estimates put the share of O&S costs at about 60-75 percent of the overall life cycle costs [DoD, 2009]). However, it does make sense both practically and strategically. From a practical perspective, it is much easier to implement the aforementioned control and oversight accountability mechanisms within the acquisition phase, with its relatively simple chain of command, tighter span of control, and shorter duration. And strategically speaking, focusing on the acquisition makes ample monetary sense: Though fewer dollars are expended during acquisition, the actions and decisions being made during this phase have a much greater impact on the life cycle cost than those being made during the O&S phase. This entire dynamic (which is really the consummate justification for systems engineering) is well depicted in the classic cost curves of Figure 1 below [Caro, 1990].

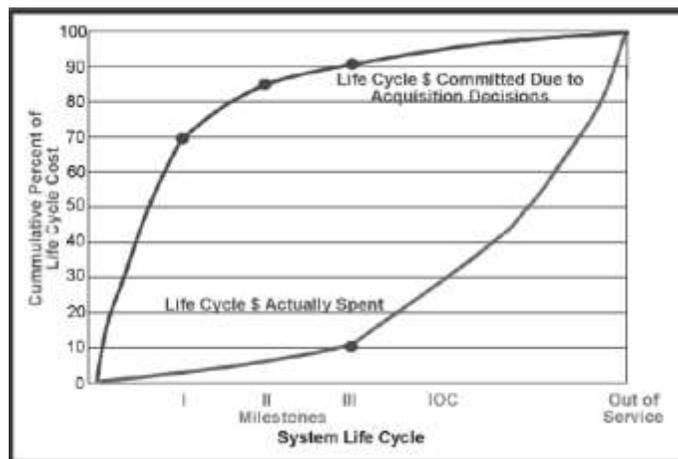


Figure 1: Relationship between phases and life cycle costs [Caro, 1990]

By virtue of its traditional focus on the acquisition component of a system's life cycle, the DoD has managed to gain a variety of valuable insights into the nature of the acquisition costs of defense, including how accurate acquisition cost estimates are and how they tend to behave over time. These insights have provided the framework for many revisions to the acquisition process and provided the opportunity for numerous improvements to the acquisition cost component of a system's LCC. Unfortunately, the same cannot be said for O&S cost projections. Despite an increased focus on O&S costs in recent years, the fact remains that the DoD simply still does not know how O&S cost estimates compare to reality. Consequently, DoD emphasis on a program's life cycle cost is effectively a hollow requirement. Without knowledge of the validity of a program's O&S cost estimates, we cannot have confidence in its LCC estimates. And without confidence in LCC estimates, any efforts to reduce LCCs are effectively nullified, and attempts to meaningfully discern the value of competing systems based on their respective LCCs are rendered futile.

If one accepts the premise that accurate LCC estimates are of vital importance to DoD decision makers, then it is imperative that the behavior of O&S cost estimates be fully characterized. And the opportunity to do so has never been better. The combination of long-cultivated O&S reporting requirements and the fact that enough time has elapsed for the resultant data to sufficiently accumulate, allow analysts—for the first time in history—to conduct a relatively comprehensive assessment of DoD O&S cost behavior.

Many senior defense acquisition officials routinely make key decisions involving weapon systems that are projected to cost billions of—or perhaps even a trillion [Hebert, 2011] dollars over their life cycle. These high-dollar decisions may involve how many units to procure, how to phase program funding, or even whether to fund a program at all. Typically, the decision will not only have major implications on the life of a given program, but it can also impact the Pentagon’s overall budget and strategic direction. In light of the looming, significant reductions to the defense budget [GPO, 2011], these program decisions are bound to become both more difficult and more important, as questions of value and affordability increasingly take center stage.

For the senior decision-maker, a principal tool for assessing the value and/or affordability of a given defense program is via long-term program cost estimates such as Life Cycle Cost (LCC) and per unit Operating and Support (O&S) cost. It is therefore essential that these estimates be reliable and accurate. But what if they are not? What if the forecasted ownership costs of a given program are far different from the actual costs? If there is a significant disconnect between estimated and actual costs, the concern naturally arises as to the utility of the estimates, and how sound are any decisions based upon them. These are not just hypothetical questions. The authors recently completed a study that shows DoD estimates of long-term program cost are often highly inaccurate and—perhaps more surprisingly—improve very little, if at all, as programs mature [Ryan et al., 2012].

This finding logically leads one to consider a more formidable challenge: How can the accuracy of DoD life cycle cost estimates be improved? In this report, we tackle the problem through a fundamentally different approach to cost estimating. We propose a technique that, in essence, models the error in the program estimate as a random variable whose value is determined by a salient group of top-level program summary indicators. This prediction of the estimate error is then used to adjust the official program estimate to a value that is, on average, significantly closer to the eventual, actual cost of the program. We refer to this technique as macro-stochastic cost estimation. The authors have borrowed the term “macro-stochastic” from the physical sciences where it is used to describe large-scale phenomenon that can only be analyzed effectively in a statistical manner, such as dynamic structural loads or earthquakes [Wijker, 2009].

3 A TAXONOMY OF METHODS FOR VALUING FLEXIBILITY

Flexibility is a difficult concept to measure and value, and yet is highly desirable in military, manufacturing, and service systems. Coupling inflexibility with environmental uncertainty or adversarial decisions can have dire consequences for public and private entities. Knowing how much flexibility to invest in for any particular system requires detailed modeling of the system environment to compute the value of flexibility. Flexibility only has value if 1) there is uncertainty about future states of the system, and 2) the flexible alternatives are responsive to the uncertain states. Uncertainty about future states could stem from unpredictable failure of system components, actions of other (possibly adversarial) decision-makers, market uncertainties (demand, price, competition), or natural uncertainties (geologic or atmospheric phenomena). The ability of flexible systems to respond in ways that capitalize on realizations of uncertainty is what creates value for the system operator. Although a number of techniques have been suggested and applied to compute the value of flexibility in various settings, these techniques appear in disparate literatures and have varying assumptions and abilities. There has been no attempt to provide guidance to a decision-maker as to what techniques apply under which circumstances, and their relative merits and trade-offs. This section attempts to fill this gap by studying game and decision theoretic, mathematical programming, dynamic programming, and differential equation formulations of the decision-maker's problem and analytical (closed-form), numerical (exact and heuristic), and simulation-based solution methods. Valuing flexibility is an optimization exercise since the value of flexibility for any system is computed from the expected value under optimal control of the flexible system. We propose a taxonomy which a decision-maker could use to identify which formulations and solution methods are most appropriate under a given set of circumstances. This taxonomy is based on salient features of a system or decision environment including the number of decision epochs, the number of alternatives, and the characterization of uncertainty. We also discuss other important criteria for valuing flexibility, such as the decision-maker's risk preferences and objective, decomposability, and the effect of unknown unknowns.

After presenting our taxonomy, we present a series of examples on valuing flexibility in the design and operation of an observation satellite in order to illustrate the formulations and solution methods considered in the taxonomy. The flexibilities we model include the ability to execute orbital transfers to observe regions providing the highest value of information, and the ability to upgrade sensor technology. We present the example in a series of cases, varying assumptions to illustrate the variety of techniques which can be used to model flexibility.

3.1 LITERATURE REVIEW – VALUING FLEXIBILITY

Flexibility is an old and well-studied concept, with entire journals and several excellent literature reviews devoted to the subject. Therefore, our goal for this section is to briefly highlight important results and illuminate the path to more detailed reviews for the interested reader. Defining flexibility is itself a non-trivial task, with many meanings both practically speaking, and in the academic literature. For the purposes of this section, flexibility refers to the ability to make decisions in the future operation of a system which respond to changes in the state of the system. Excluded in our definition of flexibility is the notion of robust decision-making, or making *a priori* decisions which attempt to ensure preferred outcomes regardless of the realization of uncertainty. Sethi and Sethi [1990] review the literature on flexibility from the economics, organizational, and manufacturing literatures and provide a classification of 11 flexibilities at the component, system, or aggregate levels of a manufacturing organization. Sethi and Sethi's classification discusses the purpose, means, and measurement of each type of flexibility. More recently, Buzacott and Mandelbaum [2008] build on the Sethi and Sethi framework to introduce the concepts of prior, state, and action flexibility, and review applications measuring or valuing flexibility. Buzacott and Mandelbaum discuss models used to represent and measure flexibility in systems and applications thereof. Although there is overlap between their models and the formulations and techniques we consider, their focus is on the measurement of flexibility, while ours is on computing its value, and they make no attempt to classify systems in relation to their models, which is the focus of our taxonomy. One of the important results in manufacturing systems is that limited flexibility can provide nearly all of the benefits of complete flexibility [Jordan and Graves, 1995]. Harvey et al., [1997] compare notions of flexibility in manufacturing systems to those in service systems, where the value of flexibility is derived from uncertain and fluid customer demands. They recommend that information technology creates more value from flexibility in service systems relative to manufacturing systems.

In addition to manufacturing and supply chains, flexible financial instruments, or options, are another primary application domain for the value of flexibility. Here, flexibility is often interpreted as an exercise right without obligation. Financial options theory is an extensive field, including the Nobel Prize winning contributions of Black and Scholes [1973], a closed-form solution to the value of a European call option. Practitioners and academic researchers have also used financial options theory to value flexibility in non-financial environments, an approach referred to as 'real options' (see Trigeorgis [1996] for a well-organized and comprehensive exposition). Bengtsson [2001] reviews flexibility and real options, based on the Sethi and Sethi [1990] classification of manufacturing flexibility. Two critical and related assumptions in applying results from financial options theory to real decision environments are the risk-attitude of the decision-maker and the completeness of markets. In the financial realm, risk-preferences can often be ignored via market completeness, however, in general decision environments, outcomes cannot be replicated with trade-able assets, and risk preferences must be taken into account. Risk preferences are often ignored since nonlinear utility functions frequently destroy closed-form solutions from options theory,

and more general approaches such as dynamic programming Dixit and Pindyck and [1994] must be employed. We now turn to constructing a taxonomy for prescribing methods to value flexibility.

3.2 CONSTRUCTING A TAXONOMY

The purpose of constructing our taxonomy is to provide guidance to a decision-maker as to which formulations and solution procedures are appropriate and well-suited to analyze the system under consideration. We define the value of flexibility analogously to the value of information [Raiffa and Schlaifer, 1961]. Any flexible system has a set of decision epochs and alternatives at those epochs, $\{A_t\}_{t \in T}$. For any system S , we define \mathcal{A} by $\cup_{t \in T} A_t$. Given two systems, S, S' , we say that S is more flexible than S' if $\mathcal{A}' \subset \mathcal{A}$, and in which case we write $S \succ S'$. This means that a decision-maker operating S can make any decision that an operator of S' could make, plus additional alternatives that S allows. Clearly this definition does not create a total ordering of systems, which is to say that given any two systems, neither may be more flexible than the other. In the case that one system is more flexible than another, we say that the systems are *comparably flexible*. For two comparably flexible systems, S, S' , we make the following definition,

Definition 1

Considering two flexible comparable systems, $S \succ S'$, the value of flexibility is defined as the expected value of the more flexible system minus the expected value of the less flexible system under optimal control of each system. That is,

$$V_F(S, S') = \mathbb{E}[V_{S'}] - \mathbb{E}[V_S]$$

Although the value of flexibility is only meaningfully computed for two comparably flexible systems, the expected value of two incomparable systems can still be compared. Here the value V_S captures both benefits and costs (of design, construction, operation, and disposal) realized throughout the lifespan of the system S . Hence, the value of flexibility computed can be used to make decisions on whether the investment in the flexibility is prudent by comparing to the cost of acquiring the extra flexibility provided by S (design, construction...). When the value of flexibility is positive, the alternatives provided by the more flexible system increases the net value of the system in expectation. In the case that the system results in multidimensional consequences which cannot be reasonably aggregated, Definition 1 is not well specified, and instead the trade-offs between the systems should be examined directly. Having defined the value of flexibility, we now outline characteristics of a system or decision environment which form the basis for our taxonomy.

3.2.1 DISTINGUISHING SYSTEM CHARACTERISTICS

The first system characteristic we consider is the number of decision epochs. In the simplest case, there is a single point in time when a decision can be made, e.g., European-style financial options. In this case, closed-form solutions for the value of the flexibility are often available. For example, when the value of the optional asset follows the log-normal distribution and other assumptions hold, the Black-Scholes formulas gives the value of the option. More generally, there may be multiple, but finitely many decision epochs. This is the case for systems with recurring decision opportunities and fixed lifespans. If the life of the system is indefinite or approximately infinite, a system with discrete decision epochs can be modeled by countably many epochs. In the extreme, systems can be modeled as having uncountably many decision epochs if decisions can be made continuously. In financial options applications, distinctions between the number of decision epochs can clearly be seen in the distinctions between European, Bermudan, and American style options. In general, systems with more decision epochs require more sophisticated formulations.

The second system characteristic we consider is the number of alternatives the system operator faces per decision epoch. By definition, any decision epoch has at least two alternatives. Systems with minimal flexibility, two alternatives at exactly one decision epoch, are rarely encountered in reality, and are structurally equivalent to European-style financial options. In general, systems need not have the same number of alternatives per decision epoch, however often holds or is assumed for convenience. Systems with just two alternatives per epoch are often well-behaved, with state thresholds delineating optimality regions for the two alternatives. Operational decisions with a range of discrete or continuous alternatives, e.g., inventory management of a continuous commodity, have finitely many, countably many, or uncountably many alternatives. As the number of alternatives per decision epoch increases, there is greater need to express the relationship between the costs, benefits, and constraints of the system and the decision alternatives through mathematical functions. The value of flexibility is always non-decreasing in the number of total alternatives the system operator faces of the lifespan of the system. Generally speaking, as the number of alternatives increases, solving for the optimal operational or control strategies becomes increasingly difficult.

The characterization of uncertainty plays an important role in the modeling, formulation, and techniques which are appropriate for valuing flexibility. Increasing the number of uncertain factors modeled significantly hampers the prospect of analytical solutions, and increases the computational burden of numerical and simulation-based solutions. The level of precision needed to characterize uncertainty parallels the timing of the decision epochs. There is no need to model the stochastic process of uncertain variables at a greater level of detail than can be utilized by the decision-maker. When detailed information on the stochastic process governing a random variable can always be reduced to the so-called calibrating distribution of the random variable at the decision epochs analytically or numerically, e.g., Boyle et al. [1989]. When limited information on uncertain variables is available, e.g., moments or data on the

distribution function, the maximum entropy principle can be used to estimate the distribution of the uncertainty. When modeling uncertainties, the assumptions of stationarity (probabilities are invariant to time shifts) and independence (realizations of one uncertainty provides no information about another) are assumptions that allow for stronger formulations and solution methods.

Other significant considerations

In addition to the three primary characteristics which form the basis for our taxonomy, there are several other factors which may inform the decision-maker's process for valuing flexibility. One such consideration is the decision-maker's objective and constraints. Particularly, whether the decision-maker's goal is to maintain a given capability while minimizing cost, maximize the rewards from a fixed cost, or most generally, maximize the net value of benefits less costs. When the goal of a decision-maker is simply to maintain a given level of capability, the value of flexibility is realized through cost savings, and valuing flexibility can be thought of as an expected total ownership cost (ETOC) or expected life-cycle cost (ELCC) exercise. In the situation that non-marketable capabilities and outputs (e.g. information, lethality...) must be valued, techniques such as Knowledge Value Added (KVA) [Housel and Bell, 2001] may be employed. In modeling the decision-maker's objective, it is important to consider all consequences of exercising flexibility, including direct and indirect financial costs, changes in capabilities and system outputs, time delays, and forgone future alternatives.

The decision-maker's risk attitude can also impact the value of flexibility and which techniques are appropriate for its computation. For example, in valuing financial instruments and options, it is assumed that there exist investments which can be used to replicate the returns of all possible outcomes, allowing value to be computed with a risk-neutral (linear) utility function. However, when valuing flexibility in general systems and decision environments, it is not possible to hedge against inherent risks, meaning that using a linear utility function may exclude important preferences of the decision-maker. Furthermore, many systems will not have frequently repeated outcomes, making the expected value less meaningful, and accounting for risk aversion more important.

One property of a system which can simplify valuing flexibility is decomposability. The operation of decomposable systems can be separated into smaller decision problems for each decision epoch or state in the original system. Decomposition often needs independent realizations of uncertainty and limited impact of the actions of the decision-maker on future states and decisions. When possible, decomposing a multi-period problem into a series of single-period problem can significantly aid a decision-maker in computing the value of flexibility.

As the lifespan of the system being evaluated increases, an important consideration becomes the degradation of the model. All models imperfectly represent reality and are inherently flawed. Useful models capture the most important characteristics of the environment to the decision-maker in a way that closely mimics reality while remaining analytically tractable are the most useful. Typically, models most similarly mimic reality at the time of their inception, and gradually lose fidelity as they age. Although models

can be updated, e.g., by adjusting parameter values, at some point, there is a paradigm shift and the model becomes fundamentally flawed. This understanding has important ramifications for the value of flexibility which is based on the ability to respond to future changes expressed through the model. When the value of a particular flexibility hinges on uncertainties and decisions in the distant future (relative to the environment under consideration), the value of flexibility should attempt to adjust for the loss of model fidelity over time. This concept is not well-addressed in the literature, and it is unclear how this reality should be accounted for, although discounting future rewards from flexible actions can crudely address this concern. Closely related is the concept of unknown unknowns. The decision-maker can model the known knowns (deterministically), the known unknowns (stochastically), but cannot model the unknown unknowns. Unknown unknowns are increasingly relevant as the system time horizon increases. Typically, unknown unknowns will decrease the value of flexibility in the same way as loss of model fidelity since flexibility is most often built to respond to the known unknowns. We now discuss the formulations and solution methods which our taxonomy considers.

3.2.2 MODEL FORMULATIONS AND SOLUTION TECHNIQUES FOR VALUING FLEXIBILITY

Here we introduce the model formulations and solution techniques at the disposal of a decision-maker for valuing flexibility which our taxonomy considers. The formulations we include in our taxonomy are: game and decision theoretic formulations, mathematical programs, dynamic programs, and differential equations formulations. Each of these formulations is supported by an enormous body of basic and applied research. We briefly describe each formulation as a tool for valuing flexibility and provide references to more detailed expositions.

Model Formulations

Games [Fudenberg and Tirole, 1991] and decision theoretic [Clemen and Reilly, 2004] formulations typically involve a small number of states and decision alternatives, or strong assumptions such as identically repeated stages or decisions. These formulations are often represented visually as decision trees or extensive form games to capture the sequence of events in the system. Risk aversion and adversarial decisions are easily incorporated into these formulations. Shachter and Mandelbaum [1996] discuss the decision theoretic approach in the context of flexibility. In systems involving decisions by multiple agents, solution concepts must consider the equilibria that are sustainable under the assumptions of information and rationality of the agents.

Mathematical programs are optimization problems defined by an objective, a set of alternatives, and a set of constraints. Math programs are classified according to the structure of these components. The alternatives and constraints combined form the feasible region, which is the set of alternatives which satisfy the constraints. When the feasible region is convex, math programs are called convex programs [Boyd and Vandenberghe, 2004], and can be further classified as linear, quadratic, or other special

cases of convex math programs. Convex programs can be solved for thousands of decision variables and constraints. Systems with integer decisions are non-convex and require greater computational resources to solve. Since the value of flexibility is tied to uncertainty and recourse decisions, stochastic programs [Birge and Louveaux, 2011; Shapiro et al., 2009] must be used to compute the expected value of flexible systems. When multiple uncertainties and states are needed to capture the system state, the number of variables and constraints needed to capture the optimal policies explodes. Stochastic programs are typically most useful when there exist mathematical relationships between the decision variables, constraints, and objectives, expressed through functions, and algorithms can capitalize on the specialized structure of the problem.

Dynamic programs capture the decision-maker's problem as the state of the system evolves, modeling a sequential series of decision epochs as stage problems. Again, since the value of flexibility is tied to uncertainty, stochastic dynamic programs (SDP) [Puterman, 2005; Bertsekas and Shreve, 1978] will be needed. The stage problems are solved by considering the current costs and rewards of decisions, as well as the impact current decisions have on future states and value that will be able to be derived from those states. Optimal policies for SDP's map the current state to optimal decisions, and are solved through the Principle of Optimality [Bellman, 1957]. For systems with a finite number of stages, backward induction can be used to solve for the optimal policy by starting with the last stage and sequentially moving to previous stages. Importantly as the level of detail in modeling uncertainty increases, the description of the optimal policy, and the size of the policy search space increase drastically due to the curse of dimensionality [Bellman and Dreyfus, 1962]. Hence the computational burden needed to solve for the optimal policy increases correspondingly. Adversarial decisions in systems modeled as dynamic programs can be incorporated into non-cooperative dynamic programming formulations [Filar and Vrieze, 1996; Basar and Olsder, 1999], although tractable formulations may require strong parametric assumptions.

Differential equation formulations are used to model systems which evolve continuously or approximately continuously. Stochastic differential equations [Oksendal, 2003] express stochastic relationships between system variables as derivatives of other variables, such as time. Well-behaved differential equations can be solved analytically, but numerical solutions [Kushner, 2000] are more generally applicable. Differential equation solution procedures are often not well-suited to scour large feasible regions, and thus are typically used in systems where the optimal decision rules are easily identified or have structural properties which can be incorporated into boundary conditions to provide an anchor point for the solution.

Solution Techniques

In order to solve these formulations to compute optimal system values, a decision-maker may turn to a variety of solution methods. We broadly group these into analytical, numerical, or simulation-based methods. Our intention is not to provide an exhaustive list of all the algorithms which can be used to solve the formulations we have

discussed, rather we generally examine the abilities and trade-offs of these classes of solutions, providing examples for each of the formulation types considered.

Analytical solutions allow the decision-maker to compute the expected value of systems via closed-form expressions. These solutions are derived from the assumptions of the model, and are typically found in simple or well-behaved environments. When they can be found, analytical solutions are easy to implement, however the fidelity of the solution is critically linked to the appropriateness of the model assumptions. For example, the Black-Scholes formulas can be easily implemented to compute the value of a European option or equivalent system, however violating critical assumptions such as no-arbitrage opportunities or the log-normal price distribution will invalidate the solution. A primary benefit of analytical solutions is the ability to perform sensitivity analyses by varying parameters or taking derivatives of the solution expression. A primary draw back of the analytical solution approach is the lack of general algorithms to identify solutions. The existence of analytical solutions is often critically tied to the tractability of parametric assumptions made of the system. Examples of analytical solutions include Cournot equilibrium quantities (games), principal-agent solutions (math programs), (S,s) policies for inventory control (dynamic programs), and the Black-Scholes formulas (differential equations).

Numerical methods bridge the gap between analytical solutions and simulation-based methods. Numerical methods are an extremely broad class of methods and algorithms, and hence applicable in the widest class of settings. Numerical methods can provide exact or approximate solutions. Backward induction (subgame perfection) is the standard algorithm for computing solutions in decision theoretic formulations (games) and finite horizon dynamic programs. Algorithms for stochastic mathematical programs include the simplex, interior-point, decomposition-based, barrier, and cutting-plane methods just to name a few [Birge and Louveaux, 2011]. The best algorithm in terms of solution speed and quality depends on the structure of the problem. Iterative procedures such as value and policy iteration and variants thereof find optimal policies and values of systems modeled as infinite horizon dynamic programs. Numerical methods for solving differential equation formulations include binomial tree methods, finite difference and finite element methods [Larsson and Thomee, 2008].

Simulation is a powerful tool for valuing flexibility, and its power increases as the ability of computational resources increases. For a given system operation policy, simulation can easily and quickly generate thousands of possible realizations of future uncertainties and policy implementations in order to estimate the expected value of the system operating under the policy. Where simulation comes up short is in its ability to identify optimal policies. Simulation is powerful enough to do so when the number of alternatives is few, however when there are decisions to be made in multiple dimensions at possibly many time epochs, simulation becomes a tool more for valuing policies, rather than optimizing. The field of simulation-based optimization seeks to provide algorithms which leverage the power of simulation to find optimal strategies and policies in high dimensional environments, e.g., in stochastic programming formulations [Shapiro, 2003; Bayraksan et al., 2011].

Although we present them distinctly, these classes of solution procedures can also be used in coordination with each other. Analytical approaches can augment numerical and simulation-based procedures by reducing the alternatives to search among for optimal policies. For example, many inventory management problems can be shown to have optimal threshold policy structures. Such analytical results can greatly reduce the search space for optimal policies. Similarly, many algorithms combine simulation-based approaches using Monte Carlo type sampling to improve the ability of numerical procedures. Table 1 presents examples or applications of each type of formulation-solution method pair. We now construct our taxonomy with these formulations and solution methods.

Formulation	Solution Technique		
	Analytical	Numerical	Simulation
Decision Theoretic Mathematical Program	Cournot Principal-Agent	Backward Induction Cutting Planes	Monte Carlo Internal Sampling
Dynamic Program	(S,s)	Value Iteration	Monte Carlo Markov Chain
Differential Equations	Black-Scholes	Finite Element	U-L Bound Convergence

Table 1: Taxonomy Formulation-Solution Pair Examples

Methodology Taxonomy

The following taxonomy is based on the previously described system characteristics, model formulations, and solution methods. The purpose of the taxonomy is not to identify the only appropriate formulation and solution method, as multiple approaches can value flexibility in any system. Rather we make recommendations in our taxonomy classifications based on the abilities and challenges of valuing flexibility via the formulations and solution methods considered. The taxonomy classes are defined by the salient characteristics as seen in Figure 2.

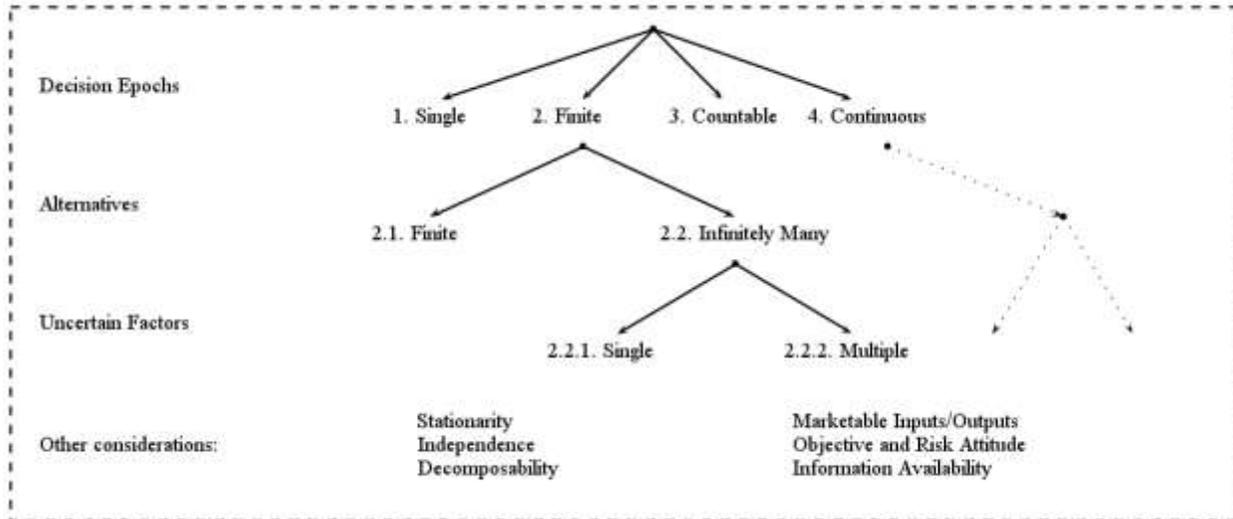


Figure 2: Taxonomy Structure - Techniques for Valuing Flexibility

We proceed through the taxonomy starting with the simplest cases for valuing flexibility, and progressing to more challenging ones. We use the number of decision epochs in the operation of a flexible system as the primary characteristic to present our taxonomy. Within each classification by the number of decision epochs, we begin by discussing evaluation of systems with a finite number of alternatives per decision epoch and a single uncertain factor before proceeding to more difficult cases.

1. Single Decision Epoch

The simplest case in our taxonomy consists of systems with a single decision epoch, with finite alternatives and a single uncertain factor. Although this class is the simplest in our taxonomy, we do not mean to imply that the systems or their optimal operation are simple, or that valuing flexibility is trivial. Rather the distinction is relative to systems where valuing flexibility is a more challenging task. When there is a single or even very few decision epochs, or when the stage decisions are decomposable, decision theoretic formulations are a natural approach. When adversarial decisions are relevant, these formulations can be extended to games theoretic formulations. As the number of alternatives or constraints increases, or if well-defined mathematical relationships exist between the system's decision variables, objectives, and constraints, mathematical programming becomes an increasingly attractive formulation. Differential equation formulations exceed the level of information that can be used in the finite decision epochs, and therefore are only reasonable if system variables are naturally modeled as stochastic processes, e.g., the price of oil.

Systems in this class of our taxonomy often yield closed-form analytical solutions, particularly when well-behaved parametric assumptions are reasonable. For example, we have already discussed the Black-Scholes solution to a European option value. Backward induction is the primary numerical solution method for decision trees and games. Mathematical programming formulations with finite alternatives will be integer formulations, and efficient solution will use specialized stochastic-integer approaches such as the disjunctive decomposition algorithm of [Sen and Higle, 2005]. Since the search space for optimal operation strategies and policies is limited, simulation-based optimization methods such as simple Monte Carlo methods or simulated annealing [Geman and Geman, 1984] can identify optimal policies and compute the value of those policies.

Infinitely Many Alternatives

With a continuous range of alternatives, solving mathematical programs for optimal strategies may become easier, particularly if functions of the decision variables which determine value or resource use are continuously differentiable. This class of problems includes two-stage stochastic linear programs with recourse, for which there are a number of solution algorithms available. Also, well-behaved functional forms are amenable to analytical solutions which utilize the Karush-Kuhn-Tucker conditions. Clearly infinitely many alternatives rules out brute force solution, so numerical solutions will take advantage of gradient and other search methods.

Multiple Uncertain Factors

When multiple uncertain factors are present and relevant in the system environment, analytical solutions will likely exist only when special relationships exist between them. For example, a compound option has the flexibility to exercise one option from a set of European options with the same expiry date, based on different underlying assets and strike prices. In this case, the multiple uncertainties are loosely linked, and an analytical solution can be obtained [Johnson, 1987]. Otherwise numerical methods, e.g., [Barraquand and Martineau, 1995], have been proposed for single alternative cases.

2. Finite Decision Epochs

As the number of decision epochs increases without decomposability, dynamic programming formulations, which are better suited to handle the sequential nature of decisions, become more attractive. When the number of alternatives is finite, and there is a single uncertain factor, optimal policies are found relatively easily via backward induction.

Infinitely Many Alternatives

As the number of alternatives increases, backward induction will require higher-powered solution algorithms to locate optimal policies. Analytical results concerning the structure of optimal policies may be able to aid in the search for optimal policies.

Multiple Uncertain Factors

Dynamic programming remains an attractive modeling approach, with backward induction for solving. However, the curse of dimensionality results in exponentially larger state spaces as the dimensions of the state space increases. This drastically increases the computational time required by deterministic numerical algorithms which seek to define optimal decisions at each state. Monte Carlo simulation-based random algorithms can deal with the curse of dimensionality in dynamic programming formulations [Rust, 1997].

3. Countable Decision Epochs

Systems with discrete but infinitely many decision epochs are well-suited to be modeled as dynamic programs when the stochastic process governing the uncertain factor is stationary (or at least ergodic). Iterative algorithms such as value and policy iteration replace backward induction as the standard solution approaches since there is no terminal state to solve for optimal policies and values.

Infinitely Many Alternatives

Similarly as before, searching through more alternatives at each decision epochs necessitates the use of more powerful algorithms, as well as possible supplement from analytical results and simulation-based methods.

Multiple Uncertain Factors

This class of problems is also amenable to dynamic programming formulations, although the curse of dimensionality is in play. This is among the hardest class of

systems to value flexibility for, particularly when the number of alternatives in each stage is large. In order to deal with integrating over multiple uncertain factors to find expectations, simulation-based techniques such as Monte Carlo Markov Chain (MCMC) techniques, including the Gibbs sampler [Geman and Geman, 1984] can be used.

Continuous Decision Epochs

When decisions can be made in continuous time, stochastic dynamic programming yields way to stochastic optimal control problems [Bertsekas, 2007], which are differential equation formulations. To capture uncertainty, systems with continuous decision epochs can be modeled by stochastic differential equations. Differential equation formulations with finite alternatives and a single uncertain factor may produce closed-form solutions such as those for optimal stopping problems and other financial options [Mckean, 1965; Geske and Johnson, 1984; Guo and Zhang, 2004]. Finite element and finite difference methods are two of the primary numerical methods used to find solutions to differential equation models [Brennan and Schwartz, 1977; Ikon and Entoivanen, 2004; Topper, 2005]. Simulation-based methods can also be used to find solutions, e.g., through the convergence of upper and lower solution bounds [Broadie and Glasserman 1997].

Infinitely Many Alternatives

As the complexity of the problem increases, closed-form solutions may still exist for the most specific and well-behaved cases, simulation is generally are more appropriate tool [Longstaff and Schwartz, 2001; Cortazar et al., 2008; Andersen and Broadie, 2004].

Multiple Uncertain Factors

Even in this complicated class of systems for valuing flexibility, closed-form solutions may exist when well-behaved parametric assumptions can be made, such as the case of the multivariate put option [Curran, 1994]. However, these may require complex machinery and mathematics. Numerical methods will be likely be of more use for this class of systems, particularly when parametric assumptions, discretization of the decision and alternative spaces, or other simplifying assumptions may be used to improve the ability of solution techniques to solve for the optimal strategies. It has been shown that in higher dimensions, the finite element method is preferred to the finite difference method [Kovalov et al., 2007].

3.3 SOFTWARE IMPLEMENTATION: FLEXIBILITY VALUATION METHOD SELECTION TOOL (FVMST)

The Flexibility Valuation Method Selection Tool (FVMST) is a decision support tool intended to aid decision makers in selecting the appropriate methodology to determine the Value of Flexibility. FVMST is a web based application that is separated into two parts. The first part, the client side, provides a simple yet sophisticated graphical user interface. The second part of the application, the server side, provides the analytical backend. The use of a client/server setup allows for the application to be cross platform compatible, easily updatable, and upgradable. Following current web development

standards, the frontend is written in HTML, JavaScript, jQuery, and CSS using the Django application framework. The analytical backend is based on the well-established and widely used Expert System Shell "C Language Integrated Production System" (CLIPS). The following sections describe FVMST in more detail.

The Application

The application consists of a single user interface (see Figure 3) which allows the user to select taxonomy options, start the computational analysis and review the results from the analysis. The layout is split into two panels. The left side consists of the input panel and the right side displays the result obtained from the analysis.

Flexibility Valuation Method Selection Tool		Solution			
Taxonomy		Solution Technique			
Decision Effects	Please select one	Formulation	Analytical	Numerical	Simulation
Alternatives	Please select one		Decision Theoretic		
Uncertain Factors	Please select one	Mathematical Programming			
Other Considerations		Dynamic Programming			
<input type="checkbox"/> (Problem Structure) <input type="checkbox"/> Fixed Cost/ Cost Minimization <input type="checkbox"/> Decomposability <input checked="" type="checkbox"/> Markable Inputs/Outputs <input type="checkbox"/> (Uncertainty) <input checked="" type="checkbox"/> Independence <input checked="" type="checkbox"/> Stationarity <input checked="" type="checkbox"/> Information Availability		Differential Equations			

Figure 3: User Interface

The Input Panel

The left section, see Figure 3, consists of three drop down menus (A) for the different taxonomies. In order to run the application successfully all taxonomy options require to be selected. The second section provides additional considerations to enable a more detailed description of the problem structure and uncertainty (B). Once the required and optional sections are completed the submit button will executes the CLIPS backend (C). If additional information is required, the interface provides pop-up help info boxes (D) for any of the selectable options.

Figure 4: Input Panel

The Solution Panel

Once the selections from the input panel are submitted the CLIPS backend analyses the data and returns the solution to the solution table (E), see Figure 5. The solution is represented in the form of a color coded key in the solution technique and formulation matrix (F). The color code ranges from red to yellow, with red being the highest recommendation. By selecting any of the matrix fields the user is provided with additional information on the selected methodology to support the decision process.

Formulation	Solution Technique		
	Analytical	Numerical	Simulation
Decision Theoretic			
Mathematical Programming		F	
Dynamic Programming			
Differential Equations			

Figure 5: Solution Panel

The CLIPS Backend

The first CLIPS software tool was developed in 1985 at the NASA-Johnson Space Center. Today it is among the most widely used expert system tool due to its fast and efficient execution algorithm implementation. This efficiency allows it to be deployed in a client/server scenario with minimal computational load on the server. This is particularly important for web based applications where the server handles multiple client requests at a time.

The concept of CLIPS is based on a rules and fact system. From the graphical user interface (GUI) the CLIPS backend receives the submitted information in the form of a list of facts. These facts are then checked against a set of rules. Contrary to a traditional loop based programming approach this allows for a much more efficient checking of conditions. In Figure 6 we show two examples rules of the analytical backend. The first rule handles the facts submitted by the GUI and separates them from the facts list. The second rule performs the actual task if the condition is true. In that case CLIPS asserts the outcome of the condition to the corresponding attribute.

```

(defrule RULES::remove-is-condition-when-satisfied
  ?f <- (rule (if ?attribute is ?value $?rest))
  (attribute (name ?attribute)
             (value ?value))
  =>
  (modify ?f (if ?rest)))
(defrule RULES::perform-rule-outcome
  ?f <- (rule (if)
             (then ?attribute is ?value $?rest))
  (test (or (eq (length$ ?rest) 0)
            (neq (nth 1 ?rest) with)))
  =>
  (modify ?f (then ?rest))
  (assert (attribute (name ?attribute) (value ?value))))

```

Figure 6: CLIPS rules

The conditional outcome is defined in a set of fact definitions. Figure 7 shows an example of the decision epoch fact definition. If the condition “selected-epoch is single” is met, CLIPS asserts the value of 1 to the attribute [x,y].

```

(deffacts the-method-rules
  ; Epochs
  ;
  (rule (if selected-epoch is single)
        (then [0,0] is 1 and
              [0,1] is 1 and
              [0,2] is 1))

```

Figure 7: CLIPS Facts Definition

Once all facts are met by firing the associated rules the last step in the rule chain is to assert the attributes to the solution. Figure 8 shows the solution rule.

```

(defrule SOLUTION-METHODS::generate-methods
  (method (name ?name)
          (grid $? ?g $?))
  =>
  (assert (solution (name method) (value ?name) (recommendation ?r))

(deffunction SOLUTION-METHODS::get-solution-methods-list ()
  (bind ?facts (find-all-facts ((?f solution)
                                (eq ?f:name method))))

```

Figure 8: CLIPS Solution Rules and Definition

In order to obtain the solution for the submitted list of facts the web application calls the solution retrieval function which in turn returns the solution list. The returned list is then displayed in the table of the GUI.

The Development Environment

The development environment is setup in such a way that it doubles as system independent development environment as well as the blueprint for the application deployment. This is achieved through the use of a virtual environment. To ensure a dynamic development environment we make use of a software framework.

The Virtual Environment

The concept of a virtual environment allows the decoupling of the development environment from the physical machine. This allows the creation and maintenance of a software environment that is independent from that of the host machine on which the virtual environment is running on. The benefit of such an approach is that dependency problems are eliminated and version requirements can be ensured for each project/application. The virtual environment used for the FVMST application is based upon the Python “virtualenv” tool.

The Web Application Framework

The Django application framework, used for the FVMST development, enables the development of an application that separates the data model with business rules from the user interface. This approach has several advantages, for instance it modularizes code, promotes code reuse, and allows multiple interfaces to be applied. Consequently the FVMST application exhibits an extensible set of capabilities, for example by coding multiple interfaces it becomes equally usable on mobile devices as well as desktop computers.

In order to ensure a web site that feels responsive, speedy, and usable, we apply the “Asynchronous JavaScript and XML” (Ajax) web development technique in the form of the jQuery library. The use of the jQuery library ensures the desired characteristics of a responsive web page by exchanging small amounts of data with the server instead of reloading the entire web site.

The File Structure

The files structure for the FVMST development environment is described in Figure 9 and Figure 10.



Figure 9: File Structure Overview

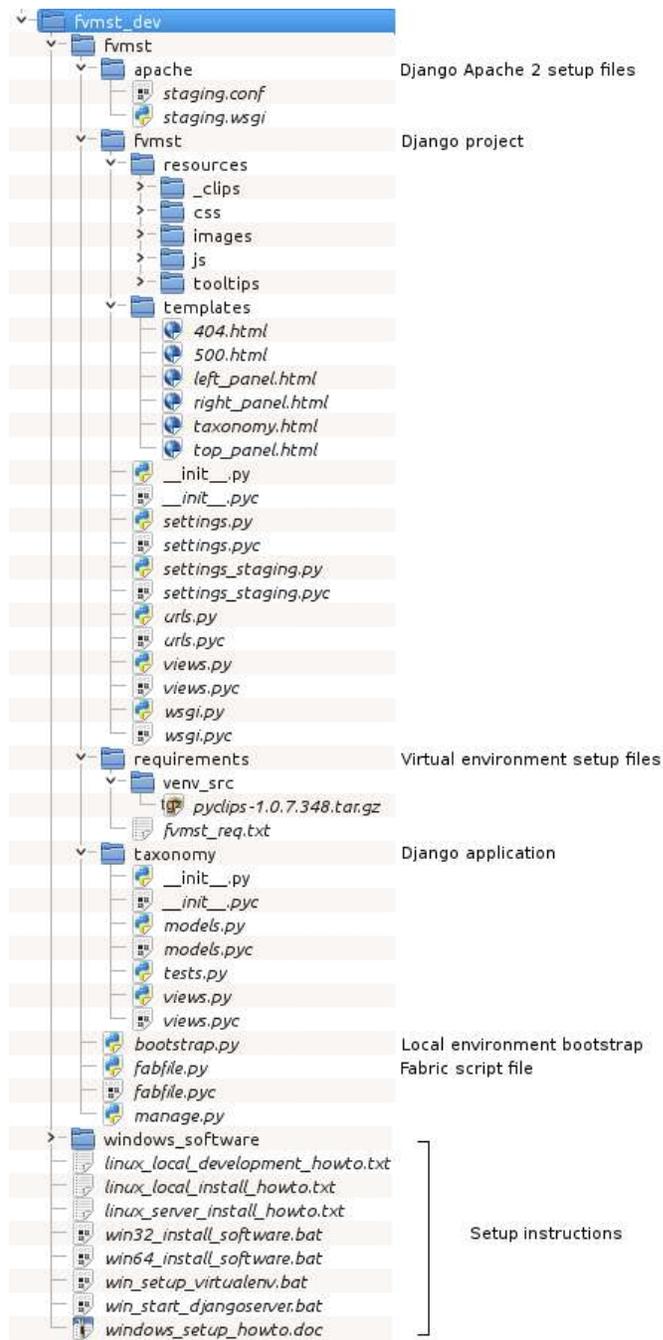


Figure 10: File Structure Details

3.4 VALUING FLEXIBILITY APPLICATION: SATELLITE DESIGN AND OPERATION

In this section, we demonstrate how the formulations and solution techniques considered in our taxonomy can be used to calculate the value of flexibility within a system. The system we consider is an observation satellite. The base scenario we use to illustrate the techniques for valuing flexibility surrounds the flexibility to transfer an observation satellite from a low Earth orbit (LEO) to a geostationary orbit (GEO). LEO satellites have altitudes in 100's of miles above the surface of the Earth, and typically revolve around the Earth several times a day [Arkali et al., 2008]. LEO satellites are therefore able to observe multiple targets, but have limited visibility of any particular target in a given day. GEO satellites are at an altitude of 22,300 miles above the surface of the Earth [Office of Satellite Operations], and rotate at the same speed as the Earth, resulting in continuous visibility over a single region. The transfer of satellites between orbits can be done fuel-optimally using a three-stage maneuver [Naidu, 1991]. Both the design and operation of satellites for public and private use are multi-billion dollar investments [Gavish, 1997]. Satellites pose many interesting and difficult problems for operations researchers and engineers, for a more thorough discussion of the types of problems and the work done in this application domain, see Fliege et al. [2012] and Gavish [1997]. While the design and operation of satellites pose many complex problems for operations researchers and engineers, we focus on a small example in order to illustrate the ability of techniques to measure the value of flexibility.

The motivation to move a satellite from LEO to GEO would be based on the high value of information from observing the GEO visible location continuously or more frequently. The value of information derived from the observation of a particular site could be estimated by the KVA approach. A high value of observation could be derived from geological events [Zhu et al., 2010], weather events, or military interest. Figure 11 depicts the decision to move a satellite from an LEO rotating the Earth to a GEO. Computing the value of flexibility must consider the expected value of optimal operation of the flexible system versus an inflexible system. Once the value of flexibility is computed the value must be compared to the additional cost of the flexible satellite in terms of development, construction, and maintenance of the capability to transfer orbits. The ability to change orbits comes at the cost of additional fuel to execute the orbital transfer maneuver. For simplicity, we assume that the fuel requirements for station-keeping are equivalent in either orbit, and that the life of the satellite is not appreciably different depending on whether the option to shift orbits is exercised, or how much time is spent in each orbit.

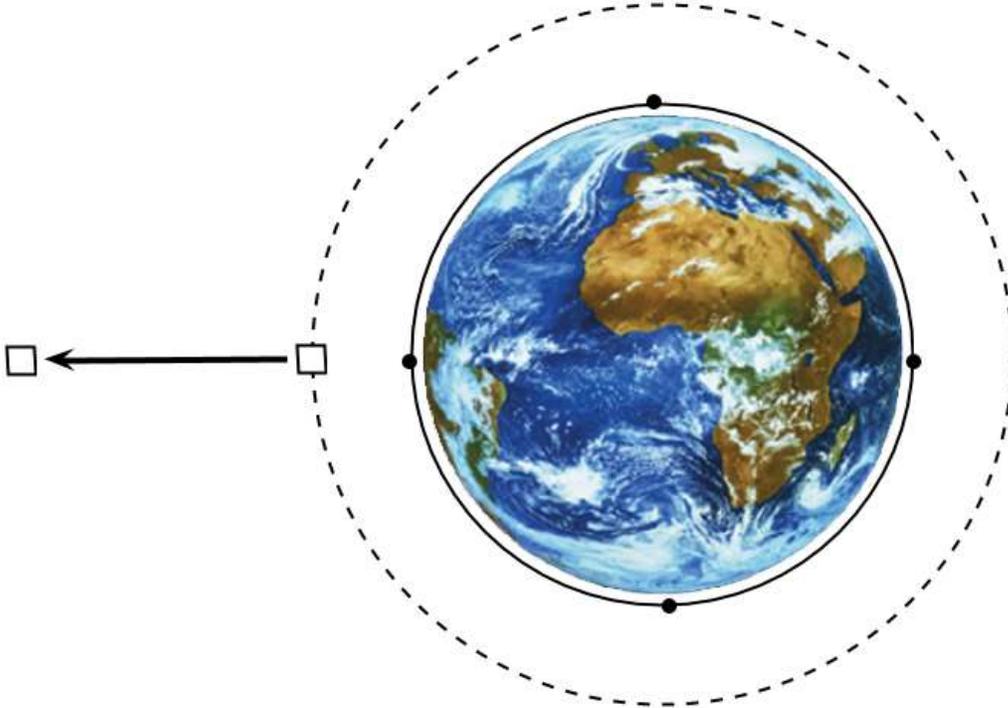


Figure 11: Orbital Transfer Decision

Notation

We adopt the following notation:

- $i \in I = \{1, \dots, n\}$, the set of observation sites,
- $t \in \mathcal{T}$, the set of time epochs,
- $j \in J = \{1, \dots, m\}$, the set of sensors,
- $o \in O$, the set of orbits the satellite could enter, where $o = I$ is a low-earth orbit which can observe all sites, and all other orbits are geostationary orbits over a single site i ,
- $(j, o) = d \in D = J \times O$, the set of decisions that can be made at each time epoch,
- $\{\pi_{ij}^t\}_{i \in I, j \in J} \in \mathbb{R}_+^{n \cdot m}$, the reward at t for each site-sensor pair
- $(o, \pi) = x \in X = (O, \mathbb{R}_+^{n \cdot m})$, the state of the satellite system, we let $o(x)$ and $\pi(x)$ denote the specific components of x ,
- L , the useful life of the satellite,
- $c(d, x)$, the cost of making decision d in state x ,
- $\delta \in (0, 1)$, the discount factor for one time epoch.

Case 1

For simplicity we let the decision epochs be equal to the time for one revolution of the satellite in LEO, r , implying $\mathcal{T} = \{1, \dots, T = L/r\}$. Let there be two observation sites ($I = \{1, 2\}$), and a single sensor. We assume that the satellite will be launched into LEO, and the only decision that can be made is to move the satellite into GEO above

observation site 1. We assume that the satellite can only move to a geostationary orbit once, and the observation site cannot be changed once the geostationary orbit is reached. Implying that $c(d, x) = M$ if $d_o \notin o(x)$.

Let d_1 denote the decision to move to GEO over site 1, and d_L denote the decision to remain in LEO. Whether in LEO or GEO, the satellite can take two images per decision epoch. The value of observing site 2 is assumed to be fixed at π_2 , whereas the value of observing site 1 follows a Wiener process, $\Delta\pi_1 = \mu_1\Delta t + \sigma_1\epsilon\sqrt{\Delta t}$, with drift μ_1 and variance σ_1^2 .

Expected Value without Flexibility:

In order to compute the value of flexibility, we compare the operation of the flexible satellite with the operation of a satellite with equivalent capabilities except that it must stay in LEO. The expected value of this inflexible satellite system is

$$V_{IF}(\pi^0) = \sum_{t=0}^T \delta^t \pi_2 + \mathbb{E}_{\pi_1^t | \pi_1^0} \left[\sum_{t=0}^T \delta^t \pi_1^t \right]$$

Replacing $\mathbb{E}[\pi_1^t | \pi_1^0]$ by $\pi_1^0 + \mu_1 r t$ yields the following expression for $V(\pi_1^0)$,

$$V_{IF}(\pi_1^0) = \pi_2 \left(\frac{1 - \delta^{T+1}}{1 - \delta} \right) + \pi_1^0 \left(\frac{1 - \delta^{T+1}}{1 - \delta} \right) + \left(\frac{\delta \mu r}{1 - \delta} \right) \left(\frac{1 - \delta^T}{1 - \delta} \right) - T \mu r \left(\frac{\delta^{T+1}}{1 - \delta} \right).$$

Note that two of these inflexible satellites equally spaced in the same orbit exactly replicate a single geostationary satellite over each site in this model.

Expected Value with Flexibility:

The optimal policy to the orbital transfer decision is a threshold policy, this fact can be easily verified by checking the conditions of Theorem 4.7.4. in Puterman [2005]. The threshold $\pi_1^*(t)$ is the level of value of observing site 1 at time t above which the decision should be made to move to GEO above site 1, and below which the decision should be made to stay at LEO. Given the current value of site 1 (π_1^t), the expected value of moving to GEO can be computed by

$$V_F^t(\pi_1^t, d_0) = \pi_1^t \left(\frac{1 - \delta^{T-t+1}}{1 - \delta} \right) + \frac{\delta r \mu}{1 - \delta} \left(\frac{1 - \delta^{T-t}}{1 - \delta} \right) - (T - t) \mu r \left(\frac{\delta^{T-t+1}}{1 - \delta} \right).$$

Whereas the value to staying in LEO can be computed by

$$V_F^t(\pi_1^t, d_1) = \pi_2 + \pi_1^t + \delta \int V_F^{t+1}(x) p_{\pi_1^t}(x) dx.$$

The threshold value is that which equates these two alternatives at time t ,

$V_F^t(\pi_1^*(t), d_0) = V_F^t(\pi_1^*(t), d_1)$. Computing these thresholds for $\pi_2 = 10, \mu = 0, \sigma = 2, r = 1, T = 10, \delta = .9$ yields the values shown in Figure 12.

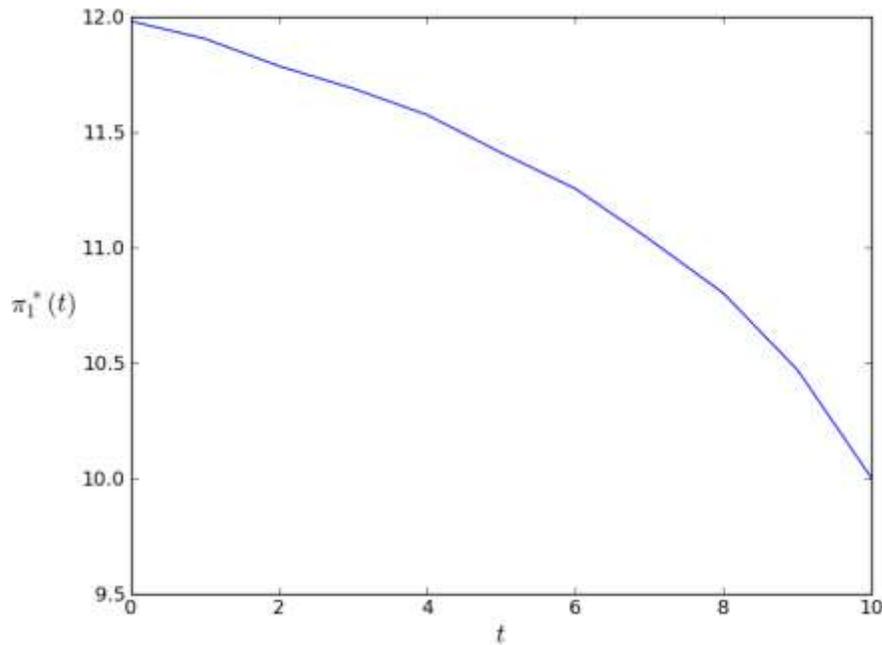


Figure 12: Optimal Threshold Values ($\pi_1^*(t)$) to Execute Orbital Shift ($\pi_2, \mu, \sigma, r, T, \delta$) = (10,0,2,1,10,.9)

The threshold for moving to GEO over site 1 in the final stage is simply π_2 since there are no consequences for future payoffs at this final stage. Moving back towards the beginning of the system, we see the threshold increases since making the decision to move to GEO over site 1 in early stages locks the satellite in over site 1 for the rest of the horizon. Using these threshold and parameter values, the expected value with and without flexibility can be calculated as a function of the initial value of observing site 1. The comparison of these is shown in Figure 13. The y-axis in Figure 13 represents the units of information provided by the respective satellites. The purpose of valuing flexibility in this manner would be to compare the additional value provided by the flexible satellite system to the additional costs associated with its development and acquisition.

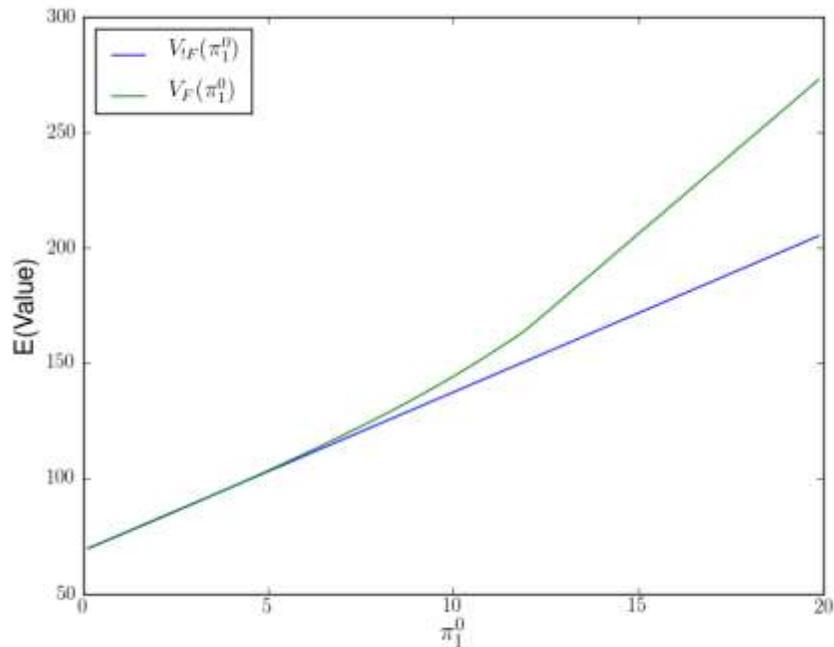


Figure 13: Expected Values with and without Orbital Shift Flexibility (F) Using Optimal Strategies

The expected values for the flexible satellite based on initial observation values are calculated system via backward induction. As mentioned previously, simulation is an effective tool for calculating the expected value of a system, but an inefficient method for calculating the optimal policy for a system. Given a threshold policy and a π_1^0 value, 10,000 sample system values can be generated in about a second by an average 2 GHz processor. Thus, given the optimal policy, the optimal value function can be estimated for a 0.1-fine grid on the initial π_1^0 range [0,20] in minutes (≈ 4.5 minutes in our tests). Figure 14 shows the simulated values under the optimal policy as well as the values of two suboptimal policies: $V'(\cdot)$ is the value of a policy where the thresholds are 25% below the optimal thresholds, and $V''(\cdot)$ is the value of a policy where the thresholds are 25% above the optimal thresholds.

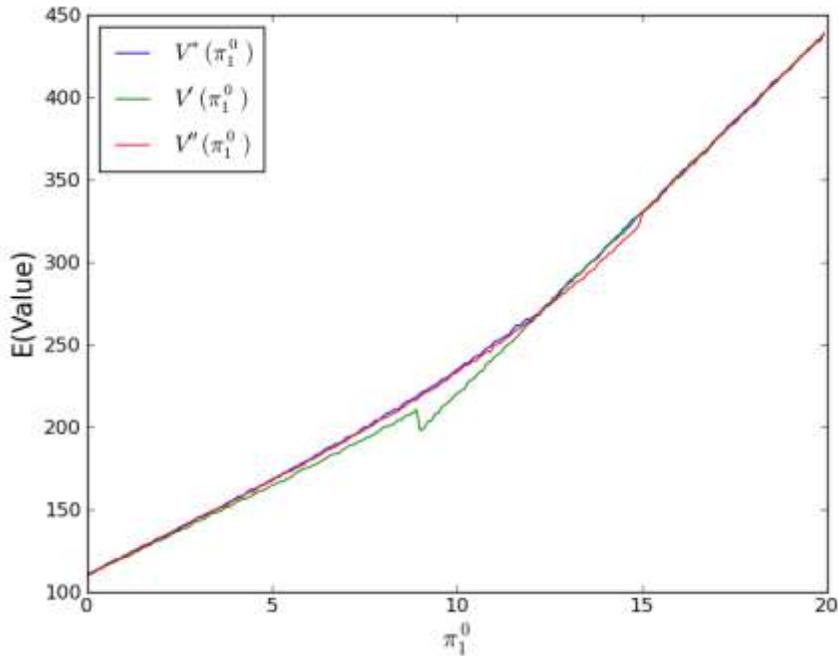


Figure 14: Simulated Expected Values for Satellite with Flexibility for Optimal and Suboptimal Policies

Case 2

Now we tweak the circumstances of the decision-environment in order to apply the Black-Scholes formula. Suppose that we send the satellite into GEO over observation site 2, but build in the capabilities to move to GEO over site 1 for one period at the end of the satellite life cycle (at time T). We again fix the value of observing site 2 (π_2), and let the value of site 1 follow a geometric Brownian motion (GBM) stochastic process with drift μ and volatility σ . Valuing this flexible satellite compared to one that is fixed over site 2 for T periods is equivalent to valuing a European call option, where we can attain payoff π_1^T if we pay the strike price, giving up π_2 . Letting the current value of observing site 1 is π_1^0 , then the value of the flexibility to move to site 1 at time T can be directly computed using the Black-Scholes equations:

$$V(\pi_1^0) = \Phi(d_1) \cdot \pi_1^0 - \Phi(d_2) \cdot \pi_2 \cdot e^{\frac{1-\delta}{\delta}T}$$

$$d_1 = \frac{\ln\left(\frac{\pi_1^0}{\pi_2}\right) + \left(\frac{1-\delta}{\delta} + \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

$$d_2 = \frac{\ln\left(\frac{\pi_1^0}{\pi_2}\right) + \left(\frac{1-\delta}{\delta} - \frac{\sigma^2}{2}\right)T}{\sigma\sqrt{T}}$$

where $\Phi(\cdot)$ is the cumulative standard normal distribution function. Figure 15 shows that the value of the flexibility increases in both the estimated volatility and initial value of information at site 1 (σ, π_1^0). When analytical solution such as the Black-Scholes formulas can be found (enabled by the simplicity or the strength of assumptions made), questions about the sensitivity of the value of flexibility can also be answered analytically. In the case of the Black-Scholes formulas, the sensitivity of the value of the option (flexibility in our case) is well-studied (see Hull [2011]).

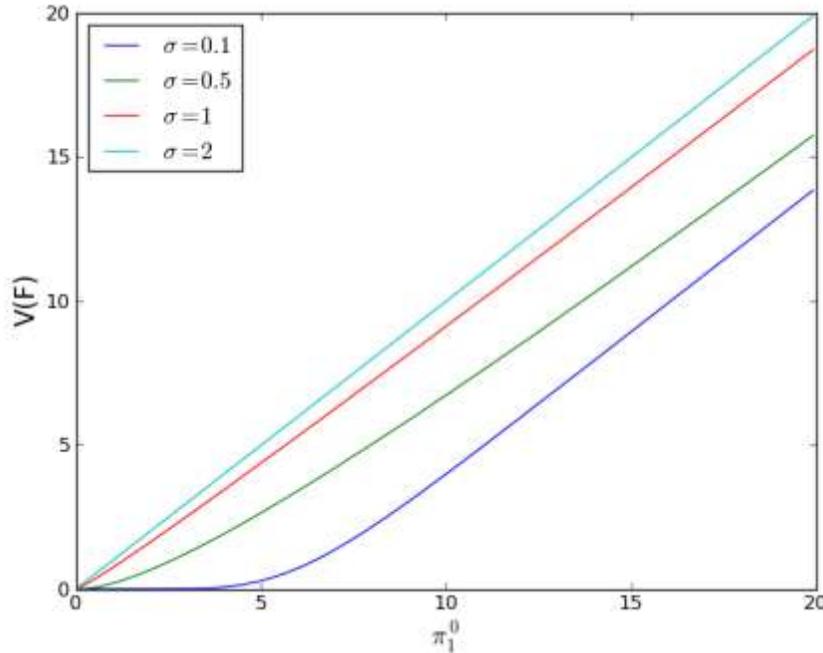


Figure 15: Value of Case 2 Flexibility by Volatility and Initial Value π_1^0

Case 3

One of the most important aspects of flexibility is the ability to respond quickly to realize value. In the third case of our satellite example, we consider two satellites with the flexibility to change sensors. In this example, we consider a GEO satellite and only one possible observation site. The motivation behind flexibility is the ability to use a standard sensor (s_1) which is cheap to use, but provides a lower quality of data, or an advanced sensor (s_2) which provides high-quality data, but is more expensive to install and operate. We assume that when observing a site with value of information π , sensor s_i can deliver a percentage of this value (γ_i) with an operational cost of c_i . This marginal operation cost can be thought of as capturing the extra resources, e.g., battery power, needed to operate the advanced sensor. We let $\gamma_1 = .75$, $\gamma_2 = .95$, $c_1 = 0$, and $c_2 = 2.5$. The value delivered from using sensor i when the value of information is π is then

$$v(\pi, i) = \gamma_i \cdot \pi - c_i$$

We compare the value of flexibility from a satellite having both sensors installed initially against a satellite with only sensor s_1 installed, but with the ability to upgrade the sensor capabilities to include s_2 . We denote the expected values of these satellites under optimal operation by $V_{initial}$ and $V_{upgrade}$. In order to isolate the value of flexibility created from the ability to realize value faster, we let the cost of installing s_2 be equal to c_u whether it is installed initially, or in an upgrade. When the operator of the upgradable satellite makes the decision to upgrade sensor technology, there is an immediate cost of c_u and a one period delay before sensor s_2 can be utilized. Once the upgrade to s_2 has been completed, the systems are identical. By setting the cost of initial installation of s_2 equal to the cost of upgrading to s_2 , we isolate the cost of time-delay in responding to the increased value of information.

Again, we formulate the value problem as a dynamic program, and solve via backward induction. Figure 16 shows the values of the two satellite systems with the difference in initial cost of c_u taken into account.

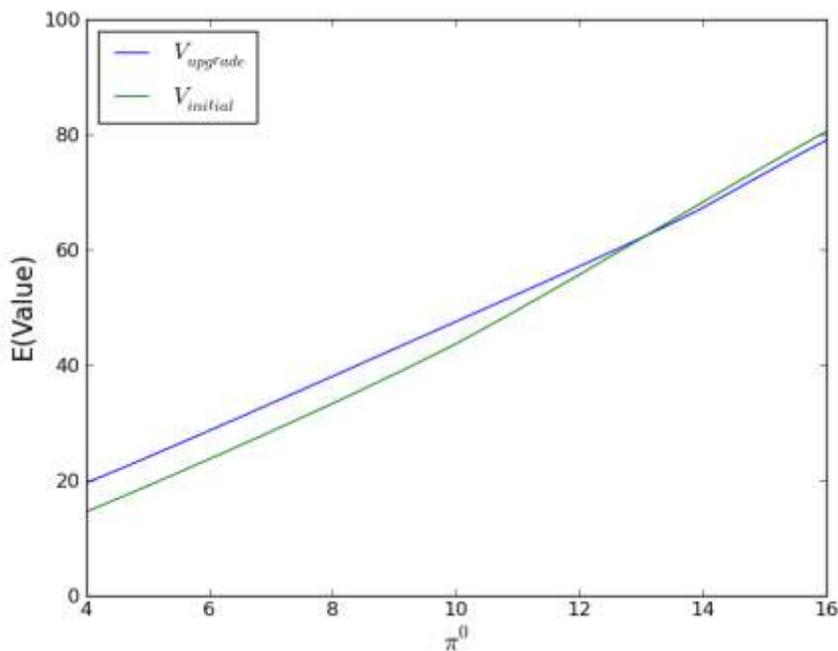


Figure 16: Value of Initial and Upgradable Sensor Flexibility

We see that when the value of information is low, installing extra flexibility initially does not provide value enough to compensate for the increased initial cost. As the value of information increases, the appeal of using sensor s_2 increases. When available either satellite chooses to use s_2 when $\pi^0 > 10$, as the increased detail is worth the additional marginal cost. Optimal control of the upgradable satellite will not make the upgrade decision until the value of information significantly exceeds this marginal threshold, although this threshold to make the upgrade decision is decreasing in the number of

periods remaining. Eventually, the value of the initially flexible satellite surpasses that of the upgradeable satellite as π_0 exceeds 13. This final case illustrates how flexibility in time to deliver capability creates value for a decision-maker.

Discussion

Flexible military, manufacturing, and service systems create value by responding to realizations of uncertainty. Valuing the flexibility in systems is necessary in order to make investment and acquisition decisions regarding flexible systems. In this section, we constructed a taxonomy of system characteristics in order to guide a decision-maker to appropriate model formulations and solution methods to compute the value of flexibility. We based our taxonomy on important characteristics of the flexible system such as the number of decision epochs, the number of alternatives, and the characterization of uncertainty. The taxonomy recommended decision theoretic and mathematical programming formulations for systems with a single, very few, or decomposable decision epochs, dynamic programming formulations for systems with discrete, recurring epochs, and differential equation formulations for systems with a continuum of decision epochs.

We used an observation satellite example to illustrate several of the methods included in our taxonomy could be used to value flexibility derived from several sources. One of the consistent messages from our examples is that the value of any particular flexibility is strongly tied to the initial conditions which determine the value added by the additional alternatives the flexibility offers.

4 EXTENDING METHODOLOGIES FOR VALUING FLEXIBILITY

4.1 CURRENT EXPECTED VALUE LIFE CYCLE COST

The question thus arises whether we can establish the merits of a capability without having to explicitly determine its value. This may be feasible through a modification to the familiar life cycle cost (LCC) model. The idea is to refine current life cycle cost calculations to better account for the value of capability opportunities that are likely to arise throughout the life of a program. Furthermore, the methodology we propose would be capable of inherently evaluating design options in aggregate, thereby rendering distinctions in capability like *flexibility* and *overcapacity* as entirely arbitrary. Before proceeding to a more comprehensive explanation, however, it may be beneficial to review the salient aspects of DOD's current LCC methodology.

Life Cycle Cost

LCC is a systematic accounting approach for aggregating all direct and many indirect costs for a given system. It includes not just total acquisition costs, but also costs related to operations, maintenance, and disposal. Importantly, LCC also accounts for risks,

generally either through sensitivity analyses or through formal quantitative risk analysis [DAU, 2010.]. For large programs, calculating the LCC is generally a tedious undertaking involving substantial time and effort. But the outcome is nevertheless generally deemed to be worth-while. As a formal measure, life cycle cost is entirely straightforward, and easily understood by the typical spate of stakeholders, to include systems engineers, users, and contractor and government managers. Moreover, by providing senior decision-makers with their single best source of estimated cost to achieve a given capability, the LCC is often instrumental in determining the ultimate fate of a program.

Formal DOD guidance calls for the LCC to be first accomplished as part of the initial Analysis of Alternatives (AoA) and is only updated as part of major milestone decision reviews. Aside from these updates, however, the system LCC is generally a static measurement. When calculated, it provides a “snapshot” estimate of total life cycle cost on the assumption that there will be no deviations from key cost, schedule, and performance parameters, which are collectively referred to as the acquisition program baseline (APB) [DAU, 2010]. Of course, one thing we know with near certainty is that there will almost always be deviations from the APB.

While the assumption of a static APB may be unwarranted, programs proceed with it anyway, presumably because the alternative of trying to account for the non-deterministic uncertainty in precisely *how* the program will deviate from the APB is simply not possible, or at least just too daunting. It can be argued, however, that even though uncertainty is—by definition—not deterministic, it may be possible to employ stochastic probability methods that can yield cost estimates that are likely to be more accurate in the long run. Although establishing the initial models to accomplish this would require significant resource investment, the possibility of more accurate LCC estimates—and the improvement in decision-making that would accompany that—promises an enormous return on such an investment.

Life Cycle Cost Under Uncertainty

Thus, there is substantial motivation to provide improved LCC estimates, at least to the level required to support decisions considering alternative flexible design options. The notion that this can be done by accounting for random events that affect the system forms the basis of life cycle cost under uncertainty (also referred to as *stochastic life cycle cost*), which was mentioned earlier as part of the discussion on value-driven design. The idea of applying this strategy to acquiring military systems appears to have been first introduced by Brown in two papers related to the F6 satellite program [Brown, Long, et al., 2007; Brown and Eremenko, 2008]. As described by Brown, stochastic life cycle cost is premised on three assertions.

- *The cost to develop, procure, and operate a system with some assured minimum capability over its lifecycle is not a deterministic value.*
- *Instead, this cost can be modeled as a random variable with a probability distribution resulting from a set of uncertainties introduced throughout the system's life.*

- *This random variable metric is a relevant basis for comparison between alternative system architectures and design choices.*

Brown is to be commended for introducing this simple but deceptively powerful notion of stochastic life cycle cost. However, the initial treatment does not develop the principle fully, nor explore its broader applicability. The type of stochastic events he considers are only those specific events that critically influence the success of a satellite system, i.e., launch failure and on-orbit component failure. Brown explicitly does not consider other aspects of life cycle uncertainty that affect virtually all programs, such as “requirements creep, funding stream volatility, technology development risk, and volatility of demand” [Brown, Long, et al., 2007]. Yet he clearly does recognize that the model could be applied to these other sources of uncertainty, noting that these variables are “left for future analysis.” To date, it does not appear that such an analysis has been accomplished by him or others.

Consequently, we propose a research strategy to logically extend this promising technique in a manner that may provide a number of potential benefits over current practices. Specifically, we intend to expand the life cycle cost under uncertainty idea to a robust and comprehensive methodology for effectively valuing system design alternatives. For the remainder of this section, we explore how such an approach could be applied to uncertainty as related to *system performance*. We expect to address its applicability to other sources of uncertainty (i.e., cost and schedule) in subsequent efforts.

Another modification to enhance the utility of the LCC concept is that it should *not* be viewed as simply a static measure only to be crafted in support of key milestones. Just as LCC is an essential decision tool for those in the role of Milestone Decision Authority (MDA) and above to gauge the value of a program, it can fulfill the same principal function to those who serve at the program manager level and below. Moreover, estimates of life cycle cost are not useful just periodically, but have ongoing utility at all stages of the program, as design decisions are continually required at various levels of the program which (to varying degrees) are likely to impact the overall system cost. And whereas early LCC values would naturally be focused on high-level architectural decisions, as the program matures, and the requirements baseline migrates from *functional* to *allocated* to *product*, the decision trade space will concomitantly shift to the more detailed design implementations. Thus, this dynamic and (probabilistically) more accurate LCC should arguably be managed, updated, and referenced as often as the program schedule.

Current Expected Value of Life Cycle Cost

To capture the utility of this improved LCC concept, we offer the appellation, CEVLCC, which stands for *Current Expected Value Life Cycle Cost*. The name is intended to convey a couple of key distinctions from the standard LCC and Brown’s stochastic LCC. The “Expected Value” phrase discriminates CEVLCC from the standard LCC as a more probabilistically accurate measurement of system cost; whereas the word “current” is intended to connote the fact that the CEVLCC would be employed as a

living, continually updated decision analysis tool. The notion that an LCC estimate might be applied dynamically, and at lower levels of system design, is distinct from Brown's view that the stochastic LCC could only be useful for "preliminary trade space exploration" and not for value determinations "below the architectural level" [Brown and Eremenko, 2008]. For clarity, here are the specific assumptions that must hold for this approach to be valid—

1. As programs mature, there will be unpredictable deviations from the APB that affect the system's LCC
2. It is possible, on average, to provide a more accurate LCC estimate through probabilistic modeling of the stochastic processes that cause deviations in the APB
3. The cost of the effort required to calculate a more accurate LCC is more than offset by the value obtained by the more accurate LCC
4. Given the CEVLCC cost accounting methodology, as long as each design meets all of its threshold requirements, then its relative value can be inferred from its cost

In addition, the proposed methodology is straightforward, consisting of the following steps:

- Establish system design options
- Construct time-phased probability distribution functions (PDFs) associated with all existing key cost, schedule, and technical performance parameters of the program
 - Assign time-phased probabilities for potential new capabilities of the system
 - Estimate standard (i.e., traditional) life cycle cost
- Estimate costs associated with modifications (consistent with PDFs) to baseline cost, schedule, and technical performance parameters
 - Estimate costs associated with the addition of new capabilities
 - Calculate CEVLCC for each system design option and select alternative with the lowest CEVLCC

Hypothetical Use Case

To appreciate the process and potential utility of CEVLCC, we illustrate its application using a hypothetical missile defense scenario. For simplicity, we will only consider technical performance as part of this analysis.

Assume we have a requirement to protect a high-value facility in a sensitive overseas location, which must conform to the following four Key Performance Parameters (KPPs).

#	Key Performance Parameter	Threshold	Objective
1	Protect facility from ballistic missile attack with X% assurance	X=95	X=99
2	Engage only missiles is >X% confident they represent an imminent threat to the facility	X=90	X=95
3	No evidence of military presence w/ in X miles of facility	X=25	X=40
4	Be able to engage X missile(s)	X=1	X=5

Table 2: KPPs for Missile Defense Scenario

The second CEVLCC assumption states that it is possible to formulate probabilistic modeling of the stochastic processes that cause deviations in the APB. One way to accomplish this is to treat the value for each performance parameter—in this case, each threshold KPP value—as a random variable, and construct the probability function. To do this with any semblance of confidence would likely require extensive empirical data from a variety of different requirement categories, program types, program levels, acquisition strategies, etc. Furthermore, the PDFs would be valid only at a single point in time, so they would need to be revised as the program matures and new information becomes available. Clearly, construction and maintenance of the CEVLCC would require significant effort; nevertheless, it could be done, and would be worth doing if the third assumption above holds. Figure 17 below provides examples of what those PDFs might look like in the case of KPPs #1 and #3:

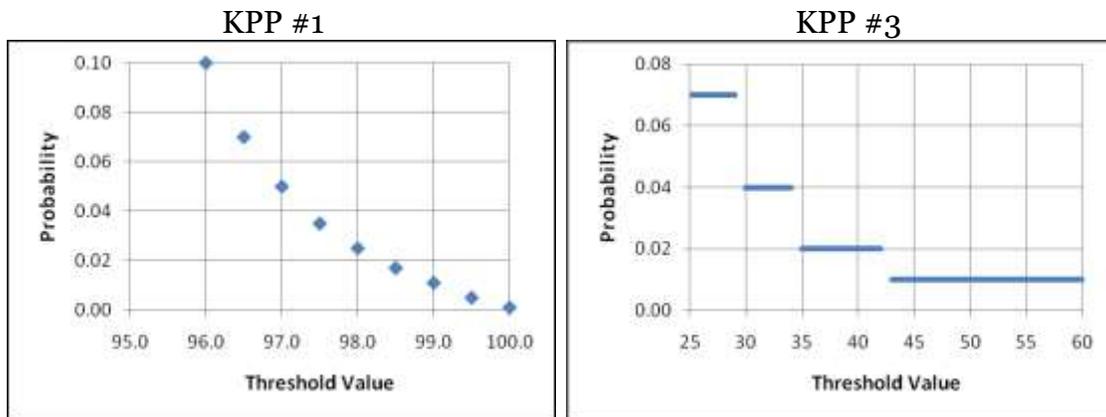


Figure 17: Notational PDFs of Missile Defense Scenario KPPs

In both cases, the x-axis is the random variable (i.e., the KPP threshold value), and the y-axis is the probability associated with a particular value of the random variable. For simplicity, we have chosen not to depict the probability that the variable will remain the same or decrease, but in a comprehensive model, these probabilities would likely need to be determined as well.

After establishing the PDFs for the parameter values of existing capabilities, we next need to account for the probability that the system will be required to support new (and obviously foreseeable) capabilities. For instance, we might conceive of the following two potential new capabilities, along with their estimated likelihoods:

- Protect against cruise missile threats (15%)
- Protect against unconventional ordnance attacks (e.g., suicide bomber) (2%)

Each of these probability functions will require a temporal dimension as well. In other words, these estimations of probability are associated with a given time horizon, and will necessarily vary depending on that horizon. For this scenario, we might estimate that if a requirement related to the first new capability (i.e., protect against cruise missiles) has not been introduced by the Preliminary Design Review (PDR), then

its likelihood of being imposed between the PDR and the Critical Design Review (CDR) is three percent, and its likelihood of being imposed between the CDR and the Test Readiness Review (TRR) is one percent, and so on. Viewed in this way, we recognize a certain similarity between these various PDFs and traditional risk burn-down plans. This is an important point, as the PDFs would need to be man-aged in a similar manner, and could reasonably be integrated with traditional risk analyses.

In both cases (i.e., the modification of existing capabilities and the addition of new capabilities), the assigned probabilities will admittedly be estimates, perhaps quite rough ones. Will they be exactly right? Absolutely not. If our stochastic models are at all valid, are they likely to be closer to reality than the assumption that nothing will change over the remaining life of the program? Almost certainly.

Next is the cost assessment step. This is executed in the context of whatever design options we have available to us at any given time. Let's assume, based on earlier assessments, that the program has chosen a defensively-oriented architecture that engages ballistic missile threats during the terminal phase. Like all true decisions, the program has made an irrevocable allocation of resources as a result, and has, to some extent, necessarily constrained their design space going forward. Nevertheless, the commitment to the terminal phase option still leaves a number of fundamental design decisions open to them. We then postulate the following list of architectural possibilities being considered by the program:

1. Terrestrial interceptor system stationed at least 25 miles from facility
2. Concealed (e.g., underground) terrestrial interceptor system
3. Airborne interceptor system
4. Terrestrial directed-energy system stationed at least 25 miles from facility
5. Concealed (e.g., underground) terrestrial directed-energy system
6. Airborne directed-energy system
7. Hardened structure that ensures survivability of facility
8. Force field

Each of these architectures has relative strengths and weaknesses based on the KPPs as writ-ten. And each of these designs has its own inherent costs to implement. All else being equal, under the traditional conception of LCC, the option above with the lowest LCC that also meets all threshold requirements is typically the one that will be selected. This is the crux of the problem, as this traditional approach does not account for the value of the flexibility embedded within certain architectural options.

The CEVLCC, however, requires that additional cost estimating be performed against the range of potential new KPP threshold values as well as the potential new capabilities. Clearly, some of these options are better poised to accommodate changes in the KPP thresholds. For instance, the concealed terrestrial architectures (i.e., options #2 and #5) will have no additional cost should there be an increase associated with the threshold value of KPP #3, where-as the non-concealed versions (i.e., options #1 and #4) would likely have an enormous cost impact. Similarly, some architectures can more easily accommodate new capabilities. If the program is directed to incorporate the capability to protect the facility against cruise missiles, then the airborne interceptor system can be

modified much more easily than the underground interceptor system (i.e., the airborne system is more *flexible*). And the hardened structure option will not have to be modified at all, as the capability to withstand the cruise missile strike was already embedded in its design (i.e., it's *overcapacitized*).

Once we've determined the estimated costs for the potential changes to the system, we calculate all of the expected values for each design with respect to each change. So suppose that for all three directed energy architectures, we estimated the following additional costs for the potential range of changes in the value of the KPP#1 threshold.

Index (<i>i</i>)	KPP#1 Threshold (X)	Additional Cost to Implement (<i>x_i</i>)	Probability (from Figure 17)
1	96.0%	\$0.0M	10.0%
2	96.5%	\$0.0M	7.0%
3	97.0%	\$0.0M	5.0%
4	97.5%	\$1.0M	3.5%
5	98.0%	\$1.0M	2.5%
6	98.5%	\$1.0M	1.7%
7	99.0%	\$3.0M	1.1%
8	99.6%	\$6.0M	0.5%
9	99.9%	\$20.0M	0.1%

Table 3: Marginal Probability Costs for Directed Energy Architectures

Using the standard formula for expected value, then architectures #4, #5, and #6 (i.e., the directed energy architectures) have an expected value of \$160k with respect to KPP #1. We repeat this process for each architecture for the remaining KPPs. We then account for the potentially new capabilities in the same manner, although the expected value calculation is a trivial weighted probability with a single term. So for each architecture, a separate CEVLCC is calculated by summing its baseline LCC, its summation of expected values for the modifiable capabilities, and its summation of expected values for the new capabilities, i.e. —

$$CEVLCC = LCC + \sum_{i=1}^m E(X)_i + \sum_{i=1}^n E(y)_i$$

m = number of modifiable capabilities

n = number of new capabilities

This leads to the fourth assumption. As long as each design meets all threshold requirements, then its relative value can be inferred from its cost. Ordinarily, this would not be valid, but given our cost formulation methodology, we implicitly accounted for those discriminators that would otherwise have contributed to the value side of the equation. Specifically, there is no need to assess how much to “credit” a particular design for its ability to exceed a KPP threshold or its capacity to accommodate future changes. Both of these inherent design values are captured (albeit in complementary fashion) in our marginal probability cost estimates within CEVLCC. In other words, if a particular design option were able to more inexpensively accommodate a capability

change—whether via flexibility or via overcapacitization—its weighted cost would be less than the competing designs, and its *value* would be greater. Note that this is why it is only valid to compare systems that meet all threshold requirements, as there needs to be a value baseline to reference. Based on these assumptions, then, we can now assert that *the system that is the best value is simply the one with the lowest CEVLCC.*

As promising as CEVLCC might be, we recognize there are also a number of potential draw-backs to this technique, most of which are tied to the model assumptions. For instance, the fundamental nature of defense acquisition may be more *chaotic* than *stochastic*, thus preventing accurate predictive modeling over a reasonable time horizon, and fully precluding analysis of unforeseeable changes (violation of assumption 2). Also, to be most effective, CEVLCC would need to be comprehensive and current, which results in a large number of permutations to account for, thus potentially making its implementation cumbersome. Even if the resource investment is deemed to be worthwhile very early in the program (i.e., when design decisions are most impacting), it is possible that the return on investment will not be sufficient to justify its use much further into the program (violation of assumption 3). Importantly, CEVLCC, as currently conceived, also cannot effectively provide a relative evaluation of design options that do not meet threshold requirement levels (violates assumption 4). Finally, the CEVLCC does not entirely sidestep the problem of valuing capability, as excess capability above the threshold often does have value that must be accounted for. This technique does not properly account for the temporal benefit that an overcapacitized solution provides, i.e., having a newly desired capability immediately (or more quickly) available vice waiting for development and implementation.

There is consensus that uncertainty is a principal reason that DOD programs continue to struggle mightily with respect to their ability to adhere to cost and schedule projections. While acquisition policies and strategies that aim to abate uncertainty are admirable and often useful, ultimately they can only help so much. Since uncertainty is a certainty, programs may be better served by infusing their systems with an inherent ability to effectively respond to uncertainty. The singular term most commonly associated with such an ability is *flexibility*. While flexibility is arguably the single best term for this concept, even it is insufficient to capture the full range of capability responsiveness we would like our systems to have. We may also need them to be *versatile* and/or *overcapacitized*. However, making a system flexible, versatile, or overcapacitized inevitably requires additional investment that must be justified. The only viable way to provide that justification is to quantify the value of the capabilities that can be more easily achieved because of the investment. For military weapons systems, this task is, at best, extremely challenging, and, at worst, simply not feasible.

Consequently, a fundamentally different approach is needed—one that does not rely on an explicit valuation of potential capabilities, and is capable of evaluating design options more strategically, thus shifting the focus from the somewhat narrow view of just flexibility, per se, to the broader view of capabilities, regardless of how they are achieved. Thus, we propose the CEVLCC, a top-down, *intrinsic* value model based on the familiar notion of life cycle cost. The idea is premised on the notion that the need for

capability changes in a program arises in a stochastic manner that can be modeled and incorporated into a continually updated, expected value model of total program cost. We believe CEVLCC potentially offers a number of advantages over current practices—

- An inherent focus on capability *in toto* that serves to automatically assimilate relevant capability concepts, such that discriminatory design considerations like *overcapacity* and *versatility* become irrelevant
- An inherent ability to incorporate cost and schedule components of a program, there-by obviating the distinction between *design* flexibility and *process* flexibility
- Being comprised of concepts already familiar to the acquisition community (i.e., life cycle cost and risk analysis), thereby greatly reducing cultural entry barriers
- Having a simple premise and an intuitive output (i.e., cost), both of which encourage adoption among stakeholders across the acquisition community
- Not being subject to criticisms specific to real options analysis
- Being able to mostly sidestep theoretical and practical challenges associated with valuing military capabilities

Currently, the CEVLCC concept is largely notional, and significant research effort remains to determine its validity and/or utility. Most of the work in this section is intended to validate, or at least characterize the limitations of, the CEVLCC assumptions. Specifically, we —

- Analyze/characterize APB behavior for historical programs and examine concomitant LCC behavior with the intent of identifying salient factors that drive perturbations.
- Construct a basic CEVLCC model based on these salient factors.
- Compare the LCC accuracy for historical programs over time to the corresponding CEVLCC, and conduct tradeoff analyses to determine when the return on investment in the CEVLCC model is no longer worthwhile.
- Identify/develop alternate methodology to address CEVLCC weakness with respect valuing existing excess capability.
- Refine CEVLCC model and validate via historical case-studies

4.2 OPERATION AND SUPPORT (O&S) COST ESTIMATION

To better understand the proposed methodology for characterizing current O&S cost estimates for DoD systems, it is necessary to be familiar with the relevant extant research and reporting mechanisms. Even assuming an awareness that the acquisition phase is regarded as more important than the O&S phase, the existing number of studies seeking to characterize O&S costs for defense systems still seems shockingly sparse. Between 1945 and 2009, there were over 130 separate studies and commissions focused on the acquisition of DoD systems, dozens of which involved the nature of acquisition cost behavior [DoD, 2009]. During this same time period, there appears not to be a single published study pertaining to how system costs behave during the O&S phase.

It would seem that WSARA has infused a greater sense of urgency into the DoD regarding the need to characterize O&S costs. Following its edict to “review existing systems and methods of the DoD for tracking and assessing O&S costs,” there has been a relative flurry of reports (four) quantitatively examining O&S costs for defense programs. Before summarizing these reports, however, a clarification regarding terminology is warranted. Within the acquisition community, “cost growth” is a well-established term with a specific meaning: the degree to which the actual costs of a system vary from the estimated (i.e., baseline) costs. When these O&S studies use the term “cost growth,” however, it has a broader meaning, denoting the difference between initial and final estimated costs in some cases, and the difference between initial and final actual costs in others. Although this conception of cost growth does allow us to gauge the stability and precision of the O&S cost estimates (or actuals), per se, it does not provide insight into the accuracy of DoD O&S cost estimates.

The most comprehensive O&S study can be found in the WSARA Product Support Assessment [DoD, 2009]. In this study, 34 weapon systems are analyzed along several proposed dimensions of sustainment effectiveness. One new measure, dubbed “Sustained Cost Management,” analyzes the year-to-year changes in actual O&S costs, which may be an effective measure of O&S costs, but does not help us achieve the goal of ascertaining the accuracy or reliability of historical O&S cost estimates.

Another broad study was conducted by the Center for Naval Analyses (CNA), also in 2009 [Choi, Alper, Gessner, et al., 2009]. The study, which involved 26 Navy programs, had the expressed and promising purpose of vetting the hypothesis that actual system O&S costs are “radically” exceeding early estimates. However, rather than comparing estimated costs to actual costs, the analysts decided—in all but three cases—to compare initial and final O&S cost estimates, choosing to assume that the final O&S cost estimate could reasonably represent actual costs¹. Furthermore, although the O&S costs of three programs were analyzed by comparing estimates to actuals, unfortunately the numerical results were not provided in the report.

In 2010, a narrower, but more in-depth study was performed by the Institute of Defense Analysis (IDA) which examined the Air Force C-17A program [Balaban, Devers, and Roark, 2010]. Like the majority of the systems in the CAN study, the C-17A analysis only involved the comparison of initial estimates to final estimates.

The fourth DoD O&S cost study was published by the Government Accountability Office (GAO) in 2010 [GAO, 2010]. This report initially sought to review O&S cost behavior of major defense programs, but ended up being more of an indictment of current Pentagon deficiencies with respect to tracking and managing O&S costs for its systems. Despite expressed reservations with the fidelity of the underlying data sources, the GAO did report O&S costing information for seven programs. For five of these systems, the analysis involved cost estimates only, but in two cases (the Air Force F-22A,

¹ CNA justified this approach by noting they were unable to obtain true actual costs for the majority of the programs in the study, and did not wish to mix methodologies.

and the Navy F/A-18E/F), estimates were compared to actuals, thereby enabling some insight into the accuracy of initial O&S cost estimates.

The salient components of these O&S studies are summarized below in Table 1. Note that only subsets of the CNA report and the GAO report employed a methodology that allows meaningful characterization of O&S cost estimates. And since there were no numerical results in the CNA report regarding these systems, those few pages of the OIG GAO report which analyze the F-22A and the F/A-18E/F would appear to represent *the only published qualitative characterization of the accuracy of O&S cost estimates for DoD systems.*

Source	Year	# of Systems	Method	Quant. Results
OSD	2009	34	“Cost Growth” in O&S Actuals	n/a
CNA	2009	23	“Cost Growth” in O&S Estimates	n/a
CNA	2009	3	O&S Estimates vs. O&S Actuals	No
IDA	2010	1	“Cost Growth” in O&S Estimates	n/a
GAO	2010	5	“Cost Growth” in O&S Estimates	n/a
GAO	2010	2	O&S Estimates vs. O&S Actuals	Yes

Table 4: Summary of existing O&S studies on defense systems

4.2.1 O&S REPORTING

Although the enactment of WSARA has undoubtedly served as a catalyst for increased focus on O&S cost issues in the DoD, it is also the case that an empirical analysis of O&S costs for defense systems would simply not have been possible until relatively recently. To conduct the analysis requires three elements: a valid source of predicted costs from the acquisition phase, a valid source of actual costs from the O&S phase, and enough elapsed time for a large number of programs to have accrued representative data from both phases.

The obvious source for obtaining official estimates of O&S costs for major defense programs is the Selected Acquisition Report (SAR). SARs are required to be submitted at least annually for all major defense programs until they have been 90 percent acquired (further evidence of the imbalanced emphasis the DoD places on the acquisition phase) [DAU, 2011]. Starting in 1989, programs were required to include as part of each SAR a “full life-cycle cost analysis,” and soon thereafter (about 1990 in most cases), programs began providing estimates of Annual Unitized O&S cost (i.e., average O&S cost per unit, per year). Starting in 2001, most programs also began providing an estimate of Total O&S cost.

Not long after O&S estimates were being required in the SARs, the DoD mandated that each military service maintain an historical database of actual O&S costs for its systems. The effort became known as Visibility and Management of Operating and Support Costs (VAMOSC), with each service component managing its own version.

Though the primary focus of VAMOSC is on future planning and the development of O&S estimates, the nature of the database allows for actual O&S costs to be broken out by weapon system and by year [DoD, 1992; DAU, 2011]. Accordingly, VAMOSC data can serve as “ground truth” for actual system O&S costs, thus enabling an accuracy assessment of the O&S cost estimates found in the SARs.

So with a consistent, time-phased, and reliable source for predicted O&S costs in one hand, and a viable source for obtaining actual O&S costs segregated by system in the other, all that remains is allowing enough time to pass for a sufficient amount of data to be collected for a valid comparison between the two cost figures. Since programs can take many years to develop and field, it is not unreasonable to suspect that it could take a couple of decades to amass sufficient data for a substantive analysis. In fact, the authors have screened all the data, and found that two decades (1991-2010) is enough time to obtain valid O&S cost estimates and actuals for over three dozen major defense programs.

4.3 O&S COSTS METHODOLOGY

At this point, the basic methodology should be evident. Our first step is to annotate the predicted O&S cost estimate (or estimates in those cases where multiple measures of O&S cost are provided) from every SAR for a given system. We next use the VAMOSC data to establish the actual O&S costs for that system. Finally, we compare the actual O&S cost of the system to what it was predicted to be each year during its acquisition phase and characterize the accuracy of that estimate over time. The principle is simple enough, but there are a number of obstacles that complicate the proposed analytical methodology. In fact, multiple authoritative sources are dubious that such an analysis is feasible [GAO, 2010]. Consequently, the remainder of this section is largely devoted to a discussion of how these obstacles can be overcome or mitigated.

The first challenge is one of system selection. There are hundreds of candidate programs that have submitted SARs; however, we have a rather stringent set of selection criteria. First, we eliminate any program that does not provide valid O&S cost estimates. This necessarily includes any program that stopped submitting SARs prior to about 1991 (recall that O&S cost estimates started first being reported around 1990). This also includes a surprising number of programs that either did not comply with the DoD requirement to provide O&S cost estimates, or that provided estimates that were clearly erroneous.

Next, we remove programs for which we cannot obtain valid—and relatively stable—actual O&S costs. Obviously, this encompasses programs cancelled prior to becoming operational, but it also necessarily includes programs for which there are too few years of actual O&S cost data or too few units operationally deployed to have a reasonable expectation of having achieved steady state O&S costs. Also, similar to the difficulties encountered with the SARs, many programs must be eliminated because the VAMOSC data is unavailable or invalid. Another reason for excluding a program is that the level of

cost reporting for the VAMOSC data is incommensurate with the SAR data; this is often the case with major system modifications, which may warrant SAR reporting, per se, but are not sufficiently delineated in VAMOSC with respect to the specific scope or phasing of the modification to allow for meaningful comparisons.

The next major challenge is reconciling discrepancies that arise in the cost element structure between the predicted costs in the SAR and the actual costs from VAMOSC. In some cases, VAMOSC is missing data for entire cost categories identified in the SAR. For instance, the Naval VAMOSC system is unable to allocate to specific weapon systems those expenditures in the category of “indirect costs” (e.g., personnel medical care and base services and support). In order to resolve the mismatch between predicted and actual O&S costs for this cost element, we normalize the Navy data by decrementing the amount of “indirect costs” from all the SAR estimates. In theory, this introduces an additional component of uncertainty into the analysis related to the relative accuracy of cost element categories; however, any ostensible effect is likely to be minimal as the missing VAMOSC costs represent, on average, less than ten percent of the projected O&S costs.

For Navy systems, the issue of missing cost elements in the VAMOSC actuals can be readily mitigated, and, generally speaking, there is no analogous problem for the Air Force VAMOSC system. The Army VAMOSC data, however, is another matter. For Army programs, specific weapon system data is simply not allocated for most of the O&S cost categories, to include indirect costs, personnel costs, contractor costs, and sustaining support costs. This deficiency in the Army cost accounting systems is so significant that it precludes inclusion of any Army programs in this study (many of these deficiencies with the Army VAMOSC system are documented in the GAO [2010] report).

Another challenge related to this proposed methodology is the fact that, for many systems, the level of reporting between the estimates and the actuals is not commensurate. In some cases, a single O&S cost estimate is provided for the system, but actuals for that same system must be aggregated from multiple sources. This generally occurs for one of three reasons. One, it is a joint program and the VAMOSC cost data spans multiple service components. Two, the core mission of the system changes over its life such that costs are accrued via different databases. Three, system variants are combined in the SAR, but segregated in the VAMOSC systems. In all three cases, it is generally possible to consolidate the cost actuals such that estimates and actuals represent an “apples-to-apples” comparison.

The complementary problem can also occur, whereby the estimates are broken out by variant (whether in a single SAR or multiple SARs), but the actuals are not segregated. This tends to pose a greater challenge as we must calculate a composite figure from the multiple O&S cost estimates in order to enable a valid comparison to the actuals. For these cases, we calculate the composite O&S cost estimate by weighting the relative number of units for each variant. This is a reasonably valid approach in most cases, but is arguably less so if the temporal phasing of the deployment of the variants is too great, or the actual relative proportion of the variants varies substantially from what

was planned. In sum, while incommensurate treatment of system variants can certainly complicate the analysis, it can be accounted for, and it hardly represents a significant methodological barrier.

There are two measures of O&S cost that can be assessed via this proposed methodology: *Total O&S cost* and *Annual Unitized O&S cost*². Each measure—within the context of this study—has its own strengths and weaknesses. The Total O&S cost is a readily intuitive metric that offers (when summed with total acquisition cost) a direct means of establishing an estimate of system LCC, the most comprehensive and facile cost indicator for system value assessments. Although Total O&S costs are not explicitly stated for any system until 2001, we can sometimes infer an estimate for these earlier SARs. Note this is possible only if certain other information is provided in the SAR, such as the Annual Unitized O&S cost and the assumed operational service life. It should be noted that this inference does represent an overly simplistic view of Total O&S costs (e.g., it neglects, ramp-up/ramp down periods, attrition, refurbishments, etc.), but this is the basic method employed in the vast majority of later SARs to estimate Total O&S costs. So while the inferential calculations we employ may be simplistic, they are at least consistent.

There are, however, two notable problems in using Total O&S costs as a metric. One is that it is highly correlated to unit quantities. This can lead to particularly high degrees of estimate variance over time; or alternatively, it can provide a ready-made strategy for programs to effectively mask increasing estimates of Annual Unitized O&S costs, by reducing fielded quantities until Total O&S costs meet cost goals (this is the same gamesmanship strategy which has been curtailed on the acquisition side via the concept of the “unit cost breach”). The second problem with examining Total O&S costs is we cannot possibly know the true actual costs at this time. With two exceptions, none of the programs in this study has reached end of life, so the portion of Total O&S costs that has not yet been incurred must be estimated by prorating actual costs to date across the remainder of the expected operational service life, potentially creating another source of uncertainty.

The Annual Unitized O&S cost may not provide direct insight into the system LCC, but the data tends to be more broadly available both on the estimate and actuals side. Estimates of Annual Unitized O&S cost often begin around 1991, and actual Annual Unitized O&S costs can be easily calculated for any year in which actual O&S costs are reported and the number of operational units can be determined. Of note, however, obtaining operational unit counts is not always as straightforward as one would expect. For aviation and maritime systems, unit counts are readily available in the VAMOSC databases, though in some cases—depending on the SAR assumptions—we use the available operational inventory counts vice the actual full inventory counts. For non-aviation and non-maritime systems, however, actual counts are not available in the VAMOSC databases. In these cases, alternative sources (e.g., the managing

² A third metric is also theoretically possible: Operating cost per unit of time (i.e., cost per flight/steaming/driving hour). However, we elected to omit this metric for this study due to the relative paucity of predicted data points and the lack of inconsistency among programs in how precisely to employ this type of metric.

organizations) must be queried to obtain historical fielded quantities. Yet even with these unit count challenges, Annual Unitized O&S costs (as opposed to Total O&S costs) often provide a more valid comparative measure across similar contemporary systems or antecedent systems. Of course, in some cases, such as when the system consists of very few units or the system has a monolithic nature, the relevance of a unitized cost metric is diminished or lost.

Perhaps surprisingly, inflation adjustment may represent the largest practical challenge with this proposed methodology. Clearly, to make valid comparisons between projected costs and actual costs, comparisons must be made in the same base year dollars. In theory, there are two approaches: either *inflate* the SAR estimates or *deflate* the VAMOSC actuals. In practice, the former option is not feasible, as the SAR cost categories do not cleanly map to specific inflation categories. And since the level of fidelity in the SAR O&S estimates does not allow insight into the specific inflation categories, it is not possible to accurately inflate these numbers based solely on the SAR data. Therefore, in all cases, we have chosen to deflate the VAMOSC data. This turns out to be a highly intricate task for most systems, as inflation factor adjustments can be quite abstruse and the specific calculations vary significantly based on service component and system type. Furthermore, since we cannot inflate the projected O&S costs, we cannot compare SAR estimates that are provided in different base years, even for the same program. This has the unfortunate effect of creating artificially separate analytical units (e.g., F-22 base year 1990 costs and F-22 base year 2005 costs), thus slightly diminishing our ability to compare systems and ascertain statistical patterns.

There were many other minor problems involved with this methodology that are too subtle to address substantively in this report. Many of these smaller issues pertain to abnormalities discovered in the SAR O&S estimates of nearly every program. Examples include inconsistencies between tabular values and the narrative text, unstated assumptions, varying metric parameters, and incorrect units of measurement. In most cases, logical inferences were made and documented, such that the problematic estimate could be included in the study (though sometimes it was assigned a lower weighting in the analysis to reflect reduced confidence in the data). In other cases, however, the necessary inference could not be reasonably justified, in which case that particular SAR data was excluded from the analysis.

4.3.1 ENHANCED TOC-PL MODELING WITH MONTE CARLO ANALYSIS

Phase 1 of this project detailed a general TOC model for valuing flexibility of product lines. In the current phase, this model was enhanced to handle specific DoD application domains, and added initial Monte Carlo simulation capabilities. We began incorporating the life cycle cost ratios for Operations and Support (O&S) shown in Table 5 and Table 6. The hardware O&S cost distributions were derived from [Redman et al., 2008] and software from [Koskinen, 2010].

System Type	O&S Costs
Missiles (average)	12%
Ships (average)	60%
Aircraft (F-16)	78%
Ground vehicles (Bradley)	84%

Table 5: Hardware Life Cycle Cost Ratios

System Type	O&S Costs
Business, Command-Control	75-90%
Complex platforms as above	50-80%
Simple embedded software	10-30%

Table 6: Software Life Cycle Cost Ratios

Setting the life cycle cost ratios as a function of system type in the tables impacts the general TOC Product Line model inputs for Ownership Time and Annual Change Cost. The user chooses a system type and ownership time, which invokes a calculated annual change costs for the relevant domain.

Flexibility Case Study for a Specialized DoD Application Domain

This case study illustrates a domain-specific analysis for a missile system with a demonstration of Monte Carlo simulation. The initial case study was for a general system, but in this scenario the user specifies a missile system for O&S life cycle cost defaults.

A missile product line development with three year ownership time is being evaluated. The user chooses the Missile System Type, and sets Ownership Time to 3 years. With these inputs, the pre-calculated Annual Change Cost = $12\%/3 \text{ years} = 4\%$. The tool results are in **Figure 18**.

Shown also are the optional Monte Carlo results from varying the relative cost of developing for flexibility. The means are listed with the ROI distribution graph. While this example demonstrates the Monte Carlo capabilities for a single parameter, all input parameters are open to variation for more sophisticated Monte Carlo analysis.

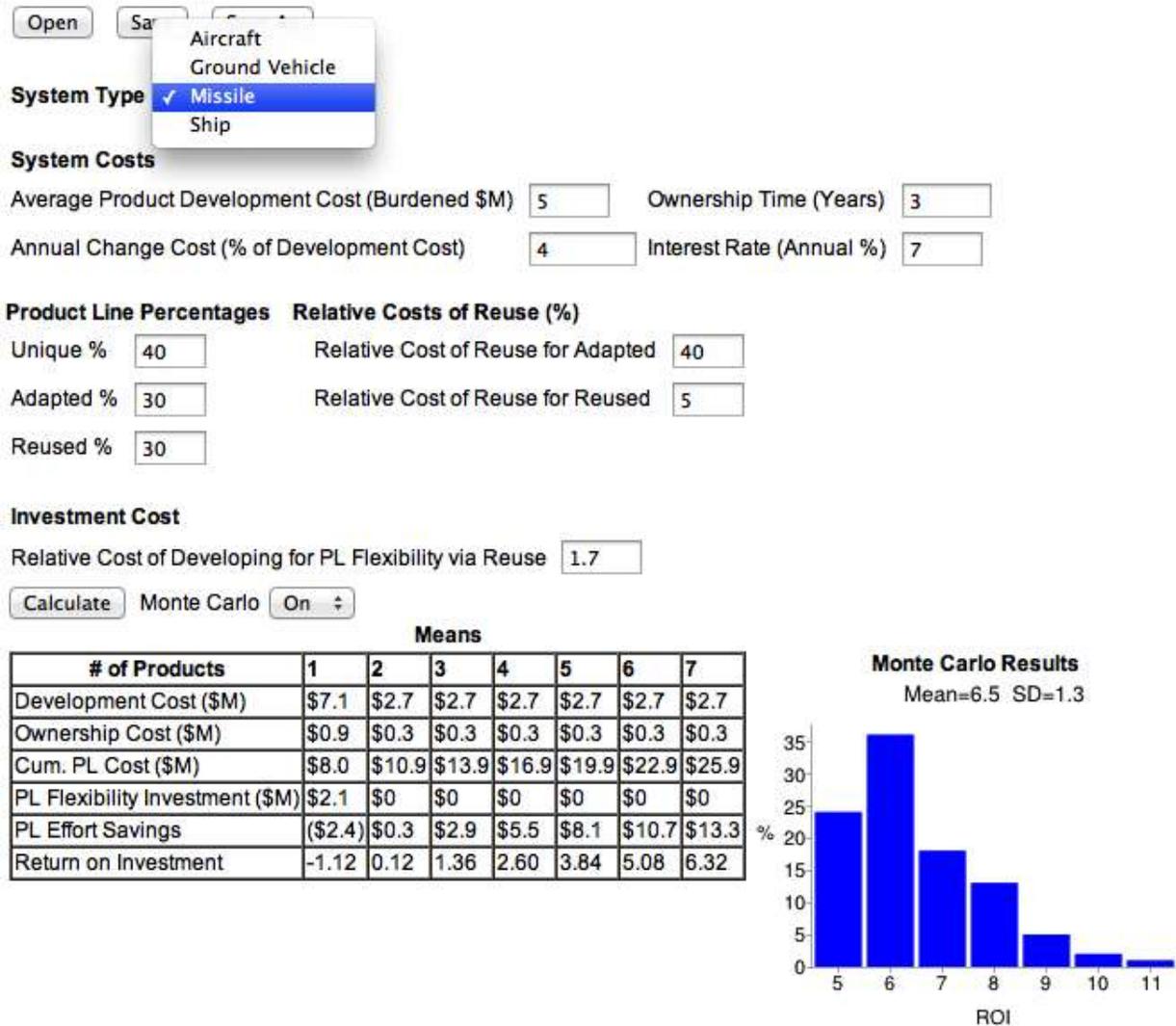


Figure 18: DoD Application Domain and Monte Carlo TOC-PL Results

4.3.2 MIXED MODEL APPROACH

The model presented in this section, are based on longitudinal data [Ryan et al., 2012], which is to say that the source data consists of repeated measurements on different subjects over time. Importantly, the nature of longitudinal data precludes the possibility of assuming an identical and independent distribution (i.i.d.) of the random variables. Because the data is clustered into programs, with repeated measurements of each program over time, there necessarily exists a correlation between the repeated measurements for a given program—and therefore the statistical errors of the observations—that must be accounted for in the statistical analysis. Further, one expects these correlations to be greater for data points close in time, such as for successive SARs from the same program. This means that the statistical errors will be correlated as well.

Importantly, the fact that we expect correlated errors for the programs in this study invalidates the underlying assumptions of simple analysis of variance and regression models, namely i.i.d. observations. To compensate for this, we instead employ *mixed model* techniques for the data in this study. Mixed models use both fixed (i.e., entire population) effects and random (i.e., subject-specific) effects within the same analysis. The key distinction between mixed models and simple regression models is that the former can produce valid models even if the subject observations are not independent. In essence, mixed models allow the data to exhibit inherent correlations and non-constant variability that arise from the program-specific effects. This allows one to effectively model not only the measures of central tendency for the data, but also the covariance structure attributable to the repeated measurements [Diggle, Liang, and Zeger, 1994; Verbeke and Molenberghs, 2000].

Relative to the standard General Linear Model, the use of a mixed model for this analysis provides several advantages, primarily relating to flexibility. A mixed model allows the use of input variables even if data is missing for one or more of the subjects (i.e., programs). Mixed models can also automatically accommodate for unequal spacing of the repeated measurements (i.e., ensure minimum variance), which is a characteristic of this data set. In addition, the mixed model allows more efficient and direct modeling of the within-subject covariance structure for the entire dataset, as opposed to unique covariances for every data pair. Finally, the results from the mixed model can be readily extended to outcomes that do not conform to a normal distribution. In this study, we have assumed the cost estimate errors are normally distributed (i.e., the solution to the mixed model equations is a maximum likelihood estimate where the distribution of the errors is normal), but the mixed model can accommodate nonlinear approaches, should they be considered more appropriate [Patetta, 2002].

To put this in mathematical terms, the GLM in matrix form is given as

$$y = X\beta + \varepsilon \quad (1)$$

where

y = the observed data vector, where $E(y) = X\beta$ and $var(y) = \sigma^2 I$
 X = the fixed effect design (i.e., model) matrix
 β = the vector of fixed effect parameter estimates (same for all subjects)
 ε = the vector of residual errors, where $E(\varepsilon) = 0$ and $var(\varepsilon) = \sigma^2 I$

For the mixed model version, a random-effects term is added

$$y = X\beta + Z\gamma + \varepsilon \quad (2)$$

where

Z = the random effect design (i.e., model) matrix
 γ = the vector of random effect parameter estimates (varies by subject)

In addition,

$$E \begin{bmatrix} \mathcal{Y} \\ \varepsilon \end{bmatrix} = 0 \text{ and } \text{var} \begin{bmatrix} \mathcal{Y} \\ \varepsilon \end{bmatrix} = \begin{bmatrix} \mathbf{G} & 0 \\ 0 & \mathbf{R} \end{bmatrix} \Rightarrow \text{var}(y) = \mathbf{Z}\mathbf{G}\mathbf{Z}' + \mathbf{R} \quad (3)$$

where

\mathbf{G} = the random effects covariance matrix

\mathbf{R} = the fixed effects covariance matrix

One of the key inputs for a mixed model analysis is what structure should be used for the random covariance matrix, \mathbf{G} . For this data set, since we tend to observe high correlations in the response variables reported in successive SARs, but increasingly less correlation as the time between SARs grows larger, a covariance structure that captures diminishing levels of correlation is desired. Therefore, a sensible choice for model development is the autoregressive (AR) structure, which has homogeneous variances and correlations that will decline exponentially with temporal distance [Wolfinger, 1993]. Multiple other covariance matrix structures were also examined, but overall model performance was best using first-order autoregressive, i.e., AR(1).

To obtain the estimates of \mathbf{G} and \mathbf{R} , we solve for the values that optimize an objective function, in this case the Restricted Maximum Likelihood (REML) criterion. The method for computing the denominator degrees of freedom for the tests of fixed effects was Kenward-Roger. Thousands of model iterations were executed to find the best set of variables from Table 7 to use in each model: The Bayesian Information Criterion (BIC) was used as the primary method of discrimination between potential models. All model analysis was accomplished using SAS version 9.3 (<http://www.sas.com/software/sas9/>).

UNCLASSIFIED

#	Program Name	SubProgram Name	Lead Component	System Type	SAR Years	# of SARs	LCC SARs	AUC SARs
1	AIM-9X	AIM-9X (Navy)	Navy	Munition	1996-2001	6	6	6
2	AMRAAM (AF)	AMRAAM (AF)	Air Force	Munition	1988-1992	5	3	3
3	AMRAAM (Joint)	AMRAAM (Joint)	Air Force	Munition	1992-2010	18	18	18
4	AOE 6	AOE 6	Navy	Maritime	1988-1997	11	7	7
5	AV-8B	AV-8B REMAN.	Navy	Aviation	1994-2002	10	NA	10
6	C-130J	C-130J	Air Force	Aviation	1996-2010	13	12	12
7A	C-17A	C-17A (BY1981)	Air Force	Aviation	1987-1994	10	8	8
7B	C-17A	C-17A (BY1996)	Air Force	Aviation	1995-2010	14	14	14
8	C/MH-53E	C/MH-53E	Navy	Aviation	1987-1994	9	1	5
9	CVN 68 (74/75)	CVN 68 (74/75)	Navy	Maritime	1987-1998	13	2	2
10	CVN 68 (76)	CVN 68 (76)	Navy	Maritime	1994-2002	9	4	4
11	DDG 51	DDG 51	Navy	Maritime	1987-2010	25	20	20
12	E-2C	E-2C	Navy	Aviation	1994-2006	14	4	4
13	F-14D	F-14D	Navy	Aviation	1987-1993	9	5	9
14	F-16C/D	F-16C/D	Air Force	Aviation	1987-1994	8	4	4
15A	F-22	F-22 (BY1990)	Air Force	Aviation	1991-2004	16	4	16
15B	F-22	F-22 (BY2005)	Air Force	Aviation	2005-2010	6	6	6
16	F/A-18C	F/A-18C	Navy	Aviation	1987-1994	10	NA	7
17A	F/A-18E/F	F/A-18E/F (BY1990)	Navy	Aviation	1992-1999	9	9	9
17B	F/A-18E/F	F/A-18E/F (BY2000)	Navy	Aviation	2000-2010	10	10	10
18	GLOBAL HAWK	GLOBAL HAWK	Air Force	Aviation	2001-2010	11	NA	11
19	JASSM	JASSM	Air Force	Munition	1999-2009	12	11	12
20A	JPATS	JPATS (BY1995)	Air Force	Aviation	1995-1999	5	NA	5
20B	JPATS	JPATS (BY2002)	Air Force	Aviation	2001-2010	9	8	9
21	JSOW	JSOW	Navy	Munition	1997-2010	14	14	14
22A	JSTARS	JSTARS (BY1983)	Air Force	Aviation	1989-1996	10	8	10
22B	JSTARS	JSTARS (BY1998)	Air Force	Aviation	1997-2003	6	3	6
23	KC-135R	KC-135R	Air Force	Aviation	1987-1994	8	NA	5
24	LHD 1	LHD 1	Navy	Maritime	1987-2005	18	15	15
25	LPD 17	LPD 17	Navy	Maritime	1996-2010	16	16	16
26A	MH-60R	MH-60R (BY1993)	Navy	Aviation	1994-2005	14	12	12

26B	MH-60R	MH-60R (BY2006)	Navy	Aviation	2006-2010	5	2	2
27	MH-60S	MH-60S	Navy	Aviation	1998-2010	17	17	17
28	MHC 51	MHC 51	Navy	Maritime	1991-1998	8	8	8
29	PREDATOR	PREDATOR	Air Force	Aviation	2009-2010	2	2	NA
30	SSGN	SSGN	Navy	Maritime	2002-2007	6	6	6
31	SSN 21	SSN 21	Navy	Maritime	1987-1999	15	11	11
32	SSN 774	SSN 774	Navy	Maritime	1995-2010	16	16	16
33	STRAT. SEALIFT	STRAT. SEALIFT	Navy	Maritime	1993-2001	11	NA	11
34A	T-45TS	T-45TS (BY1984)	Navy	Aviation	1987-1993	10	5	5
34B	T-45TS	T-45TS (BY1995)	Navy	Aviation	1994-2007	14	12	13
35	T-AKE	T-AKE	Navy	Maritime	2001-2010	10	10	10
36	T-AO 187	T-AO 187	Navy	Maritime	1987-1994	8	4	4
Total						470	317	392

Table 7: MDAP Data Used For Model Development

Independent Variables and the Unit of Analysis

As noted earlier, a central assumption of the macro-stochastic cost estimating approach is that there exists a relatively small set of high-level program parameters that, in aggregate, significantly relate to the LCC and AUC estimate errors observed for a given program. Table 8 lists and defines all of the independent variables we evaluated as potential fixed or random effect parameters for both the LCC and AUC cost models. All variables in this table are based on information available in the program SARs. Some of the variables are taken directly from the SAR, some are calculated based on information available in different sections of a single SAR, and some are calculated from information available across successive SARs. All cost figures are in native (i.e., SAR-specific) base year (BY) dollars, with the exception of variable #14. Although there are only 50 variables listed in Table 8, the inclusion of “trending versions” of several variables (see footnote #2) brings the total count to 252.

#	Variable Name	Msmnt. Level	Description (Values)
1	Program Year	Interval	Number of years since Milestone B (II) or program initiation
2	DoD Component	Nominal	Lead acquisition service component (“AF” or “Navy”)
3	Joint	Nominal	Are units being procured for more than one service? (“Yes” or “No”)
4	System Type	Nominal	Type of system (“Aviation,” “Maritime” or “Munition”)
5	Acq Phase	Nominal	Acquisition phase (“Development” or “Production”)
6	Acq Type	Nominal	Type of acquisition (“New,” “Modification,” or “Variant”)
7	Maturity	Ordinal	Program maturity level; categories based on <i>Expended</i> (#18).
8	Total Dev APBs	Interval	Cumulative number of development APBs to date
9	Avg Dev APBs	Interval	Average number of development APBs per year
10	Total Prod APBs	Interval	Cumulative number of production APBs to date
11	Avg Prod APBs	Interval	Average number of production APBs per year
12	Prime Contractor	Nominal	Contractor for 3 largest active contracts (“Boeing,” “GD,” “Lockheed-Martin,” “Northrup-Grumman,” “Raytheon,” “Other,” or “Multiple”)
13	Acq Cost Est	Interval	Current estimate of total acquisition cost
14	Acq Cost Est, BY10 ^{3,4}	Interval	Current estimate of total acq cost standardized to BY10 dollars
15	AUC Est ^{3,4}	Interval	Current estimate of annual unit O&S cost
16	O&S Cost Est ^{3,4}	Interval	Current estimate of total O&S cost
17	LCC Est ^{3,4}	Interval	Current estimate of total LCC cost
18	Expended	Interval	Percentage of <i>Acq Cost Est</i> (#13) expended to date
19	Funding Years	Interval	Current total planned funding years of program
20	PAUC Change, Dev	Interval	Percentage change in Program Acquisition Unit Cost (PAUC) from Development baseline
21	PAUC Change, Prod	Interval	Percentage change in Program Acquisition Unit Cost (PAUC) from Production baseline
22	APUC Change, Dev	Interval	Percentage change in Average Procurement Unit Cost (APUC) from Development baseline
23	APUC Change, Prod	Interval	Percentage change in Average Procurement Unit Cost (APUC) from Production baseline
24	CV, Engr ³	Interval	Total cost variance (CV) to date in engineering category as % of <i>Acq Cost Est</i> (#13)
25	CV, Est ³	Interval	Total CV to date in estimating category as % of <i>Acq Cost Est</i> (#13)
26	CV, Quan ³	Interval	Total CV to date in quantity category as % of <i>Acq Cost Est</i> (#13)
27	CV, Total ³	Interval	Total CV to date in all CV categories as % of <i>Acq Cost Est</i> (#13)
28	CV, Total-Quan ³	Interval	Total CV to date in all CV categories (except Quantity) as % of <i>Acq Cost Est</i> (#13)
29	Breaches, Sched	Interval	Cumulative number of schedule breaches to date
30	Breaches, Perf	Interval	Cumulative number of performance breaches to date
31	Breaches, Cost	Interval	Cumulative number of cost breaches to date
32	Breaches, UC	Interval	Cumulative number of unit cost breaches to date
33	Breaches, Total	Interval	Cumulative total of all breaches to date

³ Includes trend versions of variable to date, i.e., minimum, maximum, range, mean, weighted mean (by Program Year), standard deviation, and the slope of the regression line.

⁴ One or more transformations applied (i.e., unitary normalization, scalar reduction, square root, and natural log) to better achieve model stability, interpretability, and/or to capture nonlinear relationships.

34	Breach, N-M	Nominal	Has program incurred a Nunn-McCurdy breach? (“Yes” or “No”)
35	CDR-MSII	Interval	Time between Critical Design Review (CDR) and Milestone II
36	CDR-PDR	Interval	Time between CDR and Preliminary Design Review (PDR)
37	LRIP-MSII	Interval	Time between Low Rate Initial Production (LRIP) and Milestone II
38	MSIII-MSII	Interval	Time between Milestone III and Milestone II
39	IOC-MSIII	Interval	Time between Initial Operating Capability (IOC) and Milestone III
40	IOC-MSII	Interval	Time between Initial Operating Capability (IOC) and Milestone II
41	Reqmnts, New	Interval	Cumulative # of new requirements added to performance baseline
42	Reqmnts, Deleted	Interval	Cumulative # of existing requirements removed from performance baseline
43	Reqmnts, Total	Interval	Total number of requirements in current performance baseline
44	Reqmnts, Obj	Interval	Percentage of total requirements to date in which objective value was made more stringent
45	Reqmnts, Thresh	Interval	Percentage of total requirements to date in which threshold value was made more stringent
46	Reqmnts, Change	Interval	Percentage of total requirements to date in which threshold or objective value was modified
47	Procure, Plan ³ ₄	Interval	Current total planned procurement quantity
48	Procure, Change ³ ₄	Interval	Percentage change in <i>Procure, Plan</i> (#47) relative to baseline
49	Procured	Interval	Percentage of <i>Procure, Plan</i> (#47) currently procured
50	Unit Acq Ratio	Interval	Ratio of <i>AUC Est</i> (#15) to <i>Acq Cost Est</i> (#13)

Table 8: Listing of Independent Variables Evaluated in Error-Correction Models

Table 8 is also interesting for what is not included. Defense acquisition professionals and cost estimators alike are keenly interested in the cost impacts of a number of strategic policies related to procurement. Three of the most intriguing—and controversial—relate to acquisition strategy (e.g., traditional vs. evolutionary), contracting strategy (fixed-price vs. cost-reimbursement), and sustainment strategies (organic vs. contractor). Although each of these policy topics could potentially serve as an excellent macro-level predictor of cost estimating accuracy, we were unable to incorporate variables related to any of these topics.

The fundamental obstacle in all three cases was being able to effectively quantify these variables in the context of fluctuating and disparate acquisition efforts. Consider, for instance, an evolutionary acquisition strategy, which may not be implemented until late in the program when technical maturity is sufficient or may only be applied to a particular element of the system in development. It may also be that an evolutionary strategy is abandoned midway through development or blended with more traditional practices into a hybrid approach. This is just one example, but these types of subtleties tend to dominate these three important procurement policy topics, thus regrettably precluding definitive categorization.

The last item involving methodology that the reader should be aware of pertains to the unit of analysis, which is equivalent to the *model subject*. This is a subtle, but critical, analytical element that changes throughout model development, characterization, and validation. We begin with the *SubProgram* as the unit of analysis, but then switch to a broader subject defined as the *Program Category*. This

transformation is crucial to infusing predictive capability into the macro-stochastic cost model. During validation, however, the unit of analysis reverts to the full *Program* in order to present model performance in a context most likely to resonate with target users. This nonstandard progression regarding the unit of analysis (i.e., model subject) is explained in greater detail at each step of model characterization.

4.3.3 THEORETICAL MACRO-STOCHASTIC MODELS

The first task is to assess the theoretical premise of a macro-stochastic cost model. A reasonable suspicion would be that the nature of cost estimating errors for defense programs—along with the underlying uncertainty which drives them—is inherently chaotic, such that attempting to characterize these errors via a stochastic process is misguided at best. Thus, the fundamental question that must be answered at the outset is whether there is any meaningful correlation between the variables in Table 8 and the level of accuracy in a given SAR's cost estimate. We believe that the data shown in Figure 19 through Figure 22 offers a compelling response to this question.

Figure 19 is a plot of the percentage error in the empirical LCC estimates for all of the MDAPs considered. Overall, the data exhibits a high level of dispersion. Although the mean error across all programs is only -4.7 percent, the mean magnitude of the errors (i.e., the mean absolute value of the errors) is over 22 percent. The magnitude error does appear to reduce slightly as time increases, suggesting that the accuracy of LCC estimates may be improving slightly as program acquisition matures. However, as noted in the characterization, this is likely an artifact of the acquisition cost component of the LCC converging to a known value by the end of the acquisition phase [Ryan et al., 2012]. When examining total O&S cost, per se, there is no significant improvement in LCC estimating accuracy as time goes on.

Figure 20 plots the results from a macro-stochastic mixed model that attempts to predict the error in each SAR and then compensate for it. The subject of this model is the SubProgram (for reasons explained in the characterization, the SubProgram—vice the Program—is the more appropriate unit of analysis). Designating the SubProgram (or the Program, for that matter) as the model subject is a logical choice, but it has important implications to model utility to be discussed shortly.

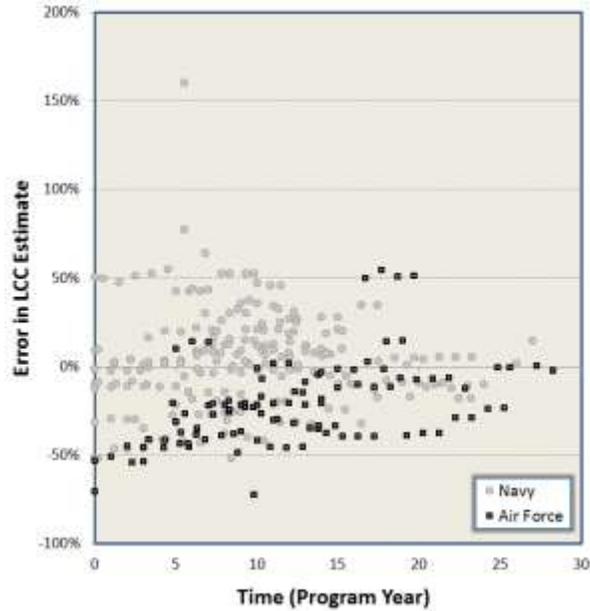


Figure 19: Error in LCC Estimate as a Function of Time (Empirical Data)

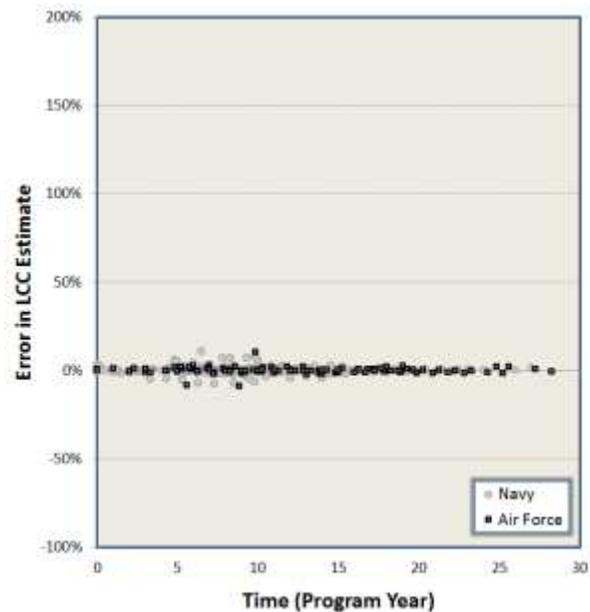


Figure 20: Error in LCC Estimate as a Function of Time (Theoretical Macro-Stochastic Model)

This so-called “theoretical macro-stochastic cost model” depicted in Figure 20 consists of just three variables: Procure, Change (#48 in Table 8), the standard deviation of the natural logarithm of Acq Cost Est, BY10 (#14), and the natural logarithm of LCC Est (#17). All three variables are modeled as fixed effects, while the first two—along with an intercept term—are also modeled as random effects. The way to interpret this result is that the broad pattern (i.e., the fixed effects) of life cycle cost estimating errors in all Navy and Air Force MDAPs can be captured by examining the extent of procurement quantity changes to date, the variability of the acquisition cost estimates to date, and the current LCC estimate. Further, each program has its own

pattern of errors (i.e., the random effects) driven by the procurement quantity changes and the variability of the acquisition cost estimate to date, as well as a unique starting point as defined by the intercept term.

Figure 21 and Figure 22 are the AUC versions of Figure 19 and Figure 20. Figure 21 shows that empirical AUC estimating accuracy for MDAPs is considerably worse than LCC estimating accuracy, with the magnitude of the errors and accompanying standard deviation almost twice as high. Figure 22 depicts the same data using a macro-stochastic model, and again, only three variables are used. This time the variables are the Unit Acq Ratio (#50), the standard deviation of the natural logarithm of Acq Cost Est, BY10 (#14), and the natural logarithm of AUC Est (#15). As before, the first two variables are modeled as both fixed and random effects, and the model includes a random intercept term. The model subject remains the SubProgram.

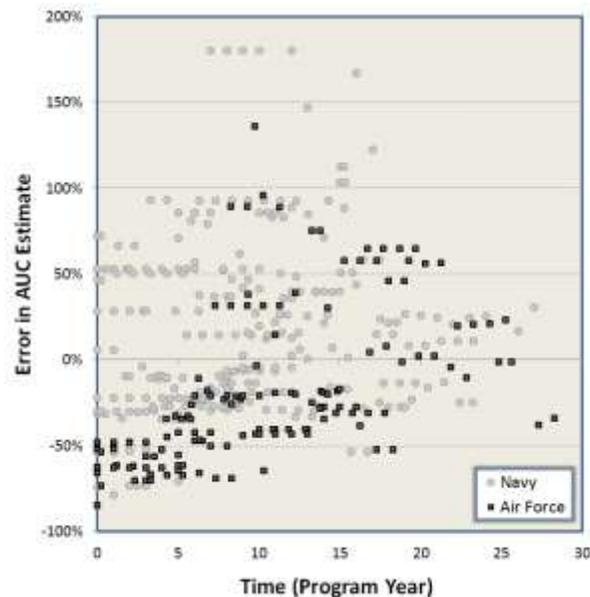


Figure 21: Error in AUC Estimate as a Function of Time (Empirical Data)

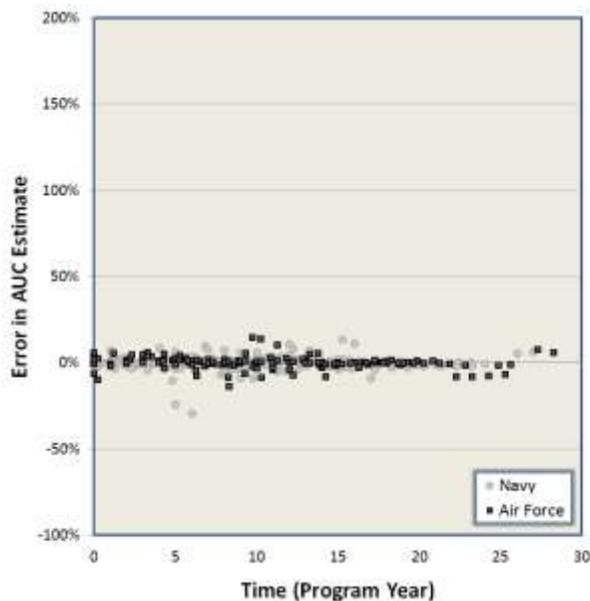


Figure 22: Error in AUC Estimate as a Function of Time (Theoretical Macro-Stochastic Model)

In both cases, the theoretical macro-stochastic model performs impressively, driving down the magnitude of the mean error in the original prediction to a little over one percent in the case of LCC estimates, and just over two percent in the case of AUC estimates. Since the result is represented in percentage terms, it is easy to lose context of the amount of money involved. But these potential improvements in estimating accuracy typically represent billions of dollars. Since the mean magnitude error in the original LCC estimate is over 20 percent, a program estimated to cost \$30.0 billion over its life cycle could be expected to actually cost somewhere in the range of \$24.0 to \$36 billion, a \$12 billion range. On the other hand, the macro-stochastic model might predict a life cycle cost of \$34.0 billion, but its equivalent expected error range would only be \$800 million. Clearly, such a massive reduction in cost uncertainty would be of tremendous benefit to defense acquisition officials.

In one respect, this significant estimating improvement is an extremely important result. Figure 20 and Figure 22 are remarkable because they show the tremendous potential utility of the macro-stochastic cost modeling approach. With a highly parsimonious model, the model is able to predict the actual LCC and AUC estimating errors for all of the programs in this study with exceptional accuracy. Moreover, the random (subject-specific) effects are very powerful, strongly suggesting there is a unique pattern for each unit of analysis. This result is especially impressive given that there are over 35 SubPrograms in both models, over half of which consist of at least 10 data points (i.e., SARs) that must be “fitted.”

However, in another—arguably more relevant—respect, this finding is of little utility. The problem with the preceding approach is that it is inherently a post-hoc analysis. This is why we refer to this model as “theoretical.” One cannot expect that the exact cost estimating error patterns of these programs will occur again. So although using the

SubProgram as the model subject may reveal powerfully descriptive random effects, the theoretical macro-stochastic model has no meaningful predictive capability.

The fact remains, however, that we now have some measure of confidence in the principle of macro-stochastic cost estimating of DoD programs. The challenge becomes how to translate this technique into a useful prognostic model.

Program Categories

In order to construct a predictive macro-stochastic model, the authors have devised a template-based solution involving the creation of Program Categories. This approach aims to achieve a better balance between model accuracy and utility by structuring the data into broader categories comprising multiple programs and using criteria that apply to both current and foreseeable programs. In this way, the Program Category supplants the SubProgram as the model subject and the unit of analysis.

To use a stock market analogy, the Program Category notion is the equivalent of forecasting an individual company's performance based on the business sector to which it is assigned. In the absence of company-specific performance indicators (which would be preferred, but may not be available until too late), we assume that the company's future performance will roughly conform to the average pattern of all the other companies in the same sector. A key to making this approach work, of course, is ensuring that companies (i.e., programs) are assigned to representative sectors (i.e., categories).

Indeed, establishing the exact Program Categories and ontological criteria was one of the most challenging aspects of model development. Our first goal was to be able to employ the model as early as possible, so the criteria used to assign a program to a particular Program Category had to be clearly discernible at the outset of a program. Second, we wanted the Program Category criteria to be simple and logical, easily derived from the list of independent variables in Table 8. Third, we sought to have each category consist of programs similar enough to one another that the new model subject (i.e. the Program Category) would continue to exhibit statistically significant subject-specific patterns that could be captured by the random design matrix of the mixed model. (Given the complex interactions between various fixed and random effect model terms and the constituent covariance matrices, identifying meaningfully similar programs is often far from clear).

In addition, the total number of program categories needed to be carefully considered as it represented another source of tension between accuracy and utility. If we create too few categories (i.e., many programs in a single category), the power of the mixed model is bound to be diminished as there will likely be little in the way of subject-specific effects to model. If we create too many categories, then we run the risk of building a model that is still too program-specific. In other words, if we have a large number of categories with a few number of programs in each, then we cannot—without additional data—have confidence that we have identified a valid Program Category that will effectively subsume a future program of interest.

We evaluated many different categorization structures defined via various variables and attribute thresholds, as well as varying numbers of categories. In the end, we empirically determined that the best balance of performance and utility was achieved through seven Program Categories defined by the following three variables: DoD Component (#2), System Type (#4), and Program Size based on Acq Cost Est, BY10 (#14). Although the Program Category criteria were the same for both the AUC and LCC model, the specific programs and SAR counts are slightly different due to differences in data availability. Table 9 and Table 10 show the Program Category structure and program assignments for each model.

PCat	DoD Comp	System Type	Size (Mean Acq Cost Est, BY10)	SARs	# of Programs	Assigned Programs
1	AF	Aviation	Small (\leq \$18.0B)	33	4	C-130J, JPATS, JSTARS, PREDATOR
2	Navy	Aviation	Small (\leq \$18.0B)	53	5	C/MH-53E, E-2C, MH-60R, MH-60S, T-45TS
3	Both	Aviation	Large ($>$ \$18.0B)	60	5	C-17A, F-16C/D, F-22, F-14D, F/A-18E/F
4	Navy	Maritime	Small (\leq \$8.5B)	41	7	AOE 6, CVN68 (74/75), CVN68 (76), MHC 51, SSGN, T-AKE, T-AO 187
5	Navy	Maritime	Medium (\$8.5B – \$30.0B)	42	3	LHD 1, LPD 17, SSN 21
6	Navy	Maritime	Large ($>$ \$30.0B)	36	2	DDG 51, SSN 774
7	Both	Munition	All	52	5	AIM-9X, AMRAAM-AF, AMRAAM-JT, JASSM, JSOW
TOTALS				317	31	

Table 9: Summary of LCC Macro-Stochastic Cost Model Program Categories (Pcats)

PCat	DoD Comp	System Type	Size (Mean Acq Cost Est, BY10)	SARs	# of Programs	Assigned Programs
1	AF	Aviation	Small (\leq \$18.0B)	58	5	C-130J, GLOBAL HAWK, KC-135A, JPATS, JSTARS
2	Navy	Aviation	Small (\leq \$18.0B)	68	6	AV-8B, C/MH-53E, E-2C, MH-60R, MH-60S, T-45TS
3	Both	Aviation	Large ($>$ \$18.0B)	83	6	C-17A, F-16C/D, F-22, F-14D, F/A-18C, F/A-18E/F

4	Navy	Maritime	Small (\leq \$8.5B)	52	8	AOE 6, CVN68 (74/75), CVN68 (76), MHC 51, SSGN, STRAT. SEALIFT, T-AKE, T-AO 187
5	Navy	Maritime	Medium (\$8.5B – \$30.0B)	42	3	LHD 1, LPD 17, SSN 21
6	Navy	Maritime	Large ($>$ \$30.0B)	36	2	DDG 51, SSN 774
7	Both	Munition	All	53	5	AIM-9X, AMRAAM-AF, AMRAAM-JT, JASSM, JSOW
TOTALS				392	35	

Table 10: Summary of AUC Macro-Stochastic Cost Model Program Categories (PCats)

Note that while the acquiring service component and the system type would not be expected to change during a program's life, the size of the program does change as acquisition cost estimates vary—sometimes significantly—over time. The dependence of the Program Category assignment on acquisition cost estimates introduces the possibility that a program's category assignment might change at some point in development. For the programs in our data set, this did not happen, but it could for some future program. If this were to occur, it's not clear whether that means the differently-sized program is in fact behaving more like the programs in its newly assigned category, or whether the size thresholds we have established here would need to be modified.

In addition, the fact that a surface maritime system (i.e., DDG 51) and a submarine system (i.e., SSN 774) are grouped together into a single category is likely to aggrieve the traditional cost estimator (as presumably would the grouping of fixed and rotary-wing aviation systems). Although both the surface vessel and the submarine are maritime systems, the Navy cost estimator knows that there are key cost-impacting differences between how each type of program is acquired and operated. With respect to the modeling approach pursued here, the point to keep in mind is that the pattern of program costs for similar systems is a fundamentally different phenomenon than the pattern of program cost errors. It is the latter that is relevant to our approach, and using this metric, the groupings in Table 9 and Table 10 proved to be the most effective.

4.3.4 A PROGNOSTIC MACRO-STOCHASTIC MODEL

By restructuring the data from individual programs into Program Categories, we can now use the model to make predictions. Given the assumption that future programs are essentially like the programs in this data set, then as long as a future program can be assigned to one of the existing categories, the macro-stochastic model can be reasonably applied at any time after program initiation to predict the expected error in the program's cost estimate and, by extension, predict the actual LCC or AUC.

This improved utility has come at a cost, however. The powerful program-specific trends depicted in Figure 20 and Figure 22, which consisted of only three independent

variables, are diluted by the amalgamation with other—albeit similarly behaving—programs. In essence, the new model subject of Program Category requires that the random effects design matrix (Z) compromise between multiple, different program trends, resulting in reduced model performance. Or, to continue the market analogy, a particular company's performance is not likely to exactly follow the average of its assigned sector: There will be important company-specific deviations. Fortunately, we can restore a large degree of expected model performance though the inclusion of additional variables.

The final LCC macro-stochastic model incorporated 12 variables (to include five random variables) from Table 8 and the final AUC macro-stochastic model incorporated 14 (to include six random variables). The selected fixed and random variables, along with their estimated parameter values, are listed in Table 11 through Table 14. Since the random variables vary by Program Category, they are specified in their own tables. The reader should be cautious in making inferences based on relative parameter estimate values as not all variables are normalized, and the relationship between parameters is complicated by the inclusion of both fixed and random effects.

#	LCC Model Variable	Random Effect?	Fixed Effect Estimate
1a	<i>DoD Component (#2) – Navy</i>	No	0.4157
1b	<i>DoD Component (#2) – Air Force</i>	No	0.0000
2a	<i>Acq Type (#6) – New</i>	No	0.2132
2b	<i>Acq Type (#6) – Modification</i>	No	0.2183
2c	<i>Acq Type (#6) – Variant</i>	No	0.0000
3	Weighted Mean of Normalized <i>Acq Cost Est, BY10 (#14)</i>	Yes	3.2555
4	Std. Dev. of Natural Log of <i>Acq Cost Est, BY10 (#14)</i>	No	9.3392
5	Natural Log of <i>LCC Est (#17)</i>	No	7.1928
6	Mean of Natural Log of <i>LCC Est (#17)</i>	No	-4.8595
7	Maximum <i>CV, Est (#25)</i>	Yes	-0.7387
8	Slope of Regression Line of <i>CV, Quan (#26)</i>	Yes	0.2188
9	Standard Deviation of <i>CV, Total (#27)</i>	No	-1.6512
10	Range of <i>CV, Total-Quan (#28)</i>	Yes	0.7593
11a	<i>Breach, N-M (#34) – No</i>	Yes	-3.1063
11b	<i>Breach, N-M (#34) – Yes</i>	Yes	-3.1440
12	Std. Dev. of Square Root of <i>Procure, Change (#48)</i>	No	-0.3264

Table 11: LCC Macro-Stochastic Model Variables and Fixed Effects Parameter Estimates

#	LCC Model Variable	Random Effect Estimate						
		PCat	PCat	PCat	PCat	PCat	PCat	PCat
		1	2	3	4	5	6	7
1	Wtd. Mean of Normalized Acq Cost Est, BY10	- 3.5641	4.4451	- 2.0476	- 3.4837	3.2109	- 3.4782	4.9177
2	Maximum CV, Est	0.238 4	- 2.0194	-1.1120	0.5359	1.0697	0.7222	0.5652
3	Slope of Regression Line of CV, Quan	0.398 9	0.1223	-1.1514	- 0.4866	2.3672	- 0.6491	- 0.6013
4	Range of CV, Total-Quan	-0.1111	1.0039	0.3164	1.8285	- 2.2551	- 0.2410	- 0.5417
5 a	Breach, N-M – No	0.492 4	- 0.0670	- 0.4322	0.2778	- 0.5395	- 0.0921	0.3606
5 b	Breach, N-M – Yes	0.5163	- 0.0541	- 0.4083	0.5018	0.0801	0.0442	- 0.7883

Table 12: LCC Macro-Stochastic Model Random Effects Parameter Estimates by Program Category (PCats)

#	AUC Model Variable	Random Effect?	Fixed Effect Estimate
1a	DoD Component (#2) – Navy	No	0.4687
1b	DoD Component (#2) – Air Force	No	0.0000
2a	Acq Phase (#5) – Development	Yes	1.6995
2b	Acq Phase (#5) – Production	Yes	1.6816
3a	Acq Type (#6) – New	No	0.3993
3b	Acq Type (#6) – Modification	No	-0.1132
3c	Acq Type (#6) – Variant	No	0.0000
4	Mean of Scaled Acq Cost Est, BY10 (#14)	Yes	0.4536
5	Natural Log of AUC Est (#15)	No	0.6391
6	Mean of Natural Log of AUC Est (#15)	No	-0.5730
7	Maximum CV, Engr (#24)	Yes	1.4515
8	Weighted Mean of CV, Est (#25)	No	1.5208
9	CV, Quan (#26)	No	0.8438
10	Mean CV, Total (#27)	No	-1.2817
11	Wtd. Mean of Natural Log of Procure, Plan (#47)	Yes	0.1570
12	Mean of Square Root of Procure, Plan (#47)	No	0.2402
13	Wtd. Mean of Square Root of Procure, Change (#48)	Yes	-0.1111
14	Unit Acq Ratio (#50)	Yes	5.8501

Table 13: AUC Macro-Stochastic Model Variables and Fixed Effects Parameter Estimates

#	AUC Model Variable	Random Effect Estimate						
		PCat	PCat	PCat	PCat	PCat	PCat	PCat
		1	2	3	4	5	6	7
1 a	<i>Acq Phase (Development)</i>	-1.2914	0.4545	0.3163	0.0417	-	-	0.7601
1 b	<i>Acq Phase (Production)</i>	-1.2125	0.5695	0.2867	-	-	-	1.0289
2	Mean of Scaled <i>Acq Cost Est, BY10</i>	-	0.3191	-	-	0.4997	-	0.5919
3	Maximum <i>CV, Engr</i>	-	-	1.8679	-	-	-	0.002
4	Wtd. Mean of Natural Log of <i>Procure, Plan</i>	0.0549	-	0.2851	0.2652	-	0.4902	-
5	Wtd. Mean of Square Root of <i>Procure, Change</i>	0.4584	0.1701	0.3614	0.0257	-	-	0.1188
6	<i>Unit Acq Ratio</i>	-	-	1.0105	-	-	1.4998	0.4168
		2.0932	0.2902		0.0981	0.4456		

Table 14: AUC Macro-Stochastic Model Random Effects Parameter Estimates by Program Category (PCat)

Note that over half of the final variables from both models are capturing, in some manner, program trends to date regarding the estimated cost and/or production quantity. These variables may capture trends either directly by what is being measured (e.g., Nunn McCurdy Breach, Cost Variance, etc.) or indirectly via changes in a given variable to date (e.g., standard deviation, mean, etc.). Regardless, a consequence of this predominance of trending variables is that a program should have at least one previous SAR on which to construct a trend value: Without a previous SAR, we find that model performance diminishes considerably. In practice, this results in a small impact on the utility of the model in that it is not suitable for use until the second SAR, which is nominally one year after program initiation.

Figure 23 and Figure 24 show, respectively, the performance of the LCC and AUC prognostic macro-stochastic models where the subject equals the Program Category. Each model is capable of predicting the accuracy of a current LCC or AUC point estimate at any point in a program's life where at least two SARs are available, and then compensating for that error to provide a statistically more accurate estimate. Although model performance is not as impressive as it was for the theoretical model (where the model subject was SubProgram), it is still far better than current estimate performance. The mean magnitude error in the prognostic LCC macro-stochastic model is more than a fourfold improvement of the empirical estimate; for the AUC model, the improvement is over fivefold.

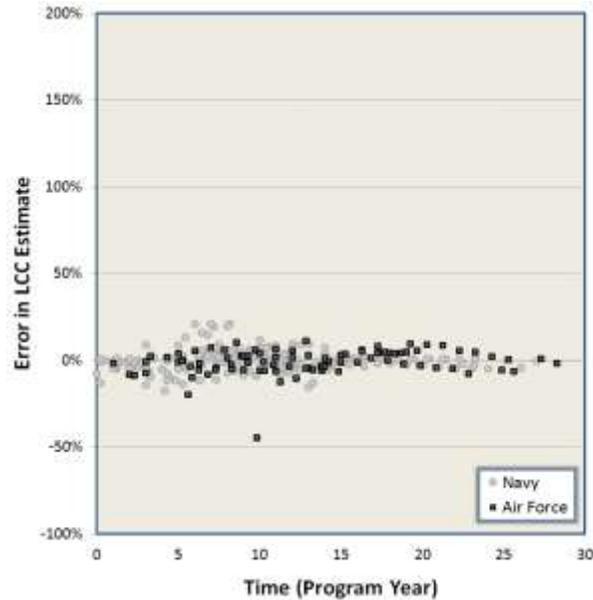


Figure 23: Error in LCC Estimate as a Function of Time (Prognostic Macro-Stochastic Model)

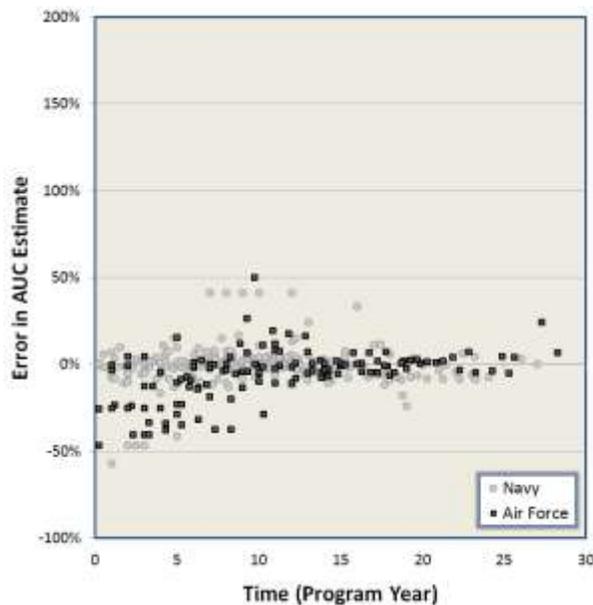


Figure 24: Error in AUC Estimate as a Function of Time (Prognostic Macro-Stochastic Model)

The performance shown in Figure 23 and Figure 24 was achieved under conditions in which the training data set and the test (i.e., validation) data set were equivalent. Thus, it is reasonable to suspect that actual model performance against future programs will be reduced [Larson, 1931; Hart and Wehrly, 1986]. In order to validate the model, we need to test its performance against data that is not available to the model. It is obviously not desirable to wait several years for new program data to become available, but the current size of the data set is an impediment to a standard data partitioning techniques (i.e., dedicated training and test data sets). With respect to validation, the

most logical unit of analysis is the program, as that is the fundamental entity for cost estimation and cost accrual accounting in the DoD. For both the LCC and AUC model, however, we have fewer than three dozen programs available for analysis, hardly sufficient to execute a robust validation involving separate training and test data sets.

This leads us to cross-validation. However, the specific method of cross-validation for the macro-stochastic model is more complicated than it might at first seem. The non-i.i.d. nature of the data also invalidates standard cross-validation techniques: omitting an observation (i.e., one SAR) does not remove the associated information due to correlations with other observations from that subject [Opsomer, Wang, and Yang, 2001; Arlot, 2010]. Suggested techniques to work around this problem include modified cross-validation [Chu and Marron, 1991], h-block cross-validation [Burman, 1994], and sequential validation [Bengio and Chapados, 2003].

Unfortunately, none of these techniques are well suited for the structure of the MDAP data. Not only is the correlation distance (i.e., the strength of the correlation) highly dependent on the program, but several programs have an insufficient number of SARs to faithfully implement the given technique. For instance, in the case of h-block cross-validation, determining the theoretically appropriate size of h in our data set is not clear, but it must be relatively large, and any value of h greater than two could eliminate as many as six programs from the validation.

As a result, we have implemented a tailored version of Leave One Out Cross Validation (LOOCV). Ordinarily, the “one” in LOOCV refers to a single observation, which is held out from the data set and used for validation after the model is trained on the remaining data. This process is then repeated for every data observation. Given the correlations within a program, we have redefined the “one” to denote an entire Program. This is an appealing strategy for two reasons. First, this is the level at which the correlations exist, so omitting an entire Program is the only assured method for fully eliminating the correlations. Second, despite restructuring the data into Program Categories, principal cost estimating interest remains with the Program, so that is the appropriate level for assessing model performance. Thus, for validation purposes, the entire Program (not just the SubProgram) becomes the unit of analysis and the observation left out.

After removing a given Program from the data set, we train the model using the remaining data and use the omitted Program as the test set. Then we record how the model performed against that Program. We repeat this process for every Program in the data set. This results in 30 separate validations for the LCC model and 35 for the AUC model (the C/MH-53E program cannot be validated because it only has one valid LCC SAR), which are then amalgamated into a single summary of overall validated model performance. This is a particularly rigorous validation as no information regarding the program to be tested remains embedded in the model. Also note that the Program Category structure still applies. This means that when validating certain programs (particularly the large and medium maritime categories) very few programs remain in

the category to form the basis of the Program Categorization parameters (refer to Table 9 and Table 10). Nevertheless, the validated version of each model performs well.

Figure 25 shows the resultant validated performance of the macro-stochastic prognostic LCC model based on our tailored LOOCV technique. This performance reflects model-corrected LCC estimates for every program with at least two valid SAR-derived LCC estimates. The analogous results for the AUC version of the model can be seen in Figure 26. As one would expect, model performance has diminished relative to the non-validated version of the model, but it still remains significantly better than empirical performance. The mean magnitude error in the validated LCC macro-stochastic model is 2.1 times better than the empirical estimate; for the AUC model, the model is 2.6 times better.

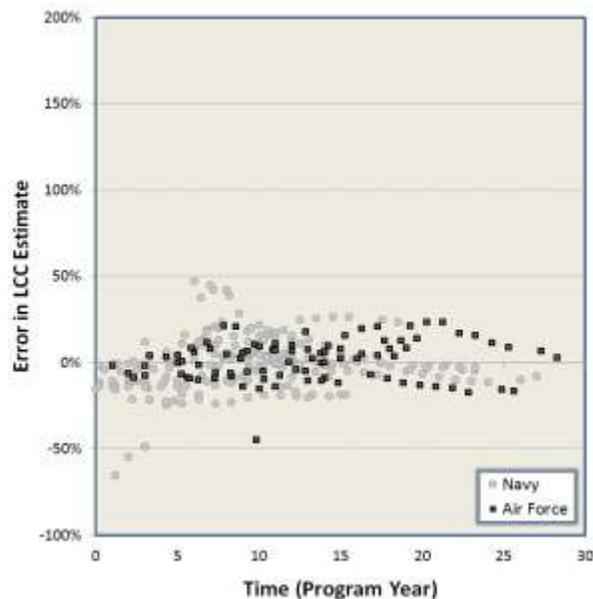


Figure 25: Error in LCC Estimate as a Function of Time (Validated Prognostic Macro-Stochastic Model)

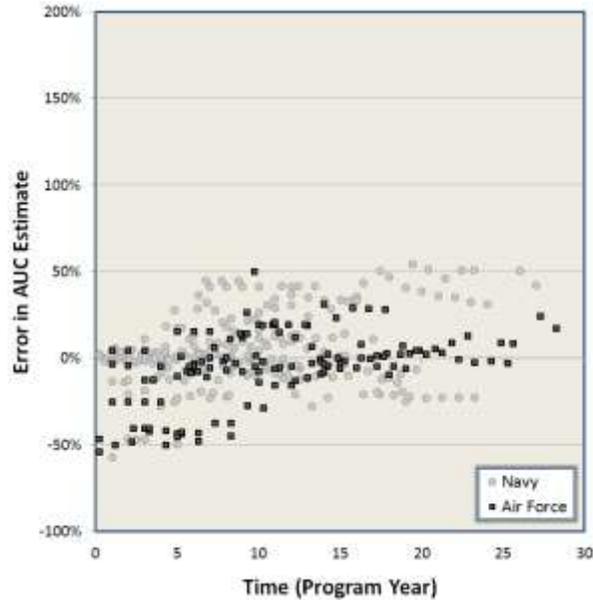


Figure 26: Error in AUC Estimate as a Function of Time (Validated Prognostic Macro-Stochastic Model)

Figure 27 compares the mean magnitude error per SAR in the empirical data to that of both the AUC and LCC validated models across all programs. For reference, performance of the non-validated version of each model is also shown. To ensure a fair comparison, all SARs omitted from the macro-stochastic models (i.e., initial SARs) were also omitted from the empirical data, which is why the mean magnitude errors for the empirical data are slightly different from those shown in Figure 19 and Figure 21.

Figure 28 shows another measure of model effectiveness, which is essentially “head-to-head” performance of each macro-stochastic model to the empirical estimates. This program-by-program comparison shows that the validated LCC model performs better (i.e., has an overall lower error across all the SARs of a given program) in 23 of 30 cases. The validated AUC model performs better for 31 of the 35 programs.

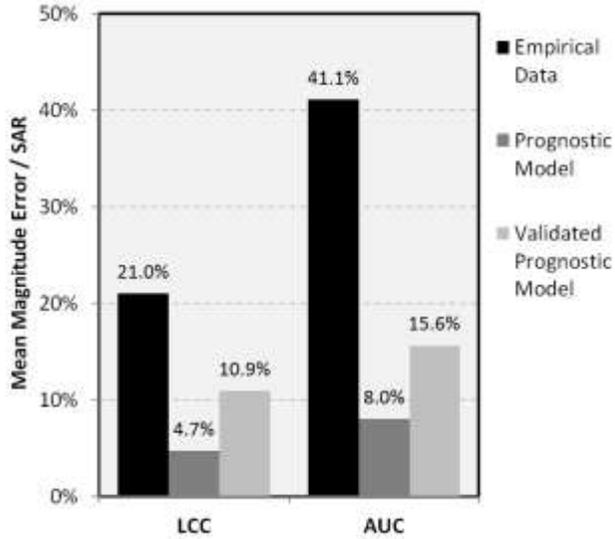


Figure 27: Model Performance as Measured by Mean Magnitude Error per SAR

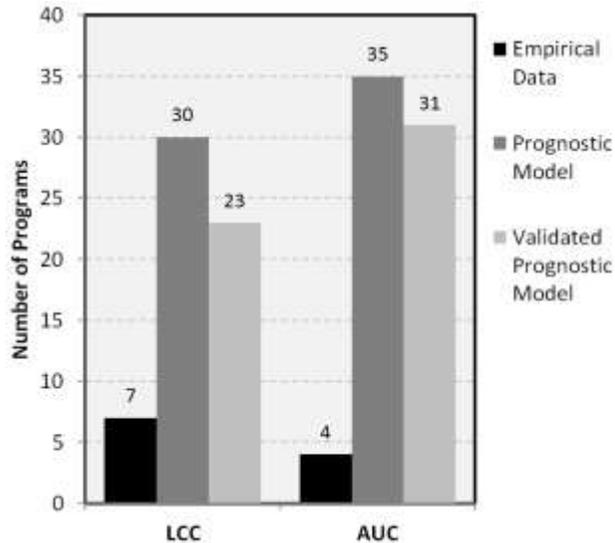


Figure 28: Model Performance as Measured by Number of Programs with Lower Overall Error

Issues and Concerns

Not surprisingly, the macro-stochastic model will sometimes predict an error estimate that overcorrects the program estimate, such that an underestimate becomes an overestimate, and vice versa. This is a natural consequence of that fact that the model is attempting to minimize variance around a “perfect” estimate (i.e., zero error), which means that it implicitly regards an overestimate as equally undesirable as an underestimate. This can (and does) create the following type of situation: The original estimate is 20 percent too high (or too low), but the model-corrected estimate is 10 percent too low (or too high). The question arises of whether we would be better off budgeting 20 percent too much or 10 percent too little. Although both underestimates

and overestimates are undesirable from a budget planning perspective, there are situations where one type of error may be preferred to the other. The macro-stochastic model could certainly be tailored to reflect such preferences through a zero error offset.

Somewhat related to the issue of overcorrections are the occasional instances where the model predicts an extremely large estimate error. While these predictions of massive errors—once applied to the original estimated cost—sometimes produce a more accurate estimate, they can also lead to unrealistic results, such as when the model predicts that the actual LCC or AUC has been underestimated by more than 100% (unless one wishes to advocate the possibility that Pentagon programs could turn a profit!). To avoid these types of nonsensical outcomes, we have embedded a threshold mechanism into the prognostic model such that the original estimates—regardless of what error the model predicts—are not corrected by more than a factor of two. In other words, the prediction of actual cost after correction for the model-predicted error will never be more than double the empirical estimate, nor less than half. In principle, the threshold could be much higher, but this level seemed appropriate from a practical standpoint. Although the program LCC and AUC estimates are sometimes inaccurate by a factor greater than two, corrections that require more than doubling or halving of the program estimate would—even if valid—likely be regarded with justifiable skepticism. Note that while thresholding did provide an improvement to overall model performance, the effect was marginal, and it was not implemented often. The threshold constraint affected the output in 26 of 709 cases (3.7 percent), and nearly half of these instances occurred on a single program (C-130J).

Another potential concern is long-term model reliability. As discussed in the previous section, the current iteration of both macro-stochastic models relies on official program estimates to produce its own estimate. This fact introduces an inherently recursive—and potentially unstable—element to longer-term model use. We know that senior defense leaders make key decisions based on the traditional program cost estimates, and that these estimates are often highly inaccurate. The nature of those decisions—and thus the ultimate trajectory of certain types of programs—may be substantively different if the decision-maker has access to more accurate cost estimates. For instance, programs that would otherwise be cancelled might instead be funded, and vice versa. This in turn, could create a negative feedback loop where cost estimate trajectories of certain program categories no longer conform to the patterns that characterize the programs that we have seen to date, thereby reducing the predictive capacity of the macro-stochastic model. Though highly speculative, this argument points to the need for continued refinement of the model as more data becomes available.

Perhaps the most significant barrier to macro-stochastic model implementation relates to the fact that it represents a fundamentally different approach to DoD cost estimating. In particular, it could be viewed in many respects as inherently non-transparent. In contrast to a traditional bottoms-up cost estimate, the specific drivers of the macro-stochastic cost estimate are not traceable, nor fully explainable. Users could be inclined to view this type of model as a “black-box,” where the output may in fact be probabilistically more accurate, but the internal workings are inscrutable. Nevertheless, the results presented here are compelling: Independent cross-validation verifies the

improvements in long-term DoD cost estimates that may be achieved by adjusting the cost estimates using the model-predicted error.

In practice, the most important caveat to using this model pertains to the Program Category structure. This construct was a strategy employed to transform the theoretical macro-stochastic model into a useful prediction tool. However, it is only a valid construct to the extent that current programs are representative of future programs, and those future programs really do “fit” into one of these established categories. Expanding on this point, the number of programs in Program Categories five (medium maritime) and six (large maritime) are fewer than we would prefer. By only having two to three constituent programs, we run the risk identified early on, i.e., that the defined Program Category may not be sufficiently representative of the next program to be assigned. Therefore, users of the current iteration of the macro-stochastic model may wish to be more wary when employing the model against these two Program Categories. This concern can be significantly mitigated only with the passage of time and the inclusion of more data.

Finally, a methodological note of caution. The specific model variables selected as well as the parameter estimates are based on the results of the previously completed characterization study [Ryan et al., 2012]. Therefore, we recommend that potential users familiarize themselves with that study in order to understand the potential issues and biases documented there before employing the macro-stochastic model. If the specific findings of the characterization study are not valid, then the specific variables and parameter values of this model are not likely to be valid either. Note, however, that concerns about the methodology of the characterization study would not be expected to weaken the underlying premise of macro-stochastic cost estimation; it would only affect its specific formulation.

Discussion

Although the validation results of the LCC and AUC macro-stochastic prognostic models yield sizeable errors, we find that overall accuracy for both models is significantly better than what was achieved in the original SAR estimates. The predicted LCC value from the validated macro-stochastic model had a mean absolute error of just under 11 percent compared to a 21 percent error in the historical program estimates. Given that the total LCC across all of the programs we evaluated was approximately \$800 billion, the model-predicted LCC estimates represent an improvement in estimating performance of about \$80 billion, or an average of \$2.6 billion per program. For the AUC estimates, the improvement was even greater. The mean magnitude error of the historical cost estimates was over 40 percent, while the model estimates had a mean magnitude error of less than 16 percent. Again, this translates to cost fidelity improvements measured in billions of dollars.

Improvements in the mean errors tell part of the story, but program-by-program performance is also important, and the macro-stochastic models performed well by this metric as well. If the macro-stochastic prognostic model presented here had been used to estimate LCC costs for every SAR of the programs in the LCC data set (aside from the

first), the model-based estimate would have had a lower overall error than the original estimate for 23 of 30 programs (77 percent). In the case of the AUC estimates, the model would have performed better for 31 of 35 programs (89 percent).

Not only can the original program estimate be improved dramatically using a macro-stochastic derived correction factor, but it can also be accomplished with minimal effort. The specific variables that feed each model are easily derived from data routinely available in the program's SARs. Program values observed for these variables can be transcribed into the model formula at any point after the program's second SAR, and a macro-stochastic estimate derived in just a few hours.

The fact that trending variables were found to be statistically significant predictors of LCC and AUC estimate errors is an intriguing result, but difficult to fully explain. Recall that the original estimates developed by the program had access to all of the same information (and far more) available to us in the SAR. Thus, any cost-impacting changes to the program should have been incorporated into the latest SAR estimate. It may be that the full cost implications of certain types of baseline changes are not fully understood until later in the program. Or it may be that certain types of historical program instability are likely to persist and/or permeate other elements of the program in ways that distort expected costs. In any case, the prominence of the trending variables make it tempting to conclude that change and cost instability tends to beget further change and cost instability. But this interpretation is too simplistic and frankly not warranted based on the data. Instead, our interpretation is more nuanced: Certain types and degrees of change in certain types of programs do tend to affect the accuracy of the current cost estimates in relatively predictable ways.

Perhaps of equal interest to the parameters included in the model are those that were omitted, i.e., those that never significantly contributed to model performance. Notable non-contributors were Joint (#3), APB-related variables (#8-11), Prime Contractor (#12), PAUC/APUC-related variables (#20-23), Requirement Changes (#41-46), Program Year (#1), Maturity (#7), and Expended (#18). The last three are perhaps the most surprising, as one would expect that variables that capture program age would be a good indicator of cost estimate accuracy (with the presumption that estimate accuracy improves as programs mature). Since they were not, this serves as additional evidence of the finding presented in the characterization, i.e., that LCC and O&S cost estimates for MDAPs are improving very little, if at all, over time [Ryan et al., 2012].

We believe that the LCC and AUC macro-stochastic cost models presented here are ready for trial use. However, it is important to understand a fundamental constraint on their intended implementation. Note that both the LCC and AUC models require as an input all of the subject program's respective cost estimates to date (see Table 11 and Table 13). This means, for one, that the output of the macro-stochastic model would generally not be suitable for internal program use. Unless perhaps implemented as a final validation check, awareness of the macro-stochastic output could influence the official SAR cost estimates, which in turn, would likely bias the output of the macro-stochastic model. This is because the macro-stochastic model implicitly relies on the

continuation of current cost estimating practices; any deviation from these practices, to include modifying the estimate based on the results of the macro-stochastic model, could fundamentally change the stochastic nature of this key input variable.

The dependence of the macro-stochastic cost model on the program's cost estimate also means that it is not meant to be used in lieu of existing program estimates. The traditional cost estimate may be perfectly accurate given the current baseline, which is an important input, per se, for senior decision-makers. The macro-stochastic model, on the other hand, is intended to be a complementary data point—it provides leadership the equivalent of a stochastic cost vector, i.e., a probabilistic indication of where program costs are likely to end up.

As a consequence of these implementation constraints, the authors envision that these models could be most effectively employed by cost validation entities outside the acquisition chain of command. An independent cost estimate is required for all MDAPs, which is provided by either the service cost agency or the Office of the Secretary of Defense, Cost Assessment and Program Evaluation (OSD/CAPE). Either of these entities may find the output of the macro-stochastic model highly useful when conducting their independent analyses. The Defense Acquisition Executive (DAE) and/or the Defense Acquisition Board (DAB) are also potential consumers, as they each require independent cost estimates as part of their review process, and the macro-stochastic model estimate could serve as an alternate source of realistic cost validation [GAO, 2009; DAU, 2012].

Another potential user of this type of model would be the service component acquisition portfolio manager, who is often required to manage the execution of several similar defense systems. The macro-stochastic model may be especially suitable in this case, as the portfolio manager is likely to be responsible for multiple systems from the same Program Category, and more accurate insights into overall portfolio cost commitments could be invaluable. Moreover, using the model for several contemporaneous programs would reduce the susceptibility of the predicted values being skewed by statistical outliers. Although the macro-stochastic model may certainly be applied to—and has been validated against—individual programs, one would expect it to perform more consistently when multiple programs are being simultaneously evaluated. This suspicion can be partially confirmed by examining aggregated program performance at the Program Category level. Although the results are not presented here due to space considerations, we did find that both models provided significantly improved estimates across every Program Category.

Despite the fact that DoD cost estimating practices have become increasingly sophisticated, the actual program cost estimates that are produced remain poor, at least when compared to the final, actual costs of the program. Our hypothesis is that this deficiency is largely due to the fact that current cost estimating techniques must assume a fixed program baseline. As a way around this unrealistic assumption, we have proposed a fundamentally different approach to cost estimating that attempts to capture this uncertainty by modeling the error in the program estimate as a random variable.

We found that the value of this variable is largely unique to a given program—and even a group of programs, to some extent—and could be predicted reasonably well through a relatively small number of top-level program summary indicators gleaned from the annual SARs.

The macro-stochastic model represents an intriguing option for vetting program estimates of Life Cycle Cost and Annual Unit O&S Cost. It not only appears to provide cost estimates that are significantly more accurate than those reported in the original SAR estimates, but the amount of effort needed to construct the estimates is minimal. Although the current version of the macro-stochastic model is not suited for replacing existing program cost estimates, the authors believe it could be extremely useful to independent costing entities outside the acquisition chain of command who are seeking a more realistic assessment of system value or program affordability.

5 APPLICATION: VALUING FLEXIBILITY IN SHIP MAINTENANCE

Ship maintenance programs play a critical role in meeting DoD readiness and cost saving objectives. The ship maintenance process alone accounts for billions of dollars in the U.S. Navy's annual budget. SHIPMAIN, and its derivatives, was an initiative designed to improve ship maintenance cost benefits performance within the Navy by standardizing processes in order to take advantage of learning curve cost savings. However, these process improvement initiatives have not yet realized the normal cost-reduction learning curve improvements for common maintenance items for a series of common platform ships. One explanation is that the initial instantiation of SHIPMAIN did not include two recommended technologies, 3D LST and CPLM, that were deemed necessary by the creators of SHIPMAIN, for ensuring the success of the new standardized approach (i.e., normal learning curve cost savings). Previous research [Ford, Housel, and Mun, 2011] indicated that adding these technologies may help SHIPMAIN, or its derivatives, to capture the potential saving. But the technologies have not been implemented to date in the ship maintenance processes.

However, Damen, a large ship building and service firm has incorporated similar technologies and is developing others to improve its operations. In addition, the Royal Dutch Navy performs all of its own ship maintenance in a single yard and operation and the potential benefits of similar technologies are extrapolated and compared with similar projections for the U.S. Navy ship maintenance processes in the current study. These organizations provide a source of relatively reliable data on operations that are comparable to those performed by the U.S. Navy.

5.1 PROBLEM DESCRIPTION

Previous research on the potential use of 3D LST and CPLM technology in U.S. Navy ship maintenance [e.g. Komoroski, 2005; Ford, Housel, and Mun, 2011] required the estimation of impacts on processes due to technology adoption. Changes such as reengineering ship maintenance processes, the sizes of reductions in cycle times, and workforce requirements are examples of model portions that required modelers to make assumptions about the potential impacts of these technologies in modeling projected results.

While the previous work has provided defensible estimates of potential improvements (in returns on investment: ROI) and cost savings, the validity and usefulness of these models has been limited by the lack of comparative data on ship maintenance processes, technology investments, and their potential impacts on performance.

To be valuable, the data source or sources for this work had to have several critical similarities with U.S. naval ship maintenance processes. The data source had to consider technological innovation and the adoption of advanced technologies to be an important part of their naval maintenance acquisition strategy. The data source or sources had to be large enough to support continuous ship maintenance operations because the intermittent stopping and restarting of operations would not be consistent with important assumptions of the modeling approach. Finally, the data source had to be accessible, willing to share the data, and allow us to obtain new data required for our modeling approach. Damen Industries and the Royal Dutch Navy (RDN) met most of these criteria and their subject matter experts (SMEs) were willing to share their data and experiences. The current work addresses the following questions:

- How are the Dutch using and preparing to adopt advanced technologies such as 3D LST and CPLM in ship building and maintenance?
- What are the potential changes in ROIs provided by the adoption of these advanced technologies?
- How do those potential returns compare with projected estimates of returns on technology adoption of 3D LST + CPLM in the U.S. Navy?

5.2 BACKGROUND AND METHODOLOGY

The traditional ROI equation is typically expressed as: (Revenue-Investment)/Investment, which represents the productivity ratio of output (i.e., Revenue in ROI ÷ Input or Investment Cost in ROI). Accomplishing this analysis in a nonprofit environment presents challenges because there is no actual revenue generated. Cost savings from reductions in manpower requirements (i.e., time allocated to employee workload for various tasks) is available to provide the impact on the denominator of the ship maintenance efforts. However, the Knowledge Value Added (KVA) methodology also allows for generation of a quantifiable surrogate for revenue in the form of common units of output described in terms of units of learning time.⁵ Specifically, the KVA methodology allowed the study team to quantify the knowledge embedded in the new processes to use in generating common units of output estimates.

The KVA analysis provided the basic ROI estimates critical in forecasting the future value of various automation options within an optimized portfolio over time using the Integrated Risk Management (IRM) framework and supporting tool set.

The research team collected data on Dutch ship operations as described below and used it to build three types of computer simulation models of ship maintenance and technology adoption: knowledge value added (KVA) models of return on technology investments in those operations, system dynamics models (based on the KVA preliminary ROI results) of ship maintenance operations, and integrated risk management models of implementation plans for technology adoption.

⁵ KVA can provide other means for describing outputs in common units, such as lines of code (controlling for complexity per line of code) and process instructions (controlling for complexity per instruction) as well as other means.

The results were then analyzed and compared with the prior study modeling results of U.S. Navy ship maintenance and technology adoption. In what follows, we review the three approaches to projecting the potential cost benefits of adopting the technologies beginning with a general review of the KVA, SD, and IRM approaches. This is followed by the projected results from applying these approaches to assess the impacts of the two technologies. A comparison of the Dutch and U.S. naval maintenance results is provided followed by the results of the IRM forecasts.

Knowledge Value Added Knowledge value added (KVA) measures the value provided by human capital assets and IT assets by an organization, process, or function at the subprocess level. It monetizes the outputs of all assets, including intangible knowledge assets. Capturing the value embedded in an organization’s core processes, employees, and IT enables the actual cost and revenue of a product or service to be calculated.

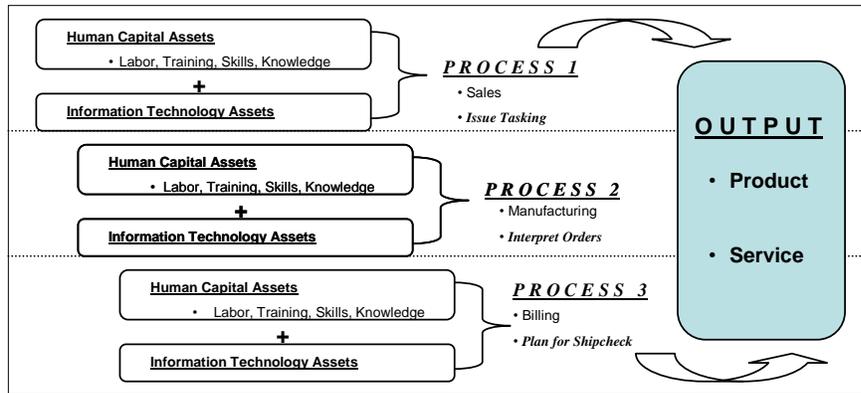


Figure 29: Measuring Output

Traditional Accounting		KVA Process Costing		
Explains what was spent	Compensation	\$5,000	Review Task	\$1,000
	Benefits/OT	1,000	Determine Op	1,000
	Supplies/Materials	2,000	Input Search Function	2,500
	Rent/Leases	1,000	Search/Collection	1,000
	Depreciation	1,500	Target Data Acq	1,000
	Admin. And Other	900	Target Data Processing	2,000
	Total	\$11,400	Format Report	600
			Quality Control Report	700
			Transmit Report	1,600
		Total	\$11,400	
		Explains how it was spent		

Figure 30: Comparison of Traditional Accounting versus Process-Based Costing

Total value is captured in two key metrics: return on investment (ROI) and return on knowledge (ROK). While ROI is the traditional financial ratio, ROK identifies how a specific process converts existing knowledge into producing outputs so decision makers can quantify costs and measure value derived from investments in human capital assets.

A higher ROK signifies better utilization of knowledge assets. If IT investments do not improve the ROK value of a given process, steps must be taken to improve that process's function and performance.

Metric	Description	Type	Calculation
Return on Knowledge (ROK)	Basic productivity, cash-flow ratio	Subcorporate, process-level performance ratio	(Outputs-Benefits in Common Units) Cost to Produce Output
Return on Investment (ROI)	Same as ROI at the sub-corporate, process level	Traditional investment finance ratio	(Revenue-Investment Cost) Investment cost

Table 15: KVA Metrics

The goal is to determine which core processes provide the highest ROIs and ROKs, and to make suggested process improvements based on the results. In the current work KVA is used to measure the benefits of technology adoption in Dutch ship maintenance. This analysis provides a means to check the reliability of the prior study estimates of the potential ROI core process improvements from using CPLM + 3D LST in ship maintenance core processes in the U.S. Navy yards.

System Dynamics

The system dynamics methodology applies a control theory perspective to the design and management of complex human systems. System dynamics combines servo-mechanism thinking with computer simulation to create insights about the development and operation of these systems. It is one of several established and successful approaches to systems analysis and design [Flood and Jackson, 1991; Lane and Jackson, 1995; Jackson, 2003]. Forrester [1961] develops the methodology's philosophy, and Sterman [2000] specifies the modeling process with examples and describes numerous applications. System dynamics is used to build causal (vs. correlation) based models that reflect the components and interactions that drive behavior and performance. The methodology has been extensively used to explain, design, manage, and thereby improve the performance of many types of systems, including development projects. The system dynamics perspective focuses on how the internal structure of a system impacts system and managerial behavior and thereby performance over time. The approach is unique in its integrated use of stocks and flows, causal feedback, and time delays to model and explain processes, resources, information, and management policies. Stocks represent accumulations or backlogs of work, people, information, or other portions of the system that change over time. Flows represent the movement of those commodities into, between, and out of stocks. The methodology's ability to model many diverse system components (e.g., work, people, money, value), processes (e.g., design, technology development, production, operations, quality assurance), and managerial decision-making and actions (e.g., forecasting, resource allocation) makes system dynamics useful for modeling and investigating military operations, the design of materiel, and acquisition.

When applied to acquisition programs, system dynamics has focused on how performance evolves in response to interactions among development strategy (e.g., evolutionary development vs. traditional), managerial decision-making (e.g., scope developed in specific blocks), and development processes (e.g., concurrence). System dynamics is appropriate for modeling acquisition because of its ability to explicitly model critical aspects of development projects. System dynamics models of development projects are purposefully simple relative to actual practice to expose the relationships between causal structures and the behavior and performance that they create. Therefore, although many processes and features of system design and participants interact to determine performance, only those that describe features related to the topic of study are included in system dynamics models.

System dynamics has been successfully applied to a variety of development and project management issues, including rework [Cooper, 1993a,b,c; Cooper & Mullen, 1993], the prediction and discovery of failures in project fast-track implementation [Ford & Sterman, 2003b], poor schedule performance [Abdel-Hamid, 1988], tipping point structures in projects [Taylor and Ford, 2008, 2006], contingency management [Ford 2002], resource allocation [Joglekar and Ford 2005; Lee, Ford and Joglekar 2007], and the impacts of changes [Rodriguez & Williams, 1997; Cooper, 1980] and concealing rework requirements [Ford & Sterman, 2003a] on project performance. See Lyneis and Ford [2007] for a review of the application of system dynamics to projects and project management.

System dynamics has also been applied to military systems, including planning and strategy [Melhuish, Pioch, et al. 2009, Bakken and Vamraak 2003, Duczynski 2000, McLucas, Lyell, et al. 2006], workforce management [Bell and Liphard, 1978], technology [Bakken, 2004], command and control [Bakken and Gilljam, 2003; Bakken, Gilljam et al., 2004], operations [Bakken, Ruud et al., 2004; Coyle and Gardiner, 1991], logistics [Watts and Wolstenholme, 1990], acquisition [Ford, Housel, and Dillard, 2010; Ford, Housel, and Mun, 2011; Ford and Dillard, 2009a,b, 2008; Bartolomei 2001; Homer and Somers, 1988] and large system programs [Cooper, 1994; Lyneis, Cooper, and Els, 2001]. Coyle [1996] provides a survey of applications of system dynamics to military issues. In the current work system dynamics is used to model ship maintenance operations to generate realistic forecasts of performance. The recent work by Ford, Housel, and Mun [2011] and Ford, Housel, and Dillard [2010] is particularly relevant to the current research because it successfully demonstrated the ability of system dynamics to be integrated with KVA analysis for DoD acquisition.

Integrated Risk Management

Integrated Risk Management (IRM) is an 8-step quantitative software-based modeling approach for the objective quantification of risk (cost, schedule, technical), flexibility, strategy, and decision analysis. The method can be applied to increase the flexibility for program management; resource portfolio allocation; return on investment to the military (maximizing expected military value and objective value quantification of nonrevenue government projects); analysis of alternatives or strategic flexibility options; capability analysis; prediction modeling; and general decision analytics. The method and toolset provide the flexibility to consider hundreds of alternatives with

budget and schedule uncertainty, and provide ways to help the decision maker maximize capability and readiness at the lowest cost. This methodology is particularly amenable to resource reallocation and has been taught and applied by the authors for the past 10 years at over 100 multinational corporations and over 30 projects at the U.S. Department of Defense (DoD).

IRM provides a structured approach that will yield flexibility through via a rapid, credible, repeatable, scalable, and defensible analysis of cost savings and total cost of ownership while ensuring that vital military capability, i.e., value, are not lost in the process. The IRM + KVA+SD methods do this by estimating the value of a system or process in a common and objective way across various alternatives and providing the return on investment (ROI) of each in ways that are both comparable and rigorous. These ROI estimates across the portfolio of alternatives provide the inputs necessary to predict the value of various flexibility options. IRM incorporates risks, uncertainties, budget constraints, implementation, and life-cycle costs, reallocation options, and total ownership costs in providing a defensible analysis describing management options for the path forward. This approach identifies risky projects and programs while projecting immediate and future cost savings, total life-cycle costs, flexible alternatives, critical success factors, strategic options for optimal implementation paths/decisions, and portfolio optimization. Its employment presents ways for identifying the potential for cost overruns and schedule delays and enables proactive measures to mitigate those risks. IRM provides an optimized portfolio of capability options or implementation options for ship maintenance while maintaining the value of strategic flexibility.

In this project, IRM provides a way to differentiate among various flexibility alternatives for implementation of 3D Pdf and Logistics from a CPLM suite with respect to ship maintenance processes, and to postulate where the greatest benefit that could be achieved for the available investment from within the portfolio of alternatives. As a strategy is formed and a plan developed for its implementation, the toolset provides for the flexibility constraints of risk factors, such as schedule and technical uncertainty, and allows for continuous updating and evaluation by the Program Manager to understand where these risks affect flexibility and to make informed decisions accordingly.

A Basic Formulation

Regardless of the specific framework employed for decision-making, there are a couple of mandatory elements as part of any conceivable valuation approach. One is the cost of the investment, and the other is the return on that investment. With respect to decisions related to design flexibility, this same basic approach applies, but must be tweaked somewhat. To begin with, an initial investment is required to implement a more flexible de-sign, which we refer to as the investment cost. Complicating our formulation, however, is the fact that the return on that investment is not directly linked to the value of the flexibility. *A flexible design does not have intrinsic value*; instead, it is the concomitant capability associated with that flexibility that has value (to digress the argument further, it is really the military outcome that can be achieved via a given capability that has value). Therefore, the value component of the decision formula is the probabilistic benefit that a particular capability may be realized with fewer resources (e.g., time or money) than had we not chosen to make that initial flexibility investment.

In addition, though, our notion of flexibility may require some additional cost later to actually implement the capability, which is also dependent on the probability that the capability will be affected. Finally, the very act of investing in flexibility (e.g., adding brackets to a tank chassis) may adversely affect other performance attributes (e.g., speed, maintainability, etc.) such that we need to include another value term to potentially decrement lost value associated with the flexibility investment.

Momentarily setting aside the aforementioned concerns related to NPV, we could conceptually (and neglecting time-value of money considerations) formulate the decision to invest in flexibility as follows—

$$NPV = V(cap) * \mathbb{P}(cap) - C(flex_{inv}) - V(flex_{inv}) - C(cap_{imp}) * \mathbb{P}(cap) \quad (1)$$

$V(cap)$ = Value of capability

$\mathbb{P}(cap)$ = Probability capability will be needed

$C(flex_{inv})$ = Direct cost of initial investment in flex design option

$V(flex_{inv})$ = Reduction in value by investing in flexibility, i.e., indirect costs

$C(cap_{imp})$ = Cost to implement capability

Note that the $V(cap)$ term—much like the $V(flex_{inv})$ term—needs to include any value decrements associated with adverse consequences to other performance attributes that are incurred by implementing the capability (e.g., adding armor may make a combat vehicle more survivable, but will likely reduce its speed and maintainability).

Assuming that all terms are commensurable measures (i.e., monetized), this formulation indicates that if the NPV is greater than zero, the investment in the flexible designs option is worth pursuing. While constructing equation (1) is relatively straightforward, assigning values to each of these terms is where the challenge arises. The two cost terms, while seldom trivial, are likely the most easily obtained, as we have ample experience in estimating the cost of engineering solutions. Establishing a valid probability term is more difficult as it is inherently linked to uncertainty; however, it at least can be rationally estimated. The real dilemma is associated with the value terms, which are extremely difficult—if not impossible—to meaningfully quantify in the context of defense acquisition. This is the crux of the problem.

Value of Capability

In order to make meaningful value judgments, we must establish a utility function that will quantify the value of capability in some ratio-level comparable units. While this is relatively routine for profit-driven commercial systems, it will necessarily be more challenging for military systems, as the utility function will almost certainly not involve a monetizable metric like earnings. Instead, for example, we would need to somehow devise a function (or more likely, a series of functions) for determining the utility of an extremely wide range of military capabilities, such as being able to resist jamming or increase an airplane's top speed from Mach 2.0 to Mach 2.5.

In principle, there is a solution. Under the neoclassic economic definition of value, an item's value can be established by determining a customer's *willingness to pay*. Thus, we can surmise that the value of a particular military capability can be determined by ascertaining the maximum amount the government is willing to give up (of some measurable resource) to obtain the capability (i.e., *the value of a given capability to the government = the maximum cost the government is willing to pay for the capability*). The devil is in the details, however. Assigning a numerical value to the right side of this equation is a daunting endeavor. The most obvious approach would be to use the dollar amount budgeted by the government. But this is problematic for a multitude of reasons. Consider that the actual system cost may include a number of other scarce resources (e.g., time, critical skills, and facilities) that are not captured in the government budget. Technically, economic cost includes the loss of opportunities as well, so we would also need to account for the cost of losing or vitiating other capabilities by virtue of the fact that we are committing resources to this capability. Once again, though, we would face the dilemma of assigning a value to a capability, with only budgets to guide us, so our original problem is further complicated because it is now recursive.

In addition, even if we were to accept that budgeted costs will be adequate, there is no reason to believe this represents the *maximum* cost the government is willing to pay. Firstly, the government may, in principle, be willing to budget more for a particular capability, but has reason to believe that a lower amount will suffice. The problem is that the government generally establishes its program budgets based on expected actual costs, not the perceived value of the program or resulting capability set. Secondly, budget allocation processes are notoriously volatile, subject to any number of political and bureaucratic vagaries that have nothing to do with the merits of a particular program or capability. Thus, one year's total budget allocation for a given program may be substantially different from the next year's allocation for the same program, though there was no change in its perceived value. And, of course, the value of capability is a function of time, which really goes to the crux of the problem!

For the sake of argument, let's assume that we can tolerate a lower fidelity estimate, and we can convince ourselves that the budgeted costs represent all costs with sufficient accuracy, and that these costs also represent the maximum cost the government is willing to pay. Unfortunately, there is still another practical obstacle to establishing a specific dollar amount corresponding to the value of a capability. The fact is that defense budgets can rarely be traced so cleanly to desired system capabilities, and certainly not at the levels of precision that would be required to make this a viable approach for detailed design decisions. Imagine a \$10.0 billion program to develop an aircraft with various capabilities related to range, reliability, speed, maneuverability, lethality, etc. The notion that we could indicate exactly how much of that \$10.0 billion investment the government is willing to spend to achieve a speed of Mach 2.5 may not be feasible, and is certainly not the basis on which government program budgets are allocated or managed today.

Clearly, using budgeting information to infer the value of capabilities is full of pitfalls. An alternative approach would be to query system end users directly. The most obvious draw-back to this approach is the inherent subjectivity; even within a single user community, different users will perceive the value of a given capability differently. This would drive a comprehensive solicitation of all potential users, in combination with some (to be specified) means of aggregating and reconciling those inputs. In addition, each user's value input would need to be provided within the context of a resource-constrained environment; else value assignments would lose relative meaning. Another potential problem stems from the fact that the end-user of the capability who is most able to appreciate its value is, ironically, the least likely to have any experience with budgeting and finance, and thus may not even be able to translate the mission value into monetary terms. Similarly, the user group may simply have no direct insight into the costs associated with the capabilities it has access to due to the nature of service/capability relationships among defense organizations.

Even more fundamentally, flexible design options may have no practical meaning to the user. Since we are inherently interested in the value of *potential* capabilities—vice *validated* capabilities—the user may be unwilling and/or unable to assign any value to the capability at all. For if the potential capabilities were valued to any level of significance, and then it likely would have already been translated into a valid system requirement! Finally, many potentially flexible design decisions over the course of a program's life (particularly those that pertain to *process* flexibility) have little to no impact on end capabilities.

It can be argued that by attempting to employ both the willingness to pay and the user query methods, it may be possible to obtain a dollar range that could serve to at least bracket the value terms. However, it's not clear we could assign valid confidence values to this range or that the calculated range would be narrow enough to have practical utility. Therefore, given the difficulty of establishing the value of military capabilities, we need a more *flexible* approach to determine the value of flexibility.

Comparative Data Collection

Several sources of comparative data were utilized, including a Dutch shipbuilder (Damen) and the Royal Dutch Navy (RDN). Data on the use of technology in Dutch fleet maintenance was collected by two primary methods: 1) in-person interviews and meetings with managers of the leading corporation in Dutch shipbuilding industry (Damen) and officers and civilian employees of the Royal Dutch Navy, and 2) tours of three Dutch shipbuilding and maintenance facilities.

Meetings, semi-structured interviews, and extended discussions were held with six managers of Damen Industries and the Royal Dutch Navy in three locations over three days. At these meetings Damen managers made presentations on Damen's operations, uses of technologies, their investigations of specific technologies for potential development and adoption (including 3D LST and CPLM software), Damen's Integrated Logistics System, and information technology products under development for use in ship maintenance. Separately, a meeting and semi-structured interview was conducted

with the two Royal Dutch Navy officers responsible for ship maintenance at the RDN shipyard at Nieuwe Haven in Den Helder.

Damen's Use of Technology

The Damen Shipyards Group (<http://www.damen.nl/>) is a large Dutch shipbuilding firm with worldwide operations (11 shipyards with five outside The Netherlands). The firm was started in 1922 by Jan and Rien Damen. The firm grew substantially after Kommer Damen (the current owner) bought the firm in 1969 and introduced modular and standardized shipbuilding to the industry. The firm now employs over 6,000 persons and builds an average of 150 vessels per year. The firm obtained Damen Schelde, which focuses exclusively on naval ship design, building, and maintenance relatively recently (in 2000). Damen Schelde manufactures an average of one to two ships per year, employs about 550 people, and performs about 210million Euro/year. Damen Schelde acts as the prime contractor and integrator on its shipbuilding projects, utilizing many subcontractors. Although Damen Schelde provides ship maintenance services to its international (i.e. not Dutch) customers, it does not provide any ship maintenance services for the Royal Dutch Navy.

Damen Schelde has used an a component of a CPLM suite, i.e., Integrated Logistics System (ILS), since 2002 to manage the shipbuilding process from project initiation through the development of a logistics plan for customers. The ILS is the plan for the development of a ship and includes ship design, production, QAQC (quality assurance, quality control), training of ship operators, and coordination with customers. The ILS does not include service contracts or lifecycle costs due to the difficulty of forecasting those costs.

The focus of the ILS is to provide maximum ship operational availability, reliability, and maintainability. It does this partially by using a single point of contact within Damen throughout the project who manages an interdisciplinary team (e.g. engineering, work preparation, procurement, service). Damen Schelde currently uses a variety of information technologies to facilitate their Integrated Logistic System (ILS) approach to shipbuilding and is constantly investigating new technologies that may improve their design and manufacturing. Of particular relevance to the current work, Damen Schelde uses four separate software products to manage their shipbuilding: An advanced three dimensional CADD program for design, a CPLM product as a database for ship components⁶, an Enterprise Resource Planning (ERP) system, and a software tool for scheduling. The latter three of these systems are connected to users with a Project Information Portal developed by Damen Schelde. The informant reported that the reason that Damen developed the portal was that the CPLM product did not include adequate user interfaces. Damen Schelde has investigated and is currently investigating other technologies for potential adoption. Four technologies were described and discussed:

- 3D Laser Scanning Technology (3D LST): This technology was investigated but was assessed to currently be too immature for adoption by Damen Schelde. The

⁶ Damen Schelde did not purchase the integrated logistic package available from Siemens.

investigation included a discussion of the potential use to scan engine rooms, for floor flattening, and the use of the technology in the automobile industry. The use of 360 degree photography (often used in conjunction with 3D LST) was considered by Damen Schelde as a potential tool for training. See Komoroski [2005] for more details on 3D LST.

- **3D PDF files:** Three dimensional animated “movies” of ship building can be created in a PDF format (by Adobe Acrobat®) and sent to shipyards for use in the field by craftsmen who view the file on an electronic reader (e.g. iPad®). The files would replace flat drawings for use in construction. The file visually communicates the sequence of building (or maintenance) operations and components and operations can have notes attached to them that provide additional information (e.g. part numbers, warnings of special issues). The ability to animate these files allow engineers to visually show craftsmen sequences of operations, routes of access and egress for Line Replaceable Units (LRU⁷), and other information that is difficult or impossible to show with traditional static two dimensional drawings. The use of this technology shifts the understanding of the design intension from the designers (in the Netherlands) to the ship building yard (typically in other countries around the world). The use of visual information (the animation of steps) is expected to greatly improve communication across languages, since many of the craftsmen in Damen’s shipyards do not read English well. Damen considers improvements in information content communicated to be the primary benefit of this system (vs. cost savings). Damen Schelde is very optimistic about the potential for this technology to improve its operations and is actively working on developing it (e.g. selecting software, addressing the importing of the 3D design drawings). Generating the animated files and adding the building steps to the design files is expected to be relatively easy once the system has been developed.
- **SIGMA Shipbuilding Strategy:** This is a standardized process for creating a ship that spans from design through materials procurement, production, and testing of a ship. The key feature of the strategy is the use of modular ship sizes and systems that can be easily adapted to specific customer needs. For example, Damen Schelde has disaggregated an entire ship into five standardized modules (e.g. fore, midship, aft) with major systems located in specific sections. Each module is considered a subproject. As an example of an advantage provided by the strategy, the modules and their interfaces are designed such that the ship can be made longer by adding an additional mid section⁸.
- **Radio Frequency Identification (RFID):** This established technology is being considered for use to improve Damen supply chain management. Primary benefits are believed to be improved value of information and a reduction in durations for getting information into Damen databases (e.g. warehouse

⁷ Line Replaceable Unit is a commonly used term in manufactured devices for any modular component that is designed to be interchangeable.

⁸ This portion of the SIGMA strategy applies the Boeing strategy for the design and production of the 737 that has different lengths to shipbuilding.

contents, components on specific ships). Both passive tags and active tags are being considered.

Damen Services also develops advanced technologies for use by Damen Enterprises. Damen Services focuses on providing ongoing maintenance parts and services to Damen customers after a ship has been designed, built, and delivered, but also provides other services such as civil works (e.g. wharves and storage facilities).

The Maintenance and Spares department maintains information on ship configuration (using an ERP system), parts inventories, spare parts packages, maintenance management systems, and provides information technology support for Damen. Damen Services has grown rapidly, from 50 employees in 2008 to 250 employees in 2012. Their primary objective for customers is to reduce costs and increase operational availability. They are developing a web portal for clients that will allow clients access to Damen-held data on each of the customer's ships down to the individual component level. This will partially be accomplished with a Work Breakdown System (WBS) that disaggregates a ship or system into product parts (e.g. engine, bilge pump) and a Functional Breakdown System that disaggregates the ship into functions (e.g. port propulsion) that are met with a product part (in the WBS) and have an associated maintenance schedule, monitoring measurements and frequency, parts documentation, etc. The WBS has three levels: Subsystems (e.g. propulsion, hoisting) with a typical ship having 20-70 subsystems, Level 2-parts (e.g. pump, shaft) with about 1,000 per ship, and Level 3 parts (e.g. bolt, flange) with 70,000-80,000 per ship.

This system will be linked with an on-line parts ordering portal so that customers can order parts from Damen (similar to Amazon's on-line selling of books, etc.). Damen Services plans to use this information captured through this system (e.g. frequency of ordering of specific components) to develop maintenance optimization information. Damen Services envisions three types of maintenance: corrective maintenance (after the component needs work), preventative maintenance (based on forecasts of maintenance needs), and condition-based maintenance (based on actual conditions of components). Condition-based maintenance is an optimized version of preventative based maintenance that is currently under development. It requires sensors to collect data on component conditions that will be used to generate condition assessments.

5.3 DATA COLLECTION RESULTS - ROYAL DUTCH NAVY (RDN) FLEET MAINTENANCE

Data collection directly from the RDN was particularly valuable for at least two reasons. First, as the navy of a sovereign country with objectives that are similar those of the United States the objectives and issues of the RDN are more likely to match those of the U.S. Navy than those of some other nations. Data collection supported this assumption. For example, technology leadership, interoperability, and reliability in meeting operational needs are paramount to the RDN, and the RDN has recently experienced, and expects to continue to experience, reductions in budgets just as is the

case with the U.S. Navy. The Dutch navy continues to face budget cuts and increasing technology needs, is currently in reorganization to reduce total workforce (internal to the navy and civilian naval workforce) by 20%, and is transferring from their legacy information systems to an integrated ERP system for maintenance operations. Second, the RDN performs all of the maintenance on its fleet, thereby making it the primary data source concerning RDN fleet maintenance process performance.

The interviews with the two RDN officers in the Naval Maintenance and Service Agency provided a general introduction to the issues faced by the Dutch navy in building and maintaining its fleet. The RDN addresses its challenges by means similar to those used by the U.S. Navy, such as waiting for technology to mature (technology readiness level (TRL) ≥ 7 before adoption) and incremental capability increases based on budgets. Noticeably different, both the RDN and Damen described the critical role and standard Dutch practice of adjusting requirements to meet budgets in shipbuilding. The RDN is facing increasing pressure to control life-cycle costs in its fleet, which are largely driven by personnel and fuel. This has led them to approve significantly stricter operations manning requirements for ship design (i.e. lower maximum shipboard personnel), which has driven Damen to increase the use of automation in their ship designs.

The primary informant on RDN fleet maintenance operations provided a diagram of those operations (Figure 31) and a written description of each of the steps identified in the diagram.

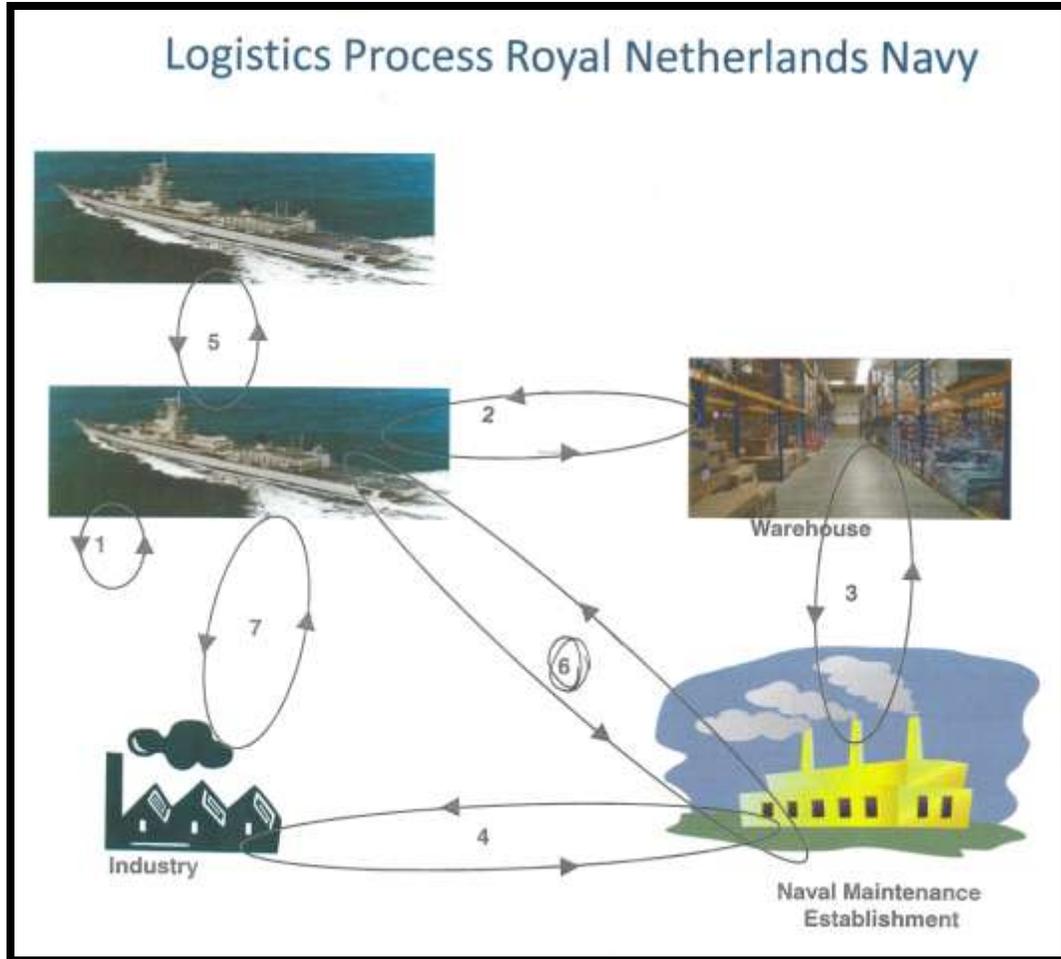


Figure 31: Diagram of Royal Dutch Navy Fleet maintenance process

The process steps shown in Figure 31 were described in writing by the informant with the following list⁹. In the list the abbreviation “LRU” stands for “Line Replaceable Unit”, a commonly used term in manufactured devices for any modular component that is designed to be interchangeable. MIL-PRF-49506, Notice 1 of 18 (Jan, 2005) “Performance Spec for Logistics Management Information” provides the following definition.

An LRU is an essential support item which is removed and replaced at the field level to restore the end item to an operational ready condition. Conversely, a non-LRU is a part, component, or assembly used in the repair of an LRU, when the LRU has failed and has been removed from the end item for repair.

Logistic Process Royal Netherlands Navy

⁹ Process step descriptions have been transcribed exactly as provided in English by the RDN, including uncommon English grammar and spelling.

- In case LRU fails the on-board personnel will replace this LRU by a spare (on-board) (OLM)
- The defect LRU will be sent to the warehouse, and a “new” LRU will be sent to the ship
- The defect LRU will be sent to the Naval Maintenance Establishment (NME) for repair. After the LRU is repaired it will be sent to the warehouse again “as good as new” (DLM)
- If the NME needs parts to repair an LRU, it will be extracted from the industry, when the NME is not able to repair this LRU it can be sent to the manufacturer. Also manpower can be hired to fix problems
- If spare is not available, sometimes it will be cannibalized from another ship
- If the on-board personnel is not able to fix the problem by themselves (due to the complexity of the failure) assistance from the NME is needed (ILM)
- If the problem is too complex for the NME also, the industry can be hired to solve this problem

The seven process steps were elaborated on by the informant. Specifically:

Step 1: Performed on board, for example to provide operational maintenance of weapons systems

Step 2: Purely a transit operation that requires only a truck driver (if ship is in port).

Step 4: Requires DLM level of training

Step 5: Requires OLM level of training

Step 6: Requires ILM level of training (=LTS+MTS+10-25 days of training)

Step 7: Requires DLM level of training

The abbreviations DLM, OLM, and ILM refer to Dutch terms for training levels. Fleet maintenance for the RDN requires a minimum of completion of education at a Lower Technical School (LTS) and a Middle/Intermediate Technical School (MTS). The Lower Technical School is typically attended between ages 12-16 and the Middle Technical School is typically attended between ages 16 and 21. After the completion of LTS and MTS, future RDN ship fleet maintenance personnel must complete at least one of three other forms of training.

OLM – 5-10 days of training

ILM – 10 -25 days of training

DLM – 15-35 days of training

Manufacturer training can take either of two alternative training paths:

LTS then MTS then either OLM or ILM or DLM

LTS then MTS then OLM then ILM then DLM

The information above was augmented by a tour of the Naval Maintenance Establishment (NME) maintenance and repair facilities. The NME provides essentially all maintenance and repair for the RDN fleet and the NME facilities can, and do, perform the required work on RDN ship components. This requires a comprehensive set of equipment and skilled personnel that cover the wide range of materials and

components. Examples of testing, maintenance, and repair capabilities seen on the tour include, but are not limited to, the repair of a wide variety of weapons systems, radar systems testing and repair, design and manufacturing of printed circuit boards, and the manufacturing of optical lens for submarine periscopes. NME holds 220,000 total items in the warehouse valued at about 500million Euros. An average Dutch naval ship contains about 60,000 components.

System Dynamics Model Structure

The system dynamics model simulates the movement of LRU among the various locations where they are used, stored, or repaired. These accumulations are referred to as stocks [Sterman, 2000]. Each flow of LRUs between two stocks represents the processing rate of one of the process steps in a Knowledge Value Added model. A simplified diagram of the stocks and flows of the model are shown in Figure 32. Boxes represent stocks, or accumulations of LRU. Each stock in Figure 32 represents a location in Figure 31, plus on-board LRU storage as a separate LRU accumulation. Arrows with valve symbols in Figure 32 represent the movement of LRUs between stocks. Numbers in parenthesis in the titles of flows represent the process steps shown in Figure 31 (ovals with arrows) and the KVA model process steps (described later).

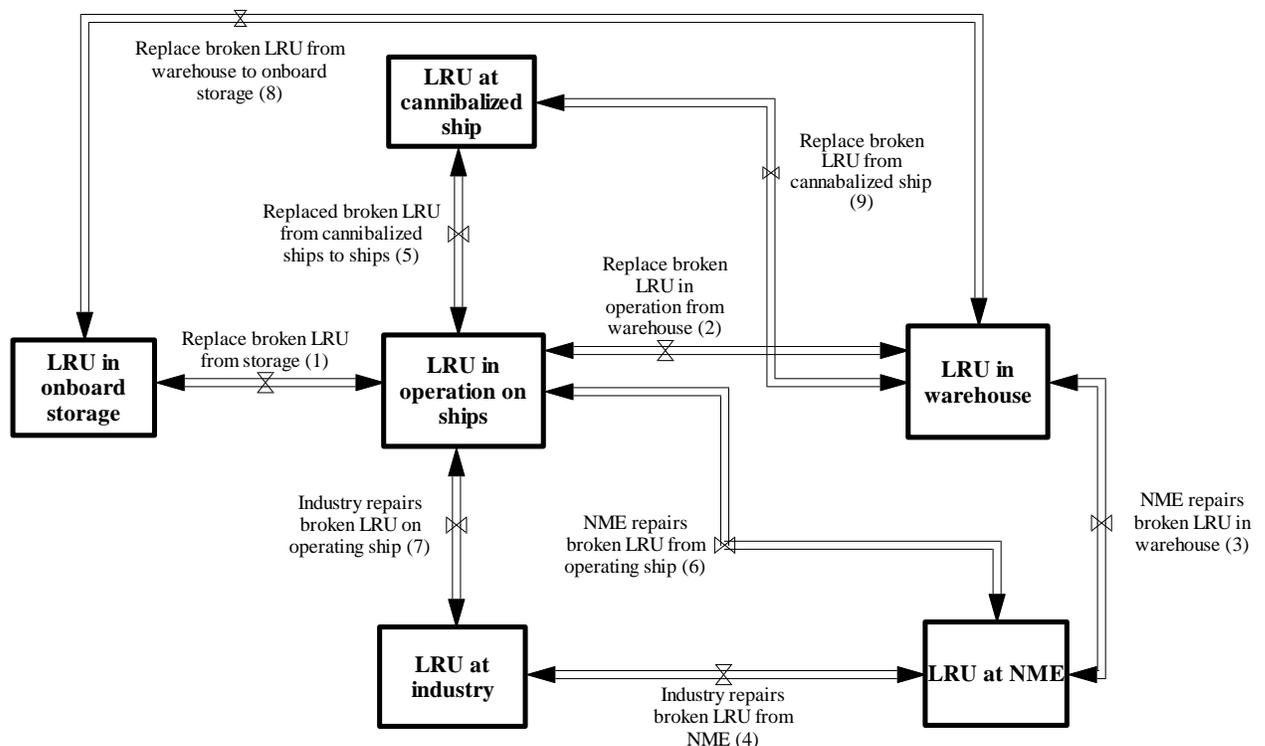


Figure 32: Royal Dutch Navy Ship Maintenance: Stocks and Flows of the System Dynamics Model

The sizes of the flows in the system dynamics model describe the rate of movement of LRUs among the stocks. Therefore the simulated flows in the system dynamics model become direct inputs to the “Times Processed per year” portion of the KVA models. Flow

rates were modeled to reflect the sequence of processes in operations. For example, in normal operations the replacement of a broken LRU in an operating ship with one from the ship's on-board storage ("Replace broken LRU from storage (1)" on left of Figure 32) would be followed by the broken LRU in storage being replaced by an operational LRU from the warehouse ("Replace broken LRU from warehouse on onboard storage (8)" at top in Figure 32). This replacement would be followed by the broken LRU being sent to the NME where it would be repaired and returned to the warehouse ("NME repairs broken LRU in warehouse (3)" on right in Figure 32). These precedencies are modeled by having the downstream process equal to its preceding process step with a delay that reflects the transit and subsequent processing time. Some flows (e.g. NME repairs broken LRU from warehouse (3)) are aggregations of multiple upstream flows. Core flows are based on the mean time between failure of LRUs and the fraction of failures addressed with each process.

The system dynamics model was calibrated to reflect RDN ship maintenance. Quantitative information on the volume of process steps performed in the maintenance of the RDN fleet was requested but was not available, primarily due to the extreme diversity of components and maintenance requirements. One RDN informant described the frequency of the maintenance operations (i.e. steps 1-7 above) as "continuous" and that frequency estimates were very difficult because of the extreme range of frequencies across component types. As an example, the informant said that work on some components happen daily while work on other types of components happens only once every few years. The informant provided the example that when a warship was at sea for 30 days only process step 1 (on-board repairs) would occur but if the ship were at port other process steps might be used. Therefore the modeler based calibration for a portion of the system dynamics model on publicly available data, data collected (e.g. numbers of LRU in NME and on board a typical ship) and estimated conditions of peacetime operations near Dutch ports. Publically available data included the types and numbers of ships in the Dutch navy [Wikipedia, 2012] (Table 16).

Ship Type	Number
Frigate	12
Landing Platform	2
Replenishment	2
Submarine	4
Mine detection	6
Dive support	4
Hydrographical survey	2
Training	2
Tugs - large	5
Tugs - harbor	7
Landing craft	17
Patrol boat - off shore	4
Patrol boat - in shore	6
Cutter	3
TOTAL	76

Table 16: Royal Dutch Navy Ship Types and Numbers

Calibration estimates were made using this information as follows. Data not documented above are modeler estimates.

Total LRU in use on all ships = 60k LRU/ship * 76 ships = 4,560k LRU on ships

Assuming one ship of each of the 14 ship types is considered “sacrificial” and used for cannibalization:

Total LRU in use on the 62 (=76-14) “operating” ships = 62ships*60k LRU/ship = 3,720k LRU

Total LRU in use on the 14 “sacrificial” ships = 4,560k – 3,720k = 840k LRU

In addition, each ship keeps 25% of its LRU in on-board ship storage:

62ships * (25%)(60k LRU/ship) = 930k LRU in storage on operating ships

14ships * (25%)(60k LRU/ship) = 210K LRU in storage on sacrificial ships

Total LRU on sacrificial ships = 840k + 210k = 1050k LRU on sacrificial ships

Warehouse initially holds one complete set for each vessel type:

60k LRU/ship * 14 ship types = 840k LRU

The number of LRU at NME is only the LRU that are currently being repaired by NME, i.e. all LRU storage occurs at the warehouse and none at NME (consistent with researcher observations).

The following fractions of broken LRUs are addressed with each solution:

25% of broken LRU are replaced with on-board replacements¹⁰
10% of broken LRU are cannibalized from other ships¹¹
35% of broken LRU are replaced with LRU in warehouse¹²
25% of broken LRU are repaired by NME directly without passing through
warehouse
5% of broken LRU are repaired directly by industry
100% TOTAL

Assume 15% of LRU repaired directly by NME need assistance from industry.

KVA Models to the Royal Dutch Navy Ship Maintenance

Four knowledge value added models were built of royal Dutch Navy Ship Maintenance:

- Baseline RDN ship maintenance processes
- Baseline RDN ship maintenance processes changed to reflect the adoption and use of a logistics package from an integrated CPLM system such as was investigated by Damen
- Baseline RDN ship maintenance processes changed to reflect the adoption and use of 3D PDF modeling managed with a CPLM system as planned by Damen
- Baseline RDN ship maintenance processes changed to reflect the adoption and use of a logistics package and 3D PDF modeling managed by an integrated CPLM system

Inputs to these models were generated as follows:

- Process Descriptions – The seven basic process steps used by the RDN to maintain the fleet were taken from data collected from RDN (Figure 31) and description provided by manager of NME. Two additional process steps (#8 and #9) were added based on the logic that broken LRU in on board storage or cannibalized ships would be replaced with operating LRU from the warehouse.
- Title of Head Process Executer – The KVA modeler matched the levels and types of training received in the different levels of training as described by the informants to the process steps based on process step requirements.
- Number of Employees - KVA modeler estimate based on manpower requirements to perform each process step in the maintenance of pumps scenario
- Corresponding Pay Grades – KVA modeler estimate of relative hourly pay rates for skill levels described by training requirements. Estimated values include labor burden and overhead.
- Rank Order of Difficulty – KVA modeler estimate based on understanding of processes from informants
- Actual Average Training Period – Based on data provided by informants (see data description above)

¹⁰ These LRU are then sent to the warehouse and replaced with an operational LRU from the warehouse.

¹¹ These LRU are then sent to the warehouse and replaced with an operational LRU from the warehouse.

¹² These LRU are then sent to the NME for repair and returned to the warehouse.

- Percentage Automation - KVA modeler estimate in base case based on understanding of processes from informants. Modeler estimate of changes due to technology adoptions based on previous KVA models of ship maintenance processes.
- Times performed in a Year – Output from system dynamics model
- Average Time to Complete – KVA modeler estimate for base case based on understanding of processes from informants . Modeler estimate of changes due to technology adoptions for other KVA models.
- Automation Tools - KVA modeler estimate for base case based on understanding of processes from informants. Modeler estimate of changes due to technology adoptions for other KVA models.

Model Simulations and Results

The system dynamics model was simulated to represent the four technology adoption scenarios described in the previous section. The output of each system dynamics model simulation was used as input to a KVA model. Those KVA models were then used to estimate the return on investment (ROI) of each process in each of the four scenarios and the cumulative ROI for each scenario. The results based on the models and their calibrations described above are shown in Table 17.

		Return On Investment (ROI)			
		Baseline	Add Logistics	Add 3D pdf	Add Logistics & 3D pdf
1	Replace LRU with on-board spare	90%	261%	501%	464%
2	Replace operating LRU with warehouse spare	90%	151%	621%	1027%
3	NME repairs warehouse LRU and returns it to warehouse	8%	65%	95%	236%
4	Manufacturer repairs LRU for NME & it returns to warehouse	31%	88%	168%	168%
5	Replace on-board LRU with LRU cannibalized from another ship	90%	151%	621%	1027%
6	NME repairs on-board LRU and returns it to ship	265%	10%	99%	192%
7	Industry repairs on-board LRU and returns it to ship	34%	178%	135%	318%
8	Replace on-board storage LRU with warehouse spare (transit only)	301%	759%	759%	759%
9	Replace cannabalized LRU with warehouse spare (transit only)	140%	329%	862%	1102%
TOTAL ALL PROCESSES		35%	77%	135%	274%

Table 17: Knowledge Value Added Model Results

Although increased throughput due to reduced processing durations (which increase the ROI numerator) can partially explain differences in the ROI in Table 17, cost reduction (which decreases the ROI denominator) is the primary driver of increases in ROI. For example, processes 8 and 9 are benefitted by reductions in rework (e.g. errors in transporting LRU) due to the adoption of a logistics package. This reduces the number of transport trips required (the function of these processes), thereby significantly reducing costs and increasing the ROI. In contrast, processes 3, 4, and 6 are highly skilled processes that are difficult to replace with technology and therefore benefit less from technology adoption than other processes. This results in a smaller increase in ROI for these processes.

Analysis of Simulation Model Results

A variance analysis was performed on the KVA model results (Table 17) to evaluate the relative impacts of the adoption of different technologies (Table 18). Returns on investment for each of the three technology adoption alternatives were compared with the baseline returns on investment to estimate improvement due to technologies (left three columns of results, Table 18). In addition the improvement from adopting both technologies over adopting only the 3D PDF technology was estimated (right column, Table 18)

	Process Description	Return On Investment (ROI)			
		Add Logistics - Improvement over Baseline	Add 3Dpdf - Improvement over Baseline	Add Logistics & 3Dpdf - Improvement over Baseline	Add Logistics & 3Dpdf - Improvement over adding only 3Dpdf
1	Replace LRU with on-board spare	171%	411%	374%	-38%
2	Replace operating LRU with warehouse spare	61%	532%	937%	406%
3	NME repairs warehouse LRU and returns it to warehouse	57%	87%	227%	140%
4	Manufacturer repairs LRU for NME & it returns to warehouse	57%	138%	138%	0%
5	Replace on-board LRU with LRU cannibalized from another ship	61%	532%	937%	406%
6	NME repairs on-board LRU and returns it to ship	-256%	-166%	-73%	93%
7	Industry repairs on-board LRU and returns it to ship	145%	101%	284%	183%
8	Replace on-board storage LRU with warehouse spare (transit only)	458%	458%	458%	0%
9	Replace cannibalized LRU with warehouse spare (transit only)	189%	721%	962%	240%
	TOTAL ALL PROCESSES	42%	100%	239%	139%

Table 18: Variance Analysis of KVA Model Results

Referring to Table 18, adding either or both of the technologies improves overall ship maintenance ROI, as indicated by the positive numbers in the last row of Table 18. Adopting 3D PDF alone improves ROI significantly more than adopting a logistics package alone (100% improvement > 46% improvement) and adding both technologies improves ROI more than adding either technology alone (239% improvement > 42% improvement or 100% improvement), suggesting that there may be synergy between the technologies. This is also supported by the 139% improvement by adding logistics if 3D PDF is already in place (lower right result in Table 18).

Adopting the technologies does not impact the ROI of individual processes equally. Among the seven core processes(#1-#7)¹³ adding only a logistics package (left column of results in Table 18) increases the “Replace LRU with on-board spare” (process #1) most, by 171%, and decreases the return of process #6, “NME repairs on-board LRU and returns it to ship” by 256%. Among the seven core processes adding only 3D PDF increases processes #2 and #5, “Replace operating LRU with warehouse spare” and “Replace on-board LRU with LRU cannibalized from another shop” most, by 532%, and decreases the return of process #6, “NME repairs on-board LRU and returns it to ship” by 166%. Among the seven core processes adding both technologies increases processes #2 and #5, “Replace operating LRU with warehouse spare” and “Replace on-board LRU with LRU cannibalized from another shop” most, by 937% and decreases the return of process #6, “NME repairs on-board LRU and returns it to ship” by 73%.

Comparison of Dutch Royal Navy and U.S. Navy Scenarios:

Previous research using the KVA approach developed estimates of returns on technology investment of a scenario in which the U.S. Navy adopts 3D laser scanning technology (3D LST) and collaborative product lifecycle management (CPLM) tools into the SHIPMAIN program. Komoroski [2005] investigated the early phases of SHIPMAIN. The relevant results are shown in Table 19.

¹³ Process #8, “Replace on-board storage LRU with warehouse spare (transit only)” supports process #1, “Replace LRU with on-board spare”. Therefore process #1 is the core process. Similarly, Process #9, “Replace cannibalized LRU with warehouse spare (transit only)” supports process #5, “Replace on-board LRU with LRU cannibalized from another ship”. Therefore process #5 is the core process.

Core Process	Process Title	"AS-IS" ROI	"RADICAL" ROI	"RADICAL" improvement over "AS-IS"
1	ISSUE TASKING	-59%	-59%	0%
2	INTERPRET ORDERS	-73%	746%	819%
3	PLAN FOR SHIPCHECK	-99%	-95%	4%
4	CONDUCT SHIPCHECK	-74%	1653%	1727%
5	REPORT ASSEMBLY	-39%	1032%	1071%
6	REVISE SCHEDULE	12%	882%	870%
7	GENERATE DRAWINGS	-54%	2977%	3031%
	TOTALS	-27%	2019%	2045%

Table 19: Preparation for Maintenance Processes - As-is and Radical ROI Differences [Komoroski, 2005]

Referring to Table 19, adding the 3D LST and CPLM technologies improves overall preparation for maintenance processes ROI, as indicated by the positive number in the lower right corner of Table 19. Adding these technologies generally improves individual processes as well, as indicated by the non-negative (and positive with one exception) numbers in the right column of Table 19. The range of improvements across individual processes is large, varying from 0% (Issue Tasking) to 3031% (Generate drawings). Cost reduction explains these differences. For example, the adoption of technology in Core Processes 4 (Conduct Shipcheck) and 7 (Generate Drawings) significantly reduce the number of people required to survey ship conditions (#4) or draft 3D drawings from the survey data (#9), resulting in large ROI if the technology is adopted.

Seaman, Housel and Mun [2007] used KVA to model the later phases of SHIPMAIN. The relevant results are shown in Table 20.

Core Process	Process Title	Annual As-Is Cost	Annual As-Is Benefits	Annual To-Be Cost	Annual To-Be Benefits	As-Is ROI	To-Be ROI
Block 250	Authorize and Issue Letter of Authorization (LOA)/Hull Maintenance Plan (HMP); Generate 2Ks	\$5,311,248	\$22,619,472	\$2,287,671	\$15,215,872	326%	565%
Block 265	Hull Installation and Risk Assessment	\$130,060,112	\$94,928,918	\$63,437,554	\$161,749,816	-27%	155%
Block 270	Authorize Installation	\$3,161,600	\$24,710,347	\$3,217,805	\$24,710,347	682%	668%
Block 280	Resolve "Not Authorized/Deferred SC	\$619,424	\$3,706,552	\$427,964	\$3,706,552	498%	766%
Block 300	Install SC	\$40,616,160	\$94,722,998	\$33,433,420	\$94,722,998	133%	183%
Block 310	Feedback: Cost, CM, Performance, Schedule, ILS	\$619,424	\$1,853,276	\$242,107	\$1,853,276	199%	665%
Block 320	Continue Installs	\$3,068,520	\$4,633,190	\$2,510,944	\$5,791,488	51%	131%
Block 330	Final Install, Closeout SC	\$309,712	\$926,638	\$304,059	\$926,638	199%	205%
Totals:		\$183,766,200	\$248,101,392	\$105,861,524	\$318,820,901	35%	201%

Table 20: Maintenance and Implementation Processes - As-is and To-be ROI Comparison

Referring to Table 20, adding the technologies also improves overall maintenance implementation process ROI, as indicated by the positive difference between the overall To-Be ROI (201%) and the overall As-Is ROI (35%) numbers in the lower right corner of Table 20. Adding these technologies also improves each of the individual processes, as indicated by the increases in the To-Be ROI values over the As-IS ROI values in Table 19. The range of improvements across individual processes is large, varying from 6% to 466% (Final install and closeout ship change), although not as wide as in the preparation for maintenance processes.

Although the same KVA modeling process was applied to ship maintenance in both of the U.S. and the Royal Dutch navies, the KVA models have important differences that complicate the comparison of their results. For example, the process steps are different and the amounts of field data available to calibrate the models differed significantly. Therefore, any comparisons can only be preliminary at this point. However, comparison can reveal some apparent similarities and differences between the scenarios that are of interest. Table 21 shows the overall baseline (existing processes) and technology-improved ROI for the two U.S. Navy scenarios and the Royal Dutch Navy scenario.

	Baseline Overall ROI	Technology-adopted Overall ROI
US Navy - SHIPMAIN (preparation for maintenance phases)	-27%	2019%
US Navy - SHIPMAIN (implementation phases)	35%	201%
Royal Dutch Navy (Damen experience extrapolation)	35%	274%

Table 21: Return on Investment: Baseline and Technology Adoption Services

The three scenarios have some similarities. All three overall returns on investment after technology adoption are positive and large. This supports the adoption of advanced technologies such as 3D laser scanning technology, 3D PDF models, and collaborative product lifecycle management to improve the efficiency of resource use. The scenarios also have potentially significant differences. The technology-adoption scenario for the preparation for maintenance phases of the U.S. scenario has a much higher overall ROI than those of the maintenance implementation phases of the U.S. scenario or the Dutch scenario (2,019% >> 201% or 274%). Several factors could explain these differences.

- The preparation for maintenance -phases of the U.S. scenario have significantly lower ROI in the As-Is (without technology) condition (-27% > 35%). This suggests that inefficiencies in the preparation for maintenance processes provided more and larger opportunities for improvement.

- The individual preparation for maintenance processes that increased the most (see Table 20) such as Generate Drawings and Conduct Shipcheck are very labor intensive and therefore costly, providing large opportunities for cost reduction through technology adoption.
- Several of the individual maintenance implementation processes (Table 21) are labor intensive but less impacted by technology (e.g. Install Shipcheck), thereby making those changes in ROI less dramatic.
- The preparation for maintenance phases of the U.S. scenario could be more optimistic in its projections than the other scenarios.
- The estimates of process changes may use different assumptions.
- Technologies adopted in the preparation for maintenance phases of the U.S. scenario may make much larger improvements in processes than those in the maintenance implementation phases of the U.S. scenario or the Dutch scenario.
- The Dutch case does not use all of the capabilities of the CPLM, thereby making it more incremental than the U.S. scenarios, where all the capabilities of the CPLM were projected to be used. Also, 3D PDF has more limited capabilities for integration with the CPLM logistics package when compared to the integration of 3D LST capabilities for broader usage in requirements analysis, planning for maintenance, and tracking of parts in the supply chain and across suppliers, contractors. This can partially explain the lower Dutch technology-adopted ROI than the U.S. preparation for maintenance ROI.
- The projections of the impacts on the maintenance implementation phases of the U.S. scenario and the Dutch scenario may be rather conservative based on research into the actual successful implementation of other modern technologies, such as RFID in inventory management and transparency. In a study of the actual use of passive RFID in two military warehouses in the Korean Air Force and Army, the actual ROIs from use of the RFID technology were more than triple the projected impact of the use of the technology in a separate study of the use of the technology in the U.S. Navy. The Korean ROIs after actual implementation of the RFID technology ranged from 610% to 576% compared to the projected returns anticipated from the implementation of the same technology in the U.S. Navy which ranged up to 133%. The implication is that actual successful implementation of information technology in a military may exceed projections of the potential impacts of the technology. It follows, that the current research on the impacts of CPLM and 3D LST or 3D PDF may be more conservative than the reality once these technologies are actually implemented on a wide scale basis.

5.4 INTEGRATED RISK MANAGEMENT MODELING AND RESULTS

IRM and Risk Simulation

Through the use of Monte Carlo simulation, the resulting stochastic KVA ROK model yielded a distribution of values rather than a point solution. Thus, simulation models analyze and quantify the various risks and uncertainties of each program. The result is a distribution of the ROKs and a representation of the project's volatility.

In real options, the analyst assumes that the underlying variable is the future benefit minus the cost of the project. An implied volatility can be calculated through the results of a Monte Carlo simulation performed. The results for the IRM analysis will be built on the quantitative estimates provided by the KVA analysis. The IRM will provide defensible quantitative risk analytics and portfolio optimization suggesting the best way to allocate limited resources to ensure the highest possible value over time.

The first step in real options is to generate a strategic map through the process of framing the problem. Based on the overall problem identification occurring during the initial qualitative management screening process, certain strategic options would have become apparent for each particular project. The strategic options may include, among other things, the option to wait, expand, contract, abandon, switch, stage-gate, and choose.

Risk analysis and real options analysis assume that the future is uncertain and that decision makers have the ability to make midcourse corrections when these uncertainties become resolved or risk distributions become known. The analysis is usually done ahead of time and, thus, ahead of such uncertainty and risks. Therefore, when these risks become known, the analysis should be revisited to incorporate the information in decision making or to revise any input assumptions. Sometimes, for long-horizon projects, several iterations of the real options analysis should be performed, where future iterations are updated with the latest data and assumptions. Understanding the steps required to undertake an integrated risk analysis is important because it provides insight not only into the methodology itself but also into how it evolves from traditional analyses, showing where the traditional approach ends and where the new analytics start.

The risk simulation step required in the IRM provides us with the probability distributions and confidence intervals of the KVA methodology's resulting ROI and ROK results. Further, one of the outputs from this risk simulation is volatility, a measure of risk and uncertainty, which is a required input into the real options valuation computations. In order to assign input probabilistic parameters and distributions into the simulation models, we relied on the U.S. Air Force's Cost Analysis Agency (AFCAA) handbook as seen in Figure 31.¹⁴ In the handbook, the three main distributions recommended are the Triangular, Normal, and Uniform distributions. We choose the Triangular distribution as the limits (minimum and maximum) are known, and the shape of the triangular resembles the Normal distribution, with the most likely values having the highest probability of occurrence and the extreme ends (minimum and maximum values) having considerably lower probabilities of occurrence. Also, the Triangular distribution was chosen instead of the Normal distribution as the latter's tail ends extend toward positive and negative infinities, making it less applicable in the model we are developing. Finally, the AFCAA also provides options for left skew, right

¹⁴ Available at <http://www.afhra.af.mil/factsheets/factsheet.asp?id=14009>

skew, and symmetrical distributions. In our analysis, we do not have sufficient historical or comparable data to make the proper assessment of skew and, hence, revert to the default of a symmetrical Triangular distribution. Using these AFCAA guidelines, which are presented as 15%, Mean, and 85% values (Figure 33), we imputed the corresponding minimum (min), most likely (likely), and maximum (max) values required in setting up the Triangular distributions (Figure 34).¹⁵

AFCAA Cost Risk Analysis Handbook
Table 2-5 Default Bounds for Subjective Distributions

Distribution	Point Estimate Interpretation	Point Estimate and Probability	Mean	15%	85%
Triangle Low Left	Mode	1.0 (75%)	0.878	0.695	1.041
Triangle Low	Mode	1.0 (50%)	1.000	0.834	1.166
Triangle Low Right	Mode	1.0 (25%)	1.122	0.959	1.305
Triangle Med Left	Mode	1.0 (75%)	0.796	0.492	1.069
Triangle Med	Mode	1.0 (50%)	1.000	0.723	1.277
Triangle Med Right	Mode	1.0 (25%)	1.204	0.931	1.508
Triangle High Left*	Mode	1.0 (75%)	0.745	0.347	1.103
Triangle High	Mode	1.0 (50%)	1.000	0.612	1.388
Triangle High Right	Mode	1.0 (25%)	1.286	0.903	1.711
Triangle EHigh Left*	Mode	1.0 (75%)	0.745	0.300	1.130
Triangle EHigh	Mode	1.0 (50%)	1.004	0.309	1.500
Triangle EHigh Right	Mode	1.0 (25%)	1.367	0.876	1.914

Figure 33: U.S. Air Force Cost Analysis Agency Handbook's Probability Risk Distribution Spreads

¹⁵ Using the Triangular distribution's probability density function (PDF), we simply compute the cumulative distribution function (CDF). In mathematics and Monte Carlo simulation, a PDF represents a **continuous** probability distribution in terms of integrals. If a probability distribution has a density of $f(x)$, then intuitively the infinitesimal interval of $[x, x + dx]$ has a probability of $f(x) dx$. The PDF therefore can be seen as a smoothed version of a probability histogram; that is, by providing an empirically large sample of a continuous random variable repeatedly, the histogram using very narrow ranges will resemble the random variable's PDF. The probability of the interval between $[a, b]$ is given by $\int_a^b f(x)dx$, which means that the total integral of the function f must be 1.0. The CDF is denoted as $F(x) = P(X \leq x)$, indicating the probability of X taking on a less than or equal value to x . Every CDF is monotonically increasing, is continuous from the right, and at the limits, has the following properties: $\lim_{x \rightarrow -\infty} F(x) = 0$ and $\lim_{x \rightarrow +\infty} F(x) = 1$. Further, the CDF is related to the PDF by $F(b) - F(a) = P(a \leq x \leq b) = \int_a^b f(x)dx$, where the PDF function f is the derivative of the CDF function F . Using these relationships, we can impute the min, likely, max values from the mean, and 15th and 85th percentiles that were provided by the AFCAA.

Symmetrical Spreads (Tringular Distribution)			
	Min	Likely	Max
1 High	0.1421	1.0000	1.8579
2 Above Average	0.2648	1.0000	1.7352
3 Medium	0.3875	1.0000	1.6125
4 Average	0.5103	1.0000	1.4898
5 Low	0.6330	1.0000	1.3670
6 Very Low	0.7558	1.0000	1.2443
7 Completed/NC	1.0000	1.0000	1.0000

1 +/- 3 Execution	-3	3
2 +/- 2 Execution	-2	2
3 +/- 1 Execution	-1	1
4 Completed/NC	0	0

Figure 34: U.S. AFCAA Handbook's Probability Risk Distribution Spreads

It is important to understand why it is necessary to apply uncertainty to the model. Because the KVA process provided a point value for each quantity, even though there is some uncertainty in the estimates provided by the SMEs, application of the appropriate statistical distributions of input is used to restore the real world's uncertainty to the model. Having inputs from only three experts, as opposed to hundreds of estimates, and rather than using these three discrete inputs, the analysts have applied the lessons learned in cost estimating as reflected in the Air Force handbook as a good starting point for representing the uncertainty and reflecting it in the simulations.

Next, using the developed KVA model, risk simulation probabilistic distributional input parameters are inserted into the three main variables: percentage automation, time process is executed, and average time (hours) to complete (Figure 35).¹⁶ A risk simulation of 10,000 to 1,000,000 simulation trials were run to obtain the results.¹⁷

Two sets of results important in the simulation analysis are volatility and probability confidence intervals. The simulation statistics obtained after running a simulation can be seen in Figure 36, where the main variable of interest is the coefficient of variation, which in this case is used as a proxy for volatility.¹⁸ The average volatilities are between 54% and 87%. To put this into perspective, the annualized volatility of blue chip stocks (e.g., IBM or Microsoft) is typically between 15% and 30%, whereas higher risk companies (stocks with low market to book ratios, low price to earnings ratios, or

¹⁶ The Monte Carlo Risk Simulation was performed using *Risk Simulator* (version 2012) software by Real Options Valuation, Inc. (www.realoptionsvaluation.com), and screenshots provided are with permission from the software developers.

¹⁷ Different numbers of trials were run to calibrate the precision of the model and to check for model convergence.

¹⁸ The coefficient of variation is simply defined as the ratio of standard deviation to the mean, where risks are common size. As standard deviation is the measure of the spread or dispersion of the data around its mean, it is oftentimes used as a measure of uncertainty, and when divided by the average of the distribution, it becomes a relative measure of risk, without any units. This measure of risk or dispersion is applicable when the variables' estimates, measures, magnitudes, or units differ, and can be used as a proxy for volatility of the project.

startups) have their stocks' volatilities above 50%, and highly speculative derivatives may have volatilities upwards of 100%.

The probability confidence intervals will be used and discussed in a later section within the realms of real options valuation. At this point in the analysis, a proxy for revenues and volatility has been identified, as well as the numerators and denominators for the ship maintenance program. The next step is to define or frame the alternatives and approaches to implementing 3D PDF and Logistics Team Centers, namely, strategic real options. The questions that can be answered include: What are the options involved, how should these new processes be best implemented, which decision pathway is optimal, and how much is the program worth to the DoD?

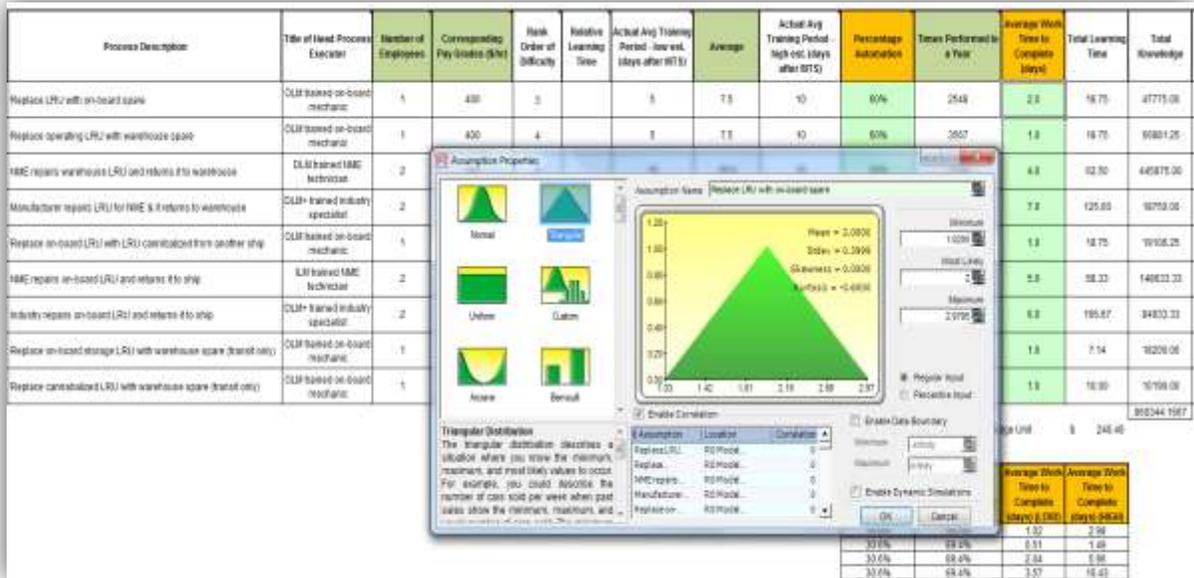


Figure 35: Risk Simulation Probability Distribution Parameters

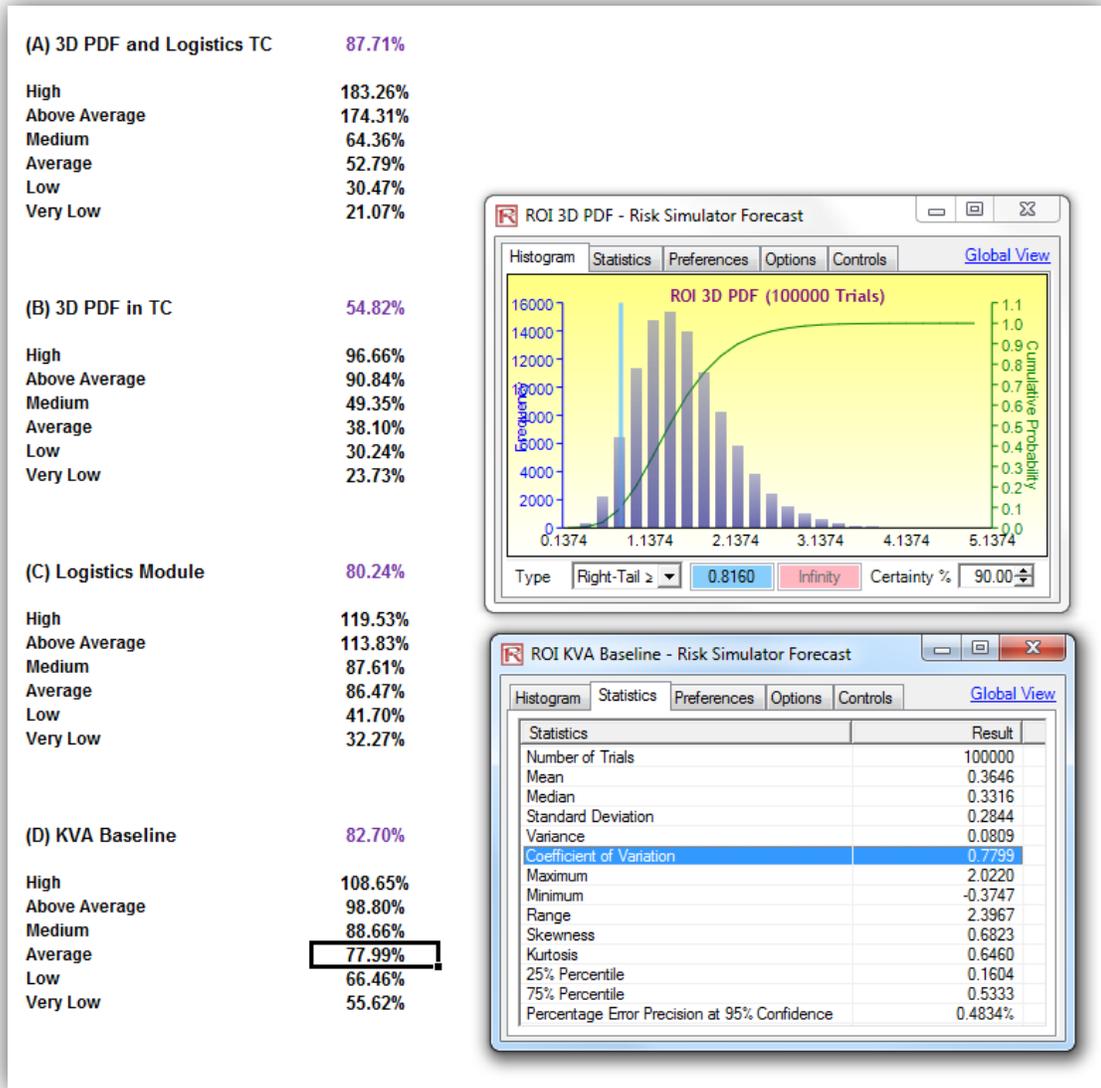


Figure 36: Risk Simulated Volatility

IRM: Why Strategic Real Options?

As described previously, an important step in performing IRM is the application of Monte Carlo risk simulation. By applying Monte Carlo risk simulation to simultaneously change all critical inputs in a correlated manner within a model, you can identify, quantify, and analyze risk. The question then is, what next? Simply quantifying risk is useless unless you can manage it, reduce it, control it, hedge it, or mitigate it. This is where strategic real options analysis comes in. Think of real options as a strategic road map for making decisions.

The real options approach incorporates a learning model, such that the decision maker makes better and more informed strategic decisions when some levels of uncertainty are resolved through the passage of time, actions, and events. The

combination of the KVA methodology, to monitor the performance of given options, and the adjustments to real options as leaders learn more from the execution of given options provides an integrated methodology to help military leaders hedge their bets while taking advantage of new opportunities over time. Traditional analysis assumes a static investment decision, and assumes that strategic decisions are made initially with no recourse to choose other pathways or options in the future. Real options analysis can be used to frame strategies to mitigate risk, value and find the optimal strategy pathway to pursue, and generate options to enhance the value of the project while managing risks. Imagine real options as your guide when navigating through unfamiliar territory, providing road signs at every turn to direct you in making the best and most informed driving decisions. This is the essence of real options. From the options that are framed, Monte Carlo simulation and stochastic forecasting, coupled with traditional techniques, are applied. Then, real options analytics are applied to solve and value each strategic pathway and an informed decision can then be made.¹⁹

IRM: Framing the Real Options

As part of the first round of preliminary analysis, Figure 37 illustrates some of the potential implementation paths for 3D PDF/Logistics TC. Clearly some of the pathways and flexibility strategies may be refined and updated over time through the passage of time, actions, and events. With the evolution of the implementation, valuable information is obtained to help in further fine-tuning the implementation and decision paths. For the preliminary analysis, the following options were identified, subject to modification:

- Option A: As-Is Base Case. The ROI for this strategic path is computed using the baseline KVA and this represents the current Royal Dutch Navy ship maintenance process – i.e., no newly added technologies.
- Option B: Execute and implement 3D PDF and Logistics package immediately across all Royal Dutch Navy ship maintenance processes. That is, take the risk and execute on a larger scale, where you would spend the initial investments and continuing maintenance expenses required and take on the risks of any potential failure but reap the rewards of the new processes' savings quickly and immediately. The analysis is represented as the current RDN process altered to reflect what we estimate to be the impacts of adopting both a Logistics package and 3D PDF models.
- Option C: This represents the current RDN process altered to reflect what we estimate to be the impacts of adopting 3D PDF models and managing them in a Team Center or similar product. This technology was chosen largely because Damen is developing and pursuing the use of this technology.
- Option D: This implementation pathway represents the current RDN process altered to reflect what we estimate to be the impacts of managing using a Logistics module in a Team Center or similar product. This technology was chosen partially because it was a technology that Damen considered but chose not to purchase.

¹⁹ The pathways can be valued using partial differential closed-form equations, lattices, and simulation. The *Real Options SLS* software, version 2012 (B), by Real Options Valuation, Inc. (www.realoptionsvaluation.com), is used to value these options with great ease. Monte Carlo risk simulations were performed using the *Risk Simulator* software, version 2012 (B), also by the same organization.

- Option E: Proof of Concept approach, that is, to execute large-scale implementation of 3D PDF and Logistics Module in TC only after an initial Proof of Concept (POC) shows promising results. If POC turns out to be a failure, we walk away and exit the program, and losses are minimized and limited to the initial POC expenses. Proceed to full implementation in POC programs first and then expand in sequential fashion to other programs, based on where best ROI estimates are shown.
- Option F: Proof of Concept on 3D PDF only. Assuming the POC works and 3D PDF is executed within a few programs successfully, the learning and experience obtained becomes valuable and allows the shipyards to expand its use into many other programs or perhaps across the Royal Dutch Navy.
- Option G: Proof of Concept on Logistics Module in TC only. Assuming the POC works and Logistics Module is executed within a few programs successfully, the learning and experience obtained becomes valuable and allows the shipyards to expand its use into many other programs or perhaps across the Royal Dutch Navy.

Figure 38 shows the preliminary input assumptions and Figure 39 shows the computed return on investment results and strategic real options results. For instance, the following inputs were assumed:

- PV Asset. This is the net total benefits or proxy revenues (numerator) obtained from the KVA analysis under each of the various options as outlined above.
- Implementation Cost. This is the total cost to implement each of the options (e.g., 3D PDF only, 3D PDF with Logistics Module TC, or Logistics Module TC only).
- Maturity. This is the time to perform the proof of concept stage, denoted in years.
- Risk-free Rate. This is the annualized U.S. Treasury rate used as a proxy of a risk-free asset. This rate is used to discount the future cash flows in the risk-neutral options model. We use a risk-free rate as the risk has already been accounted for in the risk simulation and volatility estimates. Figure 40 illustrates the U.S. Treasury security interest rates used as a proxy for the risk-free rate used in the analysis.
- Volatility. This is the annualized volatility estimate obtained from Monte Carlo risk simulation in the previous step by using the AFCAA risk spreads as a proxy.
- Dividend Rate. This variable is typically not used but is available should the need arise. Briefly, it measures the annualized percentage rate of opportunity cost of investing at a future time instead of immediately.

IRM: Strategic Flexibility Real Options Results

Figure 39 shows the results of the strategic real options flexibility values and compares them against the KVA ROI values. We see that Options B (\$154.1M at 278% ROI) and E (\$156.5M at 282% ROI) of implementing both 3D PDF and Logistics Module TC return the highest ROI and total strategic value, and both providing significant value-add above and beyond Option A's As-Is condition (\$31.9M at 35% ROI). As Options B and E are most significant, stage-gating the implementation over several phases yields a slightly higher value (Option E exceeds Option B by about \$2.4M).

In addition, Figure 41 shows the Monte Carlo Risk Simulated results on the real options values. For instance, in comparing between Options E and F, we see that Option E with the sequentially phased implementation of both 3D PDF and Logistics Module TC, there is a 94% probability that this path provides a better return than Option F. In comparing Options E with B, there is a 95% confidence that even with all the uncertainties in the collected data and risks of implementation success, uncertainties of whether the estimated returns will materialize, and so forth, we still see that there is at least a \$1.27M net advantage in going with Option E. Therefore, it is better to sequentially phase and stage gate the implementation over several years, allow the ability to exit and abandon further stages if events unfold and uncertainties become resolved in making further investment in the technology to no longer make sense.

As additional information, with KVA Baseline of Option A, we see that without doing any implementations, there is still a 4.7% probability that staying As-Is returns negative ROIs, and even in the best case analysis there is less than a 5% probability that ROI for the base case will ever exceed 93%.

The final two charts in Figure 41 show that the risk simulated real options value has an expected value (mean) of \$195M with a corresponding average ROI of 363%. Finally, Figure 42 shows the comprehensive simulated risk statistics of the various option scenarios.

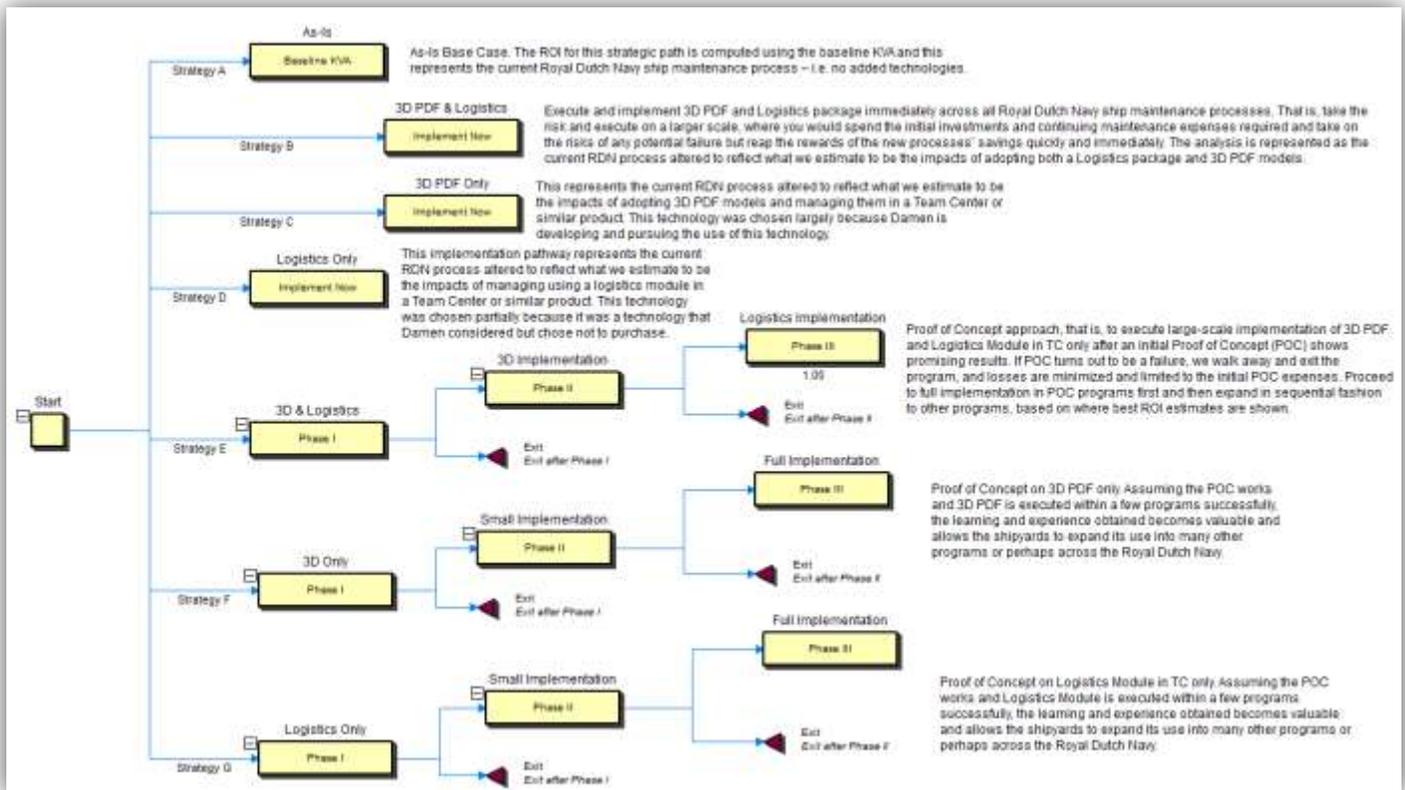


Figure 37: Strategic Real Options Implementation Pathways and Options

3D PDF AND LOGISTICS MODULE (PHASED SEQUENTIAL)

Maturity in Years	3.00	Result	\$156,569,744
Blackout Steps	0		282.88%

Underlying Asset Lattices

Lattice Name	PV Asset	Volatility %
Underlying	206904168.64	87.71

Custom Variables

Variable Name	Value	Starting Steps
Salvage	0.00	0

Option Valuation Lattices

Lattice Name	Implementation Cost	Riskfree Rate	Dividend Rate	Lattice Steps	Terminal Equation	Intermediate Equation	Intermediate Blackout Equation
Phase3	18449693	0.40	0.00	60	Max(Underlying-Cost;0)	Max(Underlying-Cost;@@)	@@
Phase2	18449693	0.40	0.00	40	Max(Phase3-Cost;0)	Max(Phase3-Cost;@@)	@@
Phase1	18449693	0.40	0.00	20	Max(Phase2-Cost;0)	Max(Phase2-Cost;@@)	@@

Figure 38: Sample Real Options Input Assumptions

ANALYSIS RESULTS

Strategy A	As-Is
Strategy B	3D PDF & LOGISTICS TC (IMPLEMENT NOW)
Strategy C	3D PDF IN TC ONLY (IMPLEMENT NOW)
Strategy D	LOGISTICS MODULE ONLY (IMPLEMENT NOW)
Strategy E	3D PDF AND LOGISTICS TC (PHASED SEQUENTIAL)
Strategy F	3D PDF IN TC ONLY (PHASED SEQUENTIAL)
Strategy G	LOGISTICS MODULE ONLY (PHASED SEQUENTIAL)

KVA ROI	KVA ROK	Strategic	Real Options	
		Real Options	ROI	Volatility
35.00%	135.00%	\$31,903,557	35.00%	82.67%
273.82%	373.82%	\$154,163,806	278.53%	87.71%
135.06%	235.06%	\$96,330,730	137.25%	54.82%
77.28%	177.28%	\$81,009,562	91.66%	80.24%
273.82%	373.82%	\$156,569,744	282.88%	87.71%
135.06%	235.06%	\$97,416,808	138.79%	54.82%
77.28%	177.28%	\$84,456,260	95.56%	80.24%

Net Differential: Strategy E over Strategy B
 Net Differential: Strategy E over Strategy F

\$2,405,938
\$59,152,936

Figure 39: Sample Real Options Values

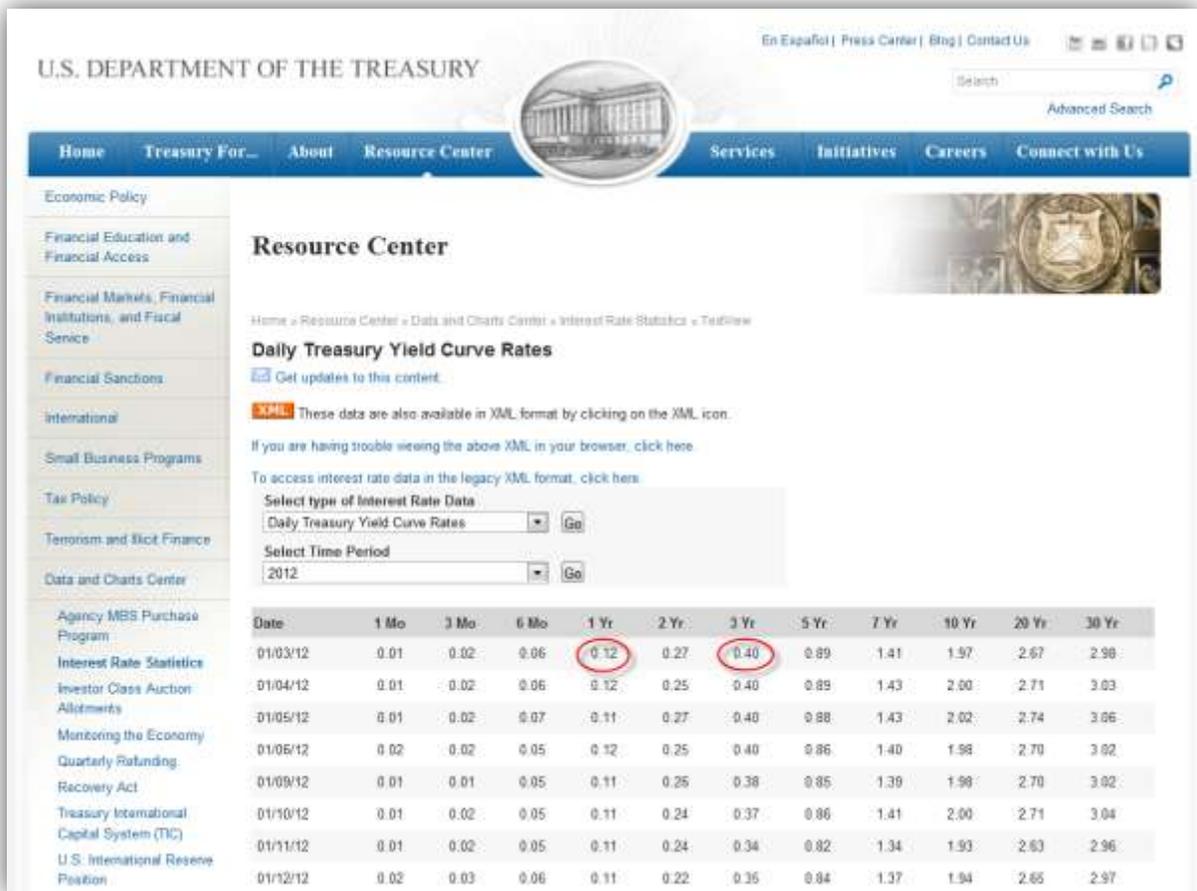


Figure 40: Risk-free Rate

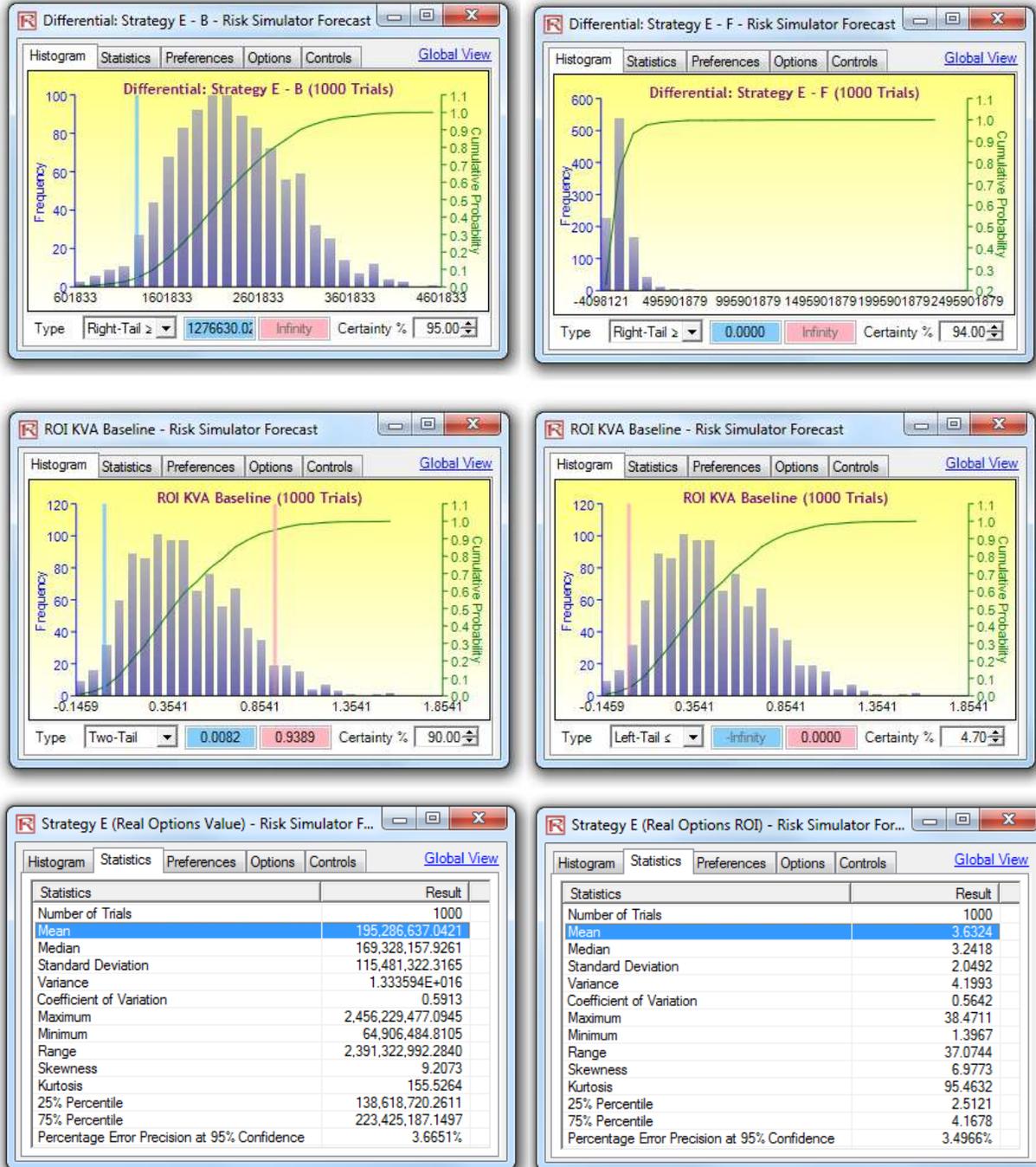


Figure 41: Risk Simulation Confidence and Percentiles

Forecast Statistics Table - Dutch Model										
Cell	Differential: Strategy E - B	Differential: Strategy E - F	ROI 3D and Logistics	ROI 3D PDF	ROI KVA Baseline	ROI Logistics	Strategy E (Real Options ROI)	Strategy E (Real Options Value)		
Name	\$Q\$31	\$Q\$32	\$U\$12	\$U\$12	\$U\$12	\$U\$12	\$R\$27	\$Q\$27		
Number of Datapoints	1000	1000	1000	1000	1000	1000	1000	1000		
Mean	2235082.3533	88115653.1929	3.5507	1.5545	0.4143	0.9721	3.6324	195286637.0421		
Median	2199188.6969	66732509.8315	3.1634	1.4801	0.3771	0.8705	3.2418	169328157.9261		
Standard Deviation	637638.6379	113301507.2564	2.0713	0.5470	0.2889	0.5666	2.0492	115481322.3165		
Coefficient of Variation	28.53%	128.58%	58.34%	35.19%	69.72%	58.29%	56.42%	59.13%		
Maximum	4505764.9505	2272222993.7653	38.4621	3.8563	1.5655	7.2963	38.4711	2456229477.0945		
Minimum	497620.8369	-64862339.3726	1.2231	0.4505	-0.1916	-0.0704	1.3967	64906484.8105		
Range	4008144.1136	2337085333.1379	37.2390	3.4058	1.7571	7.3667	37.0744	2391322992.2840		
Skewness	0.2840	8.8748	6.8191	0.6957	0.5779	3.1814	6.9773	9.2073		
Kurtosis	-0.0035	147.1703	92.2793	0.5261	0.1862	24.3862	95.4632	155.5264		
25% Percentile	1779072.9945	33251941.5657	2.4089	1.1739	0.1978	0.6140	2.5121	138618720.2611		
75% Percentile	2672672.0093	115729122.0184	4.1072	1.8937	0.6023	1.2032	4.1678	223425187.1497		
Error Precision at 95%	0.0177	0.0797	0.0362	0.0218	0.0432	0.0361	0.0350	0.0367		
5% Percentile	1275473.2163	-3291509.2148	1.7413	0.7694	0.0044	0.3367	1.8747	107163119.3176		
10% Percentile	1456789.9480	8053050.7713	1.9541	0.9093	0.0666	0.4219	2.0760	118084176.5203		
20% Percentile	1685667.7237	24600791.3515	2.2749	1.0865	0.1562	0.5642	2.3835	131076618.2203		
30% Percentile	1849736.9823	40644171.5832	2.5474	1.2254	0.2381	0.6632	2.6451	143657794.7797		
40% Percentile	2039501.7252	52639381.4116	2.8724	1.3407	0.3049	0.7735	2.9583	157896000.2912		
50% Percentile	2198268.6566	66588737.5647	3.1618	1.4795	0.3769	0.8704	3.2402	169218547.8776		
60% Percentile	2359028.5466	82692023.6386	3.4525	1.6248	0.4551	0.9886	3.5244	188851968.0234		
70% Percentile	2556015.3327	99542520.0217	3.8373	1.7996	0.5524	1.1079	3.9025	210363653.0484		
80% Percentile	2782999.6093	132791657.9897	4.4403	1.9942	0.6721	1.3175	4.4958	234621423.1570		
90% Percentile	3051915.2906	178260391.7200	5.4469	2.2839	0.8014	1.6233	5.4904	284883293.2537		
95% Percentile	3296549.5126	239364627.6082	6.3554	2.5271	0.9389	1.8812	6.3928	343000154.2593		
99% Percentile	3840681.1121	439546352.7120	10.3447	3.0736	1.1674	2.4870	10.3664	573174890.7559		
Certainty Value 0	0.0000	6.0000	0.0000	0.0000	4.7000	0.1000	0.0000	0.0000		

Figure 42: Risk Simulation Statistics and Percentiles

Discussion

New data was collected on ship maintenance processes and the use and adoption of technologies in ship maintenance by the Royal Dutch Navy and Damen Shipbuilding. The data was used to build and calibrate a system dynamics model of Royal Dutch Naval ship maintenance. Model simulations generated estimates of maintenance operations behavior that were imported into Knowledge Value Added models. Four technology adoption scenarios reflecting the potential use of two available or developing technologies were described in the KVA models. The KVA models estimate the returns on investment for individual processes and ship maintenance as a whole for each scenario. Results were analyzed to reveal relative improvement provided by individual and combinations of technologies.

Integrated Risk Management and Strategic Real Options methodologies were applied to the KVA-SD results and the results indicate that Option B had a value of \$154.1M (278% ROI) and Option E had a value of \$156.5M (282% ROI) where both options indicate that implementing 3D PDF and Logistics Module TC return the highest ROI and total strategic value, and both providing significant value-add above and beyond Option A's As-Is condition with a value of \$31.9M (35% ROI). As Options B and E are most significant, we know that implementation of 3D PDF and Logistics Module TC return the highest value, and when implemented over time in a stage-gate process over several phases, would yield a slightly higher value (Option E exceeds Option B by about \$2.4M). Therefore, we conclude that 3D PDF and Logistics Module TC implemented in a phased stage gate environment would yield the best results.

The linear ROI projections from adopting various iterations of the partial CPLM tool (i.e., only the logistics package) and the 3Dpdf tool in the Dutch context demonstrated the advantages of adopting both technologies over either technology alone compared to a baseline without either technology. Adopting 3D PDF alone improves ROI significantly more than adopting a logistics package alone (100% improvement > 46% improvement) and adding both technologies improves ROI more than adding either technology alone (239% improvement > 42% improvement or 100% improvement), suggesting that there may be synergy between the technologies. This is also supported by the 139% improvement by adding logistics if 3D PDF is already in place. These results were then used to forecast the benefits of various adoption options for the tools using the IRM methodology.

The results of the IRM analysis provided forecasts of the benefits of various options for implementing the technologies separately or in combinations. The results indicated that adoption of the technologies would provide cost-benefits far in excess of not using the technologies. The results indicated that there were marginal benefits in sequentially implementing the technologies over immediately implementing them. Given the long cycle for organizations to benefit from technology adoption, it might be better to adopt the technologies immediately.

ANALYSIS RESULTS		Strategic			Real Options	
		KVA ROI	KVA ROK	Real Options	ROI	Volatility
Strategy A	As-Is	35.00%	135.00%	\$31,903,557	35.00%	82.67%
Strategy B	3D PDF & LOGISTICS TC (IMPLEMENT NOW)	273.82%	373.82%	\$154,163,806	278.53%	87.71%
Strategy C	3D PDF IN TC ONLY (IMPLEMENT NOW)	135.06%	235.06%	\$96,330,730	137.25%	54.82%
Strategy D	LOGISTICS MODULE ONLY (IMPLEMENT NOW)	77.28%	177.28%	\$81,009,562	91.66%	80.24%
Strategy E	3D PDF AND LOGISTICS TC (PHASED SEQUENTIAL)	273.82%	373.82%	\$156,569,744	282.88%	87.71%
Strategy F	3D PDF IN TC ONLY (PHASED SEQUENTIAL)	135.06%	235.06%	\$97,416,808	138.79%	54.82%
Strategy G	LOGISTICS MODULE ONLY (PHASED SEQUENTIAL)	77.28%	177.28%	\$84,456,260	95.56%	80.24%
Net Differential: Strategy E over Strategy B				\$2,405,938		
Net Differential: Strategy E over Strategy F				\$59,152,936		

Figure 43: IRM Analysis Results

Comparing these results with the prior U.S. Navy results offers some partially confirming evidence for the prior research that projected the benefits of adopting the CPLM and 3D LST technologies for ship maintenance. There are a number of issues in making the comparisons that must be noted given the size differences of the two countries ship maintenance operations and the differences in the extent of implementation of the two types of technologies. However, the comparisons have validity when these issues are accounted for and the potential benefits of using the technologies are very high in both cases.

The scenarios have some similarities. All overall returns on investment with the technologies are positive and large. This supports the adoption of advanced technologies such as 3D laser scanning technology, 3D PDF models, and CPLM to improve the efficiency of resource use. All three scenarios also have ranges that exceed their overall

baselines with significant improvements in performance in given ship maintenance processes (U.S. Navy = 3031%>2045% , Dutch Navy = 460>166, and 444>139) significantly (by almost 50% or more). This suggests that attention must be paid to individual process steps and that the average or overall changes cannot be safely assumed to occur to all maintenance process steps.

The scenarios also have potentially significant differences. The early phase U.S. scenario has a much larger overall improvement than the later phase U.S. scenario or the Dutch scenario (2,045%>>166% or 139%). Several factors could explain this difference. The early-phase U.S. scenario could be more optimistic in its projections than the other scenarios, or the Dutch and later-phase U.S. case under estimate improvements.

6 SUMMARY

This report highlights findings from research conducted under the RT-18a: Valuing Flexibility project to identify, develop, and validate sound MPTs to enable DoD leadership and program managers to make a convincing case for investments in system flexibility when acquisition decisions are made.

In the first research thrust, a taxonomy of methods for valuing flexibility was developed. This taxonomy is designed around salient system characteristics in order to allow a decision-maker to select an appropriate method for valuing flexibility. Initial development created a software implementation of the Flexibility Valuation Method Selection Tool. Additionally, a series of cases illustrated how varying assumptions necessitates varying tools for valuing flexibility in the case of operating an observation satellite.

In the second research thrust, new tools were developed for estimating operation and support costs to improve Life Cycle Cost estimates. Validated macro-stochastic models were developed which demonstrated the ability to improve estimates of O&S costs for SARs. These models may provide significant value to decision-makers designing and investing in flexibility through improve cost estimates.

In the third research thrust, a detailed case study of valuing flexibility in ship maintenance was presented using data from both U.S. and Dutch naval experience. Knowledge Value Added and Integrated Risk Management models were used to assess technology implementation processes in order to maximize value of flexibility.

APPENDICES

APPENDIX A: SHIP MAINTENANCE RESOURCES

A.1 INFORMANTS

Sander Alles, Manager Maintenance & Spares, Damen Services, Gorinchem, The Netherlands.

Hein van Ameijden, Managing Director, Damen Schelde Naval Shipbuilding, Vlissingen, The Netherlands.

Nico de Vries, KTZT ir. Head of Corporate Planning and Strategy, Dutch Naval Maintenance and Service Agency. The Netherlands.

Bert Geisler, Business Development Director Shipbuilding, Siemens PLM Software, Hamburg, Germany.

Paul J. Kense, Advisor ILS, Naval Maintenance Establishment, Defense Materiel Organization, Royal Dutch navy, Den Helder, The Netherlands.

Desmond Kramer, Manager, Integrated Logistic Support, Engineering Department, Damen Schelde Naval Shipbuilding, Vlissingen, The Netherlands.

Randy Langmead, Director, Marine/Federal Business Development, Siemens PLM Software, Washington, D.C.

Niek Marse, Integrated Logistic System, Engineering Department, Damen Schelde Naval Shipbuilding, Vlissingen, The Netherlands.

Michael Schwind, Vice President, Federal Sector, Siemens – UGS PLM Software, Philadelphia, Pa.

Ronald van Noppen, KTZE ir. Head of Material and Logistics. Naval Maintenance and Service Agency. The Netherlands.

Frank Verhelst, Manager Project Department, Damen Schelde Naval Shipbuilding, Vlissingen, The Netherlands.

Thijs Verwoerd, Project Manager, Damen Services, Gorinchem, The Netherlands.

APPENDIX B: REFERENCES

- Abdel-Hamid, T. 1988. Understanding the “90% Syndrome” in Software Project Management: A Simulation-Based Case Study. *The Journal of Systems and Software*. 8 319-330.
- Ajah, A. and Herder, P. 2010. Addressing Flexibility During Process and Infrastructure Systems Conceptual Design: Real Options Perspective. In IEEE International Conference on Systems, Man and Cybernetics. 3711-3716.
- Andersen, L. and Broadie, M. 2004. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222-1234.
- Arkali, G., Dawande, >, and Sriskandarajah, C. 2008. Scheduling support times for satellites with overlapping visibilities. *Production and Operations Management*, 17(2):224-234.
- Arlot, S. 2010. A Survey of Cross-Validation Procedures for Model Selection. *Statistics Surveys* 4: 40-79.
- Bakken, B. E. 2004. The Atlantic Defense Technology Gap: Will It Be Closed? Proceedings of the 22nd International Conference of the System Dynamics Society. Oxford, England, The International System Dynamics Society.
- Bakken, B. T. and Gilljam, M. 2003. Dynamic Intuition in Military Command and Control : Why it is Important, and How it Should be Developed. *Cognition, Technology and Work* 5: 197-205.
- Bakken, B. T. and Vamraak, T. 2003. Misperception of Dynamics in Military Planning: Exploring the Counter-Intuitive Behaviour of the Logistics Chain. Proceedings of the 21st International Conference of the System Dynamics Society. R. L. Eberlein, V. G. Diker, R. S. Langer and J. I. Rowe. New York City, USA, The System Dynamics Society.
- Bakken, B. T., Gilljam, M., et al. 2004. Perception and Handling of Complex Problems in Dynamic Settings : Three Cases of Relevance to Military Command and Crisis Management. Proceedings of the 22nd International Conference of the System Dynamics Society. Oxford, England, The System Dynamics Society.
- Bakken, B. T., Ruud, M., et al. 2004. The System Dynamics Approach to Network Centric Warfare and Effects Based Operations : Designing a "Learning Lab" for Tomorrow. Proceedings of the 22nd International Conference of the System Dynamics Society. Oxford, England, The System Dynamics Society.
- Balaban, H., Devers, W., Roark, L. 2010. Feasibility and Advisability of Baselines for O&S Costs: C-17 Case Study, Alexandria, VA: Institute for Defense Analyses.
- Banerjee, P. and de Weck, O. 2004. Flexibility strategy—valuing flexible product options, INCOSE/ICSE Conf Synergy Between Syst Eng Project Management, Las Vegas, NV.
- Barraquand, J. and Martineau, D. 1995. Numerical valuation of high dimensional multivariate American securities. *Journal of Financial and Quantitative Analysis*, 30(3).
- Basar, T. and Olsder, G. 1999. Dynamic Noncooperative Game Theory. SIAM.
- Bartolomei, J. 2001. A System Dynamics Model of Government Engineering Support During the Development Phase of a Military Acquisition Program. Proceedings of the 19th

- International Conference of the System Dynamics Society. Atlanta, Georgia.
International System Dynamics Society.
- Baykasoglu, A. 2009. Quantifying machine flexibility, *Int J Prod Res* 47(15), 4109–4123.
- Bayraksan, G., Morton, D., and Partani, A. 2011. Simulation-Base Optimality Tests for Stochastic Programs, Chapter 3. Springer.
- Bell, J. W. Jr., and Liphard, R.E. 1978. A system dynamics model of the Department of Defense enlisted force for investigation of alternative retirement proposals. Wright-Patterson AFB, OH. (GPO Report Number: ADA065970). Air Force Institute of Technology: 165.
- Bellman, R. 1957. *Dynamic Programming*. Princeton University Press.
- Bellman, R. and Dreyfus, S. 1962. *Applied Dynamic Programming*. Princeton University Press.
- Bengio, Y. and N. Chapados. 2003. Extensions to Metric-Based Model Selection. *Journal of Machine Learning Research* 3: 1209-1227.
- Bengtsson, J. 2001. Manufacturing flexibility and real options: A review. *International Journal of Production Economics*. 74(1-3):213-224. International Conference on Industrial Engineering and Production Management. July 12-15, 1999.
- Bertsekas, D. 2007. *Dynamic Programming and Optimal Control*. Athena Scientific, 3rd edition.
- Bertsekas, D. and Shreve, S. 1978. *Stochastic Optimal Control: The Discrete-Time Case*. Academic Press.
- Birge, J. and Louveaux, F. 2011. *Introduction to Stochastic Programming*. Springer, 2nd edition.
- Black, F. and Scholes, M. 1973. The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3):637-654.
- Boden, B. and Ahlen, A. 1984. *Real Options Analysis: A Study of Implementation Impediments*. Gothenburg: Goteburg University, School of Business, Economics, & Law.
- Boehm, B., Gray, T., and Seewaldt, T. 1984. Prototyping vs. specifying: A multi-project experiment, *Proc Int Conf Software Eng*, pp. 133–145.
- Bordoloi, S., Cooper, W., and Matsuo, H. 1999. Flexibility, adaptability, and efficiency in manufacturing systems, *Prod Oper Management* 8(2), 133–150.
- Boyd, S. and Vandenberghe, L. 2004. *Convex Optimization*. Cambridge University Press.
- Boyle, P., Evnine, J., and Gibbs, S. 1989. Numerical evaluation of multivariate contingent claims. *Review of Financial Studies*, 2(2):241-250.
- Brennan, M. and Schwartz, E. 1977. Convertible Bonds - Valuation and Optimal Strategies for Call and Conversion. *Journal of Finance*, 32(5):1699-1715.
- Broadie, M. and Glasserman, P. 1997. Pricing American-style securities using simulation. *Journal of Economic Dynamics & Control*, 21(8-9):1323-1352. 1995 Derivative Securities Conference, Cornell University, Ithaca, NY, 1995.
- Brown O., Long, A., et al. 2007. System Lifecycle Cost Under Uncertainty as a Design Metric Encompassing the Value of Architectural Flexibility. In *AIAA Space 2007 Conference*. 216-229.
- Brown, O. and Eremenko, P. 2008. Application of value-centric design to space architectures: The case of fractionated spacecraft, *AIAA- 2008-7869* American Institute of Aeronautics and Astronautics, Reston, VA.

- Brown, O. and Eremenko, P. 2009. Value-centric design methodologies for fractionated spacecraft: Progress summary from Phase 1 of the DARPA System F6 Program, AIAA Space 2009 Conf Expo, Pasadena, CA.
- Burman, P. 1994. A Cross-Validatory Method for Dependent Data. *Biometrika* 81(2): 351-358.
- Buzacott, J. A. and Mandelbaum, M. 2008. Flexibility in manufacturing and services: achievements, insights and challenges. *Flexible Services and Manufacturing Journal*, 20(1-2):13-58.
- Carlson, J. and Doyle, J. 2002. Complexity and robustness, Proc Natl Acad Sci USA 99, 2538–2545.
- Caro, I. 1990. Mission Critical Computer Resources Management Guide, Fort Belvoir, Virginia.
- Cayas, C. P. 2007. Northrop wins Navy UCAS bid over Boeing. posted Aug 3, 2007. Visited March 18, 2010.
http://www.navytimes.com/news/2007/08/defense_ucas_northrop_070801/
- Chen, W. and Lewis, K. 1999. Robust design approach for achieving flexibility in multi-disciplinary design, AIAA J 37(8), 982–989.
- Choi, J. Alper, O., Gessner, M., et al. 2009. O&S Cost Forecasting: Issues and Evidence, Center for Naval Analysis.
- Christensen, D., Searle, D., and Vickery, C. 1999. The Impact of the Packard Commission's Recommendations on Reducing Cost Overruns on Defense Acquisition Contracts. *Acquisition Review Quarterly* (Summer): 251-262.
- Chu, C. and J. Marron. 1991. Comparison of Two Bandwidth Selectors with Dependent Errors. *Annals of Statistics* 19(4): 1906–1918.
- Chung, L. and Subramanian, N. 2004. Adaptable architecture generation for embedded systems, *J Syst Software* 71(3), 271–295.
- Clemen, R. and Reilly, T. (2004). Making Hard Decisions with Decision Tools Suite Update Edition. South-Western College Publishing.
- Collopy, P. 2008. Value-Driven Design. *Aerospace America* 46(12): 109.
- Collopy, P. and Hollingsworth, P. 2009. Value-Driven Design. In 9th AIAA Aviation Technology, Integration, and Operations Conference. Hilton Head, SC.
- Cooper, K. G. 1980. Naval ship production: a claim settled and a framework built. *Interfaces*, 10(6), 20-36.
- Cooper, K. G. 1993. The rework cycle: why projects are mismanaged. PM network, PMI, February, 5-7.
- Cooper, K. G. 1993. The rework cycle: how it really works... and reworks. PM network. PMI, February, 25-28.
- Cooper, K. G. 1993. The rework cycle: benchmarks for the project managers. *Project Management Journal*, PMI, 14(1), 17-21.
- Cooper, K.G. 1994. The \$2,000 hour: How managers influence project performance through the rework cycle. *Project Management Journal*. 15(1): 11-24.
- Copeland, T. and Keenan, P. 1998. How Much is Flexibility Worth. *McKinsey Quarterly* (2): 38-49.
- Cormier, P., Olewnik, A., and Lewis, K. 2008. An Approach to Quantifying Design Flexibility for Mass Customization in Early Design Stages. In ASME International Design Engineering Technical Conference. Brooklyn, NY. 203-216.

- Cortazar, G., Gravet, M., and Urzua, J. 2008. The valuation of multidimensional american real options using the lsm simulation method. *Computers & Operations Research*, 35(1):113-129.
- Courtney, V. 2007. A comparable market study of RFID for manual item-level accountability inventory and tracking systems. Master's thesis. Naval Postgraduate School.
- Coyle, R. G. and Gardiner, P.A. 1991. A System Dynamics Model of Submarine Operations and Maintenance Schedules. *Journal of the Operational Research Society* 42(6): 453-462.
- Coyle, R.G. 1996. System dynamics applied to defense analysis: A literature survey. *Defense & Security Analysis*. Vol. 212, Issue 2, pp 141-160. August 1996. Routledge Publishing.
- Curran, M. 1994. Valuing Asian and Portfolio Options by Conditioning on the Geometric Mean Price. *Management Science*, 40(12):1705-1711.
- Defense Acquisition University. 2010. Defense Acquisition Guidebook. [Online]. [cited 2010 December 21. Available from: <https://dag.dau.mil/Pages/Default.aspx>.
- Defense Acquisition University. 2011. Defense Acquisition Guidebook. [Online]. [cited 2011 Feb. Available from: <https://dag.dau.mil/Pages/Default.aspx>.
- DAU. 2012. Defense Acquisition Guidebook. Retrieved February 2012 (<https://dag.dau.mil/Pages/Default.aspx>).
- Diggle, P., Liang, K., and Zeger, S. 1994. *Analysis of Longitudinal Data*. Oxford: Clarendon Press.
- Dixit, A. and Pindyck, R. 1994. *Investment under uncertainty*. Princeton University Press.
- DoD, 1992. DoD 5000.04-M: Cost Analysis Guidance and Procedures.
- DoD, 2007. DoDD 5000.01: The Defense Acquisition System.
- DoD, 2008. DoDI 5000.02: Operation of the Defense Acquisition System.
- DoD, 2009. Weapon System Acquisition Reform Product Support Assessment, USD/ATL,.
- DoD, 2010. Tracking and Assessing Operating and Support Costs for Major Defense Acquisition Programs, Cost Assessment and Program Evaluation.
- Drezner, J., Jarvaise, J., et al. 1993. *An Analysis of Weapon System Cost Growth*. Santa Monica, CA: RAND Corporation.
- Drezner, J. and R. Krop. 1997. *The Use of Baselining in Acquisition Program Management*. RAND, Santa Monica, CA.
- Duczynski, G. 2000. Profiler : An 'Effects-Based' Military Capability Manager. Proceedings of the 18th International Conference of the System Dynamics Society. Bergen, Norway, System Dynamics Society: 59.
- Eckert, C. Clarkson, P., and Zanker, W. 2004. Change and customisation in complex engineering domains, *Res Eng Des* 15(1), 1–21.
- Ekstrom, M. and Bjornsson, H. 2005. Valuing Flexibility in Architecture, Engineering, and Construction Information Technology Investments. *Journal of Construction Engineering & Management* 131(4): 431-438.
- Filar, J. and Vrieze, K. 1996. *Competitive Markov Decision Processes*. Springer.
- Fitzgerald, G. Barad, M., Papazafeiropoulou, A., and Alaa, G. 2009. A framework for analyzing flexibility of generic objects, *Int J Prod Econ* 122(1), 329–339.
- Fliege, J., Kaparis, K., and Khosravi, B. 2012. Operations research in the space industry. *European Journal of Operational Research*, 217(2):233-240.
- Flood, R.L. and Jackson, M.C. 1991. *Creative Problem Solving: Total Systems Intervention*.

Wiley: Chichester.

- Ford, D. 2002. Achieving multiple project objectives through contingency management. *ASCE Journal of Construction Engineering and Management*. 128(1): 30-39.
- Ford, D. and Sterman, J. 1998. Modeling Dynamic Development Processes. *System Dynamics Review*. 14(1),31-68. Spring, 1998.
- Ford, D. and Sterman, J. 2003. Overcoming the 90% Syndrome: Iteration Management in Concurrent Development Projects, *Concurrent Engineering Research and Applications*. Vol. 111, No. 3, pp. 177-186.
- Ford, D. and Sterman, J. 2003. The Liar's Club: Impacts of Concealment in Concurrent Development Projects, *Concurrent Engineering Research and Applications*. Vol. 111, No. 3, pp. 211-219.
- Ford, D. N., Housel, T., and Dillard, J. 2010. System Dynamics Modeling for Improved Knowledge Value Assessment: A Proof of Concept Study. Acquisition Research Program. NPS Report No. NPS-GSBPP-10-0015. U.S. Naval Postgraduate School, August 9, 2010.
- Ford, D. N., Housel, T., and Mun, J. 2011. Ship Maintenance Processes with Collaborative Product Life Cycle Management and 3D Terrestrial Laser Scanning Tools: Reducing Costs and Increasing Productivity. *Proceedings of 2011 Acquisition Research Symposium*. U.S. Naval Postgraduate School, May 11-12, 2011.
- Ford, D.N. and Dillard, J. 2009. Modeling the Performance and Risks of Evolutionary Acquisition Defense Acquisition Review Journal. Vol. 16, No. 2. Pp. 143-158.
- Ford, D.N. and Dillard, J.T. 2009. Modeling Open Architecture and Evolutionary Acquisition in ARCI with Applications to RCIP. 6th Acquisition Research Symposium. Naval Postgraduate School. Monterey, CA.
- Ford, D.N. and Dillard, J.T. 2008. Modeling the Integration of Open Systems and Evolutionary Acquisition in DoD Programs. 5th Acquisition Research Symposium. Naval Postgraduate School. Monterey, CA.
- Forrester, J. W. 1961. *Industrial Dynamics*. Productivity Press. Portland OR.
- Fricke, E. and Schulz, A. 2005. Design for changeability: Principles to enable changes in systems throughout their entire lifecycle, *Syst Eng* 8(4), 342–359.
- Fudenberg, D. and Tirole, J. 1991. *Game Theory*. MIT Press, Cambridge, MA.
- Gavish, B. 1997. Low earth orbit satellite based communication systems – Research opportunities. *European Journal of Operational Research*, 99(1):166-179. 14th European Conference on Operational Research, (EURO XIV), Jerusalem, Israel, July 3-6, 1995.
- Gebauer, J. and Lee, F. 2007. Enterprise system flexibility and implementation strategies: Aligning theory with evidence from a case study, *Inform Syst Management* (25), 71–82.
- Geman, S. and Geman, D. 1984. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(6):721-741.
- Gershenson, J., Prasad, G., and Zhang, Y. 2003. Product modularity: Measures and design methods, *J Eng Des* 15(1), 33–51.
- Geske, R. and Johnson, H. 1984. The american put option valued analytically. *Journal of Finance*, p. 1511-1524.
- Global Security. 2010. Unmanned Aerial Vehicles (UAVs), March 22, 2010.

- <http://www.globalsecurity.org/intell/systems/uav-intro.htm>
- Government Accountability Office. 2009. DEFENSE ACQUISITIONS: Assessments of Selected Weapons Programs. Washington, D.C.: GAO.
- Government Accountability Office, 2010. DoD Needs Better Information and Guidance to More Effectively Manage and Reduce Operating and Support Costs of Major Weapon Systems, Washington, D.C.: Report No.: GAO-10-717.
- GPO. 2011. Budget Control Act of 2011. U.S. Government Printing Office. (<http://www.gpo.gov/fdsys/pkg/PLAW-112publ25/pdf/PLAW-112publ25.pdf>).
- Guo, X. and Zhang, Q. 2004. Closed-form solutions for perpetual american put options with regime switching. *SIAM Journal on Applied Mathematics*, pages 2034-2049.
- Haberfellner, R. and de Weck, O. 2005. Agile SYSTEMS ENGINEERING versus AGILE SYSTEMS engineering, Fifteenth Annu Int Symp INCOSE, Rochester, NY.
- Harvey, J., Lefebvre, L., and Lefebvre, E. (1997). Flexibility and technology in services: A conceptual model. *International Journal of Operations & Production Management*, 17(1-2):29-&.
- Hart, J. and T. Wehrly. 1986. Kernel Regression Estimation Using Repeated Measurements Data. *Journal of the American Statistical Association* 81(396): 1080–1088.
- Hastings, D. and McManus, H. 2004. Framework for understanding uncertainty and its mitigation and exploitation in complex systems, MIT Eng Syst Symp, Cambridge, MA.
- Haubelt, C., Teich, J., Richter, K., and Ernst, R. 2002. System design for flexibility, In Proc. Design Automation and Test in Europe (2002), 854–861. INCOSE, Systems engineering handbook v3.2.2, INCOSE-TP- 2003-002-03.2.2, San Diego, CA, 2011.
- Hebert, A. 2011. Lies, Damn Lies, and the Trillion-Dollar F-35. *Air Force Magazine*, July, pp. 4.
- Homer, J. B. and Somers, I. 1988. Defense Program Lifecycle Management: A Dynamic Model for Policy Analysis. Proceedings of the 1988 International System Dynamics Conference J. B. Homer and A. Ford. La Jolla, California, International System Dynamics Society: 160.
- Housel, T. and Bell, A. (2001). Measuring and Managing Knowledge. McGraw-Hill, New York.
- Housel, T. and Mun, J. 2009. SHIPMAIN: Creating flexibility with information technology options. Naval Postgraduate School Working Paper.
- Housel, T. J. and Kanevsky, V. 1995. Reengineering Business Processes: A Complexity Theory Approach to Value Added, *INFOR*, Vol. 33, pp. 248-262.
- Housel, T. J., and Kanevsky, V. 1995. A New Methodology for Business Process Auditing, *Planning Review*, May-June, pp.31–36.
- Housel, T. J., Tarantino, E., Hom, S., and Mun, J. 2008. Shipyard Planning Case, in *Making Cents Out of Knowledge Management*, Jay Liebowitz, Lanham, Maryland: The Scarecrow Press, Inc.
- Hull, J. 2011. Options, Futures, and Other Derivatives. Prentice Hall, 8th edition.
- Ikonen, S. and Toivanen, J. 2004. Operator splitting methods for American option pricing. *Applied Mathematics Letters*, 17(7):809-814.
- INCOSE. 2011. Systems Engineering Handbook v.3.2.2, INCOSE-TP-2003-002-03.2.2, San Diego, CA.
- Jackson M.C. 2003. *Systems Thinking: Creative holism for managers*. Wiley: Chichester.
- Joglekar, N. and Ford, D. 2005. Product development resource allocation with foresight.

- European Journal of Operational Research*. 160(1): 72-87.
- Johnson, H. 1987. Options on the maximum or the minimum of several assets. *Journal of Financial and Quantitative Analysis*, 22(03):277-283.
- Jones, C. A. 1997. Unmanned Aerial Vehicles (UAVs), An Assessment of Historical Operations and Future Possibilities. Research Department, Ari Command and Staff College.
- Jordan, W. and Graves, S. 1995. Principles on the Benefits of Manufacturing Process Flexibility. *Management Science*, 41(4):577-594.
- Jung, B.M. and Baek, D.H. 2009. Estimating the ROI on Implementation of RFID at the ammunition storage warehouse and the 40th supply depot: KVA methodology. Master's thesis. Naval Postgraduate School.
- Keese, D. 2007. Empirically-derived principles for designing products with flexibility for future evolution, Proc ASME Int Des Eng Tech Conf, Las Vegas, NV.
- Komoroski, C.L. 2005. Reducing Cycle Time and Increasing Value through the Application of Knowledge Value Added Methodology to the U.S. Navy Shipyard Planning Process. Report NPS-AM-06-003.
- Koskinen, J. 2010. Software Maintenance Costs, Jyväskylä: University of Jyväskylä. URL: <http://users.jyu.fi/~koskinen/smcosts.htm>
- Kovalov, P., Linetsky, V., and Marcozzi, M. 2007. Pricing multi-asset american options: A finite element method-of-lines with smooth penalty. *Journal of Scientific Computing*, 33(3):209-237.
- Krishnan, V. and Bhattacharya, S. 2002. Technology selection and commitment in new product development: The role of uncertainty and design flexibility, *Management Sci* 48(3), 313–327.
- Kumar, R. 1999. Understanding DSS Value: An Options Perspective. *Omega* 27(3): 295-304.
- Kushner, H. and Dupuis, P. 2000. Numerical methods for stochastic control problems in continuous time. Springer, 2nd Ed.
- Lafleur, J. and Saleh, J. 2009. Survey of intra- and inter-mission flexibility in space exploration systems, *Acta Astronaut* 67(1), 97– 107.
- Lane, D.C. and Jackson, M.C. 1995. Only Connect! An annotated bibliography reflecting the breadth and diversity of systems thinking. *Systems Research* 12: 217-228.
- Larson, S. 1931. The Shrinkage of the Coefficient of Multiple Correlation. *Journal of Educational Psychology* 22: 45-55.
- Larsson, S. and Thomee, V. 2008. Partial differential equations with numerical methods. Springer.
- Lee, Z., Ford, D., and Joglekar, N. 2007. Effects of resource allocation policies for reducing project durations: A systems modeling approach. *Systems Research and Behavioral Science*. Vol. 24, No. 6, pp. 551-566.
- Longstaff, F. and Schwartz, E. 2001. Valuing american options by simulation: a simple least-squares approach. *Review of Financial Studies*, 14(1):113-147.
- Lyneis, F., Cooper, K., Els, S. 2001. Strategic management of complex projects: a case study using system dynamics. *Systems Dynamics Review*. 17(3): 237 – 260.
- Lyneis, J.M. and Ford, D.N. 2007. System Dynamics Applied to Project Management: A survey, assessment, and directions for future research. *System Dynamics Review*, Vol. 23, No. 4 pp. 157-189.
- Markowitz, H. M. 1952. Portfolio Selection, *The Journal of Finance*, vol. 7, no. 1 (March), pp.

77–91.

- Marks, W. 2010. X-47B Unmanned Combat Air System Taking Shape On Board Lincoln. posted Feb., 2, 2010. Visited March 22, 2010.
http://www.navy.mil/search/display.asp?story_id=51239
- Mayer, Z. and Kazakidis, V. 2007. Decision Making in Flexible Mine Production System Design Using Real Options. 133(2): 169-180.
- McKean, H. 1965. Appendix: A free boundary problem for the heat equation arising from a problem in mathematical economics. *Industrial Management Review*, 6:32-39.
- McLucas, A. C., Lyell, D., et al. 2006. Defence Capability Management: Introduction Into Service of Multi-Role Helicopters. Proceedings of the 24th International Conference of the Systems Dynamics Society. Nijmegen, The Netherlands, The Systems Dynamics Society: 92.
- Melhuish, J., Pioch, N., et al. 2009. Improving Military Strategy using Predictive Agents with Embedded Mental Simulation Models The 27th International Conference of the System Dynamics Society, Albuquerque, New Mexico, System Dynamics Society.
- Mikkonen, T. and Pruuden, P. 2001. Flexibility as a design driver [systems analysis], *Computer* 34(11), 52–56.
- Mun, J. 2006. *Real Options Analysis: Tools and Techniques*, Second Edition. Hoboken, New Jersey: Wiley.
- Mun, J. 2010. *Modeling Risk: Applying Monte Carlo Risk Simulation, Strategic Real Options, Stochastic Forecasting, and Portfolio Optimization*, Second Edition. Hoboken, New Jersey: Wiley
- Mun, J., Housel, T., and Wessman, M. 2009. Advanced Capability Builds: Portfolio Optimization, Selection and Prioritization, Risk Simulation, KVA, and Strategic Real Options Analysis. *Naval Postgraduate School Report NPS-AM-09-144*.
- Nachtwey, A., Riedel, R., and Mueller, E. 2009. Flexibility oriented design of production systems, *Comput Indust Eng*, 720–724.
- Naidu, D. 1991. Fuel-Optimal Trajectories of Aeroassisted Orbital Transfer with Plane Change. *IEEE Transactions on Aerospace and Electronic Systems*, 27(2):361-369.
- NASA. 2008. Report of Audit and Finance Committee. NASA Advisory Council. Retrieved May 2012 (http://www.nasa.gov/pdf/314880main_AFC_KSC_NAC_Feb-5-2009.pdf).
- Nelson, M. 2010. Information Technology Portfolio Management Proof of Concept: modern portfolio theory with KVA and ROI analysis. Master's thesis. NPS. NPS report
- Nilchiani, R. and Hastings, D. 2003. Measuring flexibility in design of an orbital transportation network, Space 2003, Long Beach, CA.
- Nilchiani, R. and Hastings, D. 2006. Measuring the value of flexibility in space systems: A six-element framework, Wiley Interscience, Hoboken, NJ.
- Northrup Grumman. 2010. X-47B UCAS , visited March 18, 2010
<http://www.as.northropgrumman.com/products/nucasx47b/index.html>
- Office of Satellite Operations. 2012. Geostationary satellites. Technical report.
<http://www.oso.noaa.gov/goes/index.htm>.
- Oksendal, B. 2003. Stochastic differential equations: an introduction with applications. Springer Verlag.
- Oleson, J. 1998. Pathways to agility, Wiley, New York.
- Olewnik, A. and Kemper, L. 2006. A Decision Support Framework for Flexible System

- Design. *Journal of Engineering Design* 17(1): 75-97.
- Olewnik, A. and Lewis, K. 2006. A decision support framework for flexible system design, *J Eng Des* 17(1), 75–97.
- Olewnik, A., Brauen, T., Ferguson, S., and Lewis, K. 2001. A framework for flexible systems and its implementation in multiattribute decision making, *Proc ASME Des Eng Tech Conf*, Pittsburgh, PA.
- Opsomer, J., Y. Wang, and Y. Yang. 2001. Nonparametric Regression with Correlated Errors. *Statistical Science* 16(2): 134–153.
- OSD CAIG. 2007. Operating & Support Cost Estimating Guide. Acquisition Community Connection. Office of the Secretary of Defense Cost Analysis Improvement Group. Retrieved June 2012 (https://acc.dau.mil/adl/en-US/142233/file/27619/O_S_Cost_Estimating_Guide_Oct_2007.pdf).
- Patetta, M. 2002. Longitudinal Data Analysis with Discrete and Continuous Responses Course Notes. Cary, NC: SAS Institute.
- Pavlou, P., Housel, T. J., Rodgers, W., and Jansen, E., 2005. Measuring the Return on Information Technology: A Knowledge-based Approach for Revenue Allocation at the Process and Firm Level, *Journal of Association of Information Systems*, Vol. 6, No. 7, pp. 199-226.
- Puterman, M. 2005. Markov Decision Processes. Wiley-Interscience, New York.
- Qureshi, A., Murphy, J., et al. 2006. Principles of product flexibility, *ASME Int Des Eng Tech Conf*.
- Raiffa, H. and Schlaifer, R. 1961. Applied Statistical Decision Theory. Harvard University, Boston.
- Rajan, P., Van Wie, M., Campbell, M., and Otto, K. 2003. Design for flexibility—measures and guidelines, *Int Conf Eng Des*, Stockholm, Sweden.
- Rajan, P., Van Wie, M., et al. 2005. An empirical foundation for product flexibility, *Des Stud* 26(4), 405–438.
- Redman, Q., Crepea, A.T., Stratton, G. 2008. Weapon Design Trade Off Using Life Cycle Costs, Raytheon Corporation, URL: http://www.galorath.com/images/uploads/3_-_Quentin_Redman_-_Design_trades_using_Life_Cycle_Costs_short_version.pdf
- Rodgers, W. and Housel, T. J. 2006. Improvement of global performance measures related to intangible assets, collected papers of the 32nd European International Business Academy (EIBA) Conference (Refereed Proceedings), December 7-9, Fribourg Switzerland.
- Roser, C. and Kazmer, D. 1999. Risk effect minimization using flexible product and process design, *Proc 4th Des Manuf Conf*, Las Vegas, NV.
- Ross, A., Rhodes, D., and Hastings, D. 2008. Defining changeability: Reconciling flexibility, adaptability, scalability, modifiability, and robustness for maintaining system lifecycle value, *Syst Eng* 11(3), 246–262.
- Rust, J. 1997. Using randomization to break the curse of dimensionality. *Econometrica*, 65(3):487-516.
- Ryan, E., Jacques, D., and Colombi, J. 2011. Defining Flexibility and Flexibility-Related Terminology. In *Conference on Systems Engineering Research*. Los Angeles, CA.
- Ryan, E., Jacques, D., Ritschel, J., and Schubert, C. 2012. Characterizing the Accuracy of DoD Operating and Support Cost Estimates. *Journal of Public Procurement* (Accepted, but pending publication).

- Saleh, J. Hastings, D., and Newman, D. 2003. Flexibility in system design and implications for aerospace systems, *Acta Astronaut* 53(12), 927–944.
- Saleh, J., Mark, G., and Jordan, N. 2009. Flexibility: A multi-disciplinary literature review and a research agenda for designing flexible engineering systems, *J Eng Des* 20(3), 307–323.
- Schulz, A. and Fricke, E. 1999. Incorporating flexibility, agility, robustness, and adaptability within the design of integrated systems—key to success? IEEE, New York.
- Seaman, N. Housel, T. J. and Mun, J. 2007. The Potential Impact of Collaborative and Three-dimensional Imaging Technology on SHIPMAIN Fleet Modernization Plan, Naval Postgraduate School Report NPS-AM-07-103.
- Sen, S. and Hige, J. 2005. The C(3) theorem and a D(2) algorithm for large scale stochastic mixed-integer programming: Set convexification. *Mathematical Programming*, 104(1):1120.
- Sethi, A., and Sethi, S. 1990. Flexibility in Manufacturing: A Survey. *The International Journal of Flexible Manufacturing Systems*, 2:289-328.
- Shachter, R. and Mandelbaum, M. 1996. A measure of decision flexibility. Proceedings of the twelfth conference on uncertainty in artificial intelligence. July 21-August 03. San Mateo, CA.
- Shah, N., Viscito, L., Wilds, J., Ross, S., and Hastings, D. 2008. Quantifying flexibility for architecting changeable systems, 6th Conf Syst Eng Res, Los Angeles, CA.
- Shapiro, A. 2003. Monte Carlo sampling methods. Elsevier, Amsterdam.
- Shapiro, A., Dentcheva, D., and Ruszczyński, A. 2009. Lectures on Stochastic Programming: Modeling and Theory. SIAM - Society for Industrial and Applied Mathematics.
- Sieger, D., Badiru, A., and Milatovic, M. 2000. A metric for agility measurement in product development, *IIE Trans* 32, 637–645.
- Sivanthi, T. and Killat, U. 2008. Valuing the design flexibility of a distributed real-time embedded system, *Int Symp Indust Embedded Syst*.
- Sterman, J.S. 2000. *Business Dynamics: Systems Thinking and Modeling for a Complex World*. Irwin McGraw-Hill, New York.
- Suh, E., de Weck, O., and Chang, D. 2007. Flexible Product Platforms: Framework and Case Study. *Research in Engineering Design*. 67-89.
- Suh, N. 1998. Axiomatic design theory for systems, *Res Eng Des* 10, 189–209.
- Taguchi, G. 1993. *Taguchi on robust technology*, ASME Press, New York.
- Taylor, T. and Ford, D. 2006. Tipping Point Dynamics in Development Projects. *System Dynamics Review*. Vol. 22, No. 1. pp. 51-71.
- Taylor, T.R. and Ford, D.N. 2008. Managing Tipping Point Dynamics in Development Projects. *ASCE Journal of Construction Engineering and Management*. Vol. 134, No. 6, pp. 421-431.
- Thomke, S. 1997. The role of flexibility in the development of new products: An empirical study, *Res Policy* 26(1), 105–119.
- Topper, J. 2005. *Financial Engineering with Finite Elements*. Wiley.
- Trigeorgis, L. 1996. *Real options: Managerial flexibility and strategy in resource allocation*. MIT Press.
- Verbeke, G. and Molenberghs, G. 2000. *Linear Mixed Models for Longitudinal Data*. New York: Springer.
- Viscito, L. and Ross, A. 2009. Quantifying flexibility in tradespace exploration: Value

- weighted filtered outdegree, AIAA Space 2009 Conf Expo, Pasadena, CA.
- Wijker, J. 2009. *Random Vibrations in Spacecraft Structures Design: Theory and Applications*. Springer.
- Wikipedia. 2012. List of ships of the Royal Netherlands Navy, http://en.wikipedia.org/wiki/List_of_ships_of_the_Royal_Netherlands_Navy. Accessed August 12, 2012.
- Wolfinger, R. 1993. Covariance Structure Selection in General Mixed Models. *Communications in Statistics, Simulation and Computation* 22(4): 1079–1106.
- Younossi, O., Arena, M., et al. 2007. Is Weapon System Cost Growth Increasing? Santa Monica, CA: RAND Corporation.
- Zhu, K.-J., Li, J.-F., and Baoyin, H.-X. (2010). Satellite scheduling considering maximum observation coverage time and minimum orbital transfer fuel cost. *Acta Astronautica*, 66(1-2):220-229.

APPENDIX C: AFFORDABLE SYSTEMS: BALANCING THE CAPABILITY, SCHEDULE, FLEXIBILITY AND TECHNICAL DEBT TRADESPACE

Jo Ann Lane, Supannika Koolmanojwong, and Barry Boehm
Center for Systems and Software Engineering, University of Southern California

Introduction

The ultimate goal of today's systems and systems of systems (SoS) is to provide capabilities to the stakeholders and users of the systems. These capabilities range from "must-haves" to "nice-to-haves", often with disagreements among the stakeholders and users as to where each capability lays in this spectrum. There are many choices in developing and evolving systems to provide these desired capabilities, whether it is in the commercial space, Department of Defense (DoD) space, or other government space. These choices are typically related to development processes and product architecture decisions. Initial choices are often driven by business needs such as time to market, the desired level of performance of the capabilities, and available resources such as engineering expertise and funding. In addition, these choices often result in longer-term consequences that range from good (e.g., market share or future opportunities) to bad (e.g., missed market share, technical debt, or a failure to provide the desired capability). Other times, no particular attention is paid to these choices—they happen without much forethought, but still with the resulting longer-term consequences. Finally, there is often not an optimal set of choices, but rather the engineering team needs to evaluate the stakeholder needs and make trade decisions that sufficiently balance competing needs. This paper looks at the capability affordability tradespace of expediting systems engineering to reduce schedule and cost, encouraging flexibility in architecture decisions to support future evolution of the system, and technical debt that either results in later rework or adversely impacts future options. In addition, this paper shows how the University of Southern California (USC) Center for Systems and Software Engineering (CSSE) software and systems engineering cost models can be used in the analysis of this tradespace to show the range of options and the resulting consequences.

Background

The following discusses and characterizes each of the tradespace aspects considered in this paper.

Expedited Engineering Overview. Even though there is considerable evidence to the contrary, system sponsors and stakeholders continue to encourage developers to take shortcuts early in the development process in order to get system capabilities deployed quickly. These shortcuts are often done under the guise of agile processes with the thought that any resulting problems can be fixed later. However, the real goal is to get quality capabilities deployed quickly that do not overly constrain future evolution of the system. So the ultimate goal is to expedite "system development" which includes the upfront engineering to design a system or system capability that meets the users need

and well as build, test, and deploy that system or system capability in the shortest time possible. Ways to expedite development include:

- Minimal engineering/quick solutions with small, expert teams but that often result in increased technical debt
- Lean approaches that eliminate non-value adding activities, reduce wait times, and a “go-slow approach” at the start to establish a good foundation or architecture, well-defined interfaces, and relatively low complexity, then go fast during the build and test phases.

Flexibility Overview. The goal of “flexibility” is to focus on developing a robust foundation for the system that will provide the ability to easily modify existing system capabilities as well as expand system capabilities to meet future needs or allow the system to easily interconnect with other systems to support cross cutting capabilities in systems of systems. The challenge in pursuing system flexibility is to balance flexibility and complexity. For example, performance issues may result if system tries to be “everything for everyone”.

Technical Debt Overview. Technical debt is a term coined by Ward Cunningham to describe delayed technical work or rework that is incurred when shortcuts are taken. It is often the rework or unfulfilled outcomes/capabilities that result from insufficiently (or poorly) engineered or implemented solutions. In the case of expedited engineering, it can increase as shortcuts are employed to reduce schedule. Examples include insufficiently engineered architectures and the lack of focus on quality checks and reviews. In the case of flexibility, it may be related to unrecognized performance, security limitations, or excessive complexity that must be later addressed.

Initial Tradespace Analyses

Considerable research is currently being conducted in each of these areas (expedited engineering, flexibility, and technical debt). However, most of the research identified to date is looking at each of these areas in a stove-piped fashion, not as a set of features or outcomes to be balanced. The following describes recent research in each of these areas, some of which is currently being done through Stevens-USC Systems Engineering Research Center (SERC) tasks.

Expedited Engineering Analyses. An important aspect of the system acquisition process is understanding how long it will take to provide the new needed system or system capability. Considerable work has been done in this area to better understand system/capability development schedules, with considerable contributions from parametric cost models for software-intensive systems that are based on historical data. However, many government organizations, including the Government Accountability Office (GAO), have pointed out how much money is wasted on unsuccessful system development programs that often fail to provide the necessary capabilities or if they do provide the capabilities, are late and cost much more than expected [17]. As a result, the

Department of Defense has realized that it is often no longer cost-effective to develop new systems (or system capabilities) using traditional processes and is encouraging researchers to find ways to expedite systems engineering and development. Currently, the engineering community is embracing agile and lean processes as well as system of systems engineering to provide new capabilities through the integration and enhancement of existing systems as ways to rapidly respond to immediate needs. This, though, can lead to single-point solutions that are not flexible enough to meet the next set of needs or incur technical debt due to extensive rework and maintenance costs.

Key Approaches for Expedited Engineering	
Commercial-Off-the-Shelf (COTS) Products	} <i>Using the right people</i>
Investment in product-line architectures	
Reuse of existing systems/components	
Repurposing existing systems/components	
Value-stream focus (lean)	
Going fast in general (crisis response)	
Single purpose architecture	

Others are beginning to employ Kanban techniques to better manage work flows. To date, this has been primarily in the area of software development [1], but current research is investigating ways to employ Kanban for capability engineering work using a hierarchy of Kanbans for both systems and software engineering to improve work flow and minimize wait times once the proper foundations are in place for developing and evolving system or SoS capabilities.

A few of the larger, more success system development organizations are investing in infrastructure and product lines [9, 15] so that they can respond quickly when new opportunities or world events occur that require immediate solutions. These suppliers realize a return on their investments when they have the needed flexibility built into their system infrastructure and have minimal technical debt due to the quality of the infrastructure core built in from the start.

To better understand opportunities to expedite systems engineering and software development, one can look at the cost factors in the associated USC CSSE cost models and their influence on productivity. Figures 1 and 2 illustrate these factors for the systems engineering cost model, COSYSMO, and the software development cost model, COCOMO, respectively.

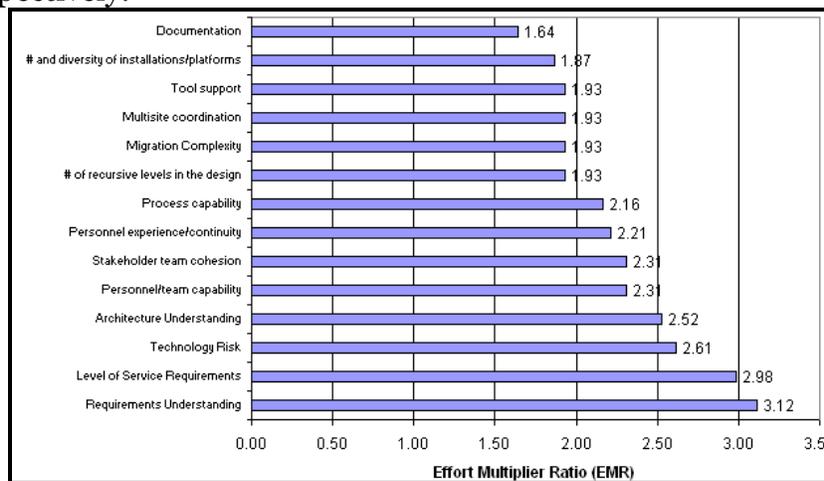


Figure 1: COSYSMO Effort Multiplier Ratio [18]

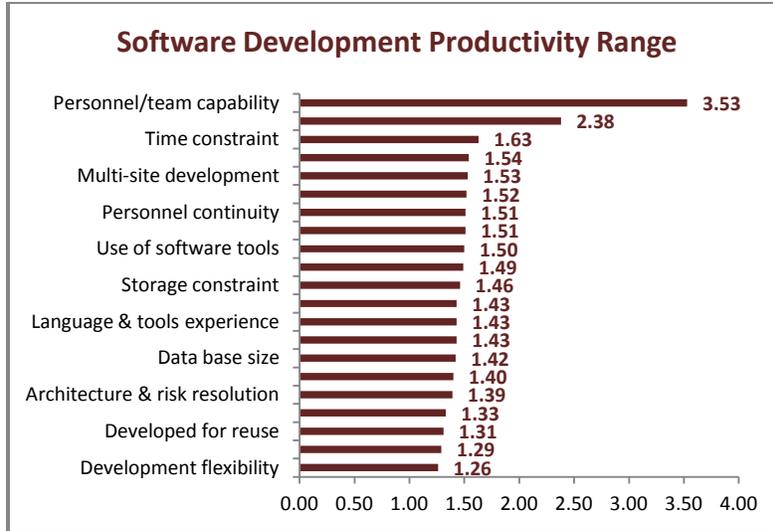


Figure 2: COCOMO software productivity ranges [2]

The greatest gains in effort savings and schedule can be made by focusing engineering improvements in the areas of the high-influence cost factors.

System Flexibility Analyses. Without a certain level of flexibility, systems can quickly become obsolete as technology and stakeholder/mission needs change. However, if too much effort is spend on developing the most flexible systems, it may be at the expense of delivery expediency and simplicity of the system core, resulting in longer development schedules for the first incarnation of the system and poorer operational performance due to the need to sacrifice single capability performance for the ability to perform multiple capabilities (or perform a single capability in multiple environments). Current SERC research in this area is focusing on evaluation of total ownership costs and options analysis. Preliminary total ownership cost analysis, illustrated in figure 3, has shown potential cost savings when upfront investments are made in architecture.

Key Approaches for Incorporating Flexibility

- Employ open architectures
- Design for reuse
- Develop/use product lines



Figure 3: TOC's for 3 Projects Relative to Baseline Costs [3]

However, as shown in table 1, many of the flexibility architecture strategies may present conflicts with other desired system characteristics such as performance, human controllability, increased development schedules and costs, and security, to name a few.

Table 1: Architecture-Based Attribute Trades with Respect to Flexibility

Architecture Strategy	Synergies	Conflicts
High module cohesion Low module coupling	<ul style="list-style-type: none"> • Interoperability • Reliability 	<ul style="list-style-type: none"> • Impacts high performance achieved via tight coupling
Service-oriented architecture	<ul style="list-style-type: none"> • Composability • Usability • Testability 	<ul style="list-style-type: none"> • Impacts high performance achieved via tight coupling • Requires additional infrastructure resulting in added costs/schedule
Autonomous adaptive systems	<ul style="list-style-type: none"> • Affordability via task automation • Response time 	<ul style="list-style-type: none"> • Excess autonomy reduces human controllability • Requires additional analysis and testing to identify emergent behaviors
Modularization around sources of change	<ul style="list-style-type: none"> • Interoperability • Usability • Reliability • Availability 	<ul style="list-style-type: none"> • Extra time on critical path of rapid fielding
Multi-layered architecture	<ul style="list-style-type: none"> • Reliability • Availability 	<ul style="list-style-type: none"> • Lower performance due to layer traversal overhead • Potential integration issues with other architecture styles
Many built-in options, entry points	<ul style="list-style-type: none"> • Functionality • Accessibility 	<ul style="list-style-type: none"> • Reduced usability via options proliferation • Harder to secure
User programmability	<ul style="list-style-type: none"> • Usability • Mission effectiveness 	<ul style="list-style-type: none"> • Full programmability causes reliability, compatibility, interoperability, safety, security risks
Spare/expandable capacity	<ul style="list-style-type: none"> • Performance • Reliability 	<ul style="list-style-type: none"> • Added cost • Usefulness may be limited by rapidly changing technologies
Product line architecture Reusable components	<ul style="list-style-type: none"> • Cost • Schedule • Reliability 	<ul style="list-style-type: none"> • Some loss of performance vs. optimized stovepipes

At the SoS level, there has been a considerable focus on how enable systems to quickly connect to perform desired cross-cutting mission capabilities, then return to their normal single-system missions. This has become more important over time as systems come together at many levels (e.g., joint services and international coalitions) and as these systems participate in multiple SoSs. Some [16] advocate for the migration to a set of standard convergent protocols to enable the needed interconnectivity and others [12] are developing techniques to evaluate and improve interoperability across a set of systems that must interoperate as an SoS.

Managing Technical Debt Analyses. Recent research by Steven McConnell [14], Practical Systems and Software Measurement (<http://www.psmc.com/>) affiliates, and others are focusing on a concept referred to as “technical debt” and ways to better manage it in the development of systems and software. Ward Cunningham [6] first coined the term and further explained it in his YouTube video [7]. Technical debt refers

to delayed technical work or rework that is incurred when shortcuts are taken. Some technical debt is reasonable given time-to-market or other urgent constraints. But over the long term, if the system is to be sustained, the technical debt must be paid back, often with interest (i.e., it is more expensive to fix than it is to do it right the first time). Deferred long-term technical debt can result in fragile, error-prone systems that take excessive time to modify, with more time spent on maintaining existing capabilities than adding or improving capabilities.

Common Causes of Technical Debt

- Pressures to compress schedule
- Lack of requirements understanding
- Lack of system understanding
- Inflexible architectures/software
- Overly complex design/implementation
- Delayed defect resolution
- Inadequate testing
- Lack of current documentation
- Parallel development in isolation

Expedited Engineering vs. Valuing Flexibility

The focus of this research effort used the COCOMO suite of cost models illustrated in Figure 4 to compute estimated effort and schedule for a given project with different project drivers such as expedited engineering and valuing flexibility.

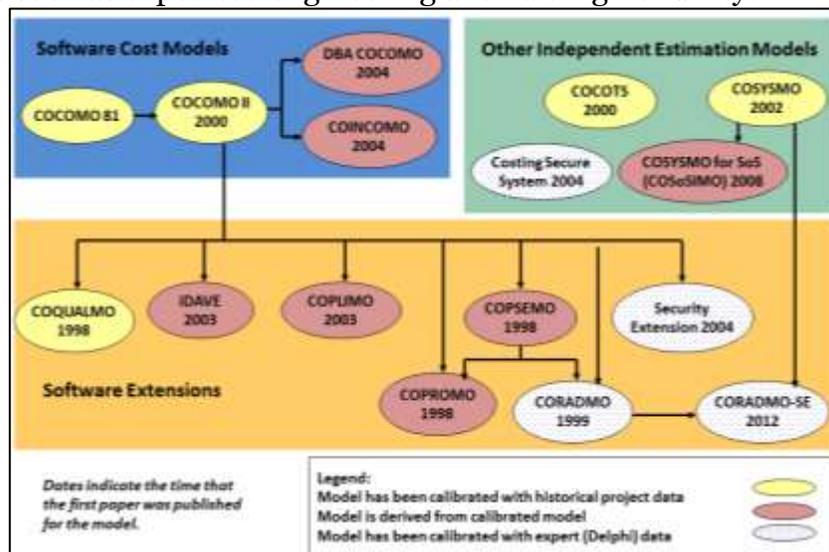


Figure 4: Overview of USC CSSE Cost Models.

For the first case, the project drivers characterized expedited engineering. For the second case, the project driver characterized the development of a flexible product. This comparison illustrates how the USC CSSE cost models can be used to calculate the associated engineering effort and schedule for each of these cases. Key to this analysis is the most recent addition to the COCOMO suite of cost models, CORADMO-SE, that can be used to calculate the savings in schedule for systems engineering. Table 2 shows the CORADMO-SE factors and associated weights. This model calculates a “rapid” factor that can be applied to the Constructive Systems Engineering Cost Model (COSYSMO) estimated schedule.

Table 2: CORADMO-SE Cost and Schedule Factors.

Accelerators/Ratings	Very Low	Low	Nominal	High	Very High	Extra High
Product Factor: Multipliers	1.09	1.05	1.0	0.96	0.92	0.87
Simplicity	Extremely complex	Highly complex	Mod. complex	Moderately simple	Highly simple	Extremely simple
Element Reuse	None (0%)	Minimal (15%)	Some (30%)	Moderate (50%)	Considerate (70%)	Extensive (90%)
Low-Priority Deferrals	Never	Rarely	Sometimes	Often	Usually	Anytime
Models vs Documents	None (0%)	Minimal (15%)	Some (30%)	Moderate (50%)	Considerate (70%)	Extensive (90%)
Key Technology Maturity	>0 TRL 1,2 or >1 TRL 3	1 TRL 3 or > 1 TRL 4	1 TRL 4 or > 2 TRL 5	1-2 TRL 5 or >2 TRL 6	1-2 TRL 6	All > TRL 7
Process Factor: Multipliers	1.09	1.05	1.0	0.96	0.92	0.87
Concurrent Operational Concept, Requirements, Architecture, V&V	Highly sequential	Mostly sequential	2 artifacts mostly concurrent	3 artifacts mostly concurrent	All artifacts mostly concurrent	Fully concurrent
Process Streamlining	Heavily bureaucratic	Largely bureaucratic	Conservative bureaucratic	Moderate streamline	Mostly streamlined	Fully streamlined
General SE tool support CIM (Coverage, Integration, Maturity)	Simple tools, weak integration	Minimal CIM	Some CIM	Moderate CIM	Considerable CIM	Extensive CIM
Project Factors: Multipliers	1.08	1.04	1.0	0.96	0.93	0.9
Project size (peak # of personnel)	Over 300	Over 100	Over 30	Over 10	Over 3	≤ 3
Collaboration support	Globally distributed weak comm. , data sharing	Nationally distributed, some sharing	Regionally distributed, moderate sharing	Metro-area distributed, good sharing	Simple campus, strong sharing	Largely collocated, Very strong sharing
Single-domain MMPTs (Models, Methods, Processes, Tools)	Simple MMPTs, weak integration	Minimal CIM	Some CIM	Moderate CIM	Considerable CIM	Extensive CIM
Multi-domain MMPTs	Simple; weak integration	Minimal CIM	Some CIM or not needed	Moderate CIM	Considerable CIM	Extensive CIM
People Factors: Multipliers	1.13	1.06	1.0	0.94	0.89	0.84
General SE KSAs (Knowledge, Skills, Agility)	Weak KSAs	Some KSAs	Moderate KSAs	Good KSAs	Strong KSAs	Very strong KSAs
Single-Domain KSAs	Weak	Some	Moderate	Good	Strong	Very strong
Multi-Domain KSAs	Weak	Some	Moderate or not needed	Good	Strong	Very strong
Team Compatibility	Very difficult interactions	Some difficult interactions	Basically cooperative interactions	Largely cooperative	Highly cooperative	Seamless interactions
Risk Acceptance Factor: Multipliers	1.13	1.06	1.0	0.94	0.89	0.84
	Highly risk-averse	Partly risk-averse	Balanced risk aversion, accept	Moderately risk-accepting	Considerably risk-accepting	Strongly risk-accepting

To use this model, one computes the expected effort and schedule using COSYSMO, then uses CORADMO-SE to determine the expedited factor which is then applied by multiplying the COSYSMO schedule (a cube-root function of effort) by the CORADMO-SE factor. A value of 1 does not change the COSYSMO estimated schedule, a value less than 1 decreases the estimated schedule, and a value greater than 1 increases the estimated schedule.

The example we use to illustrate the comparison of expedited systems engineering and valuing flexibility is an engineering division in a diversified company that focuses on defense applications. Assume that for the project of interest, there is a team of 20 systems engineers that are doing the up-front engineering for a new system using their standard sequential processes that are based on the Vee model. Their current tasks are to refine the operational concepts and requirements for the system, then develop a system architecture that satisfies the requirements. In addition, the defense customer has requested that they use more rapid processes in order to expedite the delivery of the system. Figure 5 highlights their current process assessment using the CORADMO-SE factors.

Accelerators/Ratings	Very Low	Low	Nominal	High	Very High	Extra High
Product Factor: Multipliers	1.09	1.05	1.0	0.96	0.92	0.87
Simplicity	Extremely complex	Highly complex	Mod. complex	Moderately simple	Highly simple	Extremely simple
Element Reuse	None (0%)	Minimal (25%)	Some (50%)	Moderate (75%)	Considerate (90%)	Extensive (95%)
Low-Priority Deferrals	Never	Rarely	Sometimes	Often	Usually	Anytime
Models vs Documents	None (0%)	Minimal (25%)	Some (50%)	Moderate (75%)	Considerate (90%)	Extensive (95%)
Key Technology Maturity	>> TRL 2 or >> TRL 3	1 TRL 3 or > 1 TRL 4	1 TRL 4 or > 1 TRL 5	1-2 TRL 5 or > 2 TRL 6	2-3 TRL 6	All > TRL 7
Process Factor: Multipliers	1.09	1.05	1.0	0.96	0.92	0.87
Concurrent Operational Concept, Requirements, Architecture, V&V	Highly sequential	Mostly sequential	2 artifacts mostly concurrent	3 artifacts mostly concurrent	All artifacts mostly concurrent	Fully concurrent
Process Streamlining	Heavily bureaucratic	Largely bureaucratic	Conservative bureaucratic	Moderate streamline	Mostly streamlined	Fully streamlined
General SE tool support CIM (Coverage, Integration, Maturity)	Simple tools, weak integration	Minimal CIM	Some CIM	Moderate CIM	Considerable CIM	Extensive CIM
Project Factors: Multipliers	1.00	1.04	1.0	0.90	0.93	0.9
Project size (peak # of personnel)	Over 300	Over 400	Over 50	Over 40	Over 3	≤ 3
Collaboration support	Globally distributed, weak comm., data sharing	Nationally distributed, some sharing	Regionally distributed, moderate sharing	Metro-area distributed, good sharing	Simple campus, strong sharing	Largely collocated, Very strong sharing
Single-domain MMPTs (Models, Methods, Processes, Tools)	Simple MMPTs, weak integration	Minimal CIM	Some CIM	Moderate CIM	Considerable CIM	Extensive CIM
Multi-domain MMPTs	Simple, weak integration	Minimal CIM	Some CIM or not needed	Moderate CIM	Considerable CIM	Extensive CIM
People Factors: Multipliers	1.13	1.06	1.0	0.94	0.89	0.84
General SE KSAs (Knowledge, Skills, Agility)	Weak KSAs	Some KSAs	Moderate KSAs	Good KSAs	Strong KSAs	Very strong KSAs
Single-Domain KSAs	Weak	Some	Moderate	Good	Strong	Very strong
Multi-Domain KSAs	Weak	Some	Moderate or not needed	Good	Strong	Very strong
Team Compatibility	Very difficult interactions	Some difficult interactions	Basically cooperative interactions	Largely cooperative	Highly cooperative	Seamless interactions
Risk Acceptance Factor: Multipliers	1.13	1.06	1.0	0.94	0.89	0.84
	Highly risk-averse	Partly risk-averse	Balanced risk aversion, accept	Moderately risk-accepting	Considerably risk-accepting	Strongly risk-accepting

Figure 5: Case study CORADMO assessment of current process.

To determine the “rapid” factor to apply to the COSYSMO, one first determines the product, process, project, people, and risk acceptance factors. For each factor, an organization highlights their sub-factor assessments in the CORADMO-SE table, as shown in Figure 5. Then, for each factor, the evaluator “averages” the values. This can be an “average” calculation or it can be subjectively adjusted by the evaluator. To adjust the average value, the evaluator weights the sub-factors based on his/her assessment of their importance to the organization and project. For the example presented here, the following factors are suggested by the CORADMO-SE table:

Product:	1.05
Process:	1.05
Project:	0.95
People:	0.97
Risk acceptance:	1.00

The next step is to multiply these factors together, resulting in a CORADMO-SE factor of 1.02, which is then multiplied with the calculated COSYSMO schedule. This is the current organization baseline expedited factor upon which the project would like to improve, i.e., reduce to a value below 1.0, the nominal expedited factor.

Approach 1 (Expedite SE Through Concurrent Engineering): To reduce schedule, project management decides to implement concurrent engineering of systems engineering work products. By itself, this process change might reasonably change the process factor from 1.05 to 0.87, resulting in a composite expedited factor of 0.96. However, making this single process change can impact some of the other CORADMO-SE factors. For example, introducing some new tools to support concurrent engineering while continuing to use other more traditional SE tools will have a “slow-down” effect: with the new tools, SE toolset is not as well-integrated as it was, there is additional time required to set up the new tools and train the users on their use. There is also a learning curve for the SE engineers with respect to the concurrent engineering. And finally, team compatibility can take a hit if management continues to use their more traditional processes with the concurrent engineering SE processes. So, making these adjustments to the CORADMO-SE parameters (General Tool Support H→N, General SE KSAs H→L, and Team Compatibility H→L), results in a CORADMO-SE factor of 1.05 (as compared to the initial baseline of 1.02). The CORADMO-SE model shows that changing the process can initially slow down the project until the new tools and processes are integrated and the team becomes familiar with them.

Approach 2 (Value Flexibility): If the engineering project decides to value flexibility over “expedited engineering”, we get some different results. With this approach, instead of transitioning quickly from a highly sequential process to a fully concurrent engineering process, the project decides to focus on valuing product flexibility and streamlining their largely bureaucratic processes with some concurrency. So, in the CORADMO-SE framework the following changes are made with respect to the baseline:

Element reuse:	None→Moderate
Low-priority deferrals:	Never→Usual
Models vs. documents:	Minimal→Moderate
Concurrency:	Highly sequential→Nominal
Process streamlining:	Largely bureaucratic→Moderate streamlining

The resulting CORADMO-SE factor is 0.86.

Analysis of Two Approaches: The resulting “valuing product flexibility” CORADMO-SE factor (0.86) is much better than the 1.05 value that obtained when a drastic change from a highly sequential process to a fully concurrently engineered process was proposed, requiring re-tooling and a steeper learning curve. Over time, as

the engineering team becomes more familiar with concurrent engineering processes, this will change and additional reductions in schedule will be realized. However, CORADMO-SE shows that by moving a little more slowly with process changes (e.g., some concurrency instead of full concurrency, some process streamlining, and focusing more on engineering models rather than documentation) initial results can be much more positive.

This current analysis has not yet addressed the issue of technical debt that sometimes occurs when teams overly focus on expedited engineering, taking shortcuts that impact longer term maintenance, rework, and evolvability of the system. Preliminary analysis using the COQUALMO cost model for software development shows that by valuing flexibility and not shortcutting reviews, schedule can be further reduced and the overall remaining defects are considerably smaller. One example comparing expedited vs. valuing flexibility shows that for 250,000 lines of code, when valuing flexibility fewer defects are introduced during development and only a small percentage remain at the end of development.

Future Work

The next steps are to continue to build upon two SERC tasks that are looking at ways to expedite systems engineering through lean Kanban and other approaches and a third task that is evaluating further tradespace options that include technical debt. As part of this future work, additional analysis using the USC CSSE cost models will be conducted and models refined to better support tradespace analyses as well as predict cost, schedule, and technical debt.

References

1. Anderson, D. 2010. *Kanban: Successful Evolutionary Change for Your Technology Business*. Blue Hole Press, Sequim, WA.
2. Boehm, B., C. Abts, A. Brown, S. Chulani, B. Clark, E. Horowitz, R. Madachy, D. Reifer, and B. Steece. 2000. *Software cost estimation with COCOMO II*. Upper Saddle River: Prentice Hall
3. B. Boehm, J. Lane, and R. Madachy. 2011. Total Ownership Cost Models for Valuing System Flexibility, Proceedings of the Conference on Systems Engineering Research, 14-16 April, Los Angeles, CA.
4. Boehm, B., J. Lane, S. Koolmanojwong, and R. Turner. 2010. "Architected Agile Solutions for Software-Reliant Systems" in *Agile Software Development Current Research and Future Directions*. Springer; 1st Edition.
5. Boehm, B., J. Lane, S. Koolmanojwong, R. Turner. 2010. Architected Agile Solutions for Software-Reliant Systems," Proceedings of the International Council on Systems Engineering Symposium, 12-15 July, Chicago, IL.
6. Boehm, B., W. Brown, R. Madachy, Y. Yang. 2004. "A Software Product Line Life Cycle Cost Estimation Model," *Proceedings of the 2004 International Symposium on Empirical Software Engineering*, ISESE'04, August 19-20 2004, pp. 156-164.
7. Cunningham, W. 1992. The WyCash portfolio management system, Experience Report. Oopsla'92.
8. Cunningham, W. 2009. Debt metaphor. <http://www.youtube.com/watch?v=pqeJFYwnkjE>.
9. Engineering Design Forum. 2012. http://www.phoenix-int.com/forum_2012/index.html.
10. Highsmith, J. 2000. *Adaptive software development: A collaborative approach to managing complex systems*, New York: Dorset House Publishing.
11. Koolmanojwong, S. and B. Boehm. 2010. "The Incremental Commitment Model Process Patterns for Rapid-Fielding Projects," in Proceedings of International Conference on Software Process 2010, July in Paderborn, Germany.

12. **Lane, J. 2009.** Cost Model Extensions to Support Systems Engineering Cost Estimation for Complex Systems and Systems of Systems, Proceedings of the Seventh Conference on Systems Engineering Research.
13. Lane, J. and R. Valerdi. 2011. System Interoperability Influence of System on Systems Engineering Effort, Proceedings of the Conference on Systems Engineering Research, 14-16 April, Los Angeles, CA.
14. McConnell, S. 2008. Best Practices White Paper: Managing Technical Debt. <http://www.construx.com/Page.aspx?cid=2801>.
15. Royce, W. 2004. *Software Project Management: A Unified Framework*. Addison-Wesley.
16. United States Air Force (USAF) Scientific Advisory Board (SAB). 2005. *Report on system-of-systems engineering for Air Force capability development*; Public Release SAB-TR-05-04.
17. U.S. Government Accountability Office. 2011. *Report - WEAPONS ACQUISITION REFORM: Actions Needed to Address Systems Engineering and Developmental Testing Challenges*, GAO-11-806, Sep 19, 2011.
18. Valerdi, R. 2005. *Constructive systems engineering cost model*. PhD. Dissertation, University of Southern California.

APPENDIX D: BIORELATIONAL MODELING AND ADAPTABILITY TO UNFORESEEN CHANGES

Dr. Alan Levin, Adjunct Professor, USC-CSSE

I INTRODUCTION

Adapting to the Unknowable

Today's defense systems are procured, developed, fielded, operated, and maintained under conditions of continual change and implicitly or explicitly as part of a larger system of systems (SOS). Though some degree of flexibility can be planned for, there are also unforeseen or unknowable changes. These changes throughout the life cycle are often dramatic and include new and unexpected threats, technological surprises, economic shocks, and radical mission shifts. Ironically, as we articulate a large set of possible threats, failure modes, and mission contingencies, we increase robustness with respect to the foreseen, yet we also increase system complication [1], [2]. This added complication often delays fielding, reduces operator acceptance, adds development and maintenance cost, and makes the system less adaptable to unforeseen change. The problem with true novelty is that the system may be forced into a system functional space that by definition (unforeseen) is new, and the most effective response will also be unknowable in advance. Planning for contingencies must be balanced with adapting to the unknowable.

Apollo 13

This example illustrates many of the points above. Ground control personnel and the astronauts were the key system components that adapted, modified, and quite literally re-purposed other spacecraft and ground subsystem components to accomplish unanticipated mission functions: lunar excursion module (LEM) as lifeboat, lunar descent engine for course correction and reentry deorbit, LEM battery function for life support, ground training equipment as mission simulator [3]. Certain design attributes required for manned space "bought adaptability at no additional cost." Other aspects of functional design and of implementation either helped or hindered this radical re-purposing:

- Different CO₂ filters for LEM and Command Module (CM)
- CM subsystems, including power were highly controllable and configurable by crew
- Reproduction of spacecraft systems on the ground allowed exploration of new procedures and "non-destructive" testing of candidate procedures
- Some subsystems were intended to run only once or power up only once
- Subsystem were used under far from nominal operating conditions that had never been tested and were far outside design specification
- Redundancies and backup systems provided flexibility and adaptability
-

This last point is most interesting and provides some insight into the long on-orbit lifetime and mission adaptability of space assets though they have a very long development cycle. In the case of Apollo 13, the foreseen challenges of harsh space environment, general inability to repair on-orbit, high cost of failure (lives, launch cost, national pride) resulted in redundancy and margins in the design and implementation. These margins utilized were used in unexpected and creative ways to deal with an unforeseen contingency. Adaptability to the unknowable was an unintended side effect of the redundancy and assurance needed for manned space flight.

Designing for foreseeable high impact contingencies has been very effective, and the metrics, tools, and processes are well developed. Yet as the history of spaceflight has shown, sometimes tragically, it is impossible to engineer away the unforeseen.

Biology

Biological metaphors for adapting to novelty are attractive and there is a history of biologically inspired computing and engineering methods including neural networks, genetic algorithms, artificial life, immune system security models, etc. Recent advances in the state of the art in theoretical biology, suggest new computational, architectural, analytical, and management approaches that can take us beyond biological metaphors in engineering adaptability.

II APPROACH

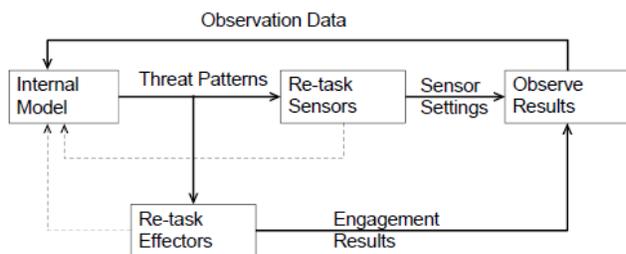


Figure 1. Sensor-Effector Loop

Figure 1 is a system block diagram of a sensor-effector control loop with internal model and a variety of feedback and feed forward loops. Our focus is on the internal models in engineered and biological systems. Typically, engineered systems have reactive internal models which, in one way or another, contain programmed responses to predefined patterns of observables. Biological systems are by nature, anticipatory and adaptive to novelty. Recent results in theoretical biology suggest how to generalize reactive engineered systems to be more anticipatory.

The features of a biological anticipatory model are:

- Relational: no pre-articulated state space
- Semiotic: syntax, semantics, interpretation
- Self referential: subjective, normative
-

A familiar example of a well studied anticipatory paradigm is the Observe, Orient, Decide, Act (OODA) loop [4]. It is the orient activity in the OODA loop that corresponds to the internal model in Figure 1. Quoting Boyd, the father of this paradigm: “The second O, orientation – as the repository of our genetic heritage, cultural tradition, and previous experiences – is the most important part of the O-O-D-A loop since it shapes the way we observe, the way we decide, the way we act.” An important point is that we cannot pre-articulate all the possible observations and actions for a tactical commander or all the possible internal/external states and system responses for any truly anticipatory system. The tactical commander brings to bear a great deal of both general and engagement specific context when she performs the second O. She may have templates and general courses of action (COAs), but her response is not merely a selection from a long list of possibilities—she is making a subjective choice of what aspects of the situation to model and how to use the assets at her disposal. The range of unforeseen external possibilities is open, as are the possibilities for re-purposing the available assets, and there is no algorithm for listing or ordering all possible ways in which a SOS or its individual systems may be used.

Of course, a simulation can always be built by constraining the internal model. Once we specify how we model the situation, we can assign a state space in a particular operating context (disallowing unforeseen contingencies). But, the creative, subjective, and normative actions of setting the context, deciding what aspects are most important, and fixing the state space must be taken by a human. Further, this constrained model of the system becomes a complicated simulation of a reactive machine rather than an anticipatory model. A machine is a system organization where the function(s) of each component are fixed, the state space of each component is fixed, and the state space of

the system is a product of the state spaces of the components. This is a very useful mathematical structure, because it promotes the “divide and conquer” synthetic approach to understanding and building systems. In our current engineering design paradigms we fix the state space of the components early in the life cycle, discover the dynamics of the system as we develop it, and we modify the system state space largely by adding components.

Biological systems have a very different sort of functional organization [5]. More properly, the adaptive, anticipatory character of biological systems is not well described using the mechanistic mathematical structure described above. It is the internal model, and in particular the self referential loops through the internal model that give rise to the subjective, context dependent, contingent character of organisms. Speaking of Darwin, Barbieri observes [6] “It was the introduction of contingency in the history of life, the idea that all living organisms, and not just humans, are subjects, individual agents which act on the world and which take care of themselves. Darwin did pay lip service to the determinism of classical physics, but what he was saying is that evolution is but a long sequence of “just so stories”, not a preordained unfolding of events dictated by immutable laws.”

Our approach is to utilize recent developments in biorelational modeling and biosemiotics to improve the adaptability of systems in a SOS context. This will increase the adaptability, flexibility, and functional utility of SOSs in the face of contingencies that were unknowable when the component systems were designed and fielded.

III Trends

Theoretical Biology

Molecular biology began with the notion that living systems were a restricted or special class of physical systems that might reveal new laws of physics [7]. Today, the dominant view in molecular biology is that the laws of physics and chemistry are sufficient to explain biological systems completely [8]. Theoretical biology has taken an alternate position that living systems are exemplars of a broader class of systems and that physics and chemistry at the molecular level are necessary, but not sufficient to effectively model living systems[5]. This is not a statement that there is anything unphysical about living systems, nor that the laws of physics as currently understood do not apply, but rather it is a statement that the mechanistic view point of physics is insufficient to explain or effectively model the subjective and contingent nature of living things.

The simplest of organisms and even metabolic reaction networks demonstrate organizational characteristics of central interest to the analysis and design of complex adaptive systems [9]. An autonomous metabolic network is anticipatory and has a resilient self-defining organization. It has the ability to replace (repair, synthesize) parts of itself as components degrade or fail.

Biorelational Modeling

Rosen developed the original biorelational analysis of metabolic organization [10] and anticipation [11]. Rosen’s M-R (metabolism-repair) model and theoretical framework for biorelational models were extended by Louie [12], who clarified the notation and underlying mathematics. Louie also extended the analysis to networks of M-R systems and articulated the category theoretic basis for biorelational models in particular and modeling in general [13].

Rosen and Louie use a directed graph notation to describe the relationship between metabolic functional components as illustrated in Figure 2. Figure 2a is a directed graph that describes a metabolic component that responds to a metabolite set A, and influences the system via metabolite set B. The solid headed arrow from A to B indicates a material flow typical of chemical reactions from metabolic reactants to products. The open headed arrow from f to A indicates an efficient implication about the reaction rate. The efficient causes are biological catalysts: enzymes or ribozymes. Every node that initiates a material flow (material edge out) must have a catalyst (efficient edge in). Of course, the enzymes and ribozymes are themselves metabolic products and reactants.

Figure 2. Biorelational Network Diagrams

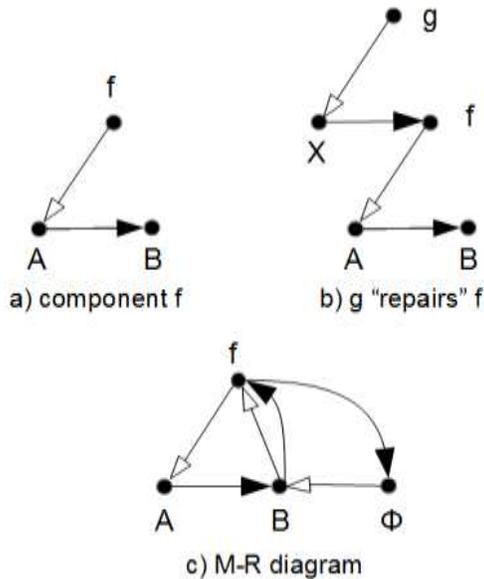


Figure 2b illustrates this with an enzyme g to catalyze replacement of enzyme f as a repair mechanism for failure or degradation f . Then we would need to postulate an enzyme h to catalyze replacement of g and so on for the replacement of each replacer. Relational models of autonomous networks avoid this potentially infinite regress using a self-referential organization of the sort shown in figure 2c. In Figure 2c, f , Φ , and B are product sets that include catalysts, and we have a closed efficient loop that creates and maintains all the catalysts necessary to repair and maintain the loop itself. Φ catalyzes the replacement of f , f catalyzes the replacement of B , and B catalyzes the replacement of Φ . This closure of efficient implication is the organizational expression of the fact that organisms, though finite, are autonomous and resilient. They are autonomous in that all the efficient implications originate within the system.

Rosen's M-R model was studied recently in simulations of autocatalytic metabolic networks [14]. The efficient edges are enzymes that catalyze the production (degradation) of other enzyme catalysts from (to) intermediates. They found multiple regions of stability where the system was able to reconstitute itself after removal of key catalysts. The system also displayed areas of bistability, so the network could implement a switching function often found in regulatory mechanisms, structural receptors and trans-membrane communication systems. They also confirmed earlier hypotheses that efficient closure in a reaction network requires multi-functional catalysts (components) coupled in a particular topology.

Another interesting result was that their autocatalytic model utilized components constructed by assembling subunits in a particular order. In other words, there appeared to be a sequence code or potential for sequence coding implicit in the model as well.

Engineering Biorelational Models

Following our earlier analysis, we can relate functional architecture in engineering design to Rosen's biorelational models [15]. A functional model is a description of how the components of a system cooperate to accomplish one or more system functions. In engineered systems, this cooperation is indicated by flows of information and/or material from one component to another. The function of each component is then identified with the transformation of inputs and production of outputs. Once again we can use a directed graph with a vertex for each component and a directed edge for each inward or outward flow. Most often, in functional architectures we make a further, often implicit, assumption that each component (vertex) has a well defined state, and that knowing the state at each vertex, along with boundary conditions is sufficient to completely determine the system behavior. For clarity we will call the directed graph under these assumptions a mechanistic model or more simply a machine. Such a model is state based, though its state space may be large and its state transition dynamics may be complicated.

Figure 3 illustrates the mechanistic directed graph network model of the control system in Figure 1. We see that there are multiple loops in the directed graph from the model through sensor tasking and observation back to the internal model and also through effector tasking and observation back to the model. The arrows represent the flow of data, information, templates, and control from one functional node to another.

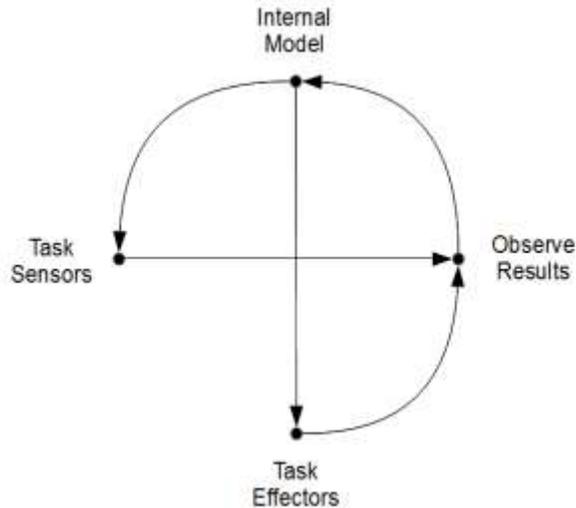


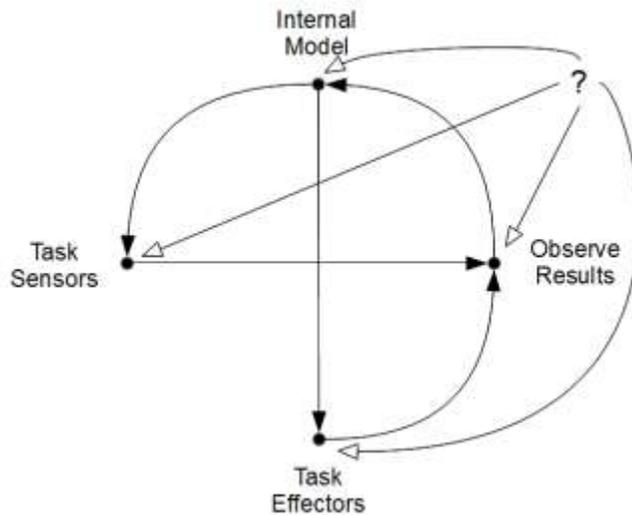
Figure 3. Sensor-Effector Graph

What's missing are the edges of efficient implication—the open headed arrows in the biorelational model. A machine or mechanistic model is a restriction of a biorelational model with a fixed state space and no efficient implication. In Figure 3, all of the nodes are either fixed in their response to inputs (completely preprogrammed) or can only have their responses changed externally. Implicitly in Figure 3, efficient implication, interpretation of function, and the functional organization of the system itself, originate externally. This is indicated by the open headed arrows in Figure 4, i.e., all the edges of efficient implication originate outside the system. They are untailed or unexplainable within the system.

This would correspond to a metabolic model in which all the product and reactant molecules had been identified, but we did not understand there was such a thing as biological catalysis. We could have a relatively complete description of the “stuff” of a metabolic pathway, including all the flows and their network topology, yet without an understanding of the efficient topology we would be at a loss to explain the functional organization and how it is maintained.

In a mechanistic model, the system state space, and the system dynamics are completely determined by the dynamics of the smallest pieces of the model and by their interaction. As we have seen all of the efficient implication comes from outside the system, in this case in the definition of subsystem dynamics and interactions. The only place where efficient loops can be closed is external to the system.

Figure 4. Biorelational Sensor-Effector Graph



Bio-semiotics

There has been significant activity in bio-semiotics in the past decade [16] influenced by progress in ecology [17], [18], system theory [19], evolutionary biology [20], and developmental biology [21]. The prefix *bio* is based on the concept (not universally accepted [6], [16]) that all semiotics must have a biological basis, and also that, no biological organism could function without being a semiotic system.

Bio-semiotics is the study of signs, their referents, and their interpretation by biological systems. It is a generalization of linguistic semiology, much as the linguistic approach was a generalization from the study of texts to the study of language. Throughout the twentieth century, the study of signs has moved from a purely literary endeavor to the study of animal communication, and more recently what might be called communication at the molecular level. The dynamic, context dependent, and self referential nature of the symbol, referent, interpreter triad is clearly relational. The meaning does not reside in any single member of the triad, but rather it is the relationship between the three that distinguishes a physical process that is a symbol or a message.

Pattee wrestled with how a physical process could become information, not in the syntactic Shannon channel capacity sense, but in the semantic sense of information about a referent. This is an important question for semiotics, biology, and as Pattee pointed out, physics and philosophy of science as well [19].

“In other words, physical laws must give the impression that events do not have alternatives and could not be otherwise (Wigner 1964) [22], while informational symbolic structures must give the impression that they could be otherwise, and must have innumerable ways of actually being otherwise. Semiotic events are based on an endless choice of alternatives, not only in symbol sequences but also in codes that interpret the symbols. It is just those innumerable alternatives, selected by heritable propagation, that are the prerequisites for evolution as well as for creative thought.”

Pattee saw this as an epistemic disagreement between the knowledge provided by physical laws vs. knowledge provided by symbolic rules. It's not that new physical laws are necessary, but rather that additional description is necessary. Physical laws are always accompanied by constraints and initial conditions, which by their nature are neither universal nor readily compressible like physical laws. Further, whether one is modeling using physical laws or symbolic rules, there is always a need to draw the distinction between subject and object [23]. He viewed the necessary separation between the observer and the observed, the controller and the controlled, the knower and the

known, and even the mind and the brain as generalizations of a broader problem, the epistemic cut. Von Neumann [24] raised this issue in his treatment of the measurement problem in physics:

“Measurement implies choosing explicitly where to place the cut and what to measure. In other words, we must be able to selectively measure something without having to measure everything. It follows that to give a functional result, a number of observable degrees of freedom in the measuring device must be selectively ignored. As a consequence of this loss of detailed information, measurement can be described only as an intrinsically irreversible process. That is, the record of the measurement must come after the measurement, and the process cannot be meaningfully reversed. On the contrary, all the most detailed fundamental physical laws are time-symmetric or reversible. Because of this necessary loss of information, a single formal description of the measurement device cannot be complete if the process is to function as a measurement.”

Barbieri [6], [8] has made an extensive study of biological codes and the adapter molecules that translate between different organic codes. Transfer RNA (t-RNA) is the classic example of an adapter molecule in the translation of a nucleic acid sequence into a protein with a functional three dimensional structure. Barbieri postulated and documented a range of adapter molecules and extended Pattee’s analysis with the insight that in addition to laws and constraints, a physical theory or dynamical model is also driven by observables. The different dynamic processes associated with different adapter molecules and different codes can involve new observables as well. Barbieri identifies copying, manufacturing, organizing, communicating, and interpreting processes with associated codes, adapter molecules, and observables. This is usually called code semiotics as distinguished from Pattee’s physical semiotics or Favareau’s sign semiotics.

Biosystem Dynamics

Ulanowicz [17] after studying biomass and energy flow in ecosystem for several decades, realized that while there were analogies between his ecosystem network analysis and information theoretic analysis of communication networks, something was amiss in how the notion of information was applied to organic system organization and flow [25]. Purely syntactic communication channel analysis explicitly defines the symbols and the symbol probability distribution externally--it has nothing to say about the meaning of the symbols. In fact, Shannon was quite explicit about removing any semantic content [26]. In the ecosystem, the symbols, their referents, and their interpretation are all defined within the ecosystem, yet these informational rules are not determined by physical laws, but as we have seen, are more akin to initial conditions, boundary conditions, or constraints on the dynamics of the system. Ecosystem models allow us to explore linkage between non-equilibrium thermodynamics, information theory, and biosemiotics. In the same sense that thermodynamics is a macroscopic theory at a higher level of abstraction than the equations of motion of individual particles, the biosemiotic issues in an ecological network are not the specific metabolic dynamics of each organism, but the broader organization of steady state flows. Ecosystems adapt their network flows in anticipation of external change, and by comparing the graphs of different steady state flow networks, we can measure the degree of adaptation and the remaining flexibility in the system. These are precisely the sort of topological metrics needed in a biorelational analysis of a network.

Deacon [19], [20], [27] has recently developed a theoretical construct that extends the thermodynamic definition of work to spatial organization of flows (morphodynamics), and functional organization of processes (teleodynamics). Sustained gradients or maintained asymmetries that keep the thermodynamic level from reaching equilibrium are the basis for doing work to create constraints that structure flow at the morphological level. Similarly, at the morphodynamic level, work may be available to the extent that flows have not reached a steady state. In this case, the work of several linked morphodynamic processes can couple to create constraint at a higher level sustaining a self-stabilizing teleodynamic process—a process that is dynamic, interpretive, and capable of recognizing conditions that promote or degrade its stability. It is at this level that a system first becomes capable of semiosis and information can actually be about something:

”What emerges in new levels of dynamics is not any new fundamental law of physics or any singularity in the causal connectedness of physical phenomena, but rather the possibility of new forms of work, and thus new ways to achieve what would not otherwise occur spontaneously. In other words, with the emergence of new forms of work,

the causal organization of the world changes fundamentally, even though the basic laws of nature remain the same.” [20]

Like Ulanowicz, Deacon argues that information is relational rather than physical, though it depends on physical objects and processes to carry and interpret it. Information depends critically on constraint, that which is absent, as well as what is present. Constraints and relations though not material themselves, have causal consequences on matter. The significant step forward is Deacon’s generalization of thermodynamic constraint and work to the dynamics of morphological and self-sustaining interpretive processes.

There has also been progress in non-equilibrium thermodynamics by Niven [28] who gives a better description of steady state flow systems and clarifies several problems with earlier descriptions of self organizing systems, dissipative systems, and chaotic systems. In particular, Niven's approach is effective for modeling and analyzing Deacon's morphodynamic work. Niven defines a potential similar to the familiar Gibbs analysis of chemical potential. Gibbs potential is G/T , where G is the Gibbs free energy and T is the temperature. This potential is the difference between the entropy change for a system and the entropy change of the surroundings during a change of state. For a closed system, this potential is related to the maximal work that the system is capable of doing as it approaches equilibrium, and it is zero at equilibrium. A similar potential can be applied to a flow system by considering the entropy flow of the system and the entropy flow of the surroundings. With an externally maintained flow, the potential is related to the maximal work per unit flow (power-like measurement) that the system is capable of doing as it approaches steady state.

The system entropy flow may positive or negative depending on the flow to the surroundings. This had been an ongoing point of confusion between maximizing and minimizing entropy production as a driving force in dissipative systems and self organization. Further, Niven points out that steady state flow systems are capable of having multiple steady states corresponding to an initial externally maintained flow state, so these systems show multi-stability and the available work depends on the steady state to which the system is heading.

IV GAPS

Modeling Issues

There are both opportunities and issues to address when we apply biorelational modeling to SOS engineering (SOSE). A closed loop of efficient implication is self-referential and quite different than a closed material, informational loop, or algorithmic loop. Materially it would just indicate a cycle familiar in both metabolism (citric acid cycle) and engineering (heat engine). Algorithmic or informational cycles can be recursive to a finite depth, but are once again familiar in engineering applications. Consider the M-R system of figure 2c. The efficient implication loop (open headed arrows) can be read as f catalyzes B , B catalyzes Φ , and Φ catalyzes f . The closed efficient loop would be impredicative or paradoxical in a purely syntactic model. In engineering terms it would read f builds (programs, implements) B builds Φ builds f . This seems to imply circular definition in a purely syntactic mechanistic model. For instance, f is the specification (information to implement) B is the specification of Φ is the specification of f --it would likely be considered an error to be corrected [29].

Ambiguity as to what the system will do in a novel situation is a key characteristic of adaptive systems that biorelational models describe quite naturally. In biology, a closed loop of efficient implication is necessary to describe how meaning (semantic content) is generated within the system, for example, the meaning of the genetic code is interpreted via protein synthesis. In our OODA loop example, the semantic content is generated by the commander who interprets the meaning of the observables. We can always approximate the orient activity in the OODA loop or the internal model of any anticipatory system after the fact with pure syntax. For a specified set of system functions and environmental context there is a mechanism that models the commander’s action. This is how foreseen contingencies are programmed into systems today. But such simulation is always based on constraint and interpretation external to the system.

As we have seen, when we reduce command decisions to detailed procedural descriptions or algorithmic courses
Contract Number: H98230-08-D-0171

TO 0014 RT-18a

of action, we lose the commander's most important capability: the ability to adapt to novel situations and use the system in new ways. Similarly, if we limit our analysis of a SOS to consider only a pre-articulated state space in a fixed functional architecture, we lose the ability of the SOS to flexibly adapt to the commanders' needs when unknowable changes occur.

The major gaps in applying biorelational modeling to engineering applications are:

- Modeling SOS efficient topology
- Reconfiguring SOS internal models
- Establishing SOS architectural and cost metrics for adaptability
-

Future Directions

We propose two biorelational research thrusts: biorelational causal loop analysis, and a cloud computing approach to internal model adjustment in real time. Biorelational modeling methods allow us to analyze the topological structure of a SOS using graph analysis and constraint analysis to identify the loops of efficient implication. This allows us to quickly identify partitioning and reconfiguration possibilities and to map out regions of functional capability that are or are not "reachable" for SOSE rapid response. Further, it gives us a modeling context for applying biosemiotic and dynamic metrics to measure adaptability and flexibility. As before, we cannot pre-articulate the ways in which a system or SOS can be utilized in all situations, but we can develop in advance, the topological categories or the functional symmetry classes into which a functional architecture can be classified based on functional roles and causal constraints. This is an analysis method that can potentially have significant impacts on SOS acquisition, development, and operational management in the future.

Although anticipatory internal models cannot in principle "contain" responses to unknowable observable combinations, within a domain of any actual observables, a state based model can be constructed to approximate the anticipatory model once the SOS is configured or reconfigured. Given defined functional topology and a range of observable expectations, a new internal model can be built and applied. Real time model design in a novel context is potentially very compute intensive. What is required is a large number of simulation runs to optimize response of some number of system components by adjusting their internal models under the influence of new but previously unexpected observational results. Cloud computing is a very natural technology to explore for this application, since it allows us to focus significant computing power quickly to any portion of a distributed SOS for simulating and revising internal models

These two research thrusts support better engineering of SOS real time response to unforeseeable changes in mission, threat environment, and technological opportunity.

V Application Impact

Software cost modeling has successfully provided quantitative measures for schedule impact of system engineering effort as a function of program size, degree of assurance, and rapidity of change as shown in figure 5 [30]. Before the success of these models, these issues were seen as purely qualitative management issues. As we collect data to quantitatively model our notional sense of where the "sweet spot" of architecture and risk reduction might be, we gain the ability to plan and manage more effectively.

Flexibility/adaptability metrics may be handled in a similar way once we properly define appropriate metrics. In figure V1, the sweet spot is a balance between architecture and risk reduction cost vs. rework expense to achieve the required reliability and assurance.

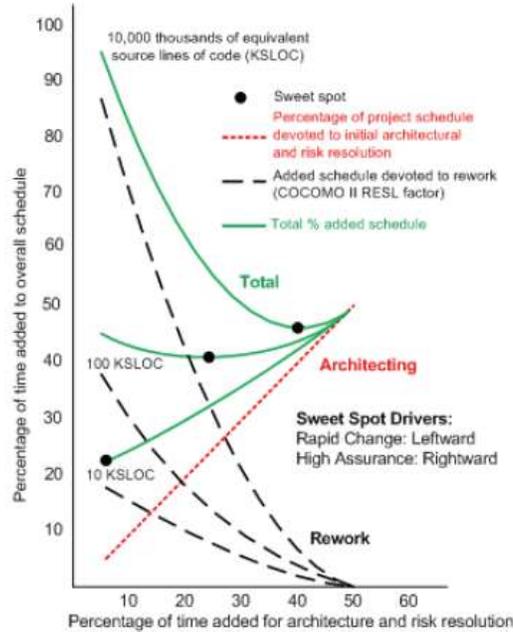


Figure 5. How Much System Engineering is Enough?

The corresponding Figure 6 for adaptability, has sweet spots that balance the cost of architecting for functional reconfiguration vs. off line time and expense during operations to achieve adaptation to unforeseen change. There are several shifts in perspective in figure 6. Instead of measuring complexity by SW size, we measure complexity by a Ulanowicz network functional measure of the system architecture in support of required functional roles in its SOS environment. In the diagram, the multi-mission system in the SOS environment (System I) has the highest Ulanowicz or functional complexity metric and the single mission stand alone system (System III) has the lowest complexity metric.

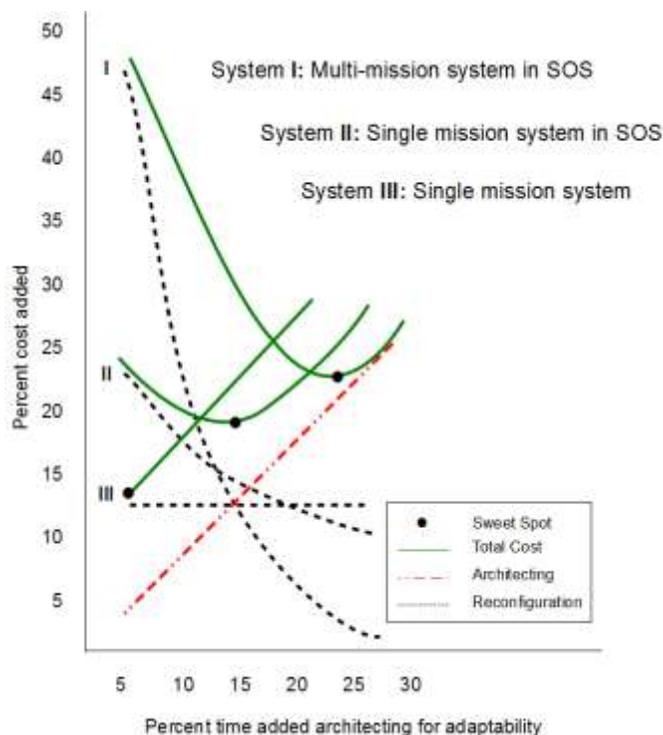


Figure 6. How Much Does Adaptability Cost?

These metrics are quite independent of software complexity metrics, though depending on partitioning during development, the functional complexity can drive software complexity. Instead of tracking rework necessary in system development and integration, we measure the cost of reconfiguring the system in operation. The sweet spot in Figure V2 is now a trade off between the added expense of designing and implementing a system architecture with functional redundancy vs. cost of lost mission effectiveness and re-engineering of of adapting to unforeseen SOS operations.

We can make these insights quantitative using the strong analogy between reliability, availability, maintenance (RAM) and reconfiguration, flexibility, adaptability (RFA). RAM policy, metrics, and tools are well developed. Further, RAM analysis is well linked to system cost analysis and performance requirements throughout the life cycle [31], [32] RAM metrics have been developed for HW and SW systems with appropriate coupling to logistics.

Table 1. lists key features of RAM and RFA. We can leverage RAM methods, tools, and processes to rapidly establish RFA methods to estimate, measure, and manage the cost of designing and implementing SOS with flexibility and adaptability to unforeseen changes. The contrasts between RAM and RFA are also important. RAM is something that belongs to each subsystem as implemented, and is generally aggregated for systems and multiple systems in a straight forward calculation. RFA is relational and is based on the functional architecture and topology, i.e., how system (or SOS) components cooperate to accomplish a mission function. RAM redundancy can and has provided a level of functional adaptability. However, this adaptability presumes that prior failures have not “consumed” backup resources and that the architecture does not preclude the necessary reconfiguration.

For instance, a multi-functional subsystem with redundancy, or multiple redundant resources distributed over multiple subsystems, provide implied functional adaptability margin. Further functional adaptation margin can be “manufactured in operation” if there is a sufficiently radical change in mission so that formerly mission critical functions can be cannibalized to achieve new mission priorities. The Apollo 13 example illustrated how unintended functional redundancy was created in an unforeseeable situation when the LEM was re-purposed for life support

during trans lunar flight once the lunar landing was no longer part of the mission.

RAM (reliability, availability, maintainability)	RFA (reconfigurability, flexibility, adaptability)
Single point of failure	Single functional pathway
Implementation redundancy	Functional redundancy
Hot backup	Duplicate components
N of M backup	Duplicate multi-function components
MTBF	Mean time between functional adaptations
MTR	Mean time to adapt
LRU	Line functional unit (facilitated variation)
Availability	Adaptability (Ulanowicz)

Table 1. RAM vs. RFA

RFA is always measured in terms of a new system context relative to a prior or reference functional architecture. However, should the implicit subsystem function or the overall system functional context shift in an unforeseen way, the precise RFA is indeterminate until that new functional context is defined. This reflects the impredicative or circular nature of relational modeling. This indeterminacy is neither a logical inconsistency nor a paradox--it is a proper reflection of the reality illustrated with the Apollo 13 example. In an unforeseen context, mission priorities, and system functions can change so radically that formerly mission critical subsystems are cannibalized to meet new and previously unknowable needs. Yet, relative to the base architecture we can articulate the degree of functional redundancy, and/or the amount of functional performance margin that can be given up to provide flexibility for foreseen contingencies and adaptation to unforeseen possibilities. This is very much along the lines of an RAM calculation, but from a functional point of view.

In summary, using biorelational modeling we can more effectively design, cost, and implement engineered systems that can adapt to unforeseen contingencies. The existing gaps between current practice in engineered systems and state of the art in theoretical biology can be addressed with modest research investments. Further, it appears that rapid reprogramming of SOS internal models is an excellent application for cloud computing. Finally, the precedent of successful SW cost estimation, and the effectiveness of RAM methodologies gives us a reasonable basis for estimating cost, schedule, and operational impact of designing for adaptability.

REFERENCES

- [1] D. Alderson and J. C. Doyle, "Contrasting Views of Complexity and Their Implications For Network-Centric Infrastructures," *IEEE TRANSACTIONS ON SYSTEMS MAN AND CYBERNETICS PART A SYSTEMS AND HUMANS*, vol. 40, no. 4, pp. 839–852, 2010.
- [2] Carlson JM and Doyle J, "Complexity and robustness.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, pp. 2538–45, 2002.
- [3] G. Kranz, *Failure Is Not an Option: Mission Control from Mercury to Apollo 13 and Beyond*, 1St ed. Simon & Schuster, 2000.
- [4] J. R. Boyd, C. W. Richards, F. C. Spinney, and G. G. Richards, *Patterns of conflict*. Atlanta, Georgia: Defense and the National Interest, 2007.
- [5] R. Rosen, *Life itself: a comprehensive inquiry into the nature, origin, and fabrication of life*. New York: Columbia University Press, 1991.
- [6] M. Barbieri, "A Short History of Biosemiotics," *Biosemiotics*, vol. 2, no. 2, pp. 221–245, 2009.
- [7] E. Schrödinger, *What is life? the physical aspect of the living cell ; with Mind and matter ; & Autobiographical sketches*. Cambridge; New York: Cambridge University Press, 1992.
- [8] M. Barbieri, "The Paradigms of Biology," *Biosemiotics*, no. 1, 2012.
- [9] J.-C. Letelier, "Organizational invariance and metabolic closure: Analysis in terms of systems," *Journal of theoretical biology.*, vol. 238, no. 4, p. 949, 2006.
- [10] Rosen R, "On the dynamical realization of (M,R)-systems.," *Bulletin of mathematical biology*, vol. 35, no. 1, 1973.
- [11] R. Rosen, *Anticipatory systems : philosophical, mathematical, and methodological foundations*, 1st ed. Oxford England ;;New York: Pergamon Press, 1985.
- [12] A. Louie, *More than life itself: a synthetic continuation in relational biology*. Frankfurt ;;New Brunswick ;Piscataway NJ: Ontos Verlag ;;[Distributed in] North and South America by Transaction Books, 2009.
- [13] A. H.-Y. Louie, "Categorical system theory," National Library of Canada, Ottawa, 1982.
- [14] Piedrafita G, Montero F, Morán F, Cárdenas ML, and Cornish-Bowden A, "A simple self-maintaining metabolic system: robustness, autocatalysis, bistability.," *PLoS computational biology*, vol. 6, no. 8, 2010.
- [15] A. Levin, "A Top-Down Approach to a Complex Natural System: Protein Folding," *Axiomathes*, vol. 20, no. 4, pp. 423–437, 2010.
- [16] D. Favareau, *Essential readings in biosemiotics anthology and commentary*. Heidelberg ;;New York :: Springer., 2010.
- [17] Ulanowicz RE, "Quantitative methods for ecological network analysis.," *Computational biology and chemistry*, vol. 28, no. 5–6, pp. 5–6, 2004.
- [18] R. E. Ulanowicz, S. J. Goerner, B. Lietaer, and R. Gomez, "Quantifying sustainability: Resilience, efficiency and the return of information theory," *Ecological Complexity*, vol. 6, no. 1, pp. 27–36, Mar. 2009.
- [19] H. Pattee, "Physical and Functional Conditions for Symbols, Codes, and Languages," *Biosemiotics*, vol. 1, no. 2, pp. 147–147–168, Aug. 2008.
- [20] T. W. Deacon, *Incomplete nature : how mind emerged from matter*. New York: W.W. Norton & Co., 2012.
- [21] M. W. Kirschner, J. C. Gerhart, and J. Norton, *The Plausibility of Life: Resolving Darwin's Dilemma*. Yale University Press, 2006.
- [22] E. P. Wigner, "Events, Laws of Nature, and Invariance Principles," *Science*, vol. 145, no. 3636, pp. 995–999, Sep. 1964.
- [23] H. Pattee, "The physics of symbols: bridging the epistemic cut," *Biosystems*, vol. 60, no. 1–3, pp. 5–5–21, 2001.
- [24] J. Von Neumann, *Mathematical foundations of quantum mechanics*. Princeton [N.J.]: Princeton University Press, 1955.
- [25] R. E. Ulanowicz, "Towards Quantifying a Wider Reality: Shannon Exonerata," *Information*, vol. 2, no. 4, pp. 624–634, Oct. 2011.
- [26] C. E. Shannon and W. Weaver, *The Mathematical Theory of Communication*. University of Illinois Press, 1963.

- [27] T. Deacon, "Shannon – Boltzmann – Darwin: Redefining information (Part I)," *Cognitive Semiotics*, vol. 2007, no. 15, pp. 123–148, Jan. 2007.
- [28] R. K. Niven, "Maximum entropy analysis of steady-state flow systems (and extremum entropy production principles)," *AIP Conference Proceedings*, vol. 1443, no. 1, pp. 270–281, May 2012.
- [29] S. W. Kerckel, "Entailment of Ambiguity," *CHEMISTRY AND BIODIVERSITY*, vol. 4, no. 10, pp. 2369–2385, 2007.
- [30] B. Boehm et. al., "Systems 2020 Strategic Initiative." SERC, Technical Report SERC-2010-TR-009, 29-Aug-2010.
- [31] DOD, "Military Specification RAMC-MANUAL, Department of Defense Reliability, Availability, Maintainability, and Cost Rationale Report Manual, Revision C." 01-Jun-2009.
- [32] DOD, "Military Specification MIL-HDBK-189C, Department of Defense Handbook Reliability Growth Management, Revision C." 14-Ju