A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics

30 September 2006

by

Dr. Keebom Kang, Naval Postgraduate School

Dr. Susan M. Sanchez, Naval Postgraduate School

Dr. Kenneth H. Doerr, Naval Postgraduate School

Approved for public release, distribution unlimited.

Prepared for: Naval Postgraduate School, Monterey, California 93943
# A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics

## 1. REPORT DATE
30 SEP 2006

## 2. REPORT TYPE

## 3. DATES COVERED
00-00-2006 to 00-00-2006

## 4. TITLE AND SUBTITLE
A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics

## 5. AUTHOR(S)

## 6. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Naval Postgraduate School, Graduate School of Business and Public Policy, 555 Dyer Road, Room 332, Monterey, CA, 93943

## 7. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)

## 8. PERFORMING ORGANIZATION REPORT NUMBER

## 9. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

## 10. NUMBER OF PAGES
49

## 11. SECURITY CLASSIFICATION OF:

- a. REPORT: Unclassified
- b. ABSTRACT: Unclassified
- c. THIS PAGE: Unclassified

## 12. ABSTRACT
See report

## 13. SUBJECT TERMS

## 14. SUPPLEMENTARY NOTES

## 15. LIMITATION OF ABSTRACT
Same as Report (SAR)

## 16. NUMBER OF RESPONSES

## 17. DATE OF ABSTRACT

## 18. CERTIFIED COPY AVAILABLE FROM

## 19. NAME OF RESPONSIBLE PERSON

---

Standard Form 298 (Rev. 8-98)
Prepared by ANSI X34-18
The research presented in this report was supported by the Acquisition Chair of the Graduate School of Business & Public Policy at the Naval Postgraduate School.

To request Defense Acquisition Research or to become a research sponsor, please contact:

NPS Acquisition Research Program
Attn: James B. Greene, RADM, USN, (Ret)
Acquisition Chair
Graduate School of Business and Public Policy
Naval Postgraduate School
555 Dyer Road, Room 332
Monterey, CA 93943-5103
Tel: (831) 656-2092
Fax: (831) 656-2253
e-mail: jbgreene@nps.edu

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.acquisitionresearch.org
Abstract

We develop a simulation model to aid in identifying and evaluating promising alternatives to achieve improvements in weapon system-level availability when outsourcing logistics services for system components. Two outcomes are valued: improvements in average operational availability for the weapon system, and reductions in the probability that operational availability of the weapon system falls below a given planning threshold (readiness risk). In practice, these outcomes must be obtained through performance-based agreements with logistic providers. The size of the state space, and the non-linear and stochastic nature of the variables involved precludes the use of optimization approaches. Instead, we use designed experiments to evaluate simulation scenarios in an intelligent way. This is an efficient approach that enables us to assess average readiness and readiness risk outcomes of the alternatives, as well as to identify the components and logistics factors with the greatest impact on operational availability.

Keywords: Performance-based Logistics, Operational Availability, Outsourcing, Design of Experiments
Acknowledgements

The research was supported by NAVSEA PEO SHIPS, PEO IWS (Integrated Warfare Systems) via the Naval Postgraduate School Acquisition Research program, and the U. S. Marine Corps’ Project Albert effort. The authors are greatly thankful to RADM Jim Greene, USN (Ret), the NPS Acquisition Research Chair, for securing the sponsor funding for this research. We also wish to acknowledge Dr. Keith Snider for his efforts on behalf of the Acquisition Research Program in the Graduate School of Business and Public Policy of the Naval Postgraduate School.
About the Authors

**Keebom Kang**, Associate Professor, joined the Naval Postgraduate School in 1988, where he teaches supply-chain, logistics engineering and computer simulation modeling courses for the MBA program. His research interests are in the areas of logistics and simulation modeling in various military applications. He received his PhD in Industrial Engineering from Purdue University. He was the Co-Editor of the *Proceedings of the 1995 Winter Simulation Conference*, and the program chair for the 2000 Winter Simulation conference. He is currently the Associate Editor for *Naval Research Logistics*.

Keebom Kang  
Associate Professor  
Graduate School of Business and Public Policy  
Naval Postgraduate School  
Monterey, CA 93950-5197  
Tel: (831) 656-3106  
E-mail: kkang@nps.edu

**Susan M. Sanchez** is a Professor of Operations Research at the Naval Postgraduate School, where she holds a joint appointment in the Graduate School of Business & Public Policy and serves as the Director of the *SEED Center for Data Farming*. Her research interests include experimental design for simulation studies, data-intensive statistics, and robust selection. She has a Ph.D. in Operations Research from Cornell University. She is currently the Simulation Area Editor for the *INFORMS Journal on Computing* and the ASA representative to the WSC Board of Directors.

Susan M. Sanchez  
Professor  
Department of Operations Research  
Graduate School of Operational and Information Sciences &  
Graduate School of Business and Public Policy  
Monterey, CA 93943  
Tel: 831-656-2780  
E-mail: ssanchez@nps.edu
**Kenneth Doerr**, B.S. Indiana University, 1984; Ph.D. University of Washington, 1994, is Associate Professor of Operations Management, Graduate School of Business and Public Policy, Naval Postgraduate School. Prior to joining the faculty at the Naval Postgraduate School, he taught at the University of Miami, the University of Washington and Santa Clara University, and also held research fellowships at the University of Waterloo and the University of Cincinnati. He has worked for Shell Oil, Monsanto Corporation and Peoplesoft in manufacturing and supply chain systems. His research has appeared in several leading journals, including *Management Science*, *The Academy of Management Review*, *IIE Transactions* and *The Journal of Applied Psychology*. His research interests are in process design, work design, and information technology to facilitate operations. Dr. Doerr is originally from St. Louis, Missouri.

Ken Doerr  
Associate Professor  
Graduate School of Business & Public Policy  
Naval Postgraduate School  
Monterey, CA 93943  
Phone: (831) 656-3625  
E-mail: khdoerr@nps.edu
A Design of Experiments Approach to Readiness Risk Analysis for Performance-Based Logistics

30 September 2006

by

Dr. Keebom Kang, Naval Postgraduate School

Dr. Susan M. Sanchez, Naval Postgraduate School

Dr. Kenneth H. Doerr, Naval Postgraduate School

Disclaimer: The views represented in this report are those of the author and do not reflect the official policy position of the Navy, the Department of Defense, or the Federal Government.
# Table of Contents

1. Introduction ........................................................................................................... 1

2. Background ............................................................................................................. 3
   - Performance-Based Logistics ........................................................................ 3
   - Design of Experiments ................................................................................. 5

3. Case Study ............................................................................................................. 7

4. Results .................................................................................................................. 11
   - Average Operational Availability ............................................................... 11
   - Readiness Risk: 80th Percentile ................................................................. 16

5. Final Remarks ..................................................................................................... 19

List of References ..................................................................................................... 21
Executive Summary

We develop a simulation model to aid in identifying and evaluating promising alternatives to achieve improvements in weapon system-level availability when outsourcing logistics services for system components. Two outcomes are valued: improvements in average operational availability for the weapon system, and reductions in the probability that operational availability of the weapon system falls below a given planning threshold (readiness risk). In practice, these outcomes must be obtained through performance-based agreements with logistic providers. The size of the state space, and the non-linear and stochastic nature of the variables involved precludes the use of optimization approaches. Instead, we use designed experiments to evaluate simulation scenarios in an intelligent way. This is an efficient approach that enables us to assess average readiness and readiness risk outcomes of the alternatives, as well as to identify the components and logistics factors with the greatest impact on operational availability.

We believe that our results illustrate that this approach has the potential to significantly improve decision-making related to readiness improvement efforts for weapon system programs.
1. Introduction

Performance Based Contracts are becoming increasingly popular in both the Department of Defense and the commercial defense industry. Performance-Based Logistics (PBL) contracts are a type of performance-based contract intended to improve weapon system availability at a reduced cost.

The unique aspect of performance-based contracts is their outcome focus; the client organization specifies key performance goals and allows the vendor to determine the best way of obtaining those goals (Assistant Secretary of the Navy for Research, Development and Acquisition [ASN-RDA] 2003). Such contracts are called contra proferentem, because in contrast to typical contract law, ambiguities in the contract (in particular, lack of detail in methods for obtaining the contracted results) are construed in favor of the client organization, rather than the vendor. Indeed, the main point of performance-based contracts is to outsource not only the tasks involved in obtaining an outcome (e.g., the inventory management required to improved system availability), but also the risk associated with those tasks. In other words, the client wishes to rely on the outcomes specified in the contract, and to have the vendor bear the risks associated with insuring the delivery of those outcomes. Hence, in such contracts it is important for the client organization to evaluate not only expected outcomes, but also the associated risk (Doerr et al., 2005).

In the model we develop, system operational availability, or the average percentage of assets which are available for operations \( (Ao) \) is a valued outcome, but it does not address the risk associated with contract performance. We will use readiness risk (Kang et al., 2005) as a measure of the risk that a vendor will fail to deliver a desired threshold of operational availability, such as the probability that less than 80% of a given type of aircraft will be available for operations at any given time.
The simulation approach we describe in this paper is intended to help decision-makers develop the most effective alternatives for reducing readiness risk of a weapon system. The alternatives involve specifying component-level outcomes for one or more of four logistic elements: component-level inventory service level, reduction in component failure rate, increase in component repair rate, or reductions in component logistic delay (the time required for transportation and administrative work). Our model captures the joint affect of all of these component-level logistic elements on operational availability and calculates a lifecycle cost for each alternative. We then use a design of experiments approach developed for large-scale simulation experiments (Kleijnen et al., 2005) to sample the state space of possible alternatives in an intelligent way. Using this sampling approach, we can estimate which logistic elements and which components have the greatest potential to improve availability.

The contribution of our work lies in the integrative nature of our solution approach. We apply a recently developed method for sampling in large-scale simulation experiments and use a performance metric (readiness risk) designed for performance-based agreements.
2. Background

In this section, we will review the literature on both Performance-Based Logistics and Design of Experiments, in order to place our own work in the context of what has been done before.

**Performance-based Logistics**

There is a small but growing literature on various aspects of Performance-Based Logistics (PBL) contracting. For instance: Berkowitz et al. (2003) conduct a survey of military applications of PBL and formulate a set of best practice recommendations. Apgar and Keane (2004) describe the strategic goals of PBL and assert that the principle of specifying outcomes rather than methods is consistent with a broad long-standing military strategy known as “commander's intent.” Doerr et al. (2005) examine metrics for PBL and develop an argument for the centrality of risk measurement in such contracts. Kim et al. (2006) look at a situation in which a contractor awarded a system-level prime contract for availability improvement must negotiate with subcontractors to achieve given component-level performance. But a recent Government Accountability Office report (GAO, 2004) is critical of systems-level PBL contracts, and recommends greater emphasis on PBL contracts at the component level to better maintain control over costs and performance. As Kang et al. (2005) show, the proper valuation and management of such component-level contracts entails the development of a comprehensive model which incorporates key performance dimensions of all critical components. They demonstrate tradeoffs between readiness risk and lifecycle cost on given alternatives, with a numerical analysis using two (disjoint) simulations.

Risk-based capacity models such as the one proposed in this paper have been the subject of a great deal of research in the commercial sector (Van Miegham, 2003) and have also been applied to the acquisition of production capacity for airfoils used in military aircraft (Prueitt & Park, 2003). Risk-based capacity models deal with
technological, demand, or price uncertainty, and are not directly applicable to the valuation of logistic services or the impact those services will have on system availability. The probability that Ao will remain above a certain planning threshold (or target readiness) is what we call readiness risk. This measure is not new—it is one of many imbedded in a system used by the US Air Force for planning levels of spare-parts inventory (Slay et al., 1996). Methodologically, it is simply a type of quantile analysis. But from the warfighter’s point of view, this risk may be the key performance dimension (Eaton et al., 2006). The warfighter, after all, is less concerned with the average number of mission-capable aircraft than he is concerned with the probability that he will have enough aircraft to fly a particular mission.

Performance-based contracting changes the way risk should be valued and measured in component-level contracts to improve system availability. The impact of variance in component-level reliability (e.g., failure rates) and maintainability (e.g., repair time) on average system availability was well understood (Blanchard et al., 1996) before PBL contracts ever became popular. More recent work examines alternatives for reliability or maintenance improvement at the component level, with the primary outcome being system-level availability (Cassady et al., 2004). These authors use a cost function which assumes a continuous range of available alternatives for both reliability and maintenance, but they do not examine logistic delay (which we will show to be a critical logistic element in determining system availability), nor do they use readiness risk as an outcome measure.

Within the field of reliability engineering, reliability allocation methods seek to minimize the cost of allocating resources for component-level reliability in order to obtain a given system-level reliability requirement (Kececioglu, 1991, pp. 363-399). These procedures generally assume a continuous range of reliability is available for each component and that the cost of achieving higher reliability levels increases exponentially. This work differs from ours in that they are primarily focused on reliability (failure rates) as an outcome measure at the component and system level.
Design of Experiments

Clearly, simulation models of even relatively simple logistics systems can have a very large number of inputs—many of which may be uncertain or unknown—that potentially impact the model's performance. In the design of experiments (DOE) literature, these are referred to as factors. Factors can be qualitative or quantitative. They can include distributional models (e.g., the use of exponential, triangular, or (truncated) normal distributions for service times), parameters of these distributions (e.g., means, standard deviations, or rates), or different policy choices that determine how a subsystem within the model behaves (e.g., use of priority queues to process critical components more rapidly).

In real-world experiments, it is difficult to control more than a handful of factors at a time. This is not the case for simulation experiments, where the analyst has the ability to specify the levels (values) for all of the input factors before running the simulation. Still, once the factors and potential levels have been determined, this creates a huge number of potential scenarios (or design points). For example, if an analyst wished to explore nine factors, each at 10 levels, there are one billion ($10^9$) different scenarios that could be considered. The design might need to be replicated for stochastic simulations, because specifying all input factors does not remove randomness from the output. Such a large experiment is clearly impractical. Even if it were possible to run all scenarios in a reasonable amount of time, the volumes of output data would easily overwhelm most post-processing analytic tools, leaving the analyst limited in his/her abilities to statistically interpret the results.

Fortunately, efficient experimental designs can be used to specify a small number of suitable scenarios. The following characteristics of experimental designs are desirable (Cioppa et al., 2004; Kleijnen et al., 2005):

- the ability to examine many variables (ten or more) efficiently;
- the ability to approximate orthogonality between inputs, to facilitate response surface metamodeling;
• the existence of minimal \textit{a priori} assumptions about the response surface;

• the flexibility to allow for the estimation of many effects, interactions, thresholds, and other features of the response surface; and

• the availability of an easy method for generating the design.

Kleijnen et al. (2005) discuss situations where various classes of designs are appropriate, but there is no one-fits-all design. In our explorations of readiness risk, we want to screen many variables for importance, while simultaneously maintaining the ability to fit complex meta-models to a handful of input variables that are found to have the most impact on the responses. Given this and the above design goals, the nearly orthogonal Latin hypercubes constructed by Cioppa and Lucas (2006) are particularly useful.

We remark that the use of designed experiments for simulation models involving many factors has been successfully applied to a host of other military applications. Links to over 40 Master of Science theses by students at the Naval Postgraduate School (NPS) are available online at the \textit{SEED Center for Data Farming} web pages at <http://diana.cs.nps.navy.mil/seedlab>, along with links to papers, software, spreadsheets, and other tools to facilitate experimental design. Summaries of successful studies conducted in the US or in several allied countries are available at the Project Albert web site at <http://www.projectalbert.org>.
3. Case Study

We use the decision environment of Kang et al. (2005) in this paper, but we develop an integrative model investigate potential alternatives for development. We are interested in readiness analyses of an unmanned aerial vehicle (UAV) squadron that has 40 aerial vehicles (AV). When a critical component in an AV fails, the faulty component is removed from the AV, an RFI (ready-for-issue) spare is installed, and the faulty component is sent to the repair facility. After the repair is complete, the component becomes an RFI spare and is sent to the spare pool. When a critical component fails, and an RFI spare is not available, the AV will be grounded (and will become not mission capable, or NMC) until an RFI component is available. A failure of a non-critical component may degrade readiness, but the system is assumed to be operable (that is, mission capable (MC) or partially mission capable (PMC)). In this case study, we do not consider “cannibalization,” the swapping of a working component from one downed AV to another.

Our simulation model estimates the average operational availability and the readiness risk at various thresholds of interest. Our goal is to better understand how changes in reliabilities, number of spare parts and other logistics factors (e.g., repair times and transportation delays) affect the average operational availability and the readiness risk of the squadron.

We consider three critical components in this case study: engines, propellers, and avionic computers. We assume that the time between failures for each component follows an exponential distribution. The ranges of MTBF (mean time between failures) of the individual components are provided in Table 1, along with the ranges of the number of spare components, component repair times (in hours), and the transportation/logistics delay (in days).
Table 1. Ranges of Input Parameters

<table>
<thead>
<tr>
<th>Input Parameter</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>MTBF of Engine</td>
<td>200 – 400 hrs</td>
</tr>
<tr>
<td>MTBF of Propeller</td>
<td>150 – 300 hrs</td>
</tr>
<tr>
<td>MTBF of Avionic Computer</td>
<td>300 – 600 hrs</td>
</tr>
<tr>
<td>Spare Engines</td>
<td>1 – 20 units</td>
</tr>
<tr>
<td>Spare Propellers</td>
<td>1 – 20 units</td>
</tr>
<tr>
<td>Spare Avionic Computers</td>
<td>1 – 20 units</td>
</tr>
<tr>
<td>Repair Time for Engines</td>
<td>1 – 30 hrs</td>
</tr>
<tr>
<td>Repair Time for Propellers</td>
<td>1 – 30 hrs</td>
</tr>
<tr>
<td>Repair Time for Avionic Computers</td>
<td>1 – 30 hrs</td>
</tr>
<tr>
<td>Transportation/Administrative Delay</td>
<td>1 – 15 days</td>
</tr>
</tbody>
</table>

Several designs are possible, but we use an NOLH with 257 runs (Cioppa & Lucas, 2006). This design is capable of handling up to 29 factors without increasing the number of scenarios. It can be easily constructed by entering the low and high values in Table 1 into a spreadsheet (Sanchez, 2005). (We remark that that ten input factors could be examined using a NOLH with as few as 33 scenarios if the simulation run-time was long. Because our model runs quickly, we opt for a larger design to allow a more detailed investigation of our model’s behavior.) The input parameters for the first ten scenarios are shown in Table 2. In all, there are ten different simulation inputs used as factors for our designed experiment. In addition, there is a stochastic element that occurs due to the pseudo-random numbers generated for stochastic failure times, repair times, and transportation/administrative delay times.
Table 2. Input Parameter Settings for First 10 of 257 Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>MTBF Engines</th>
<th>MTBF Props</th>
<th>MTBF AvComp</th>
<th>Mean Engine Repair (hrs)</th>
<th>Mean Prop Repair (hrs)</th>
<th>Mean AvComp Repair (hrs)</th>
<th>Mean Trans/Admin Delay (days)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>280</td>
<td>282</td>
<td>478</td>
<td>13</td>
<td>13</td>
<td>18</td>
<td>30</td>
</tr>
<tr>
<td>2</td>
<td>223</td>
<td>210</td>
<td>662</td>
<td>19</td>
<td>15</td>
<td>11</td>
<td>29</td>
</tr>
<tr>
<td>3</td>
<td>282</td>
<td>229</td>
<td>335</td>
<td>12</td>
<td>13</td>
<td>15</td>
<td>28</td>
</tr>
<tr>
<td>4</td>
<td>281</td>
<td>174</td>
<td>420</td>
<td>13</td>
<td>15</td>
<td>15</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>273</td>
<td>234</td>
<td>464</td>
<td>9</td>
<td>19</td>
<td>16</td>
<td>23</td>
</tr>
<tr>
<td>6</td>
<td>288</td>
<td>205</td>
<td>597</td>
<td>4</td>
<td>17</td>
<td>18</td>
<td>22</td>
</tr>
<tr>
<td>7</td>
<td>208</td>
<td>242</td>
<td>432</td>
<td>3</td>
<td>11</td>
<td>11</td>
<td>23</td>
</tr>
<tr>
<td>8</td>
<td>277</td>
<td>156</td>
<td>410</td>
<td>9</td>
<td>15</td>
<td>15</td>
<td>21</td>
</tr>
<tr>
<td>9</td>
<td>215</td>
<td>227</td>
<td>579</td>
<td>13</td>
<td>6</td>
<td>16</td>
<td>22</td>
</tr>
<tr>
<td>10</td>
<td>297</td>
<td>181</td>
<td>569</td>
<td>15</td>
<td>2</td>
<td>13</td>
<td>17</td>
</tr>
</tbody>
</table>

For each scenario, the simulation model reads a row of data from the spreadsheet excerpted in Table 2. The MTBFs of three components are first read, followed by the number of spares for each component, the modes of the component repair times, and the mode for the transportation/administrative delay. The repair times are assumed to follow symmetric triangular distributions with lower and upper bounds of 0.5(mode) and 1.5(mode), respectively. The same approach is used for the repair-time distributions. The transportation and administrative delay (in days) follows a symmetric triangular with lower and upper bounds of 0.75(mode) and 1.25(mode), respectively. Flight operations are conducted 24 hours per day, seven days per week. Each air vehicle operates an average of four hours per day. The repair shop operates eight hours per day, seven days per week.
4. Results

We ran a total of 257 scenarios, each of which is simulated over a period of 1,000,000 hours—sufficiently long that we need not be concerned about initial bias. The results of the simulation are the average Ao (operational availability) and the quantiles (10%, 20%, … , 80%, and 90%) of Ao; these are automatically written onto an EXCEL spreadsheet worksheet and then imported into JMP software for further analysis. We remark that the outputs must be matched to the scenarios (specifically, the levels of each input factor must be available) in order to analyze the data. Also, for large experiments it can be very helpful to automate the process of running the simulation for different scenarios; see Kleijnen et al. (2005) or Sanchez (2006) for further discussion.

For demonstration purposes, we present only the results for the average Ao and its 80% quantile (i.e., the probability that the Ao goes below 80%). Our intent is to illustrate the types of insights that can be gained from a designed experiment approach, rather than to make inferences regarding readiness risk for a real weapons system.

**Average Operational Availability**

We begin assessing the output by looking at histograms of the simulation responses. This can be a way of “accidentally” performing verification and validation of a simulation model by revealing combinations of input-factor settings for which the model does not work properly—presenting results that may, at first glance, challenge the analyst’s intuition, or suggesting additional features that should be included in the simulation model (Kleijnen et al., 2005). Our results indicate that the average operational availability differs widely across the different scenarios. The Ao ranges from 0.599 to 0.976. The average Ao across the 257 scenarios is 0.795 with a standard deviation of 0.085. It appears that at least one of the input factors does, indeed, have a substantial influence on the system’s performance.
After confirming that the results appear reasonable, we turn to our main goals—identifying those factors and components that have the greatest impact on performance. A useful, non-parametric tool is a regression tree (Friedman, 2002), as in Figure 1. These graphics have proven beneficial in both communicating and helping analysts understand the results of thousands of runs over many factors. Regression trees are more human-readable and can be easier to understand than multiple regression models. Trees simply show the structure in the data. Initially, the data are grouped in a single cluster. All potential input factors are examined to identify how best to split them to yield two leaves so that the variability in the response within each leaf decreases and the variability in the response between the leaves increases.

Figure 1 shows the regression tree for predicting the average Ao from the 257 simulation scenarios. The dominant factor is clearly the average transportation/administrative delay. For example, the first split at the top indicates that the average Ao is 0.737 across the 138 scenarios that had a mean transportation/administration delay of eight or more days. In contrast, the average Ao was 0.862 (17% higher) among the 119 scenarios that had a mean transportation/administration delay of less than eight days. Even with only four splits, the regression tree achieves an $R^2$ value of 0.74.
Because they are easy to interpret, regression trees are useful displays for succinctly presenting the results to decision-makers. For larger trees with many leaves, it may be helpful if the leaves corresponding to favorable, intermediate, and unfavorable outcomes are colored green, yellow, and red, respectively (Cioppa et al., 2004).

Regression trees are non-parametric approaches for fitting a statistical model to the simulation output. They can clearly identify subsets of the output that behave much differently than the rest. Regression metamodels can also be valuable. They may confirm the regression tree results concerning which factor or factors have the greatest influence on the results, or they may allow more succinct descriptions of the simulation model’s performance if it can be well-described by simple polynomial metamodels.
Accordingly, we also fit regression metamodels of the Ao as a function of main effects, quadratic effects, and two-way interactions of the ten input factors. There are a total of 65 potential terms in the model (ten main effects, ten quadratic effects, and 45 two-way interactions). We use stepwise regression to identify the most important factors, then simplify the model even further by eliminating a few terms with p-values in order of magnitude higher than the others. Our final metamodel is shown in Figure 2. The adjusted R$^2$ is 0.97, indicating that the regression metamodel does an excellent job of explaining the variability in the simulation output. We tried other models as well. For example, a model with only six significant main effects (three MTBFs, the transportation/administrative delay, and mean repair times for the two least reliable components: propellers and engines) yields an R$^2$ of 0.92. This simpler model might also be used to make inferences.

The large |t_ratio| for the mean transportation/administrative delay (Figure 2) shows it to be the dominant factor, and agrees with our regression tree results. Note that the numbers of spare parts do not appear in the model. This means that raising them from their lowest levels to the highest levels in Table 1 does not lead to any appreciable improvement in the average operational availability. This suggests that it might be possible to entirely eliminate increases in spare parts as an improvement option, or even to reduce spare parts levels, without adversely affecting operational availability. Of course, such a possibility would need to be confirmed by running new scenarios and observing the output.
A plot of the residuals vs. the predicted values (not shown) indicates that there are a few outliers from this metamodel. Three points result in substantially lower operational availability than predicted. Depending on the vendor PBL contract, these could be worth a closer look.

Because it can be difficult to look at a regression equation and get an accurate sense of how the factors and interactions affect the response, interaction plots are often useful. The interaction plot for our regression metamodel appears in Figure 3. This consists of several small subplots that indicate how the predicted performance (Ao) varies as a function of pairs of input factors. For example, the subplot that appears at the center of the upper row shows the joint effect of the MTBF for aircraft engines and the (mean) engine repair hours. The flat upper line (in blue) shows that when the MTBF is 400 hours, changing the engine repair time between its low and high values (1-30 hours) has little impact on Ao. But, if the MTBF for engines is only 200 hours (lower line, in red), then longer engine repair times decrease Ao. The difference in slopes indicates an interaction between engine MTBF and repair times: the impact of high repair times is mitigated by large MTBF. An even stronger interaction is observed between MTBF and repair times for the propellers.
Transportation/administrative delays are so dominant that we re-ran the experiment after fixing the average delays to five days for all scenarios. (Note that individual delays still follow a random distribution.) These results allow us to focus on the other factor effects and interactions. A portion of the regression tree, corresponding to the better outcomes, is provided in Figure 4. Here, we see the impact of the MTBF and repair times for the least reliable component (propellers); the next component to show up in the tree is the engine, via its MTBF. The left-hand portion of this regression tree (not shown) has the same variables at each branch, although the “splits” at the branches occur at different factor levels.

![Figure 3. Interaction Profile Plot, First Experiment](image)

**Readiness Risk: 80th Percentile**

The analyses for the 80\textsuperscript{th} percentile of readiness risk are similar, with a few interesting differences from a decision-maker’s point of view (See Figures 5 and 6). Briefly, when the mean transportation/administrative delay varies between one day and 15 days (see Figure 5), it is the dominant factor in both the regression tree and the regression metamodel. The “splits” which the regression tree uses to break this
delay into different components differ slightly from those for the average Ao. For example, the best leaf for readiness risk of 80% or better corresponds to an average transportation/administrative delay of less than six days and a MTBF for propellers of at least 201 hours. The best leaf in the regression tree for average operational availability, however, corresponds to an average transportation/administrative delay of less than three days and a MTBF for propellers of at least 186 hours. These differences confirm that the measures are not substitutes for one another. The cost of reducing transportation delays, for example, from six to three days may be considerable, and may not be justified if readiness risk is the appropriate measure. Our regression tree with four splits and five leaves yields an $R^2$ of 0.78, and our regression model with seven terms (five main effects and two interactions) yields an $R^2$ of 0.92.

For the second experiment with the transportation/administrative delay fixed to five days (see Figure 6), we once again find that the least reliable components are the major determinants of performance. The results are similar to those for average Ao (Figure 4); although once again, individual regression coefficients differ from the levels at which splits occur in the regression tree—which has implications for decision-making in a performance-contracting environment.

The most significant difference between the average Ao results (reported in Figures 1 and 4) and the readiness risk results (reported in Figures 5 and 6) are in the range of outcomes and the variance in the estimated parameters.

The difference in variance in parameter estimates can be seen by looking at the coefficient of variation of the estimates reported in the leaf nodes. For example, one of the leaf nodes in Figure 1 shows a coefficient of variation of 0.047 (0.04/0.859). However, the corresponding leaf in Figure 5 shows a coefficient of variation of 0.84 (27.3/32.5). Comparing charts in general, the reader will see more relative variance in the readiness risk estimates than in the average Ao estimates. This difference is most likely an artifact of our design: we used a fixed number of runs and run sizes to estimate both the mean and the 80% readiness risk, although
it is known that tail probability estimates are less stable than estimates of the mean of a distribution. Since our readiness risk estimates are sufficiently accurate for our purposes, and because there are other advantages to maintaining an identical number of simulation runs and run sizes, we believe this difference in the quality of the estimates is justifiable.

The differences in range of outcome, however, may have important implications for decision-makers and those who assess the results of their decisions. The average Ao reported in Figure 4 was 84.8%, and the range of outcomes varied from 67.2% to 88.4%. The average readiness risk reported in Figure 6 was 28.5%, but the range of outcomes was from 13.1% to 81.2% (that is, in the leaf corresponding to the best performance cases, Ao dropped below 80% only 13.1% of the time, while in the leaf corresponding to the worst cases, Ao drops below 80% over 81% of the time). The decisions being simulated have a far greater impact on readiness risk (the risk of falling below the desired readiness threshold) than they have on average availability. There is nothing especially surprising in this, since a small change in the mean of a distribution can easily create large differences in tail probabilities. However, if readiness risk is the appropriate measure of availability, our results imply that readiness improvement efforts may have a far greater impact on readiness than what is suggested by a simple examination of average Ao.
5. Final Remarks

As we discuss earlier, the simulation model used in this paper is not intended to provide detailed insights regarding a particular real-world situation. For example, the use of exponential times between failures may not be appropriate, and the triangular service time distributions are unlikely to be accurate representations of real-world data. However, the same approach can easily be applied to simulations that are more realistic.

We believe that our results illustrate that this approach has the potential to significantly improve decision-making related to readiness improvement efforts for weapon system programs.

![Figure 4: Results for Ao, Second Experiment](Transportation/Administrative Delay not changed across scenarios)
Figure 5. Results for Readiness Risk, First Experiment

Figure 6. Results for Readiness Risk, Second Experiment
(Transportation/Administrative Delay not changed)
List of References


Initial Distribution List

1. Defense Technical Information Center
   8725 John J. Kingman Rd., STE 0944; Ft. Belvoir, VA  22060-6218
   2

2. Dudley Knox Library, Code 013
   Naval Postgraduate School, Monterey, CA  93943-5100
   2

3. Research Office, Code 09
   Naval Postgraduate School, Monterey, CA  93943-5138
   1

4. Robert N. Beck
   Dean, GSBPP
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

5. Keith F. Snider
   Associate Professor, GB/Sk
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

6. James B. Greene
   Acquisition Chair, GB/Jg
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

7. Bill Gates
   Associate Dean for Research, GB/Gt
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

8. Keebom Kang
   Associate Professor, GB/Kk
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

9. Susan M. Sanchez
   Professor, OR/Su
   555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
   1

10. Kenneth Doerr
    Associate Professor, GB/Nd
    555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
    1

11. Karey L. Shaffer
    Program Manager, Acquisition Research Program, GB/Ks
    555 Dyer Road, Naval Postgraduate School, Monterey, CA  93943-5000
    1

Copies of the Acquisition Sponsored Research Reports may be printed from our website www.acquisitionresearch.org
2003 - 2006 Sponsored Acquisition Research Products

Acquisition Case Series


Sponsored Report Series


Working Paper Series


**Acquisition Symposium Proceedings**


**Technical Reports**


**Presentations, Publications and External Forums**


Dillard, John, and Mark Nissen. “Computational Modeling of Project Organizations under Stress.” In review.


IS4710 - Qualitative Methods. This research-seminar course has integrated the results of the FY05 Dillard-Nissen research into the students' course project.


Copies of the Acquisition Sponsored Research Reports may be printed from our website [www.acquisitionresearch.org](http://www.acquisitionresearch.org)