Analysis of Consistently Poor Work Estimates in U.S. Submarine Availabilities

by

Sjaak André de Vlaming

Submitted to the Department of Mechanical Engineering
in partial fulfillment of the requirements for the degrees of

Naval Engineer

and

Master of Science in Mechanical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2018

DISTRIBUTION A. Approved for public release: distribution unlimited.

Author . . . .
Department of Mechanical Engineering
May 11, 2018

Certified by. . . . .
Themistoklis Sapsis
Associate Professor
Thesis Supervisor

Accepted by . . . .
Rohan Abeyaratne
Chairman, Department Committee on Graduate Theses

Analysis of Consistently Poor Work Estimates in U.S. Submarine Availabilities

by

Sjaak André de Vlaming

Submitted to the Department of Mechanical Engineering
on May 11, 2018, in partial fulfillment of the
requirements for the degrees of
Naval Engineer
and
Master of Science in Mechanical Engineering

Abstract

The maintenance of a warship requires an involved combination of scheduling, funding, and execution. For one finite maintenance period, known as an "availability," a small setback in one of these areas can have a significant deleterious effect on the availability as a whole. Compounded and obscured by complexity, the root causes of such setbacks may remain unresolved and recur within the same availability or in one that follows, resulting in cumulative cost increases and schedule delays.

The United States Navy has a strong incentive to better understand availability execution. In support of that objective, this thesis investigates man-hour cost data from 57 submarine availabilities across all four public naval shipyards, spanning 315 ship systems, from December 2006 to December 2017.

The results of this thesis are best understood in two parts: the first is an observation of system population characteristics, and the second is a multiple linear regression analysis. The first part identifies nine specific submarine systems for which work is consistently over- or underestimated in a majority of availabilities, and also partitions the data to gain insights about the performance of categorical subsets, such as a particular shipyard, availability type, or period in time, compared to the aggregate. These results include a "tier ranking" of the systems whose improvement would yield the greatest benefit for cost. The second part yields two different multiple regression models of the data to create revised estimates for what is known as "New Work," which is unexpected work whose scope is notoriously difficult to predict. Both models result in significantly higher error than that which exists without them, invalidating multiple linear regression analysis as a path to gaining insights about availability performance.

Thesis Supervisor: Themistoklis Sapsis
Title: Associate Professor

Acknowledgments

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 1

Introduction

This thesis quantitatively addresses platitudes that are readily acknowledged by the longtooths of the U.S. Navy submarine maintenance community. To state that "Trim and Drain are notoriously difficult systems to maintain," is straightfoward, and will likely garner little objection from experienced engineers—except perhaps when a different ship system caused a memory more acute and lasting than either Trim or Drain. However, such statements can only be confirmed locally and on a case-by-case basis with (for instance) Project Superintendents or Zone Managers. As is characteristic of highly complex and geographically disparate systems, the knowledge for availability planning and execution inheres in no single person, residing instead within Navy-as-organism. Comprising both blind and guided exploration, this thesis attempts to derive insights by aggregating large amounts of system-level data.

This thesis uses information which, if discussed too specifically, would be sensitive to national security. As a result, systems which could be named are instead identified with numerical codes known only within the Navy, and at times there may exist omissions which permit the public release of this thesis. This thesis is unclassified.

1.1 Submarine Life Cycle Maintenance Planning

Submarine maintenance involves extensive planning, and each vessel is built with a maintenance plan already established for its entire service life. Figure 1-1 illustrates a generic life cycle maintenance plan for a submarine, where different maintenance milestones are interspersed with deployment requirements.

Although the grey dots between availabilities represent discrete deployments, time at sea is not limited to deployments alone; local operations, participation in exercises, and underway training are other important "value added" operations for submarines.
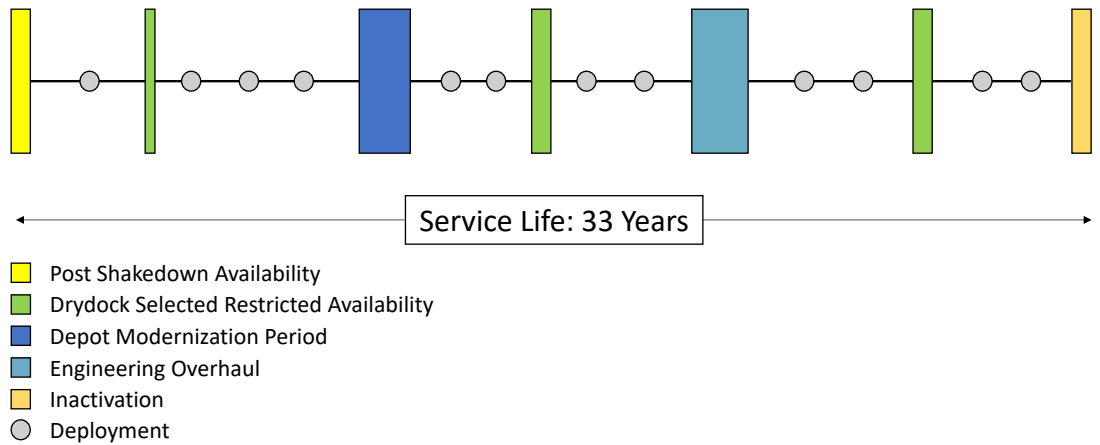
Figure 1-1: Generic Life Cycle Maintenance Plan for a Submarine

1.2 Availability Schedule Delays

If maintenance availabilities take longer than expected, the additional days spent in a drydock translate directly into fewer days at sea. Figure 1-2 demonstrates how successive schedule delays (depicted in red) across multiple availabilities can add up over the life of a submarine. In this example, the 12th availability is lost because of compression in the life cycle plan.
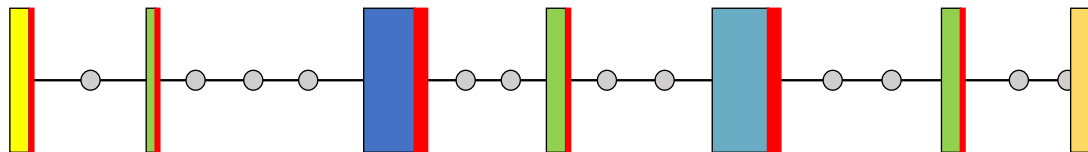
Figure 1-2: Loss of an Availability Due to Maintenance Slip

The Navy closely tracks maintenance schedule slippage because of its impact on the submarine fleet. Figure 1-3 shows the number of fleet-wide operational submarine days lost over more than a decade as a result of maintenance delays [1]. Given a procurement cost of $1.71 billion for each *Los Angeles*-class submarine[1] and $2.77 billion for each *Virginia*-class submarine,[2] these lost operational days represent performance which was paid for, but not realized [2][3].

Availability delays are not unique to the submarine community; the Vice Chief of Naval Operations testified before Congress that delayed maintenance periods are a fleet-wide issue with corrosive effects on warfighting readiness [4].

---

[1] The official cost is $900m in FY1990 dollars, which for the sake of comparison has been adjusted to FY2018 dollars.

[2] Direct FY2018 dollars used.
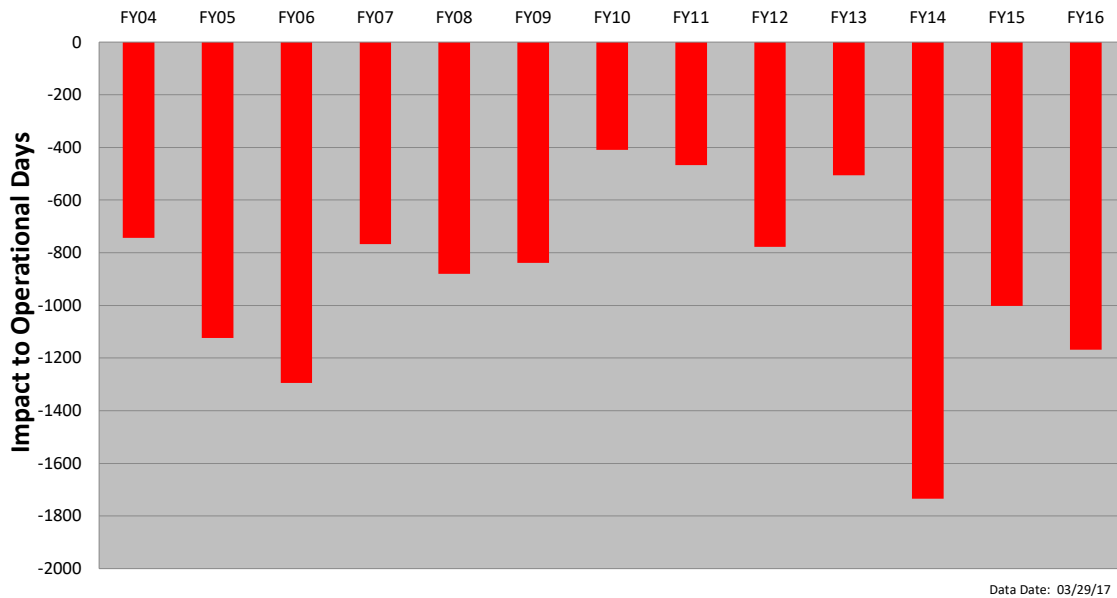
Figure 1-3: Submarine Operational Days Lost Due to Availability
Overrun, Fiscal Years 2004-2016

1.3   Causes of Schedule Delays

The issue of availability delays is best summarized by the statement of VADM Thomas Moore, Commander, Naval Sea Systems Command (COMNAVSEA), to the Senate Armed Services Committee:

"The high operational tempo in the post 9/11 era combined with re-
duced readiness funding and consistent uncertainty about when these re-
duced budgets will be approved have created a large maintenance mis-
match between the capacity in our public shipyards and the required
work... Today, despite hiring 16,500 new workers since 2012, Naval Ship-
yards are more than 2,000 people short of the capacity required to ex-
ecute the projected workload, stabilize the growth in the maintenance
backlog and eventually eliminate that backlog. This shortfall, coupled
with reduced workforce experience levels (about 50 percent of the work-
force has less than five years of experience) and shipyard productivity
issues have impacted Fleet readiness through the late delivery of ships
and submarines. The capacity limitations and the overall priority of work
toward our Ballistic Missile Submarines (SSBNs) and Aircraft Carriers
(CVN) have resulted in our Attack Submarines (SSNs) absorbing much of
the burden, causing several submarine availabilities that were originally

scheduled to last between 22 and 25 months to require 45 months or more to complete."[5]

This thesis does not purport to address operational tempo, budgeting, or manning issues; however, the need to improve availability planning and execution—by any means—is apparent.

1.4   The NAVSEA 00 Planning Summit

A month before issuing the statement quoted in Section 1.3, COMNAVSEA convened a planning summit in which stakeholders of the maintenance planning process could work together to improve availability planning and execution [6]. This Summit resulted in a set of 28 specific actions "to close the gaps and drive alignment between the maintenance planning budget process and the maintenance execution budgets" [7]. In particular, two of these action items serve as guideposts for the formulation of this thesis:

2. Analyze feedback mechanisms to improve identification of Growth/New Work.[3]

24. Review availability data and recommend appropriate upward adjustment factors to correct for known underestimation of shipyard man-day requirements caused by continued degradation when availabilities are deferred beyond class lifecycle maintenance plans.

These two items highlight the value of investigating existing shipyard data as a means to better understanding the multidimensional schedule delay problem. Analysis exists at the Ship Work Breakdown Structure (SWBS) level, which aggregates all systems under common umbrellas like "auxilliaries" or "hull/structure," but it does not investigate at a finer granularity. The comparative lack of data analysis at the system level—for specific systems such as Trim or Drain—set the course for this thesis.

---

[3]"New Work" is work that is added to an Availability Work Package (AWP) in excess of the baseline work package.

1.5 Thesis Statement

Having detailed the background and motivation for this investigation, the thesis itself is as follows:

> There are specific and actionable insights which can be derived from analyzing large amounts of data at the system level across a broad cross-section of shipyard availabilities. Furthermore, a multiple linear regression model derived from that data could be used to improve the Navy's estimates for New Work in future availabilities.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 2

Literature Review

U.S. Navy cost data from public shipyards is a specialized area of study; accessing the data can be problematic for academic investigations starting outside of the Department of Defense (DoD). Although there is a plentiful selection of defense-oriented web articles and consultant-driven publication, academic research in this field is limited. This literature review found five academic works addressing shipyard cost data in the past two decades; that prior body of research helps to provide additional context for the problem this thesis addresses. Outside the realm of shipyard analysis, there are also many examples of the efficacy of regression analysis, which inspires the hypothesis that such methods could find value in applications with shipyard data.

2.1 Cost Analysis of Naval Availabilities

It is satisfying to the engineering mind to be able to identify the root cause of a problem, to fix it, and to subsequently observe improved results—this is the quintessence of the scientific method. Because availability cost and schedule overruns are often the result of interrelated root causes, this template for a problem/solution pairing seldom obtains. Instead, high-level analysis helps to get a sense of where further attention should be directed. For instance, there are meaningful correlations between availability timeliness and ship class, cost performance ratio, and work stoppages prior to work start, but little to no relationship between availability timeliness and number of concurrent availabilities, concurrence of submarine and carrier availabilities, or quantity or length of work stoppages [8]. Given that the implementation new shipyard performance metrics results in improved performance along those metrics over time [9], there is certainly an impetus to identify correlations of interest and then ensure that established metrics track them. Put metaphorically, we seek to find the horse's mouth, and put the bit *there*.

Another approach to improving an availability's performance is to correctly anticipate its duration. This is because major availabilities occur in drydocks, which are a limited resource whose allocation problem is itself worthy of study [10]. To this end, the Navy issues Technical Foundation Papers (TFPs) which prescribe the notional duration (and cost) for the various availability types. Recent work demonstrates how difficult it is for TFPs to accurately predict the true amount of work required; even when back-fed historical data, the equations used in TFPs may yield predictions far from actual results [11]. The challenge of generating accurate TFPs is also addressed in the action items from the 2017 planning summit discussed in Section 1.4, which call for the standardization and increased frequency of TFPs [7].

This thesis follows the lines of questioning of past research in that it both attempts to derive insights about the cause-and-effect relationships that steer availability performance, as well as to develop a more accurate predictive model for availability planning.

2.2   Applications of Multiple Regression Analysis

Although there is no historical precedent for multivariate regression analysis of shipyard data, there is one instance of an analysis (both univariate and multivariate) for submarine operation and sustainment costs. The results of that study show that regression analysis does not provide suitable estimate cost models for predictors such as crew size, length overall, and submerged displacement [12].

Although literature referencing multiple linear regression for shipyard data is rare, there exist innumerable regression-based investigations of real-world phenomena. For example, regression analysis is extremely valuable in the field of analytical epidemiology [13], where regression analysis has improved understanding of the risk factors for diseases of the heart, lungs, and brain [14] [15] [16]. Other applications of multiple regression analysis have led to public health insights about causal factors of bacteria blooms in lakes [17], and to economic insights about optimal gasoline taxes [18]. These examples only graze the surface of a vast body of academic investigation using regression analysis, but they are sufficient to substantiate the interest in seeing whether the same techniques could unlock benefits for availability planning also.

Chapter 3

Methodology

A chapter devoted to thesis methodology is important to both explain the metrics being used as well as their limitations. These limitations help inform the ways in which the thesis is best pursued and provide the caveats necessary to prevent data misinterpretation.

3.1 Data Source

The data used in this thesis is furnished by the Naval Sea Systems Command 04X (SEA04X) directorate. SEA04X supports the Deputy Commander for Logistics, Maintenance, and Industrial Operations and provides oversight, operation, and advocacy for the Naval Shipyards. Appendix B-1 contains a figure of the organizational heirarchy for the public shipyards, SEA04X, and COMNAVSEA, substantiating SEA04X as the appropriate source for data in this study. Data originates from the Performance, Measurement, and Control (PMC) database, with the exception of the SSN 701 PIRA of 2009 whose data comes from the Advanced Industrial Management (AIM) database.[1] In all cases, the data is queried by means of the Navy's business objects platform.

3.2 Basis for Data Time Interval

The selection of the timeline from December 2006 to December 2017 is premised on the publication of this thesis in June 2018 and the 2006 issuance of the "Implementation of Lean Release 1.0" and "Implementation of Lean Release 2.0" memorandums

---

[1]The PMC data for this availability departs severely from all other data of its type; however, its AIM data is consistent with other completed availabilities.

from Commander, Naval Sea Systems Command. These documents issued a "...set of Corporate Improvement Initiatives and Standardization Initiatives, for standard implementation across all Naval Shipyards" [19][20]. That standardization was not immediate, and an additional Lean 3.0 memorandum further expanded Lean practices less than a year later [21]. Acknowledging that 2006 marks the beginning of a change period in availability execution (where change implies data variability that is uncorrected), the notion of a "static" period better suited for analysis is fallacious: process improvement is an ever-continuing part of shipyard availabilities. At a minimum, all availabilities studied by this thesis are subject to the same standard of execution.

One critique of this interval selection is that the Operational Interval (OPINTERVAL) for *Los Angeles*-class submarines and the Operating Cycle (OPCYCLE) for *Virginia*-class submarines were both increased from 48 months to 72 months in September of 2009 [22]. This significant change to the amount of time between major availabilities implies an unmeasured bias in cost results caused by availability planners adjusting to the "new normal." To uncover any bias that may exist in availability performance due to the Lean 1.0/2.0 initiatives or to the OPINTERVAL/OPCYCLE changes, Section 5.1.2 includes a partitioned analysis of availabilities in the first half versus the second half of December 2006 to December 2017.

A final important observation is that for the time period under consideration, all four public shipyards were operating under "Mission Funding" status as opposed to the former "Navy Capital Working Fund" vehicle. The last two shipyards (Portsmouth and Norfolk) switched funding mechanisms on October 1, 2006 [23].

3.3   Availabilities Chosen for Analysis

Only availabilities which started and finished inside the specified time interval are analyzed in this thesis. The 57 availabilities examined here fall into one of five availability types:

1. Depot Modernization Period (DMP)

2. Drydock Selected Restricted Availability (DSRA)

3. Extended Drydock Selected Restricted Availability (EDSRA)

4. Engineered Overhaul (EOH)[2]

5. Pre-Inactivation Restricted Availability (PIRA)

---

[2]The SSN 755 EOH is exempted due to the fire and untimely decommissioning of that vessel, which created an unrepresentative data set.

These availabilities can represent different durations and levels of effort; the extent to which these differences might contribute to performance are investigated in Section 5.1.2. It is important to note that even between like availabilities, work on any given system can vary widely in duration and level of effort.

3.4 Metrics

There are three major cost determinants in a shipyard availability:

1. Man-Hours: This measures how long it takes to complete a job.

2. Labor Rate: This is a cost-per-hour rate which is multiplied by man-hours to obtain the cost of labor.

3. Material Cost: This is the cost of the raw goods used to build or repair system items.

This thesis only addresses the man-hours needed to complete work on a system. This metric is useful for comparison of like jobs between different shipyards because labor rates and material costs may vary according to differences in (for instance) costs of living, labor pools, day of the week, raw good shipping rate, and dollar inflation. The man-hour metric stands independent of such variables and is persistently relevant in spite of differing locations and times.

3.4.1 Leveraging Earned Value Management

The Federal Acquisition Regulation (FAR) sets requirements for government procurement in the United States. The DoD (to include the Navy) must implement a system known as Earned Value Management (EVM) to track budgets and costs for efforts such as shipyard availabilities [24]. Figure 3-1 is a classic EVM conceptual model, with minor modifications made to suit the needs of this thesis: traditional EVM metrics refer to "cost" as opposed to "quantity" (using ACWP instead of AQWP, for instance), and some labels have been changed or removed in order to emphasize the two metrics considered in this thesis, which are Quantity at Completion (QAC) and Actual Quantity of Work Performed (AQWP).

QAC

Notionally, QAC represents an estimate made *prior* to starting work. It is a specific value that the entire work effort is expected to equal, once the task is complete. The
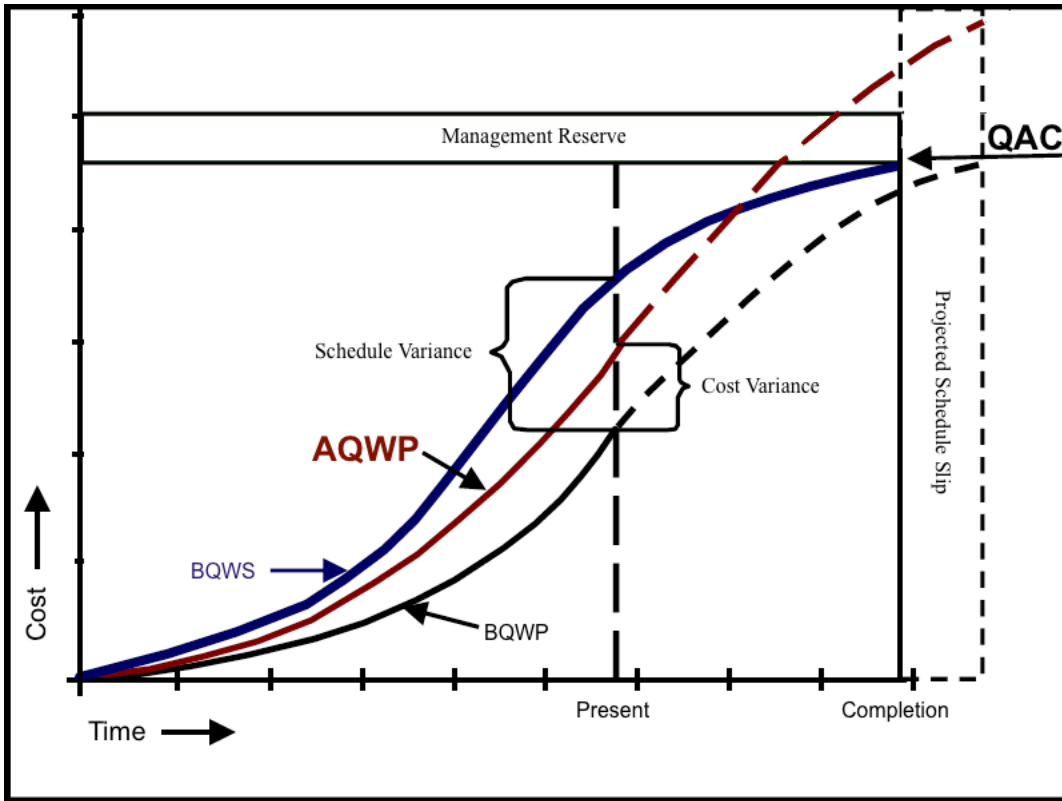
Figure 3-1: Earned Value Management Conceptual Graph

"S" shaped curve in Figure 3-1 reflects how work typically tapers at the beginning and end of a task, with maximum productivity near the midpoint, but QAC only concerns the final amount of work. It is a best-guess at the work required, and the Navy has a feedback mechanism in place whereby QAC for specific tasks gets revised based on historical data. However, estimating how many man-hours a job will require can be extremely difficult. Two jobs with the same numerical identifier can may differ in regard to the condition of the materials (for an item requiring maintenance) or to the experience of the worker.

One important characteristic of QAC is that for a completed availability (the only type considered in this thesis), the final QAC reported for a task may not be the same as the QAC which was originally estimated. This is because individual jobs can be "rebaselined," which means that the a new, larger QAC value replaces the original value. Rebaselining has the effect of artificially improving the cost performance EVM data in an availability; however, sometimes it is necessary to rebaseline expectations for work in situations where a job's scope was dramatically underreported. This thesis makes no effort to assess the number of rebaselines that occur for a task, and only uses the final value that existed at the time that the job was completed.

24

AQWP

AQWP is the other principle EVM metric considered in this thesis. It represents the final amount of work that is needed to complete a given task, after the task is complete. In Figure 3-1, AQWP is depicted as surpassing QAC, meaning that the man-hours actually required are greater than the estimate. Ideally, AQWP equals QAC at the end of a task.

As with QAC, there are some limitations to AQWP which merit discussion. AQWP is recorded after the receipt of a verbal report of job completion (or job progress) at the end of a work shift. Furthermore, one worker may report the completion of several tasks over the course of a shift, leaving it to the work supervisor to determine the allocation of an eight-hour shift over the completed tasks. Database entry of AQWP is manual. The data entry does not resemble the timestamp-driven process of a shipping company; it does not generate the "pristine" data that one might desire in an analytics-driven enterprise.

3.4.2 Treatment of Anomalous Data

In the course of researching this thesis, it became apparent that bookkeeping processes exist which allow accurate reporting of man-hours at the availability level, but which can cause curious anomalies in data at the system level. For example:

1. The SSN 767 DSRA of 2009 has an AQWP transfer charge of +48,745.7 man-hours to Ship System Index (SSI) 581, and -48,745.7 man-hours to SSI 591.

2. The SSN 753 EOH of 2014 has an AQWP of -16,073 man-hours for SSI 043.

3. The SSN 762 DSRA of 2010, SSN 771 DSRA of 2012, SSN 776 EDSRA of 2015, and SSN 767 DSRA of 2009 have small but negative AQWP values for SSIs 577, 000, 177, and 800, respectively.

In other cases, an numerical entry may exist for AQWP but not for QAC, or vice versa. Sometimes an SSI may have a charge of 0.0 hours for both QAC and AQWP. This thesis is not an assessment of QAC and AQWP data quality and it does not investigate or judge the causes of anomalies that exist in the data. The treatment for all cases of anomalous data is to delete them from the data set.

3.5 Tracking Work by Job Order Number

Every job in an availability is tracked with a Job Order Number, which is a 13-digit identifier used for planning and budgeting purposes. The generic format of the

alphanumeric identifier is:

$$12 \; ABC \; 345 \; 67 \; D \; 89$$

The "345" portion of this number is known as the "Ship System Index" (SSI) and identifies the system in which work will occur for this a job. The 315 systems analyzed in this thesis are listed by SSI in Appendix A.3; only these three digits are used to uniquely identify man-hour data throughout this thesis. This means that no inferences can be made about subcomponents of systems; instead, all subcomponent data is aggregated by SSI into two sum totals (QAC and AQWP) per system. One availability may have work in anwhere from 80-200 systems, depending on the duration and scope of the availability. Furthermore, one unique system may get maintained in two different availabilities, yet have dramatically different man-hour expenditures based on the unique states of that system. Because of this inherent variability, it is difficult to make quantitative statements about maintenance at the system level using some fixed performance metric. Instead, this thesis reframes the problem as an investigation into the Navy's ability to estimate man-hours for a task. While cross-availability performance measurement is stymied by the high degree of variability at the system level, the broader question of estimate validity adjusts with the scope of work. Thus the first objective in the thesis statement, which is intentionally vague to allow for exploration based on the constraints of available data, gains focus as a targeted investigation of estimate accuracy.

3.6   Population Observations: Eliminating Outliers

The data structure created in this thesis is a 315x57 matrix, given the 315 systems maintained across 57 availabilities. In a 315x57 matrix, there are 17,955 cells, but only 7,016 cells have cost data in them; all of the remaining cells of this matrix are not-a-number or "NaN" values. Given the QAC and AQWP anomalies addressed above, the question arises as to whether all 7,016 data entries are "believable" or otherwise grounded in reality. One way to populate the 315x57 matrix is by placing the quotient of AQWP/QAC in applicable cells and then observing trends. Section 4.1 details use of the ratio AQWP/QAC as a method for data analysis, but here we will briefly consider the methodological validity of this tool.

Of the 7,016 maintained systems, 477 entries of AQWP/QAC are larger than two, implying that the actual work required was at least double the amount predicted in the estimate. For those 477 entries, the mean is 42.7 and the standard deviation is 274.8.

For the remaining 6,539 system values (all of which fall between zero and two), the mean is 1.07 and the standard deviation is .35.

The data characteristics of the latter group are of great interest because they so closely approximate a normal, or Gaussian, distribution. The vast majority of the data is compact and consistent, while a small subset is extreme and variable. The assumption made in this thesis is that the values in that small and variable subset of data—477 out of 7,016 entries—can be rejected because either:

1. They are caused by data anomalies, or

2. They are caused by very small QAC values that divide into large AQWPs, demonstrating more about small number division than about the Navy's ability to generate reasonable estimates.

When the maximum quotient in the entire set is 3,142.7, there is a clear impetus to delete that value because it so clearly breaks the meaning inherent to the quotient comparison. And if this value is to be deleted, then perhaps too the next maximum, which is 2,860.7. And the next maximum of 2,656... Rather than wrestle with where to draw the cutoff line for aberrant data values, we instead prefer to select a value for its conceptual merits; in this case, "anything larger than two" is a good choice for exemption from the data set based on mean and standard deviation characteristics. This is not to say that there is nothing to be learned from those specific individual outliers; to the contrary, many shipyard representatives have the uncanny forensic ability to recreate the chain of events that resulted in a specific value. However, this thesis seeks broad inferences that apply to all availabilities rather than insights from case studies of outliers.

3.7   Regression Data

The second portion of this thesis explores the potential use of multiple linear regression to improve QAC estimates. This implies that the error inherent to QAC is eliminable and that large amounts of data can yield a predictive model which reduces that error. Whether or not this is true, an ill-conceived regression model will yield a useless result.

To build a good regression model, we must first recognize the relationships which exist within the data. QAC is intended to predict AQWP—this implies that the independent variable matrix must be comprised of QAC values and the dependent "solution" matrix contains AQWP values. However, there are also important relational qualities within QAC data that inform the model. All SSIs in the 000- and

900-series, as well as two specific 800-series, are contingent on the the work expected for systems 100-800. Therefore, a one-way dependency relationship exists within the data; additional 100 series work can result in a corresponding increase in 000 and 900, but not the other way around. Therefore the regression model omits 000, 900, and two of 800 series SSIs, instead investigating the QAC-to-AQWP relationship expressed by the remainder of the SSIs.

Section 3.6 explicitly addresses the treatment of anomalous data; the regression-based portion of this thesis also resolves the problem of anomalous data, albeit implicitly. First, the regression models use direct man-hour data (no division or other adulteration), eliminating the possibility of dividing by very small numbers to create inflated values in the data. Second, the regression models exempt any SSI with fewer than 14 observations, and in many cases exempted even more SSIs depending on the need to generate a meaningful set of coefficients.[3] Due to the fact that regression models are more meaningful when derived from many observations (rather than very few), this thesis employed a "pruning" process to arrive at the largest set of nonzero regression coefficients. This method is discussed in Section 4.2.1, and the result of this method is that it removes underrepresented SSIs from the model until there is enough fidelity in the data to generate a set of nonzero coefficients and their associated regression statistics. This ensures that anomalies are either eliminated from consideration, or they are aggregated with other data of the exact same SSI which results in a smoother data profile.

---

[3]SSIs with very little representation amongst the data drive the family of regression coefficients to zero, resulting in a model with no predictive power.

Chapter 4

Methods

There are two primary methods, or tools, used to test this thesis: the first is a high-level observation of the total population of availabilities; the second is the creation of two multiple regression models based on the same set of data. This chapter details the steps taken to create these tools, with the intent that this research can be independently reproduced and verified.

4.1   Population Observations

To make meaningful observations about the characteristics of submarine availability data, it first helps to consider what we expect to see in a process that produces a random result. Given the metrics QAC and AQWP discussed in Section 3.4.1, we expect that if a work estimate perfectly anticipates the work required, then $QAC = AQWP$, and subsequently:

$$\frac{AQWP}{QAC} = 1$$

Using this baseline, we want to build an intuition about the extent to which a process creates quotients which are less than one (overestimates) or greater than one (underestimates). Given that this thesis studies 57 availabilities whose work estimate quotients fall predominantly near one, the plots in Figure 4-1 show 57 points randomly distributed between zero and two, as a baseline for our intuition.

Note that the dashed line illustrates the mean for each data set. As this is an intuition-building exercise, the specific values of this data aren't as important as the general impressions that the visual plots create. Plots 1,2, and 3 in Figure 4-1 show randomly distributed points in the range [0,2], but we don't necessarily expect that work estimates will be random. To the contrary, Section 3.6 demonstrates that they
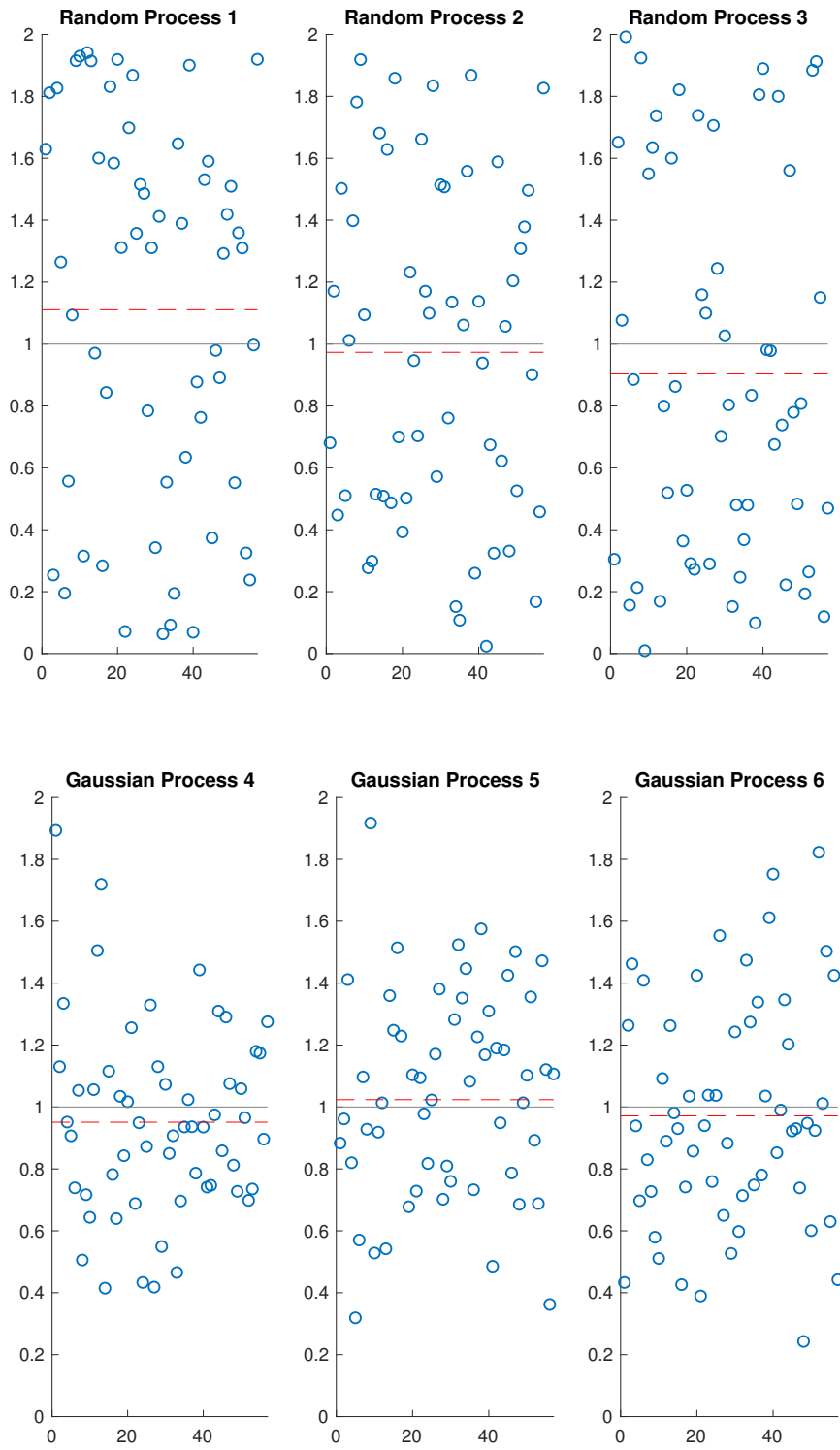
Figure 4-1: Hypothetical Process Data

are very nearly Gaussian. Plots 4, 5, and 6 of Figure 4-1 show data generated with a Gaussian mean (μ) of 1 and standard deviation (σ) of 1/3, which is very similar to our data and is centered on the range [0,2] to maintain consistency. The visual differences between plots 1-3 and 4-6 are immediately apparent, with Gaussian distributions being "clustered" around one and becoming less prevalent at the extremes of the range. Note that this is entirely by design; the Gaussian Process charts could be made to resemble the random process charts or even to display nothing (on the range [0,2]) by selecting different values of μ and σ. However, such graphs would do nothing to build our intuition as they wouldn't resemble man-hour estimate data.

Next we consider the case of actual systems which exhibit the kind of characteristics we expect from a Gaussian distribution. Figure 4-2 shows three different submarine systems—170, 426, and 513—which exhibit the kind of quotient distribution that we now expect. Note in particular that a nontrivial number of quotients fall below and above 1; this confirms the Gaussian tendency to generate comparable numbers of over- and underestimates.
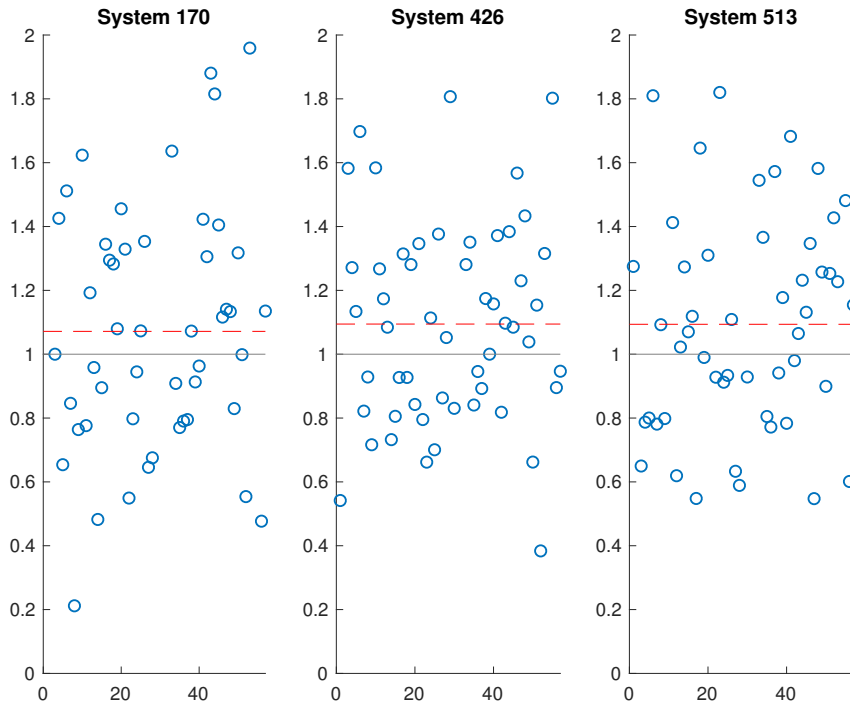
Figure 4-2: Actual Data for Three Submarine Systems

Now that a standard of normalcy is established, we ask what would constitute abberant or "interesting" data. To answer this, we must revisit the premise of this conceptual framework; namely, that the quotient of one represents an estimate that

exactly matched the actual work that was later required. Given that we can look at the estimates for systems over many availabilities, we are primarily interested in whether we *consistently* overestimate or underestimate work for a given system, in spite of a range of differences such as shipyard, availability type, etc. In other words, we want to know if a disproportionate number of quotients is greater (or less) than one. Figure 4-3 shows a scatterplot of the 315 systems, where each system is plotted according to two pieces of information:

1. Percentage of Availabilities Working the System: This counts the number of times the system was worked and divides it by the total number of availabilities considered (57). This helps us to rule out systems which do not have sufficient representation in our data set to merit strong conclusions.

2. Clustering Factor (CF): Consider the number of availabilities where a system's quotient is less than one, and divide by the number of availabilities (57). Due to symmetry, one minus this number equals the the percentage of systems with quotients greater than one; the CF counts only the largest of these two values such that the range is [0.5,1]. This value represents the extent to which system estimates are "lopsided;" if there is an equal number of over- and underestimates, the CF will equal 0.5, and if they are either all below one or all above one, the CF will equal 1.

Points existing in the top right of this plot have both a high degree of representation across all availabilities, as well as very densely clustered quotients either above or below one. Using an 80% representation rate and a CF of .85, we see that nine systems meet these criteria and constitute an excellent starting point for investigating submarine availability performance. In this thesis, systems meeting the above specified critera for representation and clustering are referred to as *consistently poorly estimated* systems. All such systems are addressed in Chapter 5.
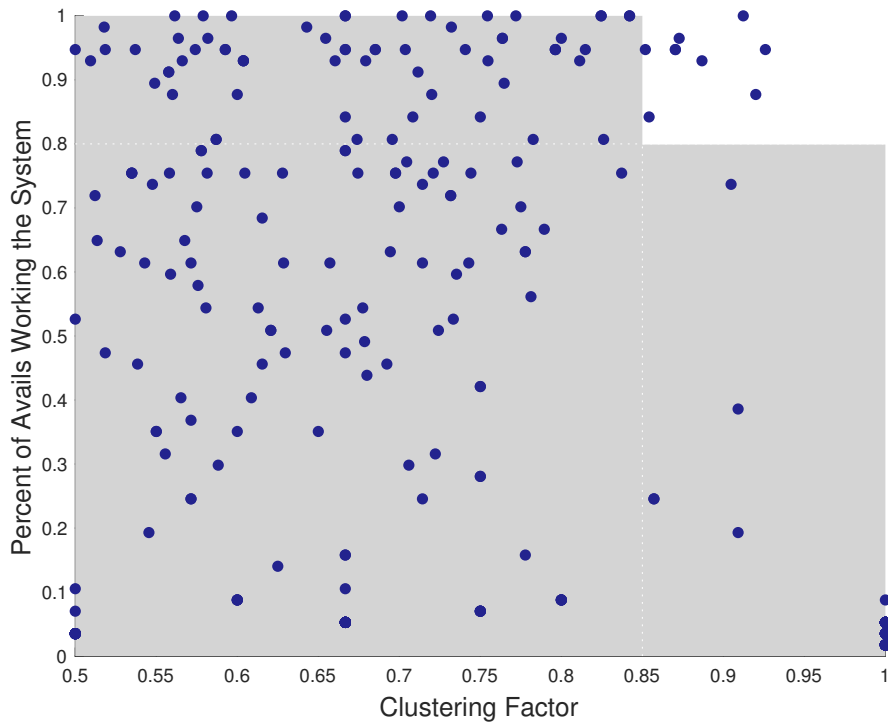
Figure 4-3: Selecting the Most "Interesting" Systems

4.2 Regression Analysis

The second method to test this thesis is a multiple regression analysis to predict New Work. A multiple regression model attempts to characterize the relationship between independent variables (systems worked) and one dependent variable (availability cost). In the context of this thesis, a successful regression model would more accurately predict AQWP for New Work than QAC. As addressed at length in Section 3.7, the model would notionally "back out" the error inherent to QAC to provide a more accurate estimate of how much work a new job will require. Therefore the hypothesis being tested is that each independent variable (expressed as QAC) has a predictive relationship with the resulting AQWP.

To test a model's ability to predict New Work, it is necessary to have documentation of the New Work that occurred in an availability, expressed both in QAC and AQWP. Only 29 availabilities between PHNS and PSNS meet this requirement; these availabilities are explicitly identified in Appendix A.2. Whereas the population observations in Section 4.1 concern a quotient, the regression models use the man-hour values of QAC and AQWP, with one minor modification: all the New Work QAC and AQWP is subtracted from the data. The intent of this is that it preserves the maxi-

33

mum number of availabilities to create the best-quality regression possible, while still providing fresh data that can be used to assess the model's efficacy. Furthermore, the ability to accurately estimate the impact of New Work on an availability is of great interest to the Navy.

Each availability constitutes an "observation" of work; more specifically, each one constitutes a family of observations on estimate/actual cost pairs for specific systems. No single availability maintains all 315 systems, and very few systems are represented in every availability. Attempting to generate a regression model for 315 independent variables with only 57 observations is not possible. Formulated mathematically, there are more variables than equations to solve for them; the system is indeterminate. However, there are two straightforward ways of reducing the number of variables: the first is to segment the availability data by SSIs; the second, to segment it by SWBS groups. In each case, the goal is to produce a set of regression coefficients which each multiply with an associated independent variable (QAC) and then sum to estimate revised AQWP. A generic model for such a relationship is:

$$AQWP_{rev} = f(QAC)$$

More specifically, for each i_{th} QAC value, the model will propose a corresponding coefficient λ_i; the sum of the products of these pairings is the revised estimate for AQWP:[1]

$$AQWP_{rev} = \sum \lambda_i * QAC_i$$

4.2.1 Regression by SSI

The original aproach for this thesis was conceived as a summation of all QAC data and all AQWP data in order to obtain a regression model, but because the system has too many variables, that approach is not valid. However, the availability cost data is granular enough that it can be segmented into subcategories to reduce the number of variables while maintaining the number of observations at 57. Instead of one massive 57x315 multiple regression model, eight smaller models can be constructed by selecting estimates for smaller batches of systems (informed by SWBS group) and by using the AQWP obtained only for those systems across all 57 availabilities. Figure 4-4 shows a miniaturized version of the 57x315 matrix developed for this research, with

---

[1]Note that the i values correspond to three digit numbers that do not increase linearly and therefore do not lend themselves to a convenient summation notation starting with $i = 1$ and proceeding to n.

all of the SWBS 200 SSIs highlighted.[2] This highlighted section, along with the corresponding AQWPs, results in a 57x41 matrix which is solvable. Repeating this method for the remaining seven SWBS groups breaks the data into smaller pieces that can be solved piecemeal.

This solution is not without problems—some systems have little representation across the 57 availabilities (this characteristic is evidenced by the horizontal white bands of empty cells in Figure 4-4). Creating a regression model with so much missing data most often results in coefficient matrices comprised entirely of zeroes. Given that a model built from many data points is more valuable than one built with few, the first step is to delete those SSIs with relatively low representation. This "pruning" process begins by deleting all systems with 14 or fewer observations. Even after doing this for each SWBS category, the derived regression coefficients have many zeroes. This poses a twofold problem:

1. The regression model produces too many zero-value coefficients, which serve no purpose in a predictive model because they eliminate (rather than scale) a baseline estimate. Therefore, systems must be removed from the model to improve its predictive power.

2. When it is time to validate this model using New Work data, deleted systems will find no expression in the model. If there is a New Work item for which there is no corresponding coefficient, this constitutes a missed opportunity to provide an improved estimate.

The solution pursued in this thesis is to iteratively observe which systems have a zero coefficient, find the one with the fewest number of observations, delete it, and re-run the regression model for the remaining systems until there are no zero-value coefficients. Chapter 5 discusses the results obtained by using the MATLAB "fitlm" function to generate the regression models.

4.2.2 Regression by SWBS

Creating a multiple regression model using SWBS groups is straightforward compared to the SSI method. As before, all New Work values are first subtracted from AQWP and QAC. This secures a testable set of New Work data which can be used to evaluate the regression model. Then, all QAC values are summed according to their first digit;

---

[2]The matrix displayed is actually 315x57, with SSIs as rows and availabilities as columns, but all data was transposed prior to regression. This figure is not scalable and is only intended to help readers understand the method.

Figure 4-4: Master Matrix with Only Series-200 SSIs Selected

for instance, the estimates for systems 111, 112, 117, ... 191 are all summed into the "100" group. Once this is complete for each of the 57 availabilities, and for SWBS 100-800, we have an 8x57 matrix of QAC values aggregated by SWBS group, and a corresponding 8x1 matrix of AQWP values. Using the "fitlm" function in MATLAB, we can generate eight regression coefficients which correspond to each SWBS group. Finally, we multiply each New Work item (from one the 29 applicable availabilities) by the corresponding coefficient to produce a revised estimate. Summing these revised estimates permits direct comparison to the actual hours attained (AQWP); Chapter 5 discusses these results.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 5

Results

5.1   Population Observations

There are nine systems across 57 availabilities for which there is consistent inaccuracy in estimating the scope of work. These *consistently poorly estimated* systems are:

<div align="center">

040	300	501
505	508	530
562	567	501

</div>

Plotting these systems about the target mean of 1 on the range [0,2] is particularly interesting, given the time we spent building our intuition in Section 4.1. Figure 5-1 shows these results. Note the extreme amount of clustering, whereby quotients fall predominantly above 1 (with the exception of system 040, whose quotients are predominantly below 1), indicating a consistent underestimation of work for these systems.

5.1.1   Prioritizing Error Severity by System

In order to prioritize where future attention should be placed, is important to identify the precedence ranking of these estimates. First, a precedence ranking must incorporate the severity of the over- or underestimate. To do this, we find the mean estimate error for a system. This is graphically represented by the distance between one and the dashed lines in Figure 5-1. Second, a precedence ranking must include QAC, which captures the magnitude of work which can be directly compared to other systems' QAC values. By normalizing the multiple of mean estimate error and QAC, we can examine a tier ranking of problematic systems. Table 5.1 traceably shows how to derive the priority ranking for the nine systems of interest.[1]

---

[1]Decimal values in Table 5.1 are truncated for legibility.
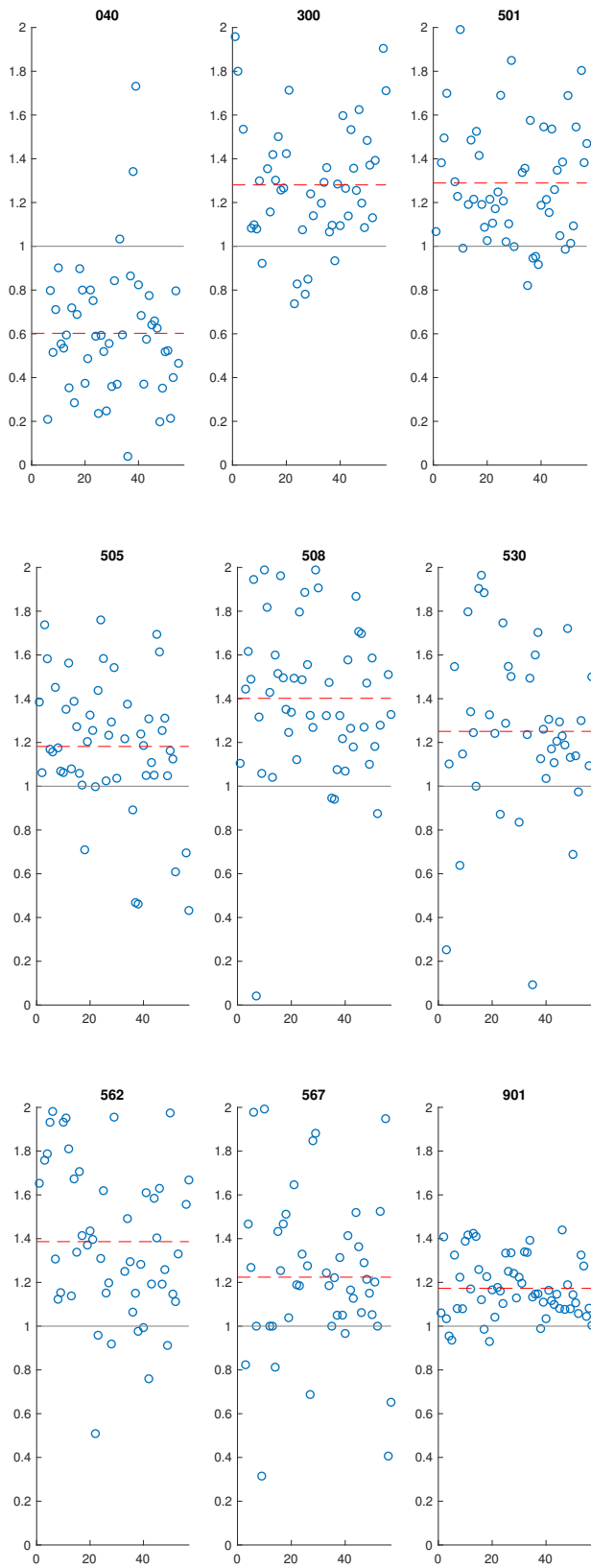
Figure 5-1: Observation Results for Nine Worst-Estimated Systems

Note that system 901 dominates the other systems due to its relatively large QAC. This is consistent across all projects and is caused because system 901 tracks project management overhead for every availability; it is always present and is always significant. We might consider excluding system 901 (and other similar SSIs) because it is not true "wrench turning" work on a system, but it nevertheless meets the criteria of consistently poor estimation, and constitutes an interesting insight.

SSI	Population Average	QAC	(QAC)*(AvgDelta)	Normalized Ordering
901	1.17	13412308	2317880.97	1.000
501	1.29	295269	85410.38	0.037
562	1.39	197434	76260.98	0.033
508	1.40	188736	75813.74	0.033
300	1.28	192106	53917.85	0.023
505	1.18	137653	25024.16	0.011
530	1.25	93134	23344.69	0.010
040	0.60	49501	19689.87	0.008
567	1.22	42439	9501.56	0.004

Table 5.1: Ranking Method for Systems with Poor Estimates

5.1.2 Using Subsets of the Data to Gain Insight

Another interesting question is how subsections of the data (availability type, ship class, shipyard, and time period) compare to the aggregate. Having identified the nine systems in Table 5.1, we ask how the family of consistently poorly estimated systems changes depending on the "slice" of the data we observe. Note that the criteria for consistently poor estimation remains the same: a system must have been worked in at least 80% of the availabilities considered, and the system's CF must be greater than or equal to 0.85. Because these thresholds assess an integer divisor and dividend, the number of poorly estimated systems that meet or exceed the thresholds is subject to "inflation."

This scenario is best interpreted as a classic "coin flip" question, where the like-lihood of attaining an AQWP/QAC quotient greater than one is 50%, and the alternative (a quotient less than one) is also 50%.[2] Consider an example evaluating the CF threshold of 85%. Using Depot Modernization Periods, which represent only four availaibilities, all four must fall either above or below one, because 3/4 is 75%

---

[2]This is not exactly accurate, since AQWP/QAC can also equal one, but there exist only 100 quotients equal to one out of the 7,016 systems worked, so this analogy is sufficient for the qualitative conclusions derived.

which fails to meet the CF threshold. In a random process, the chance of getting 4/4 "heads" flips is equal to 0.0625. If we consider the aggregate data set, we need 49/57 (or more) "heads" flips—quotients above one—to satisfy the CF threshold. This has a far lower probability of $1.3584\mathrm{e}{-}8$. This simple example illustrates the bias that small sample sizes have towards false positives; furthermore, this effect becomes more pronounced as the observed pool of availabilities shrinks. For the analysis below, data subsets of 10 or fewer are dismissed as too small to be meaningful.

1. Availability Type

    DMP: The data contains only four Depot Modernization Periods, which is too small for a meaningful comparison to the aggregate.

    DSRA: The data contains 24 Drydock Selected Restricted Availabilities yielding seven systems of interest, which is comparable to the aggregate.

    EDSRA: The data contains only four examples of Extended Drydock Selected Restricted Availabilities, which is too small for a meaningful comparison to the aggregate.

    EOH: There were 14 Engineered Overhauls, and these feature 44 flagged systems. This is a surprisingly high number, even accounting for inflation. This suggests that a disproportionately high number of poor estimates exist in EOHs.

    PIRA: There were 11 Pre-Inactivation Restricted Availabilities in the data with 12 flagged systems. This is comparable to the aggregate.

2. Submarine Class

    Los Angeles (688): The vast majority (54) of submarines examined in this thesis were *Los Angeles*-class submarines. Of interest is the fact that removing *Virginia*-class submarines added one additional system of concern to the aggregate nine, for a total of 10 flagged systems.

    Virginia (774): Only three *Virginia*-class submarines existed in this study (comprising three of the four EDSRAs). No meaningful insights can be derived from this subset.

3. Shipyard

    NNSY: Norfolk Naval Shipyard had six availabilities, which is too small for a meaningful comparison to the nine systems from the aggregate.

    PHNS: Pearl Harbor Naval Shipyard conducted 26 availabilites, which is more than any of the other shipyards. The number of systems which meet the

criteria for consistently poor estimation is 18, which is double the amount for the aggregated data. This suggests that a disproportionately high number of poor estimates exist at PHNS.

PNSY: Portsmouth Naval Shipyard followed a close second to PHNS with 22 availabilities, yet the number of systems which flag is only 13. This amount is reasonably close to the aggregate 9, taking quotient inflation into account.

PSNS: Puget Sound Naval Shipyard had three availabilities, which is too small for a meaningful comparison to the aggregate.

4. Time Period

2006-2011: The first half of the analyzed timeline has 34 availabilities with eight flagged systems, which is comparable to the aggregate.

2012-2017: The remaining 23 availabilities are captured in the second half of the timeline, with 12 flagged systems. This is comparable to the aggregate.

Figure 5-2 condenses the above information into a comparative visual. For the aggregated data, it depicts the nine consistently poorly estimated systems in black; it also depicts those same systems in black wherever they appear in smaller data partitions. Beneath the identifier for each column is a number in parenthesis which corresponds to the number of availabilities contained in that partition. All systems are ranked from highest- to lowest-priority for future investigation, consistent with the example in Section 5.1.1. This visual gives a sense of the work estimation problem considered from many different perspectives. One interesting observation of the raw data used to compile this thesis is that PHNS conducted 3/14 of the EOHs, while PNSY conducted 9/14 EOHs; this suggests that PHNS estimates are not being dominated by a disproportionate assignment of EOHs. Another worthwhile observation is that the Lean 1.0/2.0 initiatives and the OPINTERVAL/OPCYCLE changes do not result in significant differences in the number of erroneous estimates, as compared to the aggregate, over time.

Agg.	DSRA	EOH	PIRA	688	PHNS	PNSY	2006-2011	2012-2017
(57)	(24)	(14)	(11)	(54)	(26)	(22)	(34)	(23)
901	901	901	901	901	901	901	903	901
501	903	715	076	054	903	132	501	043
562	501	176	508	501	861	608	204	042
508	054	608	248	508	904	300	562	054
300	508	904	540	562	043	508	508	608
505	083	518	530	300	518	307	300	508
530	178	034	025	530	042	231	530	567
040		111	211	505	501	714	040	300
567		132	567	567	508	211		505
		054	028	040	562	540		553
		547	237		204	530		237
		552	308		515	040		040
		508			567	567		
		606			558			
		562			505			
		708			246			
		300			607			
		426			040			
		204						
		501						
		502						
		534						
		231						
		530						
		307						
		236						
		714						
		540						
		567						
		558						
		505						
		553						
		533						
		564						
		529						
		211						
		303						
		846						
		832						
		435						
		040						
		440						

Figure 5-2: Tier Rankings for Relevant Data Subsegments, Including Aggregate

5.2   Regression Analysis

5.2.1   Regression by SSI

The method described in Section 4.2.1 yields the set of regression statistics found in Appendices A.4 through A.11, which includes the regression coefficients. Those 88 coefficients comprise the following regression model by SSI:

$$AQWP|SSI = \sum \lambda_i * QAC_i$$

Expanding the model slightly for the benefit of the reader, the model is as follows:

$$AQWP|SSI = \lambda_{111} * QAC_{111} + \lambda_{131} * QAC_{131} + ...\lambda_{847} * QAC_{847}$$

This model is applied to New Work QAC, where applicable, to create a revised estimate for the impact of New Work on an availability. For New Work which has no associated regression coefficient, the QAC value is left as-is to be summed with the revised AQWP numbers (the premise is that this model will be agnostic to QAC values which have no representation in the predictive model). The regression statistics themselves merit a brief discussion, if for no other reason than to address the fact that so many p-values are "large," or at least larger than 0.05. Typically when a p-value passes that threshold, it implies that the associated regression coefficient is not statistically relevant, and that a random data set may have produced a similar coefficient. This implies that for a coefficient with a high p-value, the hypothesis in Section 4.2—that QAC has a predictive relationship with the resulting AQWP—is false. Very small p-values are indeed desireable, and it is possible that eliminating more SSIs from each "family" of models would improve the regression model. However, as addressed in Section 4.2.1, there is tension between models that have nice statistical properties and those that have enough fidelity to be of useful predictive value. In other words, removing SSIs may improve the model's statistics but neuter its ability to predict anything. As for the R^2 values, they are generally high or very high, which is desireable for predictive models, but is not a standalone metric for model suitability.

The percent error inherent to the original estimate and to the revised (regression-based) estimate can be directly compared to evaluate which estimate is better. Figure 5-3 shows the relative percent errors between the two types of estimates.

The results show that a regression model derived from SSIs does not anticipate actual work better than QAC in any of the 29 availabilities. Furthermore, in most cases the estimate is much worse than unadulterated QAC.
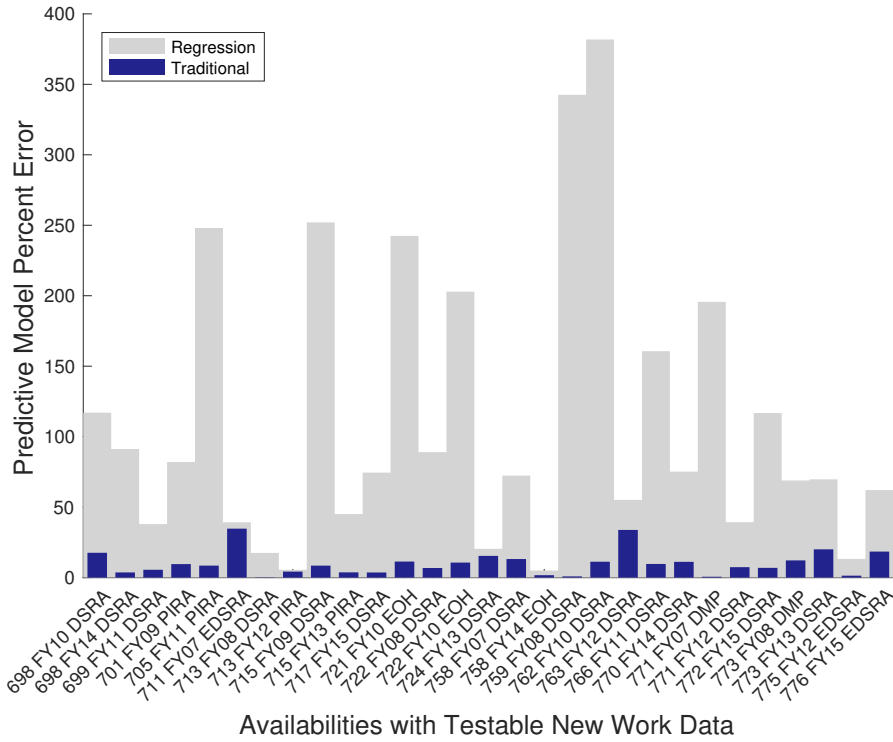
Figure 5-3: New Work Estimates Derived from a SSI Regression Model

5.2.2 Regression by SWBS

The method described in Section 4.2.2 yielded a different set of regression coefficients (and associated statistics) for the eight SWBS groups. Because this set of information is much shorter than that produced by the SSI-based model, the information is included here for convenience:

SWBS Group	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	-97613	25675	-3.8018	0.00042959	0.97
100	2.5554	0.22641	11.287	1.0264e-14	
200	0.97487	0.8411	1.159	0.25256	
300	7.8171	1.9598	3.9888	0.00024157	
400	1.3346	0.43607	3.0606	0.0037173	
500	2.0305	0.46506	4.3661	7.3306e-05	
600	-3.5105	0.97948	-3.5841	0.0008277	
700	-0.86144	1.5999	-0.53843	0.59293	
800	1.6503	0.25295	6.5241	5.1912e-08	

Applying these coefficients to the cumulative QACs by SWBS yields a revised work estimate as expressed by the following regression model:

46

$$AQWP|SWBS = \sum \lambda_i * QAC_i$$

A slightly expanded model (to aid in comprehension) is as follows:

$$AQWP|SWBS = \lambda_{100} * QAC_{100} + \lambda_{200} * QAC_{200} + ...\lambda_{800} * QAC_{800}$$

Both the original QAC and the revised, regression-based estimate can be directly compared to AQWP to derive respective error factors. Before looking at the plotted results, we note that the p-values of SWBS coefficients 200 and 700 are much higher than 0.05, indicating that randomly generated data may just as easily have produced the associated coefficients; however, the remaining coefficients all have very small p-values which suggests that they are highly statistically significant. Notionally, all of the p-values would express statistical significance, but at a minimum, this model appears to be better than anything produced under the SSI approach. Finally, the R^2 of 0.97 could mean that the model has good fit and that it has has high predictive power. Figure 5-4 shows the relative errors between the two types of estimates.
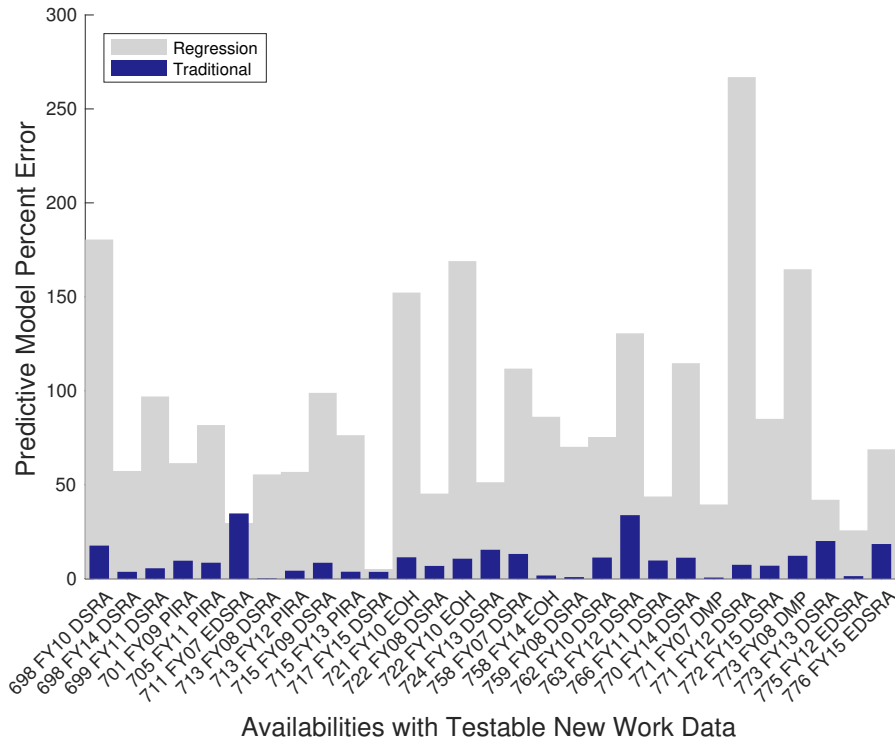

Figure 5-4: New Work Estimates Derived from a SWBS Regression Model

The results show that a regression model derived from SWBS groups anticipates actual work better than QAC in only one out of 29 availabilities. The margin of

improvement for that one availability is small; in all remaining instances, the margin by which the model is worse than QAC is substantial.

5.2.3   Understanding the Regression Results

Regression by SSI and by SWBS clearly fail to produce better estimates. During the course of this research, there arose several considerations that complicate the use of multiple linear regression for shipyard man-hour data. The following constraints explain why the regression models perform so poorly.

Observational Data is Asymmetric and Sparse

This study included 57 total availabilities, and no availability includes work for every one of the 315 systems considered in this thesis. Therefore, at the system level, the true number of observations was less than 57 most of the time. The typical range of observations for system data is between 35 and 57. The nonstandard number of observations at the system level is an unavoidable real-world constraint which has an adverse effect on the quality of regression analysis. This contrasts with the epidemiological studies cited as inspiration for this study; for instance, in determining factors that contribute to an algae bloom, one has access to pH, temperature, salinity, etc., with *every single measurement.* It cannot be overstated that the "holes" existing in shipyard cost data represent a different kind of data set than that which is produced from the sampling of lake water.

New Work is Asymmetric

One would expect that the regression coefficient for each SSI would be somewhat near one, perhaps a bit higher or a bit lower, such that work estimates are compressed or expanded slightly to arrive at a better estimate—this would be consistent with the goal of "backing out" any error inherent to a QAC estimate. Inspection of the regression coefficients in Appendices A.4 through A.11 rebuts this expectation thoroughly, as many coefficients are negative, and there exist many with magnitudes exceeding 5, 10 or even 20. For a population of New Work that includes every regressed variable, each variable's coefficient maps to a line of best fit derived from *all* variables. However, real-world availabilities typically do not have New Work in all possible SSIs. To the contrary, New Work is unevenly distributed depending on the material condition of the submarine. Consider the hypothetical case of a New Work estimate for system 178, with no other New Work for 100-series SSIs (the family of SSIs whose coefficients are derived with, and innately tied to, system 178's coefficient). The estimate

48

would be multiplied by -5.6878, and it would be severely imbalanced because of the lack of other series 100 SSIs. This process would then be repeated for the sundry of New Work items across SSIs in SWBS 200, 300, etc. The resulting estimate becomes a Frankenstein of unmatched, standalone regression coefficients which do not enjoy the "best fit" contributions of the other related coefficients in each respective SWBS family. This same kind of imbalance occurs in SWBS-derived regression estimates (which have only eight coefficients rather than 88), and is reflected in Figures 5-3 and 5-4, where regression-derived estimates are grossly more erroneous than unadulterated QAC. This issue would persist even if the regression models had used fewer SSIs to obtain "statistically pleasing" p-values (a critique acknowledged in Section 5.2.1).

Models Are Too Generic

This thesis did not investigate specialized regression models developed according to the subsets explored in Section 5.1.2. It is possible that coefficients of better fit could be attained by constraining the data set by availability type, shipyard, etc. The drawback to this method is that some of the subsets are too small to generate nonzero regression coefficients.

Models Are Overfit

The high R^2 values represented in the regression statistics could conceivably imply that the regression model is a good fit and has high predictive power; however, Figures 5-3 and 5-4 clearly show that the models have poor predictive power. As a result, the high R^2 indicates an overfit model which has too many predictor variables. Differently formulated regression problems would remove variables to improve the predictive power of the model, but doing that would not be meaningful in the context of this problem statement. For instance, a predictive model which is agnostic to inputs from SWBS 700 is of little value when New Work must occur in this group.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 6

Conclusions and Future Work

> *There are specific and actionable insights which can be derived from analyzing large amounts of data at the system level across a broad cross-section of shipyard availabilities...*

This objective was highly successful. Section 5.1 defines and ranks specific systems with consistently poor estimates. Furthermore, it also partitions the data in order to gain insights about what drives these estimates to be consistently poor. Although the SSI identification codes are never mapped to their respective systems in this thesis, discussions with individuals who have worked in this field extensively have confirmed that these systems correspond with the ship maintenance community's "tribal knowledge" of which systems are notoriously difficult. This thesis provides a quantitative substantiation of that tribal knowledge, and enhances it with a priority ranking that directs future work. An excellent path for follow-on research would be to repeat this type of analysis at the subsystem level for the nine systems identified in Section 5.1. This could determine whether individual components are driving the consistently poor estimates for the entire system.

> *...Furthermore, a multiple linear regression model derived from that data could be used to improve the Navy's estimates for New Work in future availabilities.*

Unfortunately, this objective was unsuccessful. It would have been a fantastic breakthrough to be able to provide the Navy with a process for improving New Work estimates, but insights gained while attempting to do this affirm the unsuitability of linear regression as a predictive tool for availability cost. The high degree of variability inherent to a specific availability implies the need for a flexible tool, and it may simply be that multipler linear regression is not the right tool for "solving" the

complexity of availability planning and execution. Future work in this area includes multiple linear regression of smaller, more coherent subsections of the data (such as a regression model for only one availability type), nonlinear regression models, or perhaps an investigation into the suitability of more advanced modeling techniques such as neural networks.

In addition to the value of a better estimation model for availabilities, a major concern that emerged during the course of this research is that of data quality. Compared to major public companies in the United States, the U.S. Navy does not produce the high-quality data that it needs in order to properly generate a digital model that is realistic and which can be used to optimize and anticipate future work. A documentation of the Navy's data-generation pipeline would do much to define this largely ignored institutional problem.

Appendix A

Tables

NNSY	PHNS	PNSY	PSNS
750 EOH 2012	698 DSRA 2009	691 PIRA 2007	711 EDSRA 2006
753 EOH 2014	698 DSRA 2014	700 PIRA 2009	758 DSRA 2007
756 DSRA 2007	699 DSRA 2011	700 PIRA 2014	759 DSRA 2008
764 DSRA 2009	701 PIRA 2009	706 DSRA 2007	
765 DSRA 2010	705 PIRA 2011	706 PIRA 2013	
770 DMP 2007	713 DSRA 2008	711 PIRA 2011	
	713 PIRA 2012	719 DSRA 2013	
	715 DSRA 2009	719 PIRA 2016	
	715 PIRA 2013	720 PIRA 2012	
	717 DSRA 2015	723 EOH 2008	
	721 EOH 2009	723 DSRA 2016	
	722 DSRA 2008	724 EOH 2007	
	722 EOH 2010	725 EOH 2009	
	724 DSRA 2012	751 EOH 2010	
	758 EOH 2014	752 EOH 2011	
	762 DSRA 2010	754 EOH 2013	
	763 DSRA 2012	756 EOH 2014	
	766 DSRA 2011	757 EOH 2013	
	770 DSRA 2014	760 DSRA 2008	
	771 DMP 2007	760 EOH 2015	
	771 DSRA 2012	767 DSRA 2009	
	772 DSRA 2014	772 DMP 2008	
	773 DMP 2008	774 EDSRA 2010	
	773 DSRA 2013		
	775 EDSRA 2012		
	776 EDSRA 2015		

Table A.1: Availabilities Studied in this Thesis, by Shipyard

PHNS	PSNS
698 DSRA 2009	711 EDSRA 2006
698 DSRA 2014	758 DSRA 2007
699 DSRA 2011	759 DSRA 2008
701 PIRA 2009	
705 PIRA 2011	
713 DSRA 2008	
713 PIRA 2012	
715 DSRA 2009	
715 PIRA 2013	
717 DSRA 2015	
721 EOH 2009	
722 DSRA 2008	
722 EOH 2010	
724 DSRA 2012	
758 EOH 2014	
762 DSRA 2010	
763 DSRA 2012	
766 DSRA 2011	
770 DSRA 2014	
771 DMP 2007	
771 DSRA 2012	
772 DSRA 2014	
773 DMP 2008	
773 DSRA 2013	
775 EDSRA 2012	
776 EDSRA 2015	

Table A.2: Availabilities with New Work Supporting Regression Analysis

000 001 004 006 007 013 014 015 022 025 028 032 033 034 035 036 037 040 041 042
043 044 045 051 052 054 055 061 066 071 072 073 075 076 080 082 083 084 087 089
091 100 107 111 112 117 122 123 125 126 127 128 131 132 139 140 147 150 152 156
157 165 167 170 176 177 178 180 182 183 184 185 186 187 191 201 203 204 206 207
208 209 211 212 215 216 217 218 219 223 227 229 230 231 232 233 236 237 238 239
241 242 243 244 245 246 247 248 252 253 254 255 256 258 262 264 277 300 302 303
304 307 308 309 310 311 312 314 321 323 324 325 331 332 337 342 400 404 406 407
408 409 415 416 417 418 421 422 424 425 426 427 428 429 430 432 433 435 436 437
438 439 440 441 442 443 444 446 451 452 455 457 461 462 465 469 479 481 482 487
491 493 497 501 502 503 505 507 508 509 512 513 514 515 517 518 520 521 524 527
528 529 530 531 532 533 534 538 539 540 541 542 543 546 547 551 552 553 554 555
556 558 561 562 563 564 566 567 568 569 581 582 585 592 593 594 595 597 603 605
606 607 608 609 611 612 621 623 631 633 634 635 636 638 641 650 651 655 660 665
671 691 708 709 714 715 716 717 721 751 752 754 777 778 815 821 822 823 826 827
828 829 830 831 832 835 836 837 839 840 841 843 844 846 847 848 849 853 854 860
861 901 902 903 904 905 908 920 936 938 958 959 990 998 999

Table A.3: Ship System Indices Analyzed in this Thesis

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	-19769	8816.4	-2.2423	0.031191	.971
111	-1.7706	1.104	-1.6038	0.11749	
131	2.1652	0.62152	3.4837	0.0013177	
132	1.7152	0.16494	10.399	2.1629e-12	
156	5.4956	2.7269	2.0153	0.051386	
170	-0.72837	2.0603	-0.35352	0.72576	
176	2.3567	0.22938	10.275	2.998e-12	
177	-1.6454	2.147	-0.76639	0.44844	
178	-5.6878	3.1268	-1.819	0.077232	

Table A.4: Ship System Index Regression Statistics Within SWBS 100

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	1.0208e+05	2.5664e+05	0.39777	0.75899	.945
201	23.313	43.49	0.53605	0.68674	
203	12.881	31.218	0.41262	0.75086	
204	-20.159	46.038	-0.43787	0.73726	
207	-44.664	78.038	-0.57233	0.66907	
208	-2.3396	61.01	-0.038348	0.9756	
209	3.9408	4.4156	0.89246	0.53614	
211	-6.8019	29.05	-0.23415	0.85358	
229	-47.301	195.61	-0.24182	0.84895	
230	-5.5822	46.3	-0.12057	0.92361	
231	-8.5768	59.76	-0.14352	0.90925	
233	20.311	74.283	0.27342	0.83009	
236	-33.762	59.965	-0.56303	0.67355	
237	-6.379	31.777	-0.20075	0.87388	
246	-35.911	87.621	-0.40984	0.75238	
248	-2.6382	4.3443	-0.60728	0.65256	

Table A.5: Ship System Index Regression Statistics Within SWBS 200

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	15577	5895	2.6425	0.077493	.96
300	2.9716	0.50284	5.9097	0.0096769	
302	-2.723	3.4052	-0.79966	0.48237	
303	-2.778	2.7213	-1.0208	0.38247	
304	-0.97392	6.1871	-0.15741	0.88492	
307	0.17672	0.83116	0.21262	0.84525	
308	-0.43196	1.6877	-0.25595	0.81454	
309	-6.9416	3.9012	-1.7793	0.17323	
310	10.818	4.2617	2.5384	0.084802	

Table A.6: Ship System Index Regression Statistics Within SWBS 300

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	-33759	17117	-1.9722	0.096058	.973
404	-1.7766	2.1413	-0.82966	0.43849	
407	1.1379	0.29777	3.8214	0.0087469	
415	3.6321	2.214	1.6405	0.152	
426	1.6031	2.0563	0.77963	0.46524	
429	-3.1374	2.6318	-1.1921	0.27823	
430	4.4806	1.5241	2.9398	0.025954	
435	1.8117	4.7604	0.38057	0.71663	
437	-3.1639	7.1803	-0.44064	0.67491	
438	-3.7605	4.1708	-0.90162	0.402	
440	10.646	10.121	1.0519	0.33335	

Table A.7: Ship System Index Regression Statistics Within SWBS 400

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	-70546	46872	-1.5051	0.37334	.996
501	0.60668	20.275	0.029923	0.98096	
502	-0.71165	4.216	-0.1688	0.89354	
503	12.55	19.953	0.62901	0.64255	
505	1.6594	18.412	0.090124	0.94278	
508	1.6805	8.4537	0.19879	0.87507	
513	-40.303	42.05	-0.95846	0.5135	
515	31.149	54.903	0.56734	0.67147	
518	0.54171	3.9272	0.13794	0.91274	
520	8.8227	14.075	0.62683	0.64355	
529	26.418	67.651	0.39051	0.76299	
530	-9.3367	15.747	-0.59292	0.65928	
533	-0.37154	5.6619	-0.065621	0.95828	
534	-28.09	79.736	-0.35229	0.78437	
539	-13.549	14.176	-0.95573	0.51441	
540	-16.009	22.366	-0.71578	0.60451	
542	-19.522	52.208	-0.37393	0.77219	
552	1.7395	0.76436	2.2758	0.26357	
553	36.534	42.35	0.86266	0.54685	
558	7.0646	14.495	0.48738	0.71129	
561	-3.2609	14.434	-0.22592	0.85855	
562	9.8975	8.1936	1.208	0.44022	
564	29.857	30.822	0.9687	0.51012	
567	20.59	32.182	0.6398	0.63765	
568	19.079	13.144	1.4516	0.38402	

Table A.8: Ship System Index Regression Statistics Within SWBS 500

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	16006	9961.9	1.6067	0.13213	.929
603	-23.029	51.118	-0.4505	0.65977	
605	-1.386	0.56664	-2.446	0.029433	
606	12.779	2.5936	4.9269	0.00027653	
607	1.8674	1.3148	1.4203	0.17905	
608	2.006	1.3446	1.4919	0.1596	
609	-12.062	8.7714	-1.3752	0.19232	
611	-1.7464	6.2783	-0.27817	0.78526	
612	-9.4367	7.7923	-1.211	0.24744	

Table A.9: Ship System Index Regression Statistics Within SWBS 600

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	-1484.4	3697.2	-0.40148	0.69111	.838
708	1.3207	0.58336	2.2639	0.031518	
709	1.4138	1.5446	0.91536	0.36782	
714	-0.088991	1.4999	-0.05933	0.95311	
715	1.1214	0.25476	4.4019	0.00014191	
717	2.1465	0.57603	3.7263	0.00087103	

Table A.10: Ship System Index Regression Statistics Within SWBS 700

SSI	Coefficient λ_i	Std Error	t-Stat	p-Value	R^2
(Intercept)	21472	20792	1.0327	0.41026	.995
827	1.4382	0.70806	2.0312	0.17932	
828	1.0981	0.10164	10.804	0.0084588	
830	-0.8357	3.6625	-0.22818	0.84071	
831	0.8476	0.28484	2.9757	0.096809	
839	-0.91661	1.7572	-0.52162	0.65395	
840	0.51593	1.5526	0.3323	0.77126	
841	8.5328	13.558	0.62934	0.59343	
844	0.60502	3.5103	0.17236	0.87902	
846	2.5559	6.8794	0.37153	0.74591	
847	-0.92986	3.8442	-0.24188	0.83141	

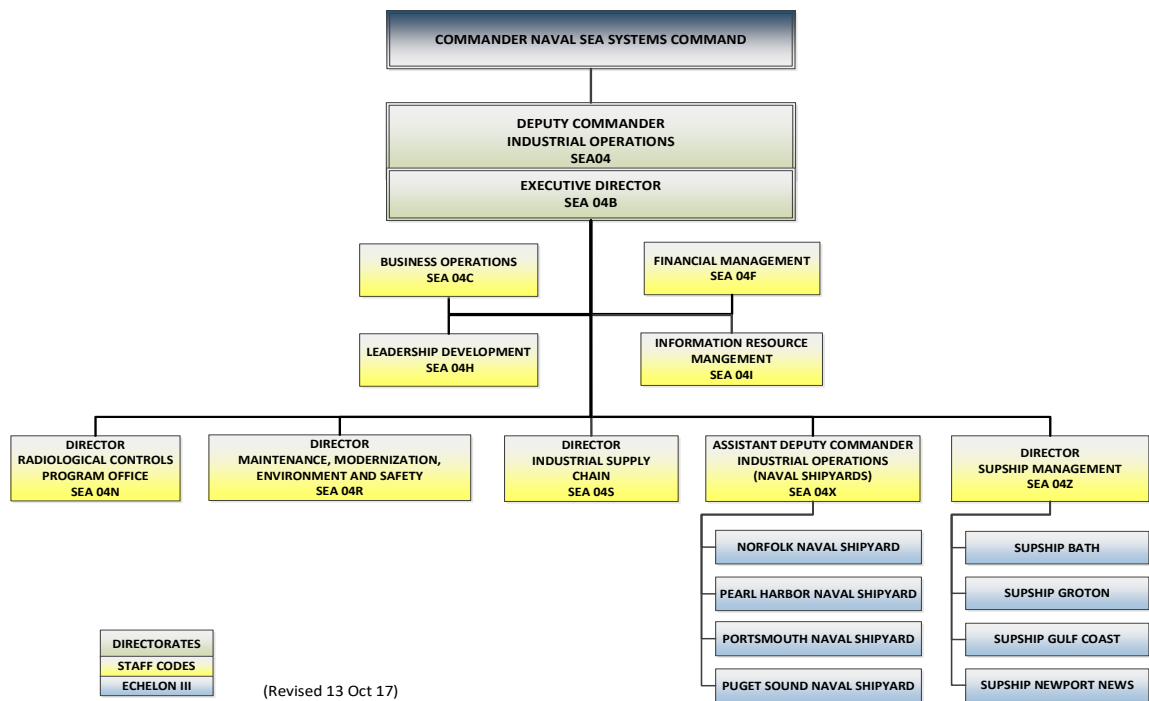Table A.11: Ship System Index Regression Statistics Within SWBS 800

Appendix B

Figures

Appendix 1-C: NAVSEA 04 Organization Chart

(Revised 13 Oct 17)

1-23

Figure B-1: NAVSEA 04 Organization Chart

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] PMS392, *03-29-17 Op Days Lost Metric.* Powerpoint slide, Washington Navy Yard, DC. Naval Sea Systems Command, Mar. 2017.

[2] "SSN-688 Los Angeles-class - Navy Ships." [Online; https://fas.org/man/dod-101/sys/ship/ssn-688.htm accessed 2018-03-29].

[3] R. O' Rourke, *Navy Virginia (SSN-774) Class Attack Submarine Procurement: Background and Issues for Congress.* CRS Report, Dec. 2017.

[4] W. F. Moran and U. S. NAVY, "Statement of Admiral William F. Moran, U.S. Navy Vice Chief of Naval Operations, Before the Subcommittee on Readiness of the Senate Armed Services Committee, on Current Readiness of U.S. Forces," p. 4, Feb. 2018.

[5] P. A. Grosklags, T. J. Moore, and U. S. NAVY, "Statement of Vice Admiral Paul A. Grosklags, Commander, Naval Air Systems Command, and Vice Admiral Thomas J. Moore, Commander, Naval Sea Systems Command, Before the Subcommittee on Readiness and Management Support of the Senate Armed Services Committee, on Depots, Shipyards, Arsenals, and Ammo Plants," p. 4, Mar. 2017.

[6] COMNAVSEA, *10 February 2017 Planning Summit.* Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser SEA 00/495, Dec. 2016.

[7] COMNAVSEA, *ACTIONS FROM NAVSEA 00 PLANNING SUMMIT.* Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser 07-021/081, Mar. 2017.

[8] J. L. Caprio and D. Leszczynski, *A Preliminary Investigation into CNO Availability Schedule Overruns.* Thesis, Monterey, California. Naval Postgraduate School, June 2012.

[9] M. J. White, *An Analysis of CNO Availability Performance Metrics and Their Relation to Availability Performance.* Thesis, Monterey, California. Naval Postgraduate School, June 2013.

[10] T. M. Nawara, *Resource Constrained Scheduling Problem at U.S. Naval Shipyards.* Thesis, Cambridge, Massachusetts. Massachusetts Institute of Technology, June 2013.

[11] R. L. Lary IV, *Analysis of SSN 688 Class Submarine Maintenance Delays*. Thesis, Monterey, California. Naval Postgraduate School, June 2017.

[12] W. J. Allison, *A Cost Model for Estimating Operating and Support Costs for U.S. Navy (Nuclear) Submarines*. Thesis, Monterey, California. Naval Postgraduate School, June 2000.

[13] R. Bender, "Introduction to the use of regression models in epidemiology.," 2009.

[14] K. Cullen, N. S. Stenhouse, K. L. Wearne, and T. A. Welborn, "Multiple regression analysis of risk factors for cardiovascular disease and cancer mortality in busselton, western australia–13-year study.," *Journal Of Chronic Diseases*, vol. 36, no. 5, pp. 371 – 377, 1983.

[15] Y.-N. Ma, J. Wang, G.-H. Dong, M.-M. Liu, D. Wang, Y.-Q. Liu, Y. Zhao, W.-H. Ren, Y. L. Lee, Y.-D. Zhao, and Q.-C. He, "Predictive equations using regression analysis of pulmonary function for healthy children in northeast china.," *Plos One*, vol. 8, no. 5, p. e63875, 2013.

[16] J. Rasero, N. Amoroso, M. La Rocca, S. Tangaro, R. Bellotti, and S. Stramaglia, "Multivariate Regression Analysis of Structural MRI Connectivity Matrices in Alzheimers Disease.," *PLoS ONE*, vol. 12, no. 11, pp. 1 – 16, 2017.

[17] M. Beaulieu, I. Gregory-Eaves, and F. Pick, "Nutrients and water temperature are significant predictors of cyanobacterial biomass in a 1147 lakes data set.," *LIMNOLOGY AND OCEANOGRAPHY*, vol. 58, no. 5, pp. 1736 – 1746, n.d.

[18] S. Li, J. Linn, and E. Muehlegger, "Gasoline Taxes and Consumer Behavior," p. 47.

[19] COMNAVSEA, *IMPLEMENTATION OF LEAN RELEASE 1.0*. Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser 04X/021, Jan. 2006.

[20] COMNAVSEA, *IMPLEMENTATION OF LEAN RELEASE 2.0*. Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser 04X/245, Sept. 2006.

[21] COMNAVSEA, *IMPLEMENTATION OF LEAN RELEASE 3.0*. Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser 04X/145, July 2007.

[22] COMNAVSEA, *USS LOS ANGELES (SSN 688) CLASS AND USS VIRGINIA (SSN 774) CLASS OPERATING INTERVALS*. Memorandum, Washington Navy Yard, DC. Naval Sea Systems Command Ser 392CM/0633, Sept. 2009.

[23] CBO, "Comparing Working-Capital Funding and Mission Funding for Naval Shipyards," Apr. 2007.

[24] GSA, "Federal Acquisition Regulation, Subpart 34.201(a)," Mar. 2005.