BENCHMARKING F-22 MISSION CAPABILITY RATES AND SORTIE OUTPUT THROUGH DATA ENVELOPMENT ANALYSIS

GRADUATE RESEARCH PROPOSAL

LeRoi G. Edwards, Major, USAF

AFIT-ENS-MS-22-M-123

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-22-M-123

BENCHMARKING F-22 MISSION CAPABILITY RATES THROUGH DATA ENVELOPMENT ANALYSIS

GRADUATE RESEARCH PROPOSAL

Presented to the Faculty

Department of Operational Sciences

Graduate School of Engineering and Management

Air Force Institute of Technology

Air University

Air Education and Training Command

In Partial Fulfillment of the Requirements for the

Degree of Master of Science in Logistics and Supply Chain Management

LeRoi G. Edwards, BS

Major, USAF

March 2022

DISTRIBUTION STATEMENT A.
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

AFIT-ENS-MS-22-M-123

BENCHMARKING F-22 MISSION CAPABILITY RATES AND SORTIE OUTPUT THROUGH DATA ENVELOPMENT ANALYSIS

LeRoi G. Edwards, BS

Major, USAF

Committee Membership:

Jeong Seong Joo, PhD
Chair

Abstract

The F-22 is combat-proven, operating in conflict areas for over a decade. Though it is the most dominant air-to-air fighter on the planet, incremental improvements to the aircraft continue to make the F-22 more lethal. Best practices to maximize the Mission Capability (MC) rate are not successfully codified and defended with data. This paper uses Data Envelopment Analysis (DEA) to identify benchmark environments where the MC rate is optimized and efficient. DEA successfully compared the relative efficiency of inputs and outputs across two units and determined the more efficient organization. Additionally, DEA provides current senior USAF leaders and tactical managers insight into performance environments where relative efficiency can be maximized to support the National Defense Strategy in a fiscally constrained environment. Finally, DEA models can be applied to analyzing additional F-22 units, other aircraft fleets, and more nuanced input/output relationships within base-level maintenance operations.

Table of Contents

List of Tables

I. Introduction

"Aircraft maintenance metrics are important." Key first words in the *Metrics Handbook for Maintenance Leaders*, which go on to explain how metrics can propel an organization to new heights if used correctly and effectively (Rainey, 2001). Performance metrics, or indicators, are the quantifiable measures that an organization uses to determine how well it meets its declared operational, strategic goals. Leading and lagging indicators effectively cause and effect, and this paper will shed light on how to view that relationship.

With warnings that metrics should be viewed in aggregate, not prescribed standards, pressures units to perform as efficiently as possible has warped the value and application of foundational aircraft maintenance metrics in the United States Air Force (USAF). This metrics handbook was published following the total quality management (TQM) initiatives of the 1990s. Despite the US Government Accountability Office's findings, "TQM had a moderate to a very positive impact on the Air Force's internal operations" (US. Government Accountability Office, 1993), Air Force leaders felt that it lacked relevance to military operations (Todorov, 2006).

The fall of 2005 introduced a new wave of process improvement and redesign, following a jarring memo published by General Michael Moseley introducing "Lean across the Air Force," challenging commanders to fundamentally change the culture and eliminate activities that do not add value to the mission (Moseley, 2005). This messaging pressed Airmen at all levels to eliminate waste around all core mission sets. This memo developed into the 2007 instruction "Air Force Smart Operations for the 21st Century (AFSO21) – CONOPS and Implementation Plan.

Performance metrics were tools during TQM and AFSO21 despite the mixed reviews of success and progress. However, performance indicators still reign supreme in leaders' eyes focused on making data-driven decisions, enabling procedures with facts versus anecdotes. Even aircraft fleets dependent on Contracted Logistics Support (CLS) are expected to adhere to the standards established by higher headquarters and require future contracts to be written concerning demonstrated metric sensitivity (ACC/A4M, 2018). The indoctrination that reportable metrics are relevant and infallible is systemic despite changing landscapes in aircraft health, data analytics, and DoD acquisition practices. Col McAneny argues that these metrics (despite warnings in handbooks) drive culture and influence behavior (2009).

Practically, maintenance managers are charged with assessing metrics and contributing to their unit meeting mission requirements while constantly searching for additional bandwidth to increase output. A commonly accepted set of reportable metrics has been used to describe fleet performance to review units. The metric division is into two categories: leading and lagging. The leading indicators express the immediate impact on a maintenance organization's ability to provide resources to conduct its operations. While lagging indicators show "firmly established trends," summed up as "the leading indicators will show a problem first, and the lagging indicators will follow" (Rainey, 2001).

The Mission Capable (MC) rate is a standard metric to determine aircraft fleet health is Mission Capable (MC) rate. This lagging indicator has highlighted the combat readiness of the Total Joint Force and its aging aircraft. In 2018, Secretary of Defense Jim Mattis ordered the Air Force and Navy to increase their MC rates to 80-percent for F-16, F-18, F-22, and F-35 fleets by September 2019. However, in 2020, Chief of Staff of the Air Force, General Charles "CQ" Brown, produced a written statement to the Senate Armed Forces Committee that the Air Force would no longer require the 80-percent threshold. However, he will still be relying on MAJCOMs to "determine required MC rates to meet readiness objectives" (2020).

Due to the small population of the F-22 and the constant weapons system upgrades, it is challenging to compare maintenance performance indicators (MPIs) to 4th generation fighter aircraft like the A-10, F-16, or F-15. However, as stated, national leaders rely on specific metrics to be a barometer of readiness for aircraft. This study provides a benchmark for Senior Leaders to understand the context where F-22 operations have existed most successfully concerning MC rate and sorties flown related to common leading indicators currently tracked as required from higher headquarters' policy. While summary in nature, the data and observations will provide insight to F-22 sustainment and provide context as to which current metrics provide more influence on fleet performance, focusing readiness efforts in today's constrained environments.

Data Envelopment Analysis (DEA) enables this study to focus on individual efficiency observations for each Data Measurement Unit (DMU). This paper focuses on DEA to use existing MPIs, to identify efficiency across combat coded F-22 squadrons within the United States Air Force. This study will advance the conversation about whether existing measurements are adequate for projecting and maintaining required readiness levels and provide a method to compare cause and effect beyond simple correlation. W.F. Bowlin (1984) explained that DEA would quantify the inefficiencies in an entity through observations and enable a heightened sense of managerial control of those variables. That control epitomizes tactical aircraft maintenance management. USAF A4/7, Lt Gen Judith Fedder explained, "We cannot understand or improve our processes unless we know what our systems are telling us. We need tools to guide our efforts and focus our resources…it is a fundamental tenet that what gets measured gets improved." (Department of the Air Force, 2018).

Research Question: How can USAF aircraft maintenance management systems be changed to improve the identification and correction of factors which limit the efficiency of F-22 military operations using established performance indicators?

II. Literature Review

The application of DEA pertaining to a 5^{th}-Generation aircraft fleet required an understanding of performance metrics as a whole. A study of related studies applying performance metrics to aircraft maintenance organizations provides context for this research. Understanding how academia and industry leverage performance indicators provide insight into best practices illustrates the cause and effect relationship that Air Force maintenance leaders expect from data (Rainey, 2001). Additionally, studying the application of DEA on large-scale government entities guarantees a balanced perspective of appropriate modeling. Finally, the comparison provided scope and language to assimilate the existing work body.

Aircraft Maintenance Related Studies

Taaffe et al. (2014) analyzed performance indicators for aircraft maintenance. These metrics were designed to ensure process control and maintain organizational success. Their team assessed Lockheed Martin's existing performance metrics. Taaffe measured each metric, using a scorecard, against five graded elements and determined its effectiveness. This research aligned with the balanced scorecard (BSC) and "SMART" metrics introduced by Kaplan and Norton in the 1990s. This approach is more qualitative than DEA; a series of surveys assisting in aligning efforts determined the calculated weight of each area's score. The qualitative analysis of critical metrics successfully revealed that not all metrics were appropriately presented, and some could be removed entirely. Because of the simplicity, the efforts could be easily replicated in other organizations.

Verhoeff et al. (2015) highlight the priority of military aviation to maximize operational readiness. They surmise that standard practice is reactive as daily management combats aircraft availability, serviceability, and the fleet's sustainability. Their flight and maintenance planning

(FMP) optimization model proactively and comprehensively addresses those variables to maximize operational capability. This novel linear program requires inputs and provides optimized outputs. The authors applied their model to Royal Netherlands Air Force aircraft over three years. Their findings indicated a defined solution for long-term flight and maintenance schedules while encompassing user-defined inputs for mathematically optimized responses to contingencies.

Werner et al. (2019) polarize maintenance requirements like the USAF in that you have scheduled, preventative, maintenance, and unscheduled, reactive maintenance. Focusing on the Remaining Useful Life (RUL), preventive maintenance projections rely on significant repositories of relevant data. Constructing models and algorithms is difficult, but they argue that it is critical to compare the estimated RUL with the actual performance of production machines, ultimately improving the simulation for future cycles. They conclude that machines can utilize modeling based on data acquisition and physics-based modeling from reoccurring inspection with existing sensor data. The digital twin would then optimize itself and provide insight into the physical counterpart. While there may not be capacity for all machines to incorporate this, new systems should require performance-based feedback and optimization strategies.

DEA Related Studies

Clarke and Goudin (1991) applied DEA to the vehicle maintenance function of a large-scale, non-profit logistics network. Logistics managers were interested in measuring the relative efficiency of the 17 vehicle maintenance shops as related to transportation and distribution activities. Their study measured available resources like workforce, material costs, the quantity of personnel, and adjust vehicle days. The model provided by Charnes, Cooper, and Rhodes, observing the relative efficiency of each shop, shaped their methodology. Unique to this study, efficiency results were presented to shop managers, who were charged with improving their scores. Their findings concluded that all 17 shops showed efficiency improvement over the four-year study. Additionally, the local managers found the DEA information understandable and actionable without formal grounding in operations research.

Han and Sohn (2011) proposed that DEA assist the Republic of Korea's Air Force (ROKAF) in proactively managing vast inventories of military assets. Each supply squadron evaluated nineteen output variables without consideration to the inputs. Thus, DEA effectively considers these units without assuming which inputs or outputs are most important. The study suggests that grouping inventory management functions would be more efficient than current practices with no grouping. An analysis compares each base individually and also provides 15 groups to evaluate. Three input and four output variables were identified through surveys conducted. It was determined that grouping the base supply systems by aircraft would be the most efficient. DEA models (CCR and BCC) supported these findings.

Charnes et al. (1985) introduced the DEA application to military aircraft maintenance to evaluate capability and efficiency. Fourteen Air Force Wings were selected and measured through input and output variables used by commanders at the time. They examined the relative maintenance efficiency of each organization. Their study enabled Air Force leaders in the Tactical Air Command to determine, objectively, whether efficiency values varied between units and their mission types. Findings indicated that the unit's efficiency scores, whose primary mission was devoted to training, were uniformly high.

Additionally, units were qualitatively compared to enable further studies to identify the sources and amounts of inefficiency. They also noted that DEA does involve optimizations. Finally, this route provides access to bodies of theory (e.g., from economics and mathematical programming) that are available both for immediate use and further extensions.

Goulany and Seroussy (1989) targeted Israeli Air Force aircraft maintenance units by evaluating efficiency through direct inputs and outputs. Rather than simply focus on a DEA-generated efficiency, the study took additional effort to consider their input and output factors as to which were most correlated with overall efficiency. The resulting research developed the hierarchical efficiency monitoring system (HEMS), which allowed the study to assess various management levels of the Israeli Air Force maintenance division. Essentially, the HEMS enabled the data to be analyzed for relative efficiency through DEA at different reference sets. These differences enabled general inferences as to efficiencies at each level of management. Further development of the HEMS would enable a sharper management control tool with periodic reviews of the pertinent output and input factors.

Sutton and Stanko (2013) highlighted the complexity of managing global assignments to the US Navy. To analyze the problem with multiple objectives, metrics, and performance indicators, they designed a DEA-based tool, the Generalized Symmetric Weight Assignment Technique (g-SWAT). This denomination enabled the minimization of overall cost while maximizing other reportable metrics. They argue that the success of their model can be applied beyond the Navy's PCS concerns, but other DoD personnel processes and in the private sector to align services with clientele specifications. Recognizing that g-SWAT and DEA are not the only viable solutions to this problem set, their findings outperformed current methods being used by the US Navy.

III. Methodology

DEA is used to measure the relative efficiencies of F-22 maintenance performance indicators on the overall MC Rate in a given month. Like all production theories, this focuses on manipulating leading and lagging indicators to represent input conversion to output. DEA is a unique application of linear programming based on the frontier methodology of Farrell (1957). Since Farrell, a significant breakthrough for advancing DEA was performed by Charnes et al. (1978) and Banker et al. (1984). DEA supports the observation of multiple input and output variables simultaneously with ambiguous preference. The measured combinations of input and output variables are called decision-making units or DMUs. Because DEA can identify relatively efficient DMU(s) among a sample of given DMUs, it is a reliable tool for comparative analysis or benchmarking.

To explore the mathematical property of DEA, let E_0 be an efficiency score for the base DMU *0* then,

$$\text{Maximize } E_0 = \frac{\left\{\sum_{r=1}^{R} u_{r0} y_{r0}\right\}}{\left\{\sum_{i=1}^{I} v_{i0} x_{i0}\right\}} \tag{1}$$

subject to

$$\frac{\left\{\sum_{r=1}^{R} u_{r0} y_{rk}\right\}}{\left\{\sum_{i=1}^{I} v_{i0} x_{ik}\right\}} \leq 1 \text{ for all } k \tag{2}$$

$$u_{r0}, v_{i0} \geq \delta \text{ for all } r, i, \tag{3}$$

where

y_{rk}: the observed quantity of output r generated by unit $k = 1, 2, \ldots, N$,

x_{ik}: the observed quantity of input i consumed by unit $k = 1, 2, \ldots, N$,

u_{r0}: the weight to be computed given to output r by the base unit 0,

v_{i0}: the weight to be computed given to input i by the base unit 0,

δ: a very small positive number.

Linear fractional programming compares the ratio of two linear functions, and that relationship can convert into a standard linear programming (LP) model. An assumption of LP is a linear relationship among variables. Consequently, an LP for DEA utilizes a constant returns-to-scale so that all observed input/output combinations can be scaled up or down relative to the other (Charnes et al., 1978).

However, when we use a piecewise LP, we can model returns-to-scale that do not possess a proportional relationship, such as an increasing, decreasing, or variable-returns-to-scale (Banker et al., 1984). Thus, different types of DEA models are available depending on returns-to-scales and various modeling approaches.

Sherman and Ladino (1995) summarize the capability of DEA in the following manner:

- Identifies the best practice DMU that uses the least resources to provide its products or services at or above the quality standard of other DMUs;
- Compares the less efficient DMUs to the best practice DMU;
- Identifies the number of excess resources used by each of the less efficient DMUs;
- Identifies the amount of excess capacity or ability to increase outputs for less efficient DMUs, without requiring added resources.

This study highlights the versatility of DEA for comparative benchmarking and quantitative feedback for organizational inefficiency. A Charnes-Cooper-Rhodes (CCR) model, a

Banker, Charnes, and Cooper (BCC) model, and a slack-based measure of efficiency (SBM) are employed. All models view the data from input and output-oriented perspectives. First, relative efficiency scores were assessed using CCR and BCC models. Next, two DMUs are highlighted, exploring the additional information provided by DEA to the information owner. Finally, slack-based measures are considered, providing actionable insight to unit maintenance managers.

Data and Variables

The Logistics, Installations, and Mission Support – Enterprise View (LIMS-EV database provided the source data analyzed for this project—the same raw data available to Senior Leaders and decision-makers in the USAF. The decision-making units (DMUs) selected for the study are monthly performances from two USAF Wings that operate a similar quantity of F-22 aircraft. Performance efficiency measurements utilize commonly stated "leading" and "lagging" indicators, providing a cause and effect context (Rainey 2001). DMUs will not refer to the wing and month utilized to preserve confidentiality but instead indexed 1 through n. We collected data from 2018-2020, incorporating aircraft inventory and seven “leading indicator” inputs, contrasting two "lagging indicator" outputs and total sorties flown during that time. Using actual data provided from LIMS-EV enables future research and analysis for tactical or academic application. The inputs and outputs used in this study, with accompanying definitions, are shown in the following table:

Table 1: Input/Output Variables and Definitions

<table>
<tr><th colspan="2">INPUTS</th><th>OUTPUTS</th></tr>
<tr><td colspan="2">BREAKS
CANN</td><td>MCHRS
12HRFIX</td></tr>
<tr><td colspan="3">Definitions of Input and Output Variables</td></tr>
<tr><td>BREAKS</td><td colspan="2">The number of code-3 breaks in a wing in a specific month.</td></tr>
<tr><td>12HRFIX</td><td colspan="2">The number of code-3 breaks in a wing in a specific month, fixed within 12 hours.</td></tr>
</table>

CANN	The total number of reported cannibalizations on an installation in a specific month.
MCHRS	The total number of hours in a specific month in which the possessed aircraft on each wing were fully or partially mission capable.

Some metrics were scaled by multiplication with a scalar to compare all values best. Scaling metrics is an acceptable practice in DEA as an optimum DMU will not vary with the units of measurement. However, the BREAKS and CANN outputs are undesirable as they track adverse effects on the fleet. To best compare the 12HRFIX and MCHRS outputs, we measure BREAKS and CANN by subtracting the LIMS-EV values from 1,000. Associating higher values with positive functions ensures that as negative instances increase, the calculated value decreases. The number 1,000 is used arbitrarily but enables calculated values to remain near 1,000. The consistent application consistently views variables with higher values as favorable.

Since the sample size influences DEA results, some discussion on the adequacy of sample size is warranted here. The sample size utilized in the present study is consistent with the various rules of thumb available in DEA literature. Cooper et al. (2007) provide two such rules: that sample size should be greater than equal to the product of inputs and outputs, and that number of observations in the data set should be at least three times the sum of the number of input and output variables. Given two inputs and two outputs, the sample size should be at least 12 DMUs. The 72 used in the present study exceeds the desirable size as suggested by the abovementioned rules of thumb to obtain sufficient discriminatory power.

Descriptive statistics of the raw data for Unit A, Unit B, and combined data are listed below. It is worth noting that Unit A has standard deviations 1.6 times those of Unit B after dropping CANN. However, even these summaries of the two units drive the question of which is more efficient and why.

Table 2: Unit A Input/Output Descriptive Statistics (36 DMUs)

	Max	Min	Mean	Std Dev
BREAKS	114.00	27.00	61.56	19.02
12HRFIX	72.00	16.00	36.78	12.46
CANN	61.00	0.00	19.08	17.01
MCHRS	22,374.57	6,322.22	14,964.84	3,678.18

Table 3: Unit B Input/Output Descriptive Statistics (36 DMUs)

	Max	Min	Mean	Std Dev
BREAKS	62.00	12.00	31.56	9.52
12HRFIX	49.00	6.00	21.94	7.48
CANN	7.00	0.00	1.08	1.67
MCHRS	31,790.13	19,668.05	25,325.22	2,537.38

Table 4: Cumulative Input/Output Descriptive Statistics (72 DMUs)

	Max	Min	Mean	Std Dev
BREAKS	114.00	12.00	46.56	21.24
12HRFIX	72.00	6.00	29.36	12.67
CANN	61.00	0.00	10.08	15.07
MCHRS	31,790.13	6,322.22	20,145.03	6,067.79

These variables have been curated and labeled by decades of anecdotal experience, as noted in the *Metrics Handbook for Maintenance Leaders*: "leading indicators are those that directly impact maintenance's capability to provide resources to execute the mission. Lagging indicators show firmly established trends" (2001). These variables will show, in part, efficiency in this accepted relationship. This vignette into the LIMS-EV database applied to the F-22 can provide insight into tactical maintenance practices, multi-year health of fleet assessment, and a perspective on the strategic application of how metrics are collected at the unit level. This data was collected at the unit level and subsequently archived. Choosing only two inputs and outputs keeps the experiment parsimonious. Future applications tailored to the MDS, sustainment climate, or management preference could be curated for a better fit. DMU efficiencies will be inherently influenced by under and over-reported metrics. It is not in the scope of this paper to provide context or alibis for those instances. Instead, it is a discussion point as to whether the data collected is value-added, reliable, or actionable in the first place.

IV. Results and Discussion

The first step was to apply CCR and BCC DEA models to each unit and their related DMUs to express each unit's efficiency. Next, we compared the relative efficiency against all DMUs, combining both units. Finally, a slack-based model was applied to each DMU, measuring individual performance, assessing the inefficiency of inputs, and identifying benchmarks for future managerial consideration. DEA enables the comparison of all DMUs simultaneously, allowing for identifying best practices. For example, those less efficient DMUs might be more closely analyzed to determine weaknesses, or the most efficient DMUs might be further researched to establish a performance benchmark.

Due to the balance of reducing resource consumption for a given goal and maximizing outputs for given resources, DEA considers input and output-oriented models. Both orientations measure the technical efficiency of each DMU. Our first DEA model, CCR, processes DMUs under an assumption that constant returns-to-scale (CRS) measuring inefficiency based on input/output configuration and the size of operation—as inputs increase, outputs increase on a linear scale. BCC, our second model, compares DMUs with variable returns-to-scale (VRS), allowing for the measurement of managerial underperformance—there are multiple efficiency edges of production, independent of scale. Comparing Units A and B through output-oriented CRS and VRS DEA models will provide concerned managers insight into under-realized productivity and wasted resources in current production efforts.

Table 5: Unit A DEA Efficiency Scores

DMU	CRS-O	VRS-O
a-1	0.9944	1.0000
a-2	0.9909	0.9934
a-3	1.0000	1.0000
a-4	1.0000	1.0000
a-5	0.9984	0.9987
a-6	0.9875	0.9905
a-7	0.9949	0.9955
a-8	1.0000	1.0000
a-9	0.9940	0.9981
a-10	1.0000	1.0000
a-11	1.0000	1.0000
a-12	0.9999	1.0000
a-13	1.0000	1.0000
a-14	0.9905	0.9941
a-15	0.9997	1.0000
a-16	1.0000	1.0000
a-17	0.9881	0.9882
a-18	0.9936	0.9987
a-19	1.0000	1.0000
a-20	1.0000	1.0000
a-21	0.9896	0.9926
a-22	0.9943	0.9965
a-23	0.9997	0.9998
a-24	0.9854	0.9918
a-25	0.9863	0.9959
a-26	0.9933	1.0000
a-27	0.9789	0.9848
a-28	0.9781	0.9866
a-29	0.9784	0.9900
a-30	1.0000	1.0000
a-31	0.9848	0.9877
a-32	1.0000	1.0000
a-33	0.9897	0.9938
a-34	0.9961	0.9984
a-35	0.9930	1.0000
a-36	0.9948	0.9990
Avg.	0.9937	0.9965
σ	0.00666	0.00456

Table 6: Unit B DEA Efficiency Scores

DMU	CRS-O	VRS-O
b-1	0.9785	0.9932
b-2	0.9817	1.0000
b-3	0.9809	0.9928
b-4	0.9807	0.9898
b-5	0.9774	0.9896
b-6	0.9751	0.9886
b-7	0.9808	0.9942
b-8	0.9816	0.9935
b-9	0.9798	0.9945
b-10	0.9845	0.9927
b-11	0.9930	0.9995
b-12	0.9965	1.0000
b-13	1.0000	1.0000
b-14	0.9866	0.9968
b-15	0.9899	0.9940
b-16	0.9868	0.9937
b-17	0.9784	0.9886
b-18	0.9754	0.9879
b-19	0.9744	0.9858
b-20	0.9792	0.9921
b-21	0.9749	0.9878
b-22	0.9749	0.9861
b-23	0.9774	0.9915
b-24	0.9855	0.9980
b-25	0.9794	0.9911
b-26	0.9776	0.9943
b-27	0.9735	0.9891
b-28	0.9743	0.9913
b-29	0.9798	0.9948
b-30	0.9693	0.9754
b-31	0.9870	0.9980
b-32	0.9790	0.9906
b-33	0.9808	0.9922
b-34	0.9738	0.9890
b-35	0.9823	0.9965
b-36	0.9798	0.9909
Avg.	0.9809	0.9923
σ	0.00645	0.00482

Table 7: Cumulative DEA Efficiency Scores

	CRS-O	VRS-O
Unit A μ	0.9937	0.9965
Unit A σ	0.00666	0.00456
Unit B μ	0.9809	0.9923
Unit B σ	0.00645	0.00482
Combined μ	0.9873	0.9944
Combined σ	0.0093	0.0052

These DMUs can be compared in multiple ways, qualitatively, depending on managerial interest. For discussion, efficiency scores were assessed as to whether F-22 maintenance metrics indicated improved relative efficiency or reduced. In all models, both units show negative linear trend lines from 2018-2020, with Unit A outperforming Unit B. These consistent trends support Unit A displaying a higher average efficiency score and a lesser standard deviation across its 36 DMUs.

To illustrate the managerial application, we will propose that Unit B outperforms Unit A:

H_o: Both Unit A and Unit B are equally efficient.

H_a: Unit B is more efficient than Unit A.

The hypothesis is rejected, where Unit B performed worse than Unit A concerning CRS and VRS, at $\alpha = 0.05$. This conclusion was reinforced when we viewed the data from an input-oriented lens. Table 8 assesses DMUs with a bilateral CRS model, measuring deviation from the most efficient DMU. This enables the DMUs to no longer be bound between 0 and 1. Through this, we retained the above hypothesis and concluded, again, that the null hypothesis must be rejected, this time at well over $\alpha = 0.01$.

Table 8 Bilateral CRS-Input Efficiency Scores

Rank	DMU	Score	Rank	DMU	Score
1	a-13	1.0622	**37**	a-28	0.9972
2	a-8	1.0382	**38**	b-11	0.9932
3	a-1	1.0323	**39**	a-29	0.9930
4	a-3	1.0308	**40**	b-15	0.9899
5	a-20	1.0246	**41**	b-31	0.9870
6	a-30	1.0232	**42**	b-16	0.9869
7	a-32	1.0226	**43**	b-14	0.9867
8	a-23	1.0224	**44**	b-24	0.9855
9	a-19	1.0221	**45**	b-10	0.9845
10	a-4	1.0209	**46**	b-35	0.9823
11	a-34	1.0185	**47**	b-2	0.9817
12	a-16	1.0148	**48**	b-8	0.9816
13	a-36	1.0142	**49**	b-3	0.9809
14	a-15	1.0138	**50**	b-33	0.9808
15	a-5	1.0135	**51**	b-7	0.9808
16	a-7	1.0131	**52**	b-4	0.9807
17	a-26	1.0128	**53**	b-9	0.9798
18	a-22	1.0117	**54**	b-29	0.9798
19	a-10	1.0116	**55**	b-36	0.9798
20	a-35	1.0115	**56**	b-25	0.9794
21	a-33	1.0111	**57**	b-20	0.9792
22	b-13	1.0109	**58**	b-32	0.9790
23	a-17	1.0098	**59**	b-1	0.9785
24	a-9	1.0097	**60**	b-17	0.9784
25	a-21	1.0087	**61**	b-26	0.9776
26	a-18	1.0085	**62**	b-5	0.9774
27	a-11	1.0080	**63**	b-23	0.9774
28	a-12	1.0075	**64**	b-18	0.9754
29	a-2	1.0071	**65**	b-6	0.9751
30	a-31	1.0062	**66**	b-22	0.9749
31	a-6	1.0051	**67**	b-21	0.9749
32	a-14	1.0046	**68**	b-19	0.9744
33	a-24	1.0016	**69**	b-28	0.9743
34	a-25	1.0007	**70**	b-34	0.9738
35	a-27	1.0003	**71**	b-27	0.9735
36	b-12	0.9978	**72**	b-30	0.9693

A tool enabled by DEA is the analysis of slack for inefficiencies within a DMU. Slacks relate to the further increases in output or reduction of input that could be gained. This specific feedback to a DMU on historical performance represents a potential improvement compared to efficient peers and provides benchmark targets. To illustrate, we can look at two DMUs where Unit A's efficiency out-performed Unit B's efficiency within the CRS-O lens. The "Observed Value" represents the value used for the DEA calculations, not the raw data pulled from LIMS-EV. The "Efficient Value" in the final column is calculated from the "Measured Value" added to the "Slack" value. When we use the SBM to evaluate the context, we can have an appropriate stratification of the DMU performance levels. Of note, DMUs A-31 and B-31 represent July 2020.

Table 9: DMU A-31 Slack
Efficiency = 0.6155

INPUTS	**OBSERVED VALUE**	**SLACK**	**EFFICIENT VALUE**
BREAKS	927	0.0%	927
CANN	1000	-0.21%	997.95
OUTPUTS	**--**	**--**	**--**
MCHRS	8504.6	124.94%	19130.32
12HRFIX	957	0.0%	957

Table 10: DMU B-31 Slack
Efficiency = 0.5796

INPUTS	**OBSERVED VALUE**	**SLACK**	**EFFICIENT VALUE**
BREAKS	955	0.0%	955
CANN	1000	0.0%	1000
OUTPUTS	--	--	--
MCHRS	8504.6	145.09%	20844.09
12HRFIX	980	0.0%	980

The significant inefficient output of MCHRS indicates that there may be more at play for this period since both DMUs were significantly inefficient compared to the entire sample.

Running the program with tailored variables could provide additional managerial insight. The amalgamation of data is a cornerstone of DEA; DMUs are measured on an equal plane, comparing relative efficiency rather than optimization based on the simple correlation.

Table 11: SBM CRS-O Efficiency Scores

Rank	DMU	Score	Rank	DMU	Score
1	a-3	1.000	**37**	b-2	0.8501
1	a-4	1.000	**38**	b-25	0.8482
1	a-8	1.000	**39**	b-36	0.8481
1	a-10	1.000	**40**	a-33	0.8456
1	a-11	1.000	**41**	a-17	0.8328
1	a-13	1.000	**42**	b-9	0.8322
1	a-16	1.000	**43**	b-35	0.8272
1	a-19	1.000	**44**	b-5	0.8243
1	a-20	1.000	**45**	a-26	0.8146
1	a-30	1.000	**46**	b-7	0.8003
1	a-32	1.000	**47**	a-6	0.79501
1	b-13	1.000	**48**	b-32	0.7919
13	a-12	0.9998	**49**	b-8	0.7919
14	a-15	0.9980	**50**	b-33	0.7852
15	a-23	0.9928	**51**	b-17	0.7777
16	b-12	0.9896	**52**	b-34	0.7746
17	b-11	0.9765	**53**	b-23	0.7709
18	a-5	0.9681	**54**	a-21	0.7689
19	b-15	0.9554	**55**	b-22	0.7524
20	b-14	0.9467	**56**	a-18	0.7511
21	a-14	0.9424	**57**	b-1	0.7498
22	b-16	0.9396	**58**	b-6	0.7353
23	a-34	0.9319	**59**	b-19	0.7059
24	b-10	0.9315	**60**	b-20	0.6913
25	a-36	0.9261	**61**	b-21	0.6887
26	a-7	0.9123	**62**	b-26	0.6878
27	a-9	0.90888	**63**	b-28	0.6767
28	a-1	0.9047	**64**	a-28	0.6736
29	b-24	0.9007	**65**	b-27	0.6612
30	a-2	0.8834	**66**	a-27	0.6592
31	a-24	0.8830	**67**	b-29	0.6315
32	a-35	0.8820	**68**	b-18	0.6310
33	a-22	0.8814	**69**	a-29	0.6269
34	b-4	0.8653	**70**	a-31	0.6155
35	a-25	0.8588	**71**	b-31	0.5796
36	b-3	0.8546	**72**	b-30	0.4327

Table 12: Summary SBM CRS-O Efficiency Scores

		Average	Max	Min	St Dev
	Eff. Score	0.8439	1.0000	0.4327	0.1311
	DMU Rank	35.58	72.00	1.00	22.27
BREAKS	**Data**	953.44	988.00	886.00	21.39
	Projection	952.71	985.18	886.00	20.83
	Diff.(%)	-0.08	0.00	-0.56	0.16
CANN	**Data**	989.92	1000.00	939.00	15.17
	Projection	989.86	1,000.00	939.00	15.17
	Diff.(%)	-0.01	0.00	-0.24	0.04
12HRFIX	**Data**	970.6389	994	928	12.7603
	Projection	970.85	994.00	928.00	12.68
	Diff.(%)	0.02	0.75	0.00	0.10
MCHRS	**Data**	14,964.84	22,374.57	6,322.22	3,704.00
	Projection	20,310.90	22,854.51	6,322.22	3,187.60
	Diff.(%)	43.75	261.50	0.00	45.75

Summarizing the data enables managers to view DMU efficiency more abstractly, potentially mitigating biases. Case in point, we see pairs of DMUs in the bottom third of the data set: X-27, X-28, X-31, etc. Identifying cross-installation trends might enable local and enterprise managers to conduct root-cause analysis with those 6 data points more quickly. This DEA analysis would provide further incentive for higher headquarters to oblige as we can see the potential return on investment (ROI) in the form of MCHRS. Bowlin (1984) expresses that because there is no inherent "market mechanism," the USAF relies on historical consumption data to determine budget and requirements. Understanding the cost and benefit of input and output would enable managers at all levels to compare the price of paying for improved supply chain responsiveness and the ROI of more mission capability with data to support those decisions.

Finally, the limitations of this analysis include the use of LIMS-EV data. The F-22 uses Integrated Maintenance Information System (IMIS) to consolidate maintenance and repair data for Lockheed Martin and USAF managers worldwide. Unfortunately, for that IMIS data to be translated into a format to be archived into LIMS-EV, it is manipulated and filtered through other USAF systems like IMDS.

In this study, we used CANN as an input. CANN may be misleading from week to week or month to month. CANN accounts for the frequency when a unit cannibalizes a part from one aircraft to fix another; however, how that action is accounted for can vary from unit to unit. Based on local policies, a more specific variable might be more appropriate to consider.

Additionally, the variables identified as an input versus output might be interchangeable based on perspective. That ambiguity would be a barrier to the sustained application of DEA for the Department of the Air Force. Aircraft Availability (AA) rate might be considered for future study iterations as it is a subset of TAI. The Department of the Air Force defines AA in AFI 3-4.21V1 (2018) as the proportion of TAI capable of meeting flying mission requirements. AA might also be considered an output, vice MC rate because it accounts for aircraft that are not impacting MC rate but are not available for a unit's use.

Maintenance managers at all levels could apply DEA to their interests. USAF performance indicators have been refined as leading and lagging; they are taught to the lowest levels as a cause-and-effect relationship (Rainey, 2001). Because of this assumed relationship, the correlation between egregious negative outputs and negative inputs is referenced to create action plans to address mission performance. Indexing relative efficiency against past performance would be a more objective perspective and analytical starting point to review performance. For USAF Maintenance Managers, metrics should be able to inform:

1. Correct decision behaviors
2. A clear understanding of process relationships
3. Process improvement

   (Department of the Air Force, 2018)

With additional analysis, it would be possible to compare "traditional" metrics and learning models against teams using DEA regularly. After months of trial, a survey and performance data could be extracted to see if DEA effectively met USAF expectations for core metrics.

Strategic Managers would use DEA to understand the cause and effect of their inputs and outputs. Understanding the data relationships would likely drive insight into new data collected and added to the model. Bowlin (1984) described non-profit organizations uniquely to know which results they desire but often struggled to evaluate effectiveness without some subjectivity. As illustrated in Tables 10 and 11, strategic decision-makers now can quantitatively compare inefficient units and have milestones for ROI in the currency of MCHRS. Conversely, they can decide if increasing efficiency is worth the money. Headquarters can make informed decisions using cost per flight hour, or sorties per MCHR, and a cost to improve the contracted logistics support. If refined, a level of budgeting awareness could be applied. Annually, the "Use-It or Lose-It" mantra is popularized as depicted in the *Military Times* (Rempfer, 2019). Bowlin (1984) comments, “If an Air Force wing has been operating inefficiently in the past; historical costs will include the cost of operating inefficiently.” Rather than reward inefficient Units with larger flying hour budgets, efforts and incentives to create efficiency may allow creative solutions to a fiscally constrained future.

Unfortunately, a significant issue for popular Air Force metrics is that they are not accurate inputs. These metrics mainly summarize outputs such as Aircraft Availability, 8-/12-HR Fix Rates, Sorties, etc. Without accurately capturing the information or resources, how can managers understand the performance (efficiency) of the maintenance processes? If inputs such as labor hours are not considered, you just focus on outputs. However, the Catch-22 is that these outputs drive the bills and are performance indicators for existing contracts. Ever-increasing sustainment costs will continue to be a considerable discussion for the lifecycle of many aircraft fleets.

V. Conclusion

While the Air Force logistics community has been pressed to Accelerate, Change or Lose and has birthed projects like Tesseract. Aircraft maintenance managers' tools to interpret production data are older than the maintained aircraft. Based on a Government Accountability Office report published, the F-22 MC rates have fallen over the last decade. Performance losses are, in part, is due to degraded stealth coatings, aircraft spare parts shortages, and more flight hours per aircraft than originally programmed (McCullough, 2020). The 5th Generation fighter fleet needs improved tools to maintain readiness until their watch is over. This study provided 5th Generation context to aircraft metrics applied to USAF maintenance operations for decades. The relative efficiency between highlighted leading and lagging indicators provides benchmark performance efficiency between F-22 units and uses fleet data to inspire additional research into areas of inefficiency.

Further DEA analysis can be used to fine-tune the cause and effect relationship valued by the USAF aircraft maintenance community or refine the data collected within LIMS-EV. Assessing fleet trends with DEA will empower our newest Airmen and long-time industry partners to preserve the F-22 fleet for conflicts to come. Explicitly comparing inputs and outputs enables tactical, unit-level, and strategic, USAF-wide, analysis, and comparison of strengths and weaknesses within any mission designation of aircraft. Maintenance managers can quickly identify which variables affect their operations significantly while more generally indexing relative performance efficiency at the tactical level. These analyses are accomplished using slack measurements and comparable efficiency scores highlighted in the research. Strategically, DEA provided a distinct vantage point towards the profusion of aircraft data and required performance reporting. Calculating inputs and outputs, coupled with relative efficiency accurately, allows the in-kind comparison of an aircraft being non-mission capable due to spare parts and the potential training (or combat) sorties. Quantifying that data enables direct comparisons for purposes of resource management and overall mission execution.

The application of DEA was limited in this study due to software limitations, data accuracy, and the impacts of non-discretionary inputs. The USAF maintains data for all aircraft fleets across many years in the LIMS-EV system. That data, however, is a reflection of various system inputs. For the F-22, information is translated through multiple Lockheed Martin software interfaces, USAF software, and ultimately to LIMS-EV. It should be noted that there is an honest effort to maintain data integrity at the unit level before reporting. However, database management is tedious and can be subject to many shortfalls in data transference. Expanding the study to larger data sets would reduce the chance of incorrect reporting. Non-discretionary inputs are variables that we did not consider in this study, which vary without management direction (e.g., favorable weather). This study is limited to discretionary inputs to allow future discussion of managerial actions available to maintenance leaders.

This study was designed to highlight efficiency trends using USAF aircraft maintenance performance metrics for the F-22 community and continue discussing improved data analytics for maintainers to keep aging fleets flying. Providing data analysis tools at the tactical and strategic levels enables those closest to the aircraft to make data-based decisions to prolong the effective employment of our airframes.

Bibliography

ACC/A4M. (2018). *ACC Instruction 21-118, Logistics Maintenance Performance Indicator Reporting Procedures.* Air Combat Command.

Banker, R., Charnes, A., & Cooper, W. (1984). Some Models For Estimating Technical and Scale Inefficiencies in Data Envelopment Analysis. *Management Science*, 1078-1092.

Bowlin, W. (1984). A Data Envelopment Analysis Approach to Peformance Evaluation In Not-For-Profit Entities With An Illustrative Application to the US Air Force. *Ph.D. dissertation.* . University of Texas at Austin.

Brown, C. (2020, May 7). *Advance Policy Questions for General Charles Q. Brown, Jr., U.S. Air Force.* Retrieved from Senate Armed Service Committee: https://www.armed-services.senate.gov/imo/media/doc/Brown_APQs_05-07-20.pdf

Charnes, A., Clark, C., Cooper, W., & Golany, B. (1985). A Developmental Study of Data Envelopment Analysis in Measuring the Efficiency of Maintenance Metrics in the US Air Forces. *Annals of Operations Research*, 95-112.

Charnes, A., Cooper, W., & Rhodes, E. (1978). Measuring the Efficiency of Decision Making Units. *European Journal of Operation Research*, 429-444.

Clarke, R. L. (1992). Evaluating USAF Vehicle Maintenance Productivity Over Time: An Application of Data Envelopment Analysis. *Decision Sciences*, 376-384.

Cooper, W., Seiford, L., & Tone, K. (2007). *Data Envelopment Analysis: A Comprehensive Text with Models, Applications, References and DEA-Solver Software (Second Edition).* New York: Springer Science & Business Media.

Department of the Air Force. (2018). *Air Force Tactics, Techniques, and Procedures: Aircraft Maintenance.* AFI 3-4.21V1. Washington: HQ USAF.

Farrell, M. (1957). The Measurement of Productive Efficiency. *Journal of the Roal Statistical Society*, 253-90.

Han, H. K., & Sohn, S. Y. (2011). DEA Application to Grouping Military Bases. *Military Operations Research*, 31-40.

McAneny, P. J. (2009). *Red Is Good. Transformational Changes for US Air Force Aircraft Maintenance.* Maxwell AFB: Air University.

Moseley, T. M., & Wyne, M. W. (2005, November 7). *"Lean across the Air Force" -- Process Development and Improvement.*

Rainey, J. C., McGonagle, R., Scott, B. F., & Waller, G. (2001). *Metrics Handbook for Maintenance Leaders.* Maxwell: Air Force Logistics Management Agency.

Sherman, H., & Ladino, G. (1995). Managing Back Productivity Using Data Envelopment Analysis. *Interfaces*, 60-73.

Sutton, W., & Dimitrov, S. (2013). The US Navy Explores Detailing Cost Reduction via Data Envelopment Analysis. *European Journal of Operational Research*, 166-173.

Taaffe, K. M., Allen, R. W., & Grigg, L. (2014). Performance Metrics Analysis for Aircraft Maintenance Process Control. *Journal of Quality in Maintenance Engineering*, 122-134.

Todorov, K. (2006, November 7). AFSO 21: Everyone benefits from playing. Moody AFB, GA.

US General Accountability Office. (1993). *TQM Implementation in the Air Force.* Washington DC: US General Accountability Office.

Werner, A., Zimmerman, N., & Lentes, J. (2019). Approach for a Holistic Predictve Maintenance Strategy by Incorporating a Digital Twin. *25th International Conference on Producting Research Manufacturing Innovation: Cyber Physical Manufacturing* (pp. 1743-1751). Chicago: Procedia Manufacturing.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*
02/18/2022

2. REPORT TYPE
Graduate Research Proposal

3. DATES COVERED *(From - To)*
04 JUL 19 - 24 MAR 22

4. TITLE AND SUBTITLE
Benchmarking F-22 Mission Capability Rates and Sortie Output Through Data Envelopment Analysis

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Edwards, LeRoi G., Maj

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Air Force Institute of Technology
Graduate School of Engineering and Management (AFIT/EN)
2950 Hobson Way
Wright-Patterson AFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER
AFIT-ENS-MS-22-M-123

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

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Distribution Statement A. Approved for Public Release;Distribution Unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
The F-22 is combat-proven, operating in conflict areas for over a decade. Best practices to maximize the Mission Capability (MC) rate are not successfully codified and defended with data. This paper uses Data Envelopment Analysis (DEA) to identify benchmark environments where the MC rate is optimized and efficient. DEA successfully compared the relative efficiency of inputs and outputs across two units and determined the more efficient organization. Additionally, DEA provides current senior USAF leaders and tactical managers insight into performance environments where relative efficiency can be maximized to support the National Defense Strategy in a fiscally constrained environment.

15. SUBJECT TERMS
F-22, Data Envelopment Analysis, Aircraft Maintenance Metrics, Health of Fleet

16. SECURITY CLASSIFICATION OF: a. REPORT	b. ABSTRACT	c. THIS PAGE	17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
					Dr. Seong-Jong Joo, AFIT/ENS
					19b. TELEPHONE NUMBER *(Include area code)*
U	U	U	UU	37	937-255-3636 x4761 seong-jong.joo@us.af.mil