Received 21 August 2020
Revised 27 October 2020
Accepted 10 March 2021

RELIABILITY PAPER

Performance measure of maintenance practices for F-16 fighter jets by data envelopment analysis

Mansik Hur
Management and Marketing, California State University Bakersfield, Bakersfield, California, USA
Seong-Jong Joo
Operational Sciences, Graduate School of Engineering and Management, Air Force Institute of Technology, Wright-Patterson AFB, Ohio, USA, and
Jaeyoung Cho
Management and Marketing, Prairie View A&M University, Prairie View, Texas, USA

Abstract

Purpose – The purpose of this study is to provide the benchmarking guideline to improve managing maintenance man-hour in producing sorties of fighter jet aircraft in the Korean Air Force by measuring relative efficiency of aircraft in a fleet.

Design/methodology/approach – Using actual operational data from 2014 to 2018, this paper measures the relative efficiencies of aircraft by data envelopment analysis (DEA) model. To analyze the efficiencies across different operational units and among aircraft with different attributes, non-parametric statistical techniques and tobit regression are employed.

Findings – The overall management of maintenance man-hour turns out to be 63.2% efficient compared to the best practice. Operational unit possessing aircraft with different configurations shows the poorest performance implying efficiency depends on the complexity of job processes. Identification of aircraft on the best-practice frontier indicates the efficiency is closely related with the reliability of components and the age of aircraft.

Practical implications – Performance of aircraft maintenance has a tremendous impact on military readiness, utilization of work force, and the quality and safety of flight operations. This study confirmed there exists plenty of room for performance enhancement and presents a reference to find a balance in the performances of maintenance practices across different installations as well as among different types of aircraft.

Originality/value – Application of DEA to engineering systems has been limited in literature, especially in military operations. Thus, this study presents meaningful guidelines to which fleet manager can refer to when establishing strategic maintenance plans to balance the utilization of manpower.

Keywords Aircraft maintenance, Efficiency, Data envelopment analysis, Performance improvement, Benchmarking

Paper type Research paper

Introduction

In production economics, efficiency is typically measured by the ratio of the amount of output produced to the amount of input used, while, in engineering, it is customarily measured in terms of energy under the famous "law of conservation of energy" stating energy produced cannot exceed the energy consumed (Bowlin, 1987). In addressing the efficiency of production organization, it is difficult to determine whether the entity is absolutely efficient, because there hardly exists a known function between the amount of various inputs and outputs. As such, we typically adopt the concept of relative efficiency and determine one is more efficient

International Journal of Quality & Reliability Management
Vol. 39 No. 1, 2022
pp. 280-296

0265-671X
DOI 10.1108/IJQRM-08-2020-0272

than the other by comparing the utilization of resources across multiple entities. This comparative analysis serves as a basis of benchmarking study.

Within a benchmarking perspective of performance management, relatively little attention has been given to the efficiency evaluation of engineering systems with common configurations such as power plants, manufacturing systems or fleet of vehicles. In this study, efficiency of fighter jet aircraft in a fleet is evaluated in terms of maintenance man-hour, then, we take the efficiency scores to draw insights on the comparative performance of maintenance operations by post-hoc analyses. It has been a common practice to address the operational efficiency of aircraft in a fleet by some ratios such as dollar per flight hour, maintenance man-hour per sortie and the number of failures per sortie. In the presence of multiple inputs and outputs, weighting scheme on relevant factors is applied for the evaluation of relative efficiencies. This method, however, is inadequate in identifying the sources and the amount of inefficiencies in the inputs and the outputs. To fill the gap, we employ a data envelopment analysis (DEA) method for the evaluation of relative efficiency of managing maintenance man-hour to produce sorties of fighter jet aircraft in the Korean Air Force.

In a typical fighter wing, a squadron is composed of 20–24 homogeneous fighter jets and all of the same type of aircraft across multiple squadrons form a fleet. All aircraft in the fleet can be considered as the same product in terms of design characteristics such as form, fit and functionality defined by a manufacturer. However, they show somewhat different levels of operational performance due to random variations in reliability characteristics built into the aircraft and variabilities in operational environment. To elaborate, expertise of maintenance crews assigned to each aircraft may differ depending on their experience and certification levels. The difference in leadership style, degree of motivation and the level of employee empowerment can create difference in quality of maintenance practice and subsequent productivity. Climate factors like temperature and humidity are different among air bases at which aircraft is stationed. All these factors eventually contribute to generate random and assignable variations in operational performance of each aircraft in the fleet. In the air force operations, maintenance function integrates various types of resources to produce aircraft sorties. However, measuring its efficiency has been considered difficult because it involves so many different kinds of factors such as manpower, various facilities, equipment, technologies, supply of spare and repair parts. Furthermore, most maintenance tasks are performed in a job-shop environment, which adds complexity in assessing its productivity. As such, this study investigates the relative efficiencies of fighter jet aircraft in terms of maintenance man-hour used. Then, findings are analyzed to draw insight on the comparative performance of maintenance management with regard to attributes of aircraft as well as practices at different installations.

The structure of this paper is as follows: we start with a brief explanation on service operations at an air force base and types of maintenance. Then, provide the survey of literature on the evaluation of efficiencies upon military applications and some engineering systems. Methodology section describes the attributes of input and output variables followed by the slack-based measure DEA modeling approach with unfavorable outcomes. Next, we present analyses of efficiency distributions with an identification of sources of inefficiencies and determines the amount of potential savings with suggestion of alternative path for improved efficiencies. The result of post-hoc analyses is summarized, and efficiency findings are confirmed by a tobit regression. We present discussion on the interpretation of results and managerial implications for fleet management followed by conclusion of this study.

Backgrounds

Maintenance operations at installations

Air force base is a service organization that produces aircraft readiness by generating sorties. Sortie is a dispatching of an aircraft to meet designated missions from take-off to landing. It is produced by means of a wide range of resources including supply of spare parts, support equipment, facilities, technical information, jet fuel and maintenance activities. Among all these support elements, maintenance occupies a significant amount of resources in terms of labor hours, e.g. approximately 50% of work force in a typical air force installation is assigned to the maintenance group. Furthermore, maintenance operations serve as a key process that integrates all the support elements and renders an aircraft availability to warfighters. Maintenance function ensures that the assigned aircraft is safe, serviceable and properly configured to meet mission needs. Maintenance tasks include, but are not limited to, inspection, repair, overhaul, modification, preservation, refurbishment, trouble-shooting, testing, analyzing condition and maintenance documentation. All levels of supervision are to place emphasis on safety, quality and timeliness in the performance of maintenance. According to internal study at Korean Air Force, the aircraft availability is significantly driven by the quality of maintenance performed.

Two types of maintenance

In military, the concept and conduct of aircraft maintenance is at three-fold level: organizational, intermediate and depot level. At each level, maintenance tasks are classified into two types: preventive and corrective. Preventive maintenance (PM) is a routine for periodically inspecting with the goal of noticing small problems and fixing them before major ones develop. The main goal behind PM is for the equipment to make it from one planned service to the next planned service without any failures caused by fatigue, neglect or normal wear (preventable items). On the other hand, corrective maintenance (CM) is carried out after failure is detected and is aimed at restoring an aircraft to a condition in which it can perform its intended function. During the study period of 2014–2018, PM (39,964,000 man-hours) were 2.8 times more than CM (14,253,000 man-hours) to the support of F-16 fleet in Korean Air Force.

Literature review

DEA is a mathematical programming technique based on the theory of linear programming. It was originally developed by Charnes *et al.* (1978) to assess the relative efficiencies of homogeneous organizational units that is called as a decision-making unit. Since its introduction as a benchmarking tool, numerous modifications followed to accommodate various real-world business cases. Application of DEA in benchmarking study includes education, health care industry, financial institutions, environmental protection and government agencies. First, we present DEA studies regarding efficiency analysis in military applications.

Charnes *et al.* (1984) paved a way for application of DEA to examine the performance of military operations, specifically, the aircraft maintenance squadrons in the US Air Force using hours of mission-capable, number of sorties flown and various maintenance tasks completed. Window analysis technique was used to overcome the shortfall in limited amount of performance data. Bowlin (1987) evaluated the operational efficiency of facility maintenance activities in the US Air Force to relate the value of efficiency analysis to budgeting and auditing agencies. To overcome the shortfalls from inadequate number of service organizations, window analysis was applied with improvements in discriminatory power of DEA method. Subsequent study by Bowlin (1989) investigated the performances of accounting and finance offices of the US Air Force to improve the operational efficiency at unit-level. Roll *et al.* (1989) measured the efficiencies of maintenance units in the Israeli Air Force with an emphasis on the choice of input factors to construct a hierarchical efficiency monitoring system. The model starts with three inputs and six outputs but reduces the number of outputs by feedback from field experts. Clark (1992) applied DEA to measure the performance of vehicle maintenance of the US Air Force during four-year period and observed the overall improvement in the relative efficiency of the participating units. The study facilitates the adoption of DEA within the organization as a benchmarking tool by drawing feedback from field-level managers through survey. Brockett *et al.* (2004) incorporated DEA method in regression model to distinguish between efficient and inefficient performance of advertising on military recruitment. Sun (2004) applied DEA to assess the performance of joint maintenance shops in the Taiwanese Army. The study examined the possibility of using DEA over time as a tool for continuous improvement. Han and Sohn (2011) proposed the application of DEA to select award-winning organization for best practice of spare parts inventory management at the Korean Air Force bases. Lu and Chen (2011) combined ratio analysis, the balance scorecard with slack-based measure of DEA model to assess the performance of financial management in Taiwanese military organization. Sutton and Dimitrov (2012) adopted a variation of DEA, the generalized symmetric weight assignment technique to solve assignment problem of sailors to jobs for the US Navy. The study presented one way for better matching soldiers and jobs to optimize the cost incurred when soldiers change their permanent units according to their career paths. Hanson (2016) adopted wages and equipment operating costs as inputs and the size and quality of several types of troops as outputs to present the analysis and productivity enhancement of operational units of Norwegian armed forces using DEA and Malmquist Index. Boehmke *et al.* (2017) using DEA assessed the performance of installation support services at major US Air Force bases to provide insights to Air Force leaders for more efficient decisions on resource allocation. Moradi and Amiri (2019) applied Malmquist Productivity Index (MPI) within the framework of DEA to assess the efficiency of health care services at military hospitals in Iranian Armed Force. The study determined the changes in average productivity of the hospitals over time and identified technology as a key factor for changes in efficiencies. Golany and Kress (2020) proposed DEA as an analytic framework for evaluating alternatives in the acquisition of defense systems in the US Armed Forces. Among major criteria used in the selection of weapon systems such as operational effectiveness, cost and risks of proposed materiel solutions, the study focuses on operational risks described as long-term readiness and sustainment of defense systems.

The use of DEA in research pertinent to engineering systems is quite limited, while very few scholars and especially engineers have implemented efficiency measurement concepts for the purpose of assessing and improving performance of engineering products. Concise summary of the applications of DEA on engineering systems can be found in Kottas *et al.* (2020). Several noticeable results are suggested. One of the earliest attempts to apply DEA on engineering systems is Bulla *et al.* (2000) who evaluated efficiencies of turbofan jet engines and compared the results with a standard engineering method for measuring efficiency. Analyses of efficiency data for 29 engines used for commercial airplane suggested how engineering parameters such as weight of engine, drag and fuel consumption can be accounted for in producing an alternative measure of performance of engineering systems. Extending Bulla *et al.* (2000), Kottas *et al.* (2020) applied two-stage network DEA to evaluate the efficiency of turbofan engines currently utilized by commercial and military aircraft. Taking technological features and progress over time, the study suggested a tool for assessing efficiency of aircraft engine in concept exploration or preliminary design stage.

Starčević *et al.* (2019) applied the analytical hierarchy process (AHP) method and DEA sequentially in equipment selection problem. The relative importance of the criteria used in evaluating potential alternatives is determined through AHP, then, the results are used as multiple outputs of DEA model for the selection of terrain vehicles needed by military units engaged in the multinational operations. Lu *et al.* (2020) applied the epsilon-based measure method that incorporates both radial and non-radial features of DEA to provide procurement decision makers with reference information in selecting fighter jet aircraft. While the selection of major weapon system requires consideration of a variety of factors to evaluate products from different manufacturers, the study takes price and take-off weight as inputs and thrust, speed and combat radius as outputs to suggest relative efficiencies of 26 fighter jet models.

While the DEA studies on military application have been constrained due to limited availability of data for the confidentiality nature of military organizations, the focus was on the performance of specific functional areas such as equipment maintenance, facility management, criteria for equipment selection and recruiting of human resource. This study, in line with Charnes *et al.* (1984), investigates the relative efficiencies of maintenance operations with regard to the most expensive type of resource–maintenance labor hour–to produce aircraft readiness.

Methodology

Multivariate regression models, analysis of variance (ANOVA) and ratio analyses do not lead to adequate explanatory models when the research objective is to identify and analyze best practice (BP) in a benchmarking study. As such, DEA model is employed to identify the BP or relatively efficient decision-making units (DMUs). DEA can also be used to classify engineering systems such as a fleet of aircraft according to the efficiency in utilization of input resources. More specifically, DEA measures the magnitude of departure from the BP frontier for each aircraft in the fleet based on the amount of labor hours used. DEA separates all aircraft that define the BP frontier from inefficient aircraft underperforming than the frontier by analyzing maintenance labor hours used to produce aircraft sorties.

In this study, we define the DMU that will be evaluated as the individual aircraft denoted by its unique tail number. As a benchmarking tool, DEA model performs a series of tasks. First, it classifies aircraft into peer groups of aircraft that are similar in their use of inputs to produce outputs. Second, it defines a BP frontier consisting of all aircraft providing their mix and volume of sorties with the least amount of resources. Third, it computes the efficiency ratings that measures how far each inefficient aircraft is from the BP frontier. Lastly, the estimate of excess resources used by inefficient aircraft is determined, which can be interpreted as potential cost savings. To form the BP production frontier, we use following input/output criteria:

(1) An aircraft is inefficient if it could produce the same sorties with fewer maintenance hours.

(2) An aircraft is inefficient if it is possible to increase the sorties without increasing the amount of maintenance hours.

This definition of best practice does not necessarily coincide with the lowest cost producers because we focus on technical efficiency rather than price efficiency. We identify pathways to reduce the amount of resources that will eventually lower cost, although the cost of maintenance hours is not available from the current accounting system.

DEA model with undesirable outputs

Failures of components, modules, sub-systems and aircraft itself are inevitable in the course of performing flight operations. In production economics, a production function is defined as an equation that shows the maximum amount of outputs that can be produced from any set of inputs given the existing practice and technology. In the presence of undesirable outputs, however, practices with more desirable outputs and less undesirable outputs relative to less input resources are recognized as efficient. To account for undesirable outputs, Tone (2001) proposed methods by modifying a slack-based measure of efficiency (SBM), which is non-radial and non-oriented and utilizes input and output slacks directly in producing an efficiency measure. Because one of two outputs, failures, is undesirable, we employ the undesirable output model by Tone, and conduct the analyses of technical efficiencies using variable return-to-scale model that filters out the scale effects.

Description of DEA model

Let us decompose the output matrix Y into (Y^g, Y^b), where Y^g and Y^b denote good (desirable) and bad (undesirable) output matrices, respectively. For a DMU (x_o, y_o), the decomposition is denoted as (x_o, y_o^g, y_o^b). We consider the production possibility set defined by:

$$P = \left\{ (x, y^g, y^b) \middle| x \geq X\lambda, y^g \leq Y^g\lambda, y^b \geq Y^b\lambda, \lambda \geq 0 \right\},$$

where λ is the intensity vector.

A DMU (x_o, y_o^g, y_o^b) is efficient in the presence of bad outputs, if there is no vector $(x, y^g, y^b) \in P$ such that $(x_o \geq x, y_o^g \leq y^g, y_o^b \geq y^b)$ with at least one strict inequality.

In accordance with this definition, slack-based measure model in Tone (2001) is modified as follows:

$$\rho^* = \min \frac{1 - \frac{1}{m}\sum_{i=1}^{m} s_{io}^- / x_{io}}{1 + \frac{1}{s}\left(\sum_{r=1}^{s_1} s_r^g / y_{ro}^g + \sum_{r=1}^{s_2} s_r^b / y_{ro}^b\right)}$$

subject to

$$\begin{aligned} x_o &= X\lambda + s^- \\ y_o^g &= Y\lambda - s^g \\ y_o^b &= Y\lambda + s^b \\ \lambda, s^-&, s^g, s^b \geq 0 \end{aligned}$$

where, X = vector for input variables

Y = vector for output variables

s^- = slack variable for input variables

s^g = slack variable for good output variable

s^b = slack variable for bad output variable.

The vector s^- and s^b correspond to excesses in inputs and bad outputs, respectively, while s^g expresses shortages in good outputs. s_1 and s_2 denote the number of elements in s^b and s^g, and $s = s_1 + s_2$. Let an optimal solution of the above program be $(\rho^*, s^{-*}, s^{g*}, s^{b*})$. Then, we can demonstrate that the DMU (x_o, y_o^g, y_o^b) is efficient in the presence of undesirable outputs if and only if $\rho^* = 1$, i.e. $s^{-*} = 0, s^{g*} = 0, s^{b*} = 0$. If the DMU is inefficient, i.e. $\rho^* < 1$, it can be improve and become efficient by deleting the excesses in inputs and bad outputs and augmenting the shortfalls in good outputs by the following projection.

$$\begin{aligned} x_o &\Leftarrow x_o - s^{-*} \\ y_o^g &\Leftarrow y_o^g + s^{g*} \\ y_o^b &\Leftarrow y_o^b - s^{b*} \end{aligned}$$

Analysis

Considering the efficiency of maintenance man-hour in producing aircraft sorties, we develop a two-inputs and two-outputs DEA model to classify 138 fighter jets in the fleet into relatively efficient ones and inefficient ones which are off the efficient frontier. Two inputs composed of preventive and corrective maintenance hours were suggested by the fleet managers at the Logistics Command in the Korean Air Force and deemed discretionary by considering its significant impact on the utilization of human resource at maintenance groups in three installations. In the output side, the number of flight sorties produced by maintenance operations serves as a direct measure of aircraft readiness; so, it is reasonably selected as an output variable. No matter how scrupulously maintenance work is conducted, failures happen on component, assembly, subsystem or system level due to the combination of the nature of inherent reliability and imperfect practice of maintenance. The frequency of failures differs with the variabilities in workmanship of maintainers and operational environment. Thus, the number of failures is included in this study as an undesirable output. In the presence of undesirable output, practice with more good output (aircraft sorties) and less bad output (failures) relative to less input resources should be recognized as efficient.

The data on maintenance man-hour and operational outcomes were obtained from the enterprise resource planning (ERP) system exclusively employed in the Korean Air Force. The data span five years from 2014 to 2018 and reflect the operational performances of 138 aircraft deployed at three geographically separated air bases. The exact number of aircraft (fleet size) which is a bit greater than 138 is not announced in this study for the confidentiality of sensitive information. Each of three units (air bases) possesses different number of aircraft and conducts both tactical and training missions in the course of air force operations. While the portfolio of mission types (e.g. air-to-air, air-to-ground or reconnaissance) and flight loads are similar among air bases, we expect differences in the frequency of component failures and the amount of maintenance man-hour needed to fix the failures. The age distribution of the aircraft in the fleet ranges from 17 years old to 34 years old with a mean of 25.6 (20 aircraft < 20 year-old; 90 aircraft < 30 year-old; 28 aircraft < 40 year-old) as a result of sequential deployment from 1980s. Whereas the initial batch (1/4 of the fleet) of aircraft was completely manufactured in the General Dynamics (now merged into Lockheed Martin Co.) facility in the US and ferried to South Korea, the remainders were assembled in South Korea under a license production agreement. Descriptive statistics of input/output variables for all DMU, the tail numbered 138 aircraft, is summarized in Table 1.

The coefficient of variation defined as the ratio of the standard deviation to the mean is less than 1 for all variables, which suggests both input utilizations and process outcomes of aircraft under study are distributed with low degree of variability.

The correlation between input and output variables is natural, but correlation between output variables is not desirable. The correlation analysis by Pearson product moment correlation coefficient in Table 2 suggests that no strong correlation exists between variables in overall.

Table 1.
Descriptive statistics of input and output variables

	PMH (preventive maintenance man-hour)	CMH (corrective maintenance man-hour)	Number of failures	Sorties flown
Max	2,614,860	288,244	206	1,418
Min	98,256	34,713	34	529
Average	289,591	103,280	93	966
St Dev	229,687	49,539	33	202

Efficiency distribution

The application of SBM model calculates the relative efficiency of each of 138 aircraft in the data set. Its distribution ranges from 0.18 to 1 and is depicted in the histogram in Figure 1. The attained efficient DMU portion is significantly low, that is, 21 out of 138 aircraft are turned out to be efficient, and consequently the achieved discriminative power deems acceptable. The arithmetic mean efficiency of all aircraft is 0.632, that is, the overall management of maintenance man-hour was 63.2% efficient compared to the best practice. More specifically, 63% (87 out of 138) of aircraft operations was less than 70% efficient, which indicates a significant potential to reduce inputs and increase outputs by more than 30%.

Ratio analysis summarized in Table 3 confirms that 21 efficient aircraft on the best practice frontier show better performances than inefficient aircraft in the fleet. Mean preventive maintenance hour (PMH) per sortie of efficient aircraft is 60% of PMH of inefficient aircraft. Mean corrective maintenance hour (CMH) per sortie of efficient aircraft is 69% of CMH of inefficient aircraft. Mean number of failures per sortie for efficient aircraft is 26% less than that number of inefficient aircraft. Among aircraft in efficiency reference set (ERS) shown in Table 4, 71% (15 out of 21 efficient) is operated at unit B, and all 21 efficient aircraft were assembled in South Korea.

	PMH	CMH	Failures	Sorties
PMH	1			
CMH	0.30	1		
Failures	0.29	0.62	1	
Sorties	0.18	0.38	0.43	1

Table 2. Correlation analysis of variables

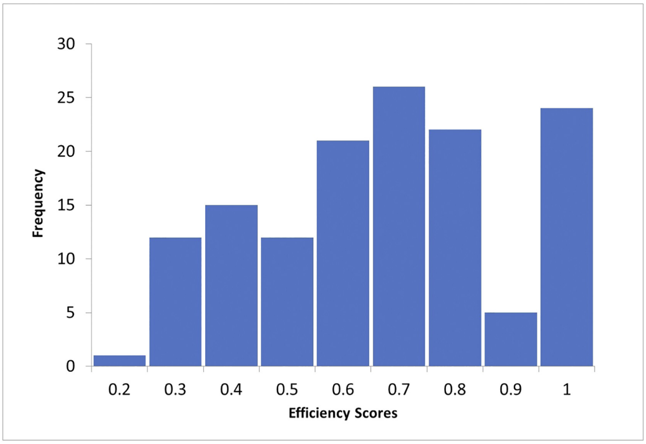

Figure 1. Histogram of efficiency distribution

	Mean PMH/sortie	Mean CMH/sortie	# Failures/1,000sortie
Efficient aircraft (21)	195	78	75
Inefficient aircraft (117)	324	113	101

Table 3. Ratio analysis: comparison of performances between efficient vs inefficient aircraft

Table 4.
Aircraft tail numbers in the efficiency reference set (ERS)

Tail No	Frequency	Unit	Assembly	Tail No	Frequency	Unit	Assembly
005	71	B	Korea	114	27	B	Korea
008	0	B	Korea	115	1	B	Korea
014	0	B	Korea	117	1	B	Korea
018	1	B	Korea	118	51	B	Korea
056	2	A	Korea	122	0	B	Korea
066	1	B	Korea	124	0	A	Korea
095	4	B	Korea	125	0	C	Korea
103	1	C	Korea	127	42	B	Korea
107	1	B	Korea	130	0	C	Korea
112	0	B	Korea	137	0	A	Korea
113	2	B	Korea				

Identification of benchmarking targets

In total, 21 aircraft are classified as efficient and they serve as benchmarking targets for each of 117 inefficient aircraft. In Table 4, 8 aircraft (Tail Number 008, 014, 112, 122, 124, 125, 130, 137) are efficient, but not referenced as benchmarking target for other inefficient aircraft with its frequency columns equal to zero. Out of 205 times that each of 13 aircraft appears in the ERS of all inefficient aircraft, 202 times are referenced by aircraft stationed at unit B. Another noticeable observation is that all 13 efficient aircraft serving as benchmarking targets were assembled at Korean manufacturer, while the data set includes 30 aircraft assembled in the US.

Potential savings and path to improvements

One of the reasons that makes DEA a popular benchmarking tool is its ability to determine the amount of potential savings in inputs or the amount of potential expansions in outputs achievable by combination of existing DMUs. When using radial DEA models such as CCR and BCC, input savings or output expansions are calculated in a proportional way, e.g. 20% of inputs should be reduced across all input factors. Furthermore, the reported efficiency scores of radial models do not consider the remaining slacks which can play an important role in evaluating managerial efficiency (Tone, 2017). On the other hand, non-radial model such as SBM directly takes the slack to the calculation of efficiency scores not relying on a proportional method. As such, this model suggests potential input savings and output expansions simultaneously in disproportionate way, which are summarized for three air bases in Tables 5 and 6.

Table 5.
Potential input savings and output expansion across air bases

	PMH	CMH	Failures	Sorties
Unit A	−57.1%	−52.3%	−43.4%	+3.5%
Unit B	−24.2%	−34.7%	−16.7%	+0.8%
Unit C	−18.7%	−31.9%	−36.7%	+0.2%
Grand total	−39.3%	−41.5%	−31.4%	+1.6%

Table 6.
Potential input savings and output expansion across aircraft by assembly location

	PMH	CMH	Failures	Sorties
Assembled at US	−67.9%	−62.5%	−50.4%	+4.7%
Assembled at Korea	−24.5%	−33.5%	−24.3%	+0.9%
Grand total	−39.3%	−41.5%	−31.4%	+1.6%

In overall, 39.3% of preventive maintenance hour (PMH), 41.5% of corrective maintenance hour (CMH) can be reduced for all 117 inefficient aircraft in comparison to 21 efficient aircraft. At the same time, 31.4% of failures can be reduced and 1.6% of sorties can be more produced to match the performances of 21 efficient aircraft on a best practice frontier. This means that man-hour can be reduced annually by more than 3 m hours in preventive maintenance and more than 1 m hours in corrective maintenance over the five-year period under study. During the same period, about 800 failures should have been reduced and about 400 sorties should have been produced more in all aircraft in the fleet if they had been managed in the same manner as the aircraft in the best practice frontier. Breakdown of projected savings and expansion reveals that unit A has the greatest potential for improvement in terms of PMH and CMH indicating 57.1 and 52.3% of respective reduction should be possible. The data also suggest the greatest amount of failures should have been reduced by 43.4% and of sortie expansion by 3.5% should be possible at unit A.

The same comparison between aircraft by different assembly locations indicates significantly greater amount of potential improvements should be possible for aircraft assembled and delivered from the US manufacturer, which represents about 22% of the fleet size.

Post-hoc analysis

Now, we turn our attention to explore the potential differences in efficiencies between different groups of aircraft, i.e. subgroups of aircraft with different attributes such as locations of assembly, configurations of aircraft, operating units and the age of aircraft. The findings of difference, if present, can serve as a new direction of fleet management to save maintenance hours, reduce failures and enhance aircraft readiness. As such, we conducted post-hoc statistical analyses. Because the efficiency scores are measured comparatively, we chose non-parametric rank-sum tests.

Efficiencies between assembly location

Our first conjecture is the efficiency difference between aircraft with different locations of assembly. Of the 138 aircraft under study, 30 aircraft were assembled at the Lockheed Martin plant in the US, while the remaining 108 aircraft were assembled at Korean indigenous aerospace company by a license production agreement. A simple ratio analysis shows the efficiency of aircraft assembled at Korea is better than the aircraft assembled in the US in terms of maintenance man-hour per sortie as well as the number of failures per sortie.

With the implications from the above ratio analysis in Table 7, hypothesis test for differences between groups is established as follows:

$$H_o : \theta_{\mathrm{US}} = \theta_{\mathrm{Korea}} \text{ (Efficiencies are equal)}$$
$$H_a : \theta_{\mathrm{US}} < \theta_{\mathrm{Korea}}$$

Test result:

	N	Median	*W*-statistic	*p*-value
Assembled at US	30	0.3295	536.0	<0.001
Assembled at Korea	108	0.6935		

The Mann–Whitney test is employed to test the difference between two groups. The reported p-value < 0.001 with a rejection of the null hypothesis at a significance level of 0.05.

We confirm that a sufficient evidence exists to conclude that the efficiency of aircraft assembled at Korea is higher than the efficiency of aircraft assembled in the US. During the interview, the fleet manager took this finding for granted in that aircraft assembled in Korea are configured with newer and more advanced mechanisms in flight control, avionics, navigation and propulsion systems leading to improved reliability and maintainability of aircraft systems.

Efficiencies between different configurations

Next, we compare the efficiencies of aircraft with different configurations. Of the 138 aircraft under study, 94 aircraft have a single seat, while the remaining 44 aircraft have two seats. Aircraft with two seats is primarily used for training mission, whereas single-seat aircraft is mainly assigned for tactical mission. We suspect that the difference in their usage may bring about the differences in operational characteristics. A simple ratio analysis shows the single-seat aircraft with caused less amount of maintenance man-hour (395 h per sortie) than aircraft with two seats (449 h per sortie). However, two-seat aircraft is slightly better in terms of the number of failures per sortie than single-seat aircraft.

Basing the implications from the above ratio analysis in Table 8, hypothesis test for differences between groups is established as follows:

$$H_o : \theta_{\text{single-seat}} = \theta_{\text{two-seat}} \text{ (Efficiencies are equal)}$$
$$H_a : \text{ Efficiencies are not equal.}$$

Test result:

	N	Median	W-statistic	p-value
Single-seat aircraft	94	0.6400	6,661	= 0.560
Two-seat aircraft	44	0.6215		

The Mann–Whitney test is employed to test the difference between two groups. The reported p-value = 0.560 with a non-rejection of the null hypothesis at a significance level of 0.05. We confirm that a sufficient evidence does not exist to conclude that the efficiency of single-seat aircraft and the efficiency of two-seat aircraft are significantly different.

Efficiencies among three operating units

Now, we look to the potential differences in efficiencies among aircraft possessed by different units. Of the 138 aircraft under study, 49 aircraft are stationed at operating unit A, 71 aircraft

Table 7.
Ratio analysis against assembly location

	Average man-hour/sortie			
Aircraft	PMH	CMH	Subtotal	Average number of failures/sortie
Assembled at US	537	152	689	0.136
Assembled at Korea	240	95	335	0.086

Table 8.
Ratio analysis against aircraft configurations

	Average man-hour/sortie			
Aircraft	PMH	CMH	Subtotal	Average number of failures/sortie
Single-seat aircraft	294	101	395	0.0989
Two-seat aircraft	328	121	449	0.0935

at unit B and 18 aircraft at unit C. A simple ratio analysis shows the average maintenance man-hour is the largest for aircraft in unit A, while no salient difference is observed between aircraft in unit B and unit C. In terms of average number of failures per sortie, the aircraft in unit B produced the fewest number of failures, while aircraft in unit A and unit C has the similar range of values.

Basing the implications from the above ratio analysis in Table 9, hypothesis test for differences between groups is established as follows:

$$H_o : \theta_{\text{unit A}} = \theta_{\text{unit B}} = \theta_{\text{unit C}} \text{ (Efficiencies are equal)}$$
$$H_a : \text{ Not all efficiencies are equal.}$$

The Kruskal–Wallis test is employed to test the difference among three groups of aircraft.

Test result:

	N	Median	Mean-rank	H-statistic	p-value
Aircraft at unit A	49	0.4020	100.1	45.7	<0.001
Aircraft at unit B	71	0.7270	50.3		
Aircraft at unit C	18	0.6115	61.8		
	138				

As in the above test result, the value of H-statistic, 45.7, is greater than critical value of 5.991 at the significance level of 0.05 with 2 degrees of freedom, so we reject the null hypothesis (p-value < 0.001) and conclude that a sufficient evidence exists to conclude that the efficiencies of aircraft are not equal among three operating units.

With the rejection of null hypothesis, we investigate the differences among units by employing Dunn's test that performs multiple comparisons of rank sums based on the z-statistics of the standard normal distribution (Dunn, 1964). Due to Dunn, we declare mean rank for group i and j to be significantly different if

$$|\overline{R_i} - \overline{R_j}| \geq z^* \sqrt{\frac{N(N+1)}{12}\left(\frac{1}{n_i} + \frac{1}{n_j}\right)} \quad (1)$$

where $\overline{R_i}$ is mean rank for group i

N is the total number of aircraft

n_i is the number of aircraft in group i.

We carry out a multiple comparison procedure at the overall significance level of 0.2 according to typical practice. We have $z^* = 1.834$ since we have total 3 comparisons. Evaluating the inequality (1) for all 3 comparison pairs indicated a significant difference exists for unit A vs unit B and unit A vs unit C. So, we argue that operating units have effect on the comparative efficiency in this study.

	Average man-hour/sortie			
	PMH	CMH	Subtotal	Average number of failures/sortie
Aircraft at unit A	436	129	565	0.1209
Aircraft at unit B	237	95	332	0.0744
Aircraft at unit C	216	96	312	0.1225

Table 9. Ratio analysis against operating units

Subsequently, to figure out the source of inefficiencies within each unit, we compare pure technical efficiency (PTE) versus scale efficiency (SE) of aircraft. According to Banker *et al.* (1984), SE is calculated by dividing aggregate technical efficiency (TE) which is obtained from constant return-to-scale (CRS) model by PTE obtained from variable return-to-scale (VRS) model. To be more specific, from the formula, TE = (PTE)(SE), it is reasoned that the same value of TE can be achieved by relatively different combination of PTE and SE. For example, TE of DMU_1 is 0.63 and it is the result of PTE of 0.9 and SE of 0.7; TE of DMU_2 is also 0.63 but it is the result of PTE of 0.7 and SE of 0.9. Though DMU_1 and DMU_2 have the same overall efficiency, TE, the sources of inefficiencies are different. Larger value of PTE implies that the technical and operational performance of DMU_1 is not far from the BP frontier, while smaller value of SE means that its overall efficiency is penalized by non-technical factors or diseconomy of scale that cannot be controlled by DMU_1. Because the sources of inefficiencies are different between DMU_1 and DMU_2, two DMUs should take different approaches to improve their performance.

As Table 10 shows SE values are larger than PTE values across all units, which implies lower PTE is driving the overall efficiencies away from BP frontier. Fleet manager has enough reasons to investigate the overall process of conducting maintenance tasks in search for the room for efficiency improvement by taking technical and operational measures at each unit. We notice the largest gap between PTE and SE is observed for unit A with almost doubled value of SE (0.905) to PTE (0.459), which is also evidenced by the simple ratio analysis suggesting the average maintenance man-hour is the largest for aircraft in unit A.

Regression analysis

Regression is a quite popular method for assessing the effect of variables on DEA efficiency. For an additional post-hoc analysis, we try tobit regression or censored regression for finding a general relationship between the variables in the DEA model and the efficiency scores (PTE). Because the efficiency scores are bounded by zero and one inclusively, it is necessary to transform the scores using natural logarithm, which allows the scores vary between negative infinity to zero and makes the transformed scores right censored only.

We choose a hierarchical regression approach that is good for evaluating variables in models. The result is summarized in Table 11. PMH in models 1 and 5 suffers the Hauck–Donner effect (Hauck and Donner, 1977), which indicates the violation of the assumption

Table 10. Average values of TE, PTE and SE against operating units

	TE	PTE	SE
Unit A	0.404	0.459	0.905
Unit B	0.640	0.738	0.869
Unit C	0.578	0.679	0.867
All units	0.548	0.631	0.882

Table 11. Tobit regression results

Variable	Model 1	Model 2	Model 3	Model 4	Model 5	Model 6
PMH	−9.29e-07***				−6.19e-07***	
CMH		−5.13e-06***			−3.41e-06***	−4.97e-06***
Failures			−0.0065***		−2.00e-03	−5.17e-03***
Sorties				0.0006**		1.38e-03***
R^2	0.2405	0.3312	0.2267	0.0628	0.4515	0.6813
LL	−83.8105	−75.6028	−83.9319	−96.2832	−63.5416	−31.9395

Note(s): Significance: "***" at 0.001; "**" at 0.01; "*" at 0.05

on parameter estimates that located in the interior of the parameter space. If an estimate is located on or near the boundary of the space, a disturbing and undesirable underappreciated effect presents. We dropped PMH from the final model or model 6 and interpreted the result of the model. Model 6 without PMH shows good values for R^2 and log-likelihood (LL). The leading signs of the coefficient estimates are as we expect. The input and undesirable output variables show negative leading signs, and the regular output or sorties display a positive coefficient. The result of model 6 can be used for predicting the effect of changes on input and output variables without running DEA models, which will be convenient for practitioners who lack the knowledge of DEA. For example, an additional sortie will increase the natural logarithm of the efficiency score by 0.00138 or an estimated efficiency score multiplied by the exponentiated value of 0.00138 (or 1.00138). That is, if the estimated efficiency score is 0.7000, the result of one sortie increase will move up the efficiency score to 0.70097. Finally, the relationship between PMH and the efficiency score is inconclusive due to the Hauck–Donner effect, which warrants further studies.

Results and discussion

Taking a benchmarking approach, this study identified efficient aircraft in the use of maintenance man-hour to produce flight sorties by measuring how far each aircraft is from the BP frontier. The achieved discriminative power of DEA model suggests good fit between the type of employed model and the applied data set. The overall efficiency of 63% indicates a plenty of room for potential reduction of inputs and increase in outputs by about 37% compared to the aircraft in best practice. Breakdown of potential improvement in Table 12 shows about 40% of maintenance man-hour can be saved in addition to 31% reduction of failures and slight increase of sorties in overall. Although all identified inefficiency cannot be eliminated due to limitations in technical, operational and managerial environment, huge amount of benefit can still be reaped only with partial improvement considering the magnitude of input resources, especially, the workforce in maintenance groups in the air force installations. That is, just 50% realization of potential savings can annually cut man-hour requirement by about 1.6 m hours in preventive maintenance and 0.6 m hours in corrective maintenance over the five-year period under study. This finding means a lot in establishing the direction for restructuring the logistics organizations in the Korean Air Force to cope with the difficulty in recruitment of soldiers arising from the rapid decline in population.

Post-hoc analysis revealed that the unit A has the greatest potential for improvement with regard to all input and output factors. Findings are consistent with the result of simple ratio analysis on the average maintenance man-hour. The interview with the chief of maintenance squadron of unit A revealed that its poor performance might be related with the age of aircraft. Note that the expected lifecycle of F-16 fighter jet is assessed as approximately 40 years by industry experts, and the mean ages of aircraft in unit A, B and C are 29.5-years, 23.0-years, and 25.1-year old, respectively. That is, the aircraft in unit A are oldest. Interestingly, the distributions of efficiencies and age show strong negative correlation with the Pearson product-moment correlation coefficient of −0.61, i.e. efficiencies decrease as aircraft becomes old. Fleet manager claims that old aircraft approaching an end-of-life phase experiences various kinds of issues ranging from parts obsolescence, old version of embedded software,

Improvement	PMH	CMH	Failures	Sorties
Overall	39.3% down	41.5% down	31.4% down	1.6% up
Unit A	57.1% down	52.3% down	43.4% down	3.5% up

Table 12. Projected potential improvements

limitation of compatibility with new support equipment and prolonged repair turn-around-time and procurement lead-time, which ultimately lead to more time and efforts in maintenance and sustainment activities. Furthermore, it turned out that the fleet of unit A were composed of two types of aircraft assembled in the US manufacturer and in Korean industry whereas unit B and C possess only one type assembled in Korea. Different types of aircraft in a fleet require a diversified support structure in support equipment, technical directives, employee expertise and layout of facilities. Integrating all these evidences, more complicated maintenance process arising from different types of aircraft and obstacles from managing aging aircraft seem to have contributed to relative inferior performance outcomes at unit A. Readers can recall the low-cost strategy of the Southwest Airline that adheres to operate just one type of aircraft, B-737 in its fleet to streamline the support structure.

The efficiency comparison of aircraft with different assembly locations indicates that significantly greater amount of potential improvements should be possible for aircraft assembled and delivered from the US manufacturer. The interview with engineers for each subsystem, e.g. fire control, navigation, avionics, hydraulics and propulsion, indicated that (1) aircraft assembled at Korea adopted relatively more recent version of components and parts in its assembly, (2) enhanced reliability contributed less frequent failures and (3) improved maintainability made easier and quicker maintenance tasks when parts fail. This finding confirms that reliability and maintenance characteristics embedded into system drive the total lifecycle cost of weapon systems.

Conclusion

Aircraft maintenance is a series of complex tasks requiring the integration of various types of resources such as materials, facilities, technology and manpower. Performance of maintenance jobs has a tremendous impact on military readiness, utilization of human resource and the quality and safety of flight operations. One of fleet manager's concerns is finding a way to strike a balance in the performances of maintenance practices across different installations as well as among different types of aircraft. As such, in this study, we analyzed the relative efficiency of fighter jet aircraft in the Korean Air Force from 2014 to 2018 with regard to expended maintenance man-hours–the most valuable and expensive input resource.

Application of slack-based measure of DEA model presented the overall efficiency of 63% across all aircraft in the fleet against best-practice frontiers, which reveals plenty of room for performance improvement. Identification of benchmarking targets for inefficient aircraft indicates the direction to which fleet manager should refer in establishing remedial action plans to enhance the overall efficiency. The plans could be practical because reduction of maintenance man-hour and increase of sorties can be sought simultaneously as opposed to the outcomes of radial models such as CCR or BCC.

The comparisons of efficiencies among aircraft with different versions suggested efficiency is highly influenced by the reliability and maintainability of components in the aircraft. While the poorest performance was shown at unit A, the survey interview with field managers discovered that complicated maintenance procedures to support two different types of aircraft based on assembly location and obstacles from aging old aircraft have causes relatively inferior efficiency in maintenance tasks. The overall findings by post-hoc analyses turned out to be consistent with the ratio analyses of utilized input and output factors and confirmed the validity of the study. Regression analysis provided additional means for estimating efficiency gains for reducing excess maintenance man-hours and by increasing flight sorties to meet the improvement targets. While the suggested savings in man-hour and potential increase in sorties may not be fully realized by remedial actions, even partial enhancement, say, 10% of reduction of input and/or expansion of output can bring about huge amount of efficiency gains considering the magnitude of labor hours at the maintenance groups in the air force installations. In the same vein, next line of study will incorporate other fleets of aircraft like F-15 and T-50 in the Korean Air Force with more input and output factors.

References

Banker, R.D. (1984), "Estimating most productive scale size using data envelopment analysis", *European Journal of Operational Research*, Vol. 17 No. 1, pp. 35-44.

Boehmke, B.C., Jackson, R.A., Johnson, A.W., White, E.D., Weir, J.D. and Gallagher, M.A. (2017), "Measuring US air force installation support activities via data envelopment analysis", *Military Operations Research*, Vol. 22 No. 1, pp. 39-58.

Bowlin, W.F. (1987), "Evaluating the efficiency of US Air Force real-property maintenance activities", *Journal of the Operational Research Society*, Vol. 38 No. 2, pp. 127-135.

Bowlin, W.F. (1989), "An intertemporal assessment of the efficiency of air force accounting finance offices", *Research in Governmental and Nonprofit Accounting*.

Brockett, P.L., Cooper, W.W., Kumbhakar, S.C., Kwinn, M.J. Jr and McCarthy, D. (2004), "Alternative statistical regression studies of the effects of joint and service specific advertising on military recruitment", *Journal of the Operational Research Society*, Vol. 55 No. 10, pp. 1039-1048.

Bulla, S., Cooper, W.W., Wilson, D. and Park, K.S. (2000), "Evaluating efficiencies of turbofan jet engines: a data envelopment analysis approach", *Journal of Propulsion and Power*, Vol. 16 No. 3, pp. 431-439.

Charnes, A., Cooper, W.W. and Rhodes, E. (1978), "Measuring the efficiency of decision-making units", *European Journal of Operations Research*, Vol. 2 No. 6, pp. 429-444.

Charnes, A., Clark, C.T., Cooper, W.W. and Golany, B. (1984), "A developmental study of data envelopment analysis in measuring the efficiency of maintenance units in the US air forces", *Annals of Operations Research*, Vol. 2 No. 1, pp. 95-112.

Clarke, R.L. (1992), "Evaluating USAF vehicle maintenance productivity over time: an application of data envelopment analysis", *Decision Sciences*, Vol. 23 No. 2, pp. 376-384.

Dunn, O. (1964), "Multiple comparison using rank sums", *Technometrics*, Vol. 2 No. 3, pp. 241-252.

Golany, B. and Kress, M. (2020), "Measuring readiness and sustainment within analysis of alternatives in military systems acquisition", Dissertation, Naval Postgraduate School, Monterey, CA.

Han, H.K. and Sohn, S.Y. (2011), "DEA application to grouping military airbases", *Military Operations Research*, pp. 31-40.

Hanson, T. (2016), "Efficiency and productivity in the operational units of the armed forces: a Norwegian example", *International Journal of Production Economics*, Vol. 179, pp. 12-23.

Hauck, W.W. Jr and Donner, A. (1977), "Wald's test as applied to hypotheses in logit analysis", *Journal of the American Statistical Association*, Vol. 72 No. 360a, pp. 851-853.

Kottas, A.T., Bozoudis, M.N. and Madas, M.A. (2020), "Turbofan aero-engine efficiency evaluation: an integrated approach using VSBM two-stage network DEA", *Omega*, Vol. 92, pp. 102-167.

Lu, W.M. and Chen, M.H. (2011), "A benchmark-learning roadmap for the military finance center", *Mathematical and Computer Modelling*, Vol. 53 Nos 9-10, pp. 1833-1843.

Lu, W., Kweh, Q.L., Nourani, M. and Shih, J. (2020), "Major weapons procurement: an efficiency-based approach for the selection of fighter jets", *Managerial and Decision Economics*, Vol. 41 No. 4, pp. 574-585.

Moradi, R. and Amiri, M. (2019), "Military hospitals efficiency evaluation: application of malmquist productivity index-data envelopment analysis", *International Journal Data Envelopment Analysis*, Vol. 7 No. 4, pp. 19-30.

Roll, Y., Golany, B. and Seroussy, D. (1989), “Measuring the efficiency of maintenance units in the Israeli Air Force”, *European Journal of Operational Research*, Vol. 43 No. 2, pp. 136-142.

Starčević, S., Bojović, N., Junevičius, R. and Skrickij, V. (2019), “Analytical hierarchy process method and data envelopment analysis application in terrain vehicle selection”, *Transport*, Vol. 34 No. 5, pp. 600-616.

Sun, S. (2004), “Assessing joint maintenance shops in the Taiwanese Army using data envelopment analysis”, *Journal of Operations Management*, Vol. 22 No. 3, pp. 233-245.

Sutton, W. and Dimitrov, S. (2012), “The US Navy explores detailing cost reduction via Data Envelopment Analysis”, *European Journal of Operational Research*, Vol. 227 No. 1, pp. 166-173.

Tone, K. (2001), “A slacks-based measure of efficiency in data envelopment analysis”, *European Journal of Operational Research*, Vol. 130 No. 3, pp. 498-509.

Tone, K. (2017), *Advances in DEA Theory and Applications: With Extensions to Forecasting Models*, John Wiley & Sons, Hoboken, NJ.

Further reading

Clarke, R.L. and Gourdin, K.N. (1991), “Measuring the efficiency of the logistics process”, *Journal of Business Logistics*, Vol. 12 No. 2, p. 17.

Department of Air Force Instruction 21-101 (2020), “Aircraft and equipment maintenance management”, available at: https://static.e-publishing.af.mil/production/1/af_a4/publication/afi21-101/afi21-101.pdf (accessed 14 February 2020).

Corresponding author

Mansik Hur can be contacted at: mansikhur@gmail.com