Big Data Opportunities in Production Records at Air Force Maintenance Depots

Mr. Braden Eichmeier
518 SWES/MXDPEA
U.S. Air Force
Hill Air Force Base, UT
braden.eichmeier.1@us.af.mil

***Abstract*—The Ogden Air Logistics Complex at Hill Air Force Base performs depot maintenance for a wide range of systems. The depot maintains a database to record work, manage logistics, forecast future workload, and facilitate report generation. This study investigates big data and machine learning applications within these data. Findings show various categories of data with sufficient quantity for big data analysis. Weaknesses of the database include infrequent and unreliable connections to legacy systems, a lack of automated user notifications, and low-quality labor hour data. Suggestions for better utilizing the current data include developing data cleaning and imputation tools to improve labor time reporting, implementing user alerts for anomalous data according to a predictive maintenance framework, developing a digital twin scheme to rapidly analyze and forecast base-wide maintenance conditions, and researching the feasibility of generating preliminary workload plans and schedules using reinforcement learning.**

I. Introduction

The Ogden Air Logistics Complex (OO-ALC) at Hill Air Force Base performs depot maintenance on various weapon systems including the F-35, the F-22, the F-16, the A-10, the C-130, the T-38, and the Minuteman III intercontinental ballistic missile (ICBM). The complex is also a leading provider of pneudraulics, secondary power systems, composites, and ICBM rocket motors for the Department of Defense (DOD) [1]. Several decades of this work have been digitally recorded in a structured query language (SQL) database.

Current data practices at OO-ALC are manually intensive and result in lost productivity. Because current processes are manually intensive, analysts require a long training period and often never become fully familiarized with the data. Recent advancements in big data analytics and machine learning can improve maintenance operations at OO-ALC. Potential opportunities focus on saving effort in monotonous tasks in order to enable airmen to focus on more challenging workloads. This saved effort will result in shorter feedback and reporting cycles, increased productivity, and higher quality analysis.

This study was conducted as part of the DAF-MIT AI Accelerator Phantom Fellowship Program. The views expressed in this article are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the U.S. Government. The inclusion of external links and references does not imply any endorsement by the author, the publishing unit, the Department of the Air Force, the Department of Defense or any other department or agency of the U.S. Government. They are meant to provide an additional perspective or as a supplementary resource.

This report presents four distinct opportunities to improve the usage of currently available data at OO-ALC. The opportunities include Improving Labor Scan Hours (Section II), Predictive Data Monitoring (Section III), Digital twinning the data using a model-based paradigm (Section IV), and improve Planning and Scheduling tools (Section V).

II. Improving Labor Scan Hours

Maintenance technicians record the time required to complete a task by scanning when they begin and end working on the task. These scan hours influence budgeting, product ordering, workload planning, and many other administrative tasks at maintenance depots. Therefore, high-quality labor hour data are necessary. Unfortunately, unstable connections to the database can set a technician's reported labor time to zero when sent to the server. A second contributor to near zero scan times is batch scanning, which is when a technician completes several tasks and scans them all at once. Further corruptions may influence technicians to match the submitted time with the expected time for a process. Analyzing the data shows up to 1/3 of completed jobs may have corrupted labor hours.

When analysts detect anomalous samples, they may consider removing those samples from the dataset and consequently miss the nuance in what the data represents [4]. Data imputation aims to salvage the remaining data by estimating missing or anomalous values. Broad categories for describing imputation techniques include classical statistics, machine learning, and deep learning [5]. Labor hour reports could benefit from anomaly detection and imputation by correcting poor scanning discipline, estimating corrupted values, and creating a more normal distribution of reported labor hours. A promising approach to correcting these data is to train an ML model using verified labor hour submissions, then use the learned model to predict anomalous entries [5]. Inputs to such a model may include production number (PDN), process identification, time of day work began, time of day work finished, employee experience, and day of the week. Other useful data can be discovered via feature selection algorithms.

III. Predictive Maintenance Data Monitoring

Predictive maintenance (PdM), also known as Condition Based Maintenance (CBM or CBM+), uses sensor data from a part or system to make maintenance decisions. These programs benefit from reducing waste and downtime by avoiding potentially unnecessary routine maintenance while reducing failures in the field. There are five capability levels of PdM [3] that categorize monitoring in terms of analysis frequency and complexity.

The database tracks the price and quantity thousands of parts. Analysts manually scan each part individually when a maintenance process begins to run over time or budget allotments. This manual process is both inefficient and ineffective and may be described as PdM Level 1. Implementing a PdM Level 3 continuous monitoring process can track the price of each part against a simple inflationary model and alert users when anomalies occur.

A second low-complexity data monitoring tool would aid in setting the expected labor hours for a task. Each year, analysts manually review 20 percent of maintenance processes and update the expected labor time of each process. This review is comparable to PdM Level 2 and may benefit from PdM Levels 3 or 4 monitoring to flag out-of-distribution processes for immediate review.

IV. Digital Twinning Maintenance Depots

Digital Twins (DT) are simulated representations of real-world entities using live sensor data. Research indicates that relevant benefits include live monitoring, scenario risk assessment, better inter-team synergy and collaborations, more efficient decision-making, and better communication [7]. These benefits stem from all the data for a given system, both live and historical, being organized in one central location. Models of the data then enable operational status visualization, expected projections, and what-if analysis.

The Under Secretary of Defense for Research and Engineering released the “Digital Engineering Strategy” in 2018 that promotes DT technology. Among other goals, this strategy calls for “[Making] use of data to improve awareness, insights, and decision making”. Properly utilizing the data will enable “DOD’s vision to build an enterprise capability that securely leverages data and analytics to enable insights and achieve faster and better data-driven decisions” [6]. The expected result of improved data visualization and modeling is an improved ability to assess current status and optimize operations in shorter periods of time.

Rapid visualization for status and reporting are weaknesses of the database’s current functionality. Generating cohesive reports typically involves combining several web application reports and SQL data pulls. These data then must be post-processed manually in a time-consuming and error-prone process that requires expert knowledge. Constructing digital twins of the assets, processes, and personnel at OO-ALC can alleviate these problems. The multi-faceted benefits of this approach include immediate data visualization, reduced experience requirements due to data centralization, and scalable model forecasting. Leadership, analysts, and planners could view all the available data pertaining to an on-base entity to rapidly make informed decisions at a glance.

V. Advanced Planning and Scheduling

Reinforcement Learning (RL) is an ML paradigm inspired by behavioral psychology where an organism learns through trial and error [8]. Recent work in RL shows promising applications in workload planning and scheduling in the Air Force. The DAF-MIT Artificial Intelligence Accelerator produced the scheduling tool Puckboard.AI that creates AI-powered training schedules for student pilots [9]. Another example is a recent data challenge hosted by the Air Force Research Lab, titled the Airlift Challenge, that explored applying RL to cargo delivery operations under uncertain conditions [2].

Given this recent success in both industry and defense, RL may be beneficial for maintenance depot workload planning. Such an application could develop preliminary workload plans that would then be fine-tuned and approved by logistics planners. This work would be near state-of-the-art and would require extensive research and development. The potential time-saving and robustness benefits merit consideration despite the expected upfront costs.

References

[1] Ogden air logistics complex, Jan 2021. URL https://www.hill.af.mil/About-Us/Fact-Sheets/Display/Article/397018/ogden-air-logistics-complex/.

[2] AFRL Public Affairs. Afrl announces airlift challenge, ai-based planning competition. *Air Force Research Laboratory*. URL https://www.afrl.af.mil/News/Article-Display/Article/3255800/afrl-announces-airlift-challenge-ai-based-planning-competition/.

[3] M. Haarman, M. Mulders, and C. Vassiliadis. Predictive maintenance 4.0 predict the unpredictable, Jun 2017.

[4] Anil Jadhav, Dhanya Pramod, and Krishnan Ramanathan. Comparison of performance of data imputation methods for numeric dataset. *Applied Artificial Intelligence*, 33(10): 913–933, 2019.

[5] Sebastian Jäger, Arndt Allhorn, and Felix Bießmann. A benchmark for data imputation methods. *Frontiers in Big Data*, 4, 2021. ISSN 2624-909X. doi: 10.3389/fdata.2021.693674. URL https://www.frontiersin.org/articles/10.3389/fdata.2021.693674.

[6] *Department of Defense Digital Engineering Strategy*. Office of the Deputy Assistant Secretary of Defense for Systems Engineering, Washington, DC, USA, 2018.

[7] Adil Rasheed, Omer San, and Trond Kvamsdal. Digital twin: Values, challenges and enablers from a modeling perspective. *IEEE Access*, 8:21980–22012, 2020. doi: 10.1109/ACCESS.2020.2970143.

[8] Richard S. Sutton and Andrew G. Barto. *Reinforcement Learning: An Introduction*. A Bradford Book, Cambridge, MA, USA, 2018. ISBN 0262039249.

[9] SrA Ashley Thrash. Beyond the board: New ai innovations in digital scheduling. *Department of the Air Force AI Accelerator*. URL https://www.aiaccelerator.af.mil/News/News-Article-View/Article/3197584/beyond-the-board-new-ai-innovations-in-digital-scheduling/.