3-26-2020

Database Analysis to Improve U.S. Transportation Command Forecasting Processes

Maxwell C. Thompson

DATABASE ANALYSIS TO IMPROVE U.S. TRANSPORTATION COMMAND FORECASTING PROCESSES

THESIS

Maxwell C. Thompson, 2nd Lieutenant, USAF

AFIT-ENS-MS-20-M-176

DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

AFIT-ENS-MS-20-M-176

DATABASE ANALYSIS TO IMPROVE U.S. TRANSPORTATION COMMAND FORECASTING

THESIS

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 Operations Research

Maxwell C. Thompson, B.S.

2nd Lieutenant, USAF

February 27, 2020

AFIT-ENS-MS-20-M-176

DATABASE ANALYSIS TO IMPROVE U.S. TRANSPORTATION COMMAND FORECASTING

THESIS

Maxwell C. Thompson, B.S.
2nd Lieutenant, USAF

Committee Membership:

Dr. Brian J. Lunday
Chair

Dr. J. O. Miller
Member

AFIT-ENS-MS-20-M-176

Abstract

The United States Transportation Command (USTRANSCOM) facilitates air, land, and sea transportation for the DOD. On a periodic basis, a myriad of different agencies within USTRANSCOM project future workload to facilitate resource planning, budgeting, and reimbursable rate identification. Within USTRANSCOM, there are a variety of databases and metrics utilized for workload forecasts; neither a standard nor a preferred technique is prescribed. Currently, USTRANSCOM faces challenges in producing accurate workload forecasts [1]. These challenges can lead to unreliable budget requests and, ultimately, hinder the effectiveness and efficiency of USTRANSCOM [1].

For the purpose of routine aircraft movements of cargo and personnel, this research seeks to answer (1) whether any data sets are dominated with respect to data quality, allowing for their removal from consideration and (2) the degree to which any data set is superlative with respect to informing high quality air workload forecasts. Furthermore, this research identifies a possible major problem with USTRANSCOM's current forecasting procedure and provides recommendations on how to best utilize the data sets readily available for use.

AFIT-ENS-MS-20-M-176

To my amazing girlfriend and friends who supported me throughout this endeavor. I appreciate you more than I can ever express.

Acknowledgements

If I have seen further, it is by standing on the shoulders of giants.

– Sir Isaac Newton

I would like to express my sincere appreciation to my advisor, Dr. Brian J. Lunday, for his guidance and support throughout the course of this thesis. The insight and experience was greatly appreciated.

Maxwell C. Thompson

Contents

List of Figures

List of Tables

DATABASE ANALYSIS TO IMPROVE U.S. TRANSPORTATION COMMAND FORECASTING

I. Introduction

Nothing happens until something moves.

-Albert Einstein

1.1 Background

The United States Transportation Command (USTRANSCOM) was established in 1987 to integrate global air, land, and sea transportation for the Department of Defense(DoD) [2]. As such, USTRANSCOM is the single manager of the United States military's transportation system. It coordinates missions worldwide using both military and commercial transportation resources to include aircraft, trains, and ships.

USTRANSCOM's activities are funded by the Transportation Working Capital Fund (TWCF). On an annual basis, USTRANSCOM determines the shipping rates that Department of Defense (DOD) customers must pay to use its services [3]. To calculate such rates, USTRANSCOM must forecast future demand for their resources. Currently, the Joint Distribution Processing Analysis Center (JDPAC) oversees demand forecasting of workload for USTRANSCOM [4].

Ideally, USTRANSCOM seeks to neither gain or lose money from the TWCF [5]. Doing so is exceedingly difficult to attain. In addition to the stochastic demands for steady-state processes, USTRANSCOM has the challenge of demands that shift with national security priorities and DOD responses to geopolitical events. More than most commercial transportation systems, future USTRANSCOM transportation demands are notably difficulty to forecast accurately.

A major part of USTRANSCOM's movement of goods happens by 'channel air.' Channel air is defined as regularly scheduled airlift for movement of passengers and cargo over designated and validated routes [6]. In this sense, channel air represents routine air transportation.

1.2 Overview

In this study, databases from across USTRANSCOM's multiple agencies are evaluated to determine the validity and efficacy of their use in USTRANSCOM forecasting processes. This analysis includes an evaluation of the quality of four databases, as well as the degree to which each of these databases accurately inform USTRANSCOM's current forecasting process.

1.3 Problem Statement

Given a set of recurring forecasts for channel (cargo and passenger) air transport workload and using a disparate set of databases for forecasts generated by different agencies within USTRANSCOM, this research seeks to both reduce the number of databases utilized to generate forecasts within JDPAC and ensure the quality of workload forecasts are robust with respect to the various forecasts required by different USTRANSCOM agencies.

1.4 Research Questions

To address the problem statement, this study answers two fundamental research questions.

1. Is any single data set used by different agencies to generate channel (cargo and passenger) air transport workload forecasts superior in *data quality*?

2. Which data set yields the best performing forecasts over the set of forecast needs within USTRANSCOM?

With respect to the first question, it is of interest to examine the quality of each data set. This requires adopting a definition of data quality and assessing it for each data set both absolutely and relative to the other data sets.

1.5 Organization of this Study

The remainder of this study is organized as follows. Chapter II reviews literature pertinent to data quality assessments, forecasting, and previous studies in support of USTRANSCOM. Chapters III and IV respectively answer the first and second research questions. Chapter V summarizes the garnered insights and recommendations for the research sponsor as well as recommendations for future research.

II. Literature Review

Research is what I am doing when I don't know what I am doing.

–Wernher von Braun

2.1 Overview

The purpose of this chapter is to analyze what information exists and is relevant to database analysis for USTRANSCOM forecasting. An initial review of scholarly articles and theses shows that much has been done to analyze USTRANSCOM forecasting. Indeed, this research stretches back to at least 1973 [7] with a large amount research [3, 5, 8] as recent as 2019 [4].

This chapter summarizes and provides an overview of literature on data quality analysis, time series modeling, and current forecasting methods. Additionally, it provides understanding into USTRANSCOM's current forecasting process as well as insights obtained from previous analytic studies on USTRANSCOM forecasting.

2.2 Data Quality Analysis

Assessing the quality of data is an important step in any data driven research. The quality of outputs can only be as good as the quality of inputs.

A paper by Ardagna et al. [9] focuses on how to provide a data quality assessment for applications aimed at analyzing 'big data' sources. The authors address the problem of assessing data quality and selecting proper data for supporting analytics inputs while respecting constraints such as execution time and cost. They propose an architecture for data quality assessment that includes their Data Quality Module and CCT (Confidence/Cost/Time) Model. The authors additionally illustrate that the sensitiveness of the data quality assessment to the confidence of analysis depends both on the data source features as well as the specific data quality dimensions.

Additionally, there are large number of factors and terms used for big data quality assessment. Abdallah [10] presents over 40 terms used to help qualify data. These terms are broken into four categories. The first of these categories is *data perspective*, which focuses on the quality of data itself. Second is *management perspective*, which deals with how management deals with data. Next is *processing perspective*, which is concerned with the purpose for which the data is being used. Finally, *user perspective* pertains to the data will be delivered and visualized. Each of these categories are relevant as this research seeks to identify quality data in the context of how it will be used and managed.

Another way in which various authors qualify data quality is through varying amount of words beginning with the letter V. Panimalar et al. [11] discuss the evolution of describing big data with V's. The authors give a background on the 3 V's, 4 V's, 5 V's, 10 V's, and 14 V's of big data before introducing three new terms which bring the count up to 17. This is excessive. It suffices to qualify data in terms of 4 V's of big data presented by IBM's Big Data & Analytics Hub: volume, variety, velocity, and veracity [12].

However, the data used in this research is quite manageable and should not be through of as 'big data.' So instead, this research focuses on Pipino [13] who presents a more general overview of data quality assessment, proposing 16 dimensions to evaluate. Pipino points out that assessing data quality requires awareness of the fundamental principles underlying the development of subjective and objective data quality metrics.

Unfortunately, there is no standardized convention for assessing data quality because data is heavily context-dependent. Hence, assessing data quality in this research is a subjective process, although one informed by the workload forecasting purpose for which the various data sets are used by USTRANSCOM.

2.3 Workload Forecasting Methods

Archer's "Forecasting Demand" [14], divides forecasting techniques into two classes: numerical methods and intuitive methods. Intuitive methods for forecasting, Archer states, are more appropriate when data is insufficient. Numerical methods are further dived into time-series and causal methods. For time-series methods, the standard approach for forecasting is using moving averages. Exponential smoothing often provides a more meaningful product because it weights past observations of data, with higher weights accorded to more recent data. However, the compound growth assumption inherent in geometric progressions degrades the performance of exponential smoothing models for medium and long range forecasting. Another approach to time-series forecasting is through temporal regression. A study by Jones et al. [15] shows that multivariate time series models can be used to reliably forecast. Finally, machine learning techniques (e.g., neural networks) have also been shown to be capable of forecasting well (e.g. see [16]).

Causal models, as Archer describes, are best approached through regression. He notes that "if forecasts are required for more than about two years ahead, it is no longer realistic to assume that the present relationships between variables will remain constant"[14]. At this point, investigation is needed to determine how parameters should be changed. It is beneficial to consider independent variables – in addition to previous observations of the dependent variable – that one thinks may directly affect the dependent variable. A transportation study by Phyoe [17] forecasts air traffic demand in the Singapore Flight Information Region by examining the relationship between air traffic and economic variables, such as gross domestic product (GDP).

The study concluded that GDP has a large influence on air traffic. The forecasting models used were exponential trend, ARIMA, and ARIMAX. ARIMAX is an ARIMA model with additional consideration of an exogenous, independent variable.

2.4 Time Series Overview

Time series models relate a sequence of observations of a dependent variable over time, projecting future values of the dependent variable. These models work best when the parameters within the time series forecasting model remain constant over time. Bowerman [18] points out that time series models can often have auto-correlated error terms. These correlations causes models to become inaccurate. However, correlated error terms can be detected using residual plots and the Durbin-Watson statistic. This auto-correlation can be addressed using the first-order auto-regressive process. Additionally, time series models lend themselves to modeling seasonal data. Time series regression can model seasonal data using dummy variables and trigonometric functions.

In "Time Series Analysis," Box [19] points out five important uses of time series and dynamic models. The first use is forecasting future values of a time series from current and past values. The second use is determining the transfer function of a system subject to inertia. The third use is the use of indicator variables in transfer function models to represent the effects of unusual intervention events on a time series. The fourth use is the determination of multivariate dynamic models to represent interrelationships among related time series variables. The final use is the design of control schemes to adjust the model for deviations.

Auto Regressive Integrated Moving Average (ARIMA) is one of the two most widely used forecasting approaches and aims to describe the auto correlations in the data [20]. The AR in ARIMA indicates that the model is a regression of the variable against itself. In an AR model, "we forecast the variable of interest using a linear combination of past values of the variable" [20]. The MA component of ARIMA indicates a moving average model that uses past forecast errors in a regression-like model [20]. "If we combine differencing with autoregression and a moving average model, we obtain a non-seasonal ARIMA model." In this context, 'integration' in ARIMA is the reverse of differencing [20].

The other most widely used forecasting approach is exponential smoothing[20]. This method provides a complimentary approach to ARIMA and is based on a description of trend and seasonality in the data [20]. "Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight" [20]. This framework can create reliable forecasts quickly for a wide range of time series. This "is a great advantage and of major importance to applications in industry"[20].

There exist a variety of univariate methods for forecasting with time-series data. Peng and Chu [21] compare six univariate models to forecast one-year throughput for three major Taiwan shipping ports. The models used were the classical decomposition model, the trigonometric regression model, the regression model with seasonal dummy variables, the grey model, the hybrid grey model, and the SARIMA model (i.e., an ARMIA model having a seasonal component). By comparing forecast accuracy using mean absolute error, mean absolute percent error, and root mean squared error, the authors found "that in general the classical decomposition model appears to be the best model for forecasting" their problem. The authors suggest that a simple method, like classical decomposition, seems to perform well even though "it is not based on formal statistical theory." Furthermore, complex methods do not necessarily provide more accurate forecasts than simpler models. The authors recommend that the first step in finding a good method for forecasting is to carefully exam the distribution of data. It is desirable to develop methods that combine methods that accurately predict the short term with methods that are more effective in forecasting the long term.

2.5 JDPAC Current Forecasting State

Currently, JDPAC uses a variety of databases at their disposal to forecast the following metrics[22]: (1) number of tails (aircraft per time period), (2) flying hours, (3) million ton miles (MTM), and (4) short tons (sTons) [15]. In the future... JDPAC would like to forecast Cubic Feet as well [22].

These forecasts are used by at least five customers to include (1) Air Mobility Command (AMC) and the following Directorates of the Joint Staff: (2) J3 - Operations, (3) J4 - Logistics, (4) J5 - Strategy, Plans, & Policy, and (5) J8 - Force Structure, Resources, & Assessment [22].

To forecasts these metrics, JDPAC follows a five-step process for generating forecasts [23]. This process can be described briefly as running multiple models for each forecast and having subject matter experts (SMEs) examine and approve the results.

Step 1 of this process is to generate multiple models in R for a given forecasting need (for example: the projected workload for a channel pair – a specific aircraft origin and destination – for a type of cargo and over a projected temporal horizon). These models include exponential smoothing and ARIMA models. The exponential smoothing models consider models both with and without seasonality patterns and trends. From these generated models, the model with the lowest Akaike's Information Criterion (AICc) is chosen as the 'best' model. This best model is used to calculate forecasts. This step is repeated for all forecasts that are needed [23].

Step 2 consists of reviewing every forecast generated in Step 1. This process is expedited with the help of Microsoft Excel VBA Macros and a custom graphical user interface (GUI). In Step 3, these forecasts are manually adjusted based on "operational assumptions" [23]. The final two steps consist respectively of approval by subject matter experts and aggregating the data for distribution to customers [23].

2.6 Existing Studies on USTRANSCOM Forecasting

A significant amount of work has been done by the United States Government Accountability Office (GOA) to identify the need for improved forecasting for airlift services. GOA's September 2018 report to Congressional committees summarizes these findings [1]. Currently, USTRANSCOM has not been supplying their forecast to the Air Force with sufficient time to support budget deliberations. As a consequence, Congress does not have sufficient information for the accurate appropriation of funds. GOA also found that USTRANSCOM has challenges producing an accurate forecast of its workload. Towards a positive outcome, "GOA found that forecast inaccuracy averaged 25 percent" and has become increasingly accurate since 2007 [1]. However, "[USTRANSCOM] lacks an effective process to gather workload projections from its customers", "no longer uses forecasting accuracy metrics", "has not established forecast accuracy goals to monitor its performance," and does not have a plan to improve the increasing inaccuracy of its forecasts [1]. These inaccurate forecasts lead to unreliable budget requests and hinder operational planning.

Bradshaw [5] reported that USTRANSCOM's inability to accurately forecast workload demand leads to inaccurate service provider rate-setting. As a result, some customers become dissatisfied when rates spike. These customers then "seek service from other competitors, which generates lost revenue, customer dissatisfaction and the inability to maximize workload to meet the readiness goals of the command"[5]. Bradshaw examined "a variety of time-series techniques applied to historical cargo and flying hour workload demand for Air Mobility Commands (AMC) contingency and special airlift assignment missions." For cargo time-series, Bradshaw's research shows that many models are statistically similar and lead to over-fitting, which in turn results in severely inaccurate forecasts for annual workload demand. Instead, the median value over multiple forecast models' predictions was found to be a more accurate indicator of annual demand. For flying hour time-series, similar patterns of over-fitting were revealed, yet a superior indicator of predictive behavior was not found. Bradshaw's research also provides a standardized way to sanitize raw data into aggregates for forecasting purposes and outlines various, alternative forecasting techniques for consideration by USTRANSCOM.

III. Data Quality Assessment

Without data, you're just another person with an opinion.

–Dr. W. Edwards Deming, American engineer & statistician

3.1 Overview

The first question this research seeks to answer is as follows: is any single data set used by different agencies to generate channel (cargo and passenger) air transport workload forecasts superior in data quality? To address this question, we begin by assessing the quality of our data based on selected principles identified in literature.

Additionally, this chapter will discuss the data used in this research as well as provide an initial data quality assessment. Understanding data is an important part in any data driven research. This chapter will help inform the forecasting and analyses in Chapter 4.

3.2 Data Preparation

Obtaining the data for this research was simple for the researcher, due to the responsive support provided by USTRANSCOM. The data was obtained from USTRANSCOM's Joint Distribution Process Analysis Center (JDPAC). JDPAC is a directorate under USTRANSCOM, and its purpose is to provide analysis and engineering support to improve the nation's ability to move and sustain the Joint Force and operate the Joint Deployment and Distribution Enterprise [6].

These channel air data used in this study were pulled from four databases and were provided by JDPAC. These databases are used for various purposes across multiple agencies. These include maintenance, financial, and operational and represent information about channel and passenger flights. A number of spreadsheets provided this data and encompassed different date ranges. Ultimately, the data was merged into six spreadsheets comprising the same scope of time.

The first database is the Global Air Transportation Execution System (GATES). GATES is an operational database that presents information on groups of cargo. This study was provided with two spreadsheets (cargo and passenger) representing data from October 2013 through April 2019.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	TM_POE	TM_POD	TonMiles	TimeSeries	FY	QTR	Mth_Name	TStons	MajCom	Display	CUSTOMER	PoeApc	PoeCo	PoeZone	PodApc	PodCountryC	PodZone	PoeToPodZone
2	101	AKT	8.4475	12/1/2017	2018	1	December	0.0025	EUCOM	WRI-RM	AIR FORCE	WRI	USA	1	RMS	GM	6	1-6
3	101	AZ1	59.55	1/1/2014	2014	2	January	0.01	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
4	101	AZ1	273.93	11/1/2013	2014	1	November	0.046	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
5	101	AZ1	21027.11	10/1/2013	2014	1	October	3.531	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
6	101	NIM	101.312	8/1/2018	2018	4	August	0.016	AFRICOM	NGU-JIB	DLA	NGU	USA	1	JIB	DJ	12	1-12
7	101	OA2	354.3225	1/1/2014	2014	2	January	0.0595	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
8	101	OA2	494.265	2/1/2014	2014	2	February	0.083	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
9	101	OA2	952.8	11/1/2013	2014	1	November	0.16	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18
10	101	OA2	1444.088	12/1/2013	2014	1	December	0.2425	CENTCOI	DOV-OA	DLA	DOV	USA	1	OA1	AF	18	1-18

Figure 1. Screenshot of GATES Spreadsheet

The second database is the Defense Enterprise Accounting and Management System's Component Billing System (DCBS). DCBS is a financial database the lists the number of short tons and number of passengers that traveled over specific channels in a given month. This study was provided with six spreadsheets representing data from October 2013 through August 2019. These spreadsheets were merged into two (cargo and passenger).

The third database is the Reliability and Maintainability Information System (REMIS) database. REMIS is a maintenance database. Unlike the previous databases discussed, REMIS is a list of flights, not individual cargo or personnel shipments. Furthermore, REMIS, as well as the next database, is not dived into cargo and passenger components. This study was provided with one spreadsheets representing data from October 2013 through March 2019.

The fourth and final database is denoted as *Hybrid* in this research and comprises a merge of selected GATES and Global Decision Support System (GDSS) database

	A	B	C	D	E	F	G
1	Fiscal Year	Month	Customer	Channel	Total Short Tons	Total Miles	Total Ton Miles
2	FY2014	2013/Oct	Air Force	ADA - AVB (I	0.045	5,785	52
3	FY2014	2013/Oct	Air Force	ADA - DOV (I	0.5195	24,335	2,529
4	FY2014	2013/Oct	Air Force	ADA - FRU (I	0.0135	7,276	25
5	FY2014	2013/Oct	Air Force	ADA - HIK (In	0.132	28,932	954
6	FY2014	2013/Oct	Air Force	ADA - IUD (Ir	0.0245	9,738	28
7	FY2014	2013/Oct	Air Force	ADA - KEZ (Ir	0.009	3,800	6
8	FY2014	2013/Oct	Air Force	ADA - KWI (I	0.0055	2,355	4
9	FY2014	2013/Oct	Air Force	ADA - LGS (Ir	0.0485	5,824	141
10	FY2014	2013/Oct	Air Force	ADA - MHZ (I	0.5595	13,856	969

Figure 2. Screenshot of DCBS Spreadsheet

P	Q	R	S	T	U	V	W	X
Mission Number	Takeoff ICAO Location	Landing ICAO Location	Start Date / Time	Stop Date / Time	Hours	Sorties	Landings	Full Stop Landings
0BB09W3	ETAR	LIPA	17 Sep 2014	17 Sep 2014	1.3	1	1	1
0BB09W3	LIPA	LTAG	17 Sep 2014	17 Sep 2014	4	1	1	1
0BB09W3	LTAG	LCRA	18 Sep 2014	18 Sep 2014	1.8	1	1	1
0BB09W3	LCRA	LTAG	18 Sep 2014	18 Sep 2014	1.6	1	1	1
0BB09W3	LTAG	ETAR	19 Sep 2014	19 Sep 2014	5.2	1	1	1
2JWRJ19(	KADW	ETAR	11 May 2017	11 May 2017	8.8	1	1	1
5JBRVM7	KPIA	MUGM	21 Jul 2015 :	21 Jul 2015 2	5.5	1	1	1
5JBRVM7	MUGM	MKJP	22 Jul 2015 :	22 Jul 2015 1	0.9	1	1	1
5JBRVM7	MKJP	MUGM	22 Jul 2015 :	22 Jul 2015 1	0.9	1	1	1

Figure 3. Screenshot of REMIS Spreadsheet

AF	AG	AH	AI	AJ	AK	AL	AM	AN	AO
ARVL_FLT_TM	Flight_Distance	Passenger_Quantity	TRVL_CTGY_CD	Best_Departure_Date	APOE_AERL_PORT_CD	APOD_AERL_PORT_CD	Pax_On	Pax_Off	PaxThru
1.00	212	2	R	2018-09-30	DHF	IUD	2	2	0
.58	168	5	R	2018-09-30	O6R	O1R	5	0	5
.42	56	1	R	2018-09-30	SDA	KEZ	0	0	1
8.72	5073	237	R	2018-09-30	RIV	KWI	237	0	237
4.33	2217	2	R	2018-09-30	SIZ	BAH	2	2	0
7.55	3210	22	R	2018-09-30	RMS	KDH	22	22	0
.58	168	2	R	2018-09-30	SDA	OT4	0	2	0
1.17	373	7	R	2018-09-30	IUD	KEZ	7	7	0
1.08	308	10	R	2018-09-30	KDH	OA1	10	10	0
3.63	1326	75	R	2018-09-30	KEZ	OA1	75	75	0

Figure 4. Screenshot of Hybrid Spreadsheet

fields, as customized by personnel within the Operations Divisions (i.e., TCAC-O) of JDPAC. This database also lists flights and includes information on the number of passengers on these flights. This study was provided with six text files representing data from October 2013 through July 2019. These text files converted to a single spreadsheet using R.

3.3 Scope and Data Description

This research examines channel flight workload data from fiscal years 2014 through 2018 (FY14-FY18). This scope was chosen based on the availability and relevance of data. Based on the available data, it was possible to use data up to March of 2019. However, FY14-FY18 was chosen to simplify understanding of inputs and outputs.

Below, Table 1 shows a brief summary of each database. These databases range by size, number of variables, purpose, and what the data entries represent.

Table 1. Cargo Database Comparison, FY14-FY18

Database	Variables	Entries	Purpose	Entries Represent
GATES	18	169,904	Operational	Groups of cargo
DCBS	7	92,516	Financial	sTons/channel
REMIS	24	40,427	Maintenance	Flights
Hybrid	41	172,333	Mixed	Flights

3.4 Assessing Data Quality

An important first step in any data-driven research is an initial assessment regarding the quality of the data provided. As discussed previously, there are a large number of terms used for data quality assessment throughout relevant literature. Pipino et al. [13] provided a fairly concise list of 16 terms. Of these, five terms were chosen for the initial data quality assessment in this research. Theses terms represent a consolidation of the terms provided by Pipino that are relevant to this study. The selected terms, as defined by Pipino, are as follows:

1. Volume: The extent to which the volume of data is appropriate for the task at hand
2. Completeness: The extent to which data is not missing data and has sufficient depth for the task at hand
3. Representation (Concise/Consistent): The extent to which data is compactly represented and is presented in a consistent format
4. Ease of Use: The extant to which data is easy to manipulate, free-of-errors, and easy to understand
5. Relevancy: The extent to which data is applicable and beneficial for the task at hand

These terms will be qualitatively assessed in the following paragraphs and given an ordinal ranking, first to fourth, based on this assessment. These rankings help to compare these databases across the five data quality parameters. However, it is important to note that these rankings reflect the data used in this research and may or may not accurately reflect the databases as whole. The spreadsheets used in this research represent pulls from these databases and, as a result, data fields that effect these rankings may be missing.

Volume

Ranking the volume of data was conducted by ranking the number of entries each database has. When considering cargo and passenger information together, GATES has the most entries, over 200,000. Next in volume is the Hybrid database which contained more than 172,000 entries. This quantity happens to be roughly the size of the GATES cargo information. However, although the Hybrid database does have less entries, it represents flights unlike GATES, which represents groups of cargo. For this reason, the Hybrid database likely represents more flights than the GATES database.

Next, the DCBS database contains 115,829 entries of cargo and passenger information. Though it has less entries, DCBS does represent a large amount of information because it represents the amount of short tons or passengers that traveled over a given channel in a given month. Finally, the REMIS database is the smallest with 40,427 entries. Like the Hybrid database, REMIS entries represent flights. However at 40,427 entries, this is less than a quarter size of the Hybrid database. The rankings for volume, as reported in Table 2, are based solely on the number of entries of data.

Completeness.

Completeness is defined above as "the extent to which data is not missing data and has sufficient depth for the task at hand." The hybrid database is the only database that has some missing fields. Just under two percent of aircraft tail numbers and flight priority codes are missing. Of the remaining 41 fields, eighteen fields are missing information for less than one percent of flights. However, these blanks are for variables that do not directly affect this research in its current form. Regarding the other databases, they do not have missing values within the entries, but it is indeterminate whether there are missing entries altogether. It is assumed in this research that the databases are otherwise complete and not missing entries.

Since the data provided was fairly clean and had few missing entries, the ranking of completeness is based on the depth of the databases. To elaborate, this depth is represented in the number of variables relevant to our task. It is readily apparent that some databases contain a lot more information than others in their number of variables. The DCBS database has the least amount of variables at 6 for passenger and 7 for cargo. GATES has many more variables at 18. REMIS has a similar amount, 24 variables. Finally, the Hybrid database has by far the most amount of variables with 41 columns. The corresponding ordinal rankings for completeness are likewise reported in Table 2.

Representation.

As described above, representation refers to "the extent to which data is compactly represented and is presented in a consistent format"[13].

The benchmark for representation in this study is the GATES database. GATES is the example in this research for not having too much and too little data.

Second in representation is the REMIS database. REMIS is similar to GATES but does present more information that is not relevant to this study.

Third in representation is the Hybrid database. As discussed previously, the Hybrid database has by far highest number of variables and consequently most amount of information per entry. This characteristic causes the database to be less concise and compact then it could be. Though the data could be presented in a more concise format, the data is still presented in a clear and consistent format. Presenting too many fields in a clear format is far better then presenting too few fields.

Finally, the DCBS database ranks last in representation. The data provided is as compact as it can be for use in this study. To elaborate, the data is already presented in origin-destination pairs for each month, which is the format needed for forecasting in Chapter 4. However, the data provided is too compact. As a result, extra work is required forecast with the data. For example, the 'Channel' column, as displayed in Figure 2, must be separated into origin and destinations before it can be used for forecasting. The fourth column of Table 2 reports the respective database rankings for representation.

Ease of Use

All of the data sets are fairly similar with respect to ease of use. The DCBS data provided was too concise, as discussed previously. Because of this, extra of work had to be done before forecasting efforts could be accomplished in R. As a result, DCBS ranks below GATES and REMIS. The latter two databases tie in their ease of use.

The Hybrid database is an exception to the data sets being similar with respect to ease of use. This database presented a challenge because the information was provided as text files. Given non-uniform delimitation of entries within the Hybrid database's text files, the files could not be imported directly to view in Microsoft Excel. For this reason it is ranked last in ease of use. The fifth column of Table 2 reports the respective database rankings for ease of use.

Relevancy.

The final of the five terms is relevancy. In terms of relevancy, all databases are applicable and beneficial for the task at hand. However, they are relevant in different ways since they provide information that differs in, and they also differ with respect to and the channel workload metrics they can forecast. For this reason, these databases are initially ranked evenly on relevancy. However, the rest of this research will derive insights regarding the relevancy of each database, as will be examined in Chapter 4.

Table 2 presents a summary of the rankings. It is readily observable that no single database is superior (or inferior) across all characteristic rankings.

Table 2. Ordinal Ranking of Database Characteristics

Database	Volume	Completeness	Representation	Ease	Relevancy
GATES	1	3	1	1	1
DCBS	3	4	4	3	1
REMIS	4	2	2	1	1
Hybrid	2	1	3	4	1

3.5 Summary

This chapter discussed the data used in this research as well as provided an initial data quality assessment. This research asks is any single data set used by different agencies to generate channel (cargo and passenger) air transport workload forecasts superior in data quality? The qualitative assessment provided in this chapter suggests that none of the data sets are superior in quantity. Understanding data is an import part in any data driven research and this chapter helps inform the next chapters forecasting efforts.

IV. Forecasting Accurately with JDPAC Databases

Never make forecasts, especially about the future.

–attributed to Samuel Goldwyn

4.1 Motivation and Overview

The initial data quality assessment is useful but insufficient for answering the research questions. To answer the research questions, the quality of forecasting must inform the recommendation. To this end, this research seeks to emulate forecasting efforts accomplished by JDPAC. These forecasts are then used to compare databases and inform our research into the possible superiority of a database and the use for each.

Sections 4.2.1–4.2.5 present the metrics forecast in the testing, how the available data is decomposed for training and testing of forecast models, what forecasting models are considered, how the testing is computationally implemented, and how the 'best' model is selected.

4.2 Common Elements in all Forecast Model Identification

All forecasting efforts were accomplished in R (3.6.1) and RStudio (1.2.1335). These programs ran on a Lenovo P920 (QEB2018B) running Windows 10 (Build 17134) with an Intel Xeon Gold 5120 CPU (2.20GHz) and 256GB of RAM.

4.2.1 Metrics to Forecast

This research seeks to replicate JDPAC forecasting efforts. Currently, JDPAC uses the databases at their disposal to forecast the following metrics [22]: (1) Number of

Tails (Aircraft per time period), (2) Flying Hours, (3) Ton Miles (TM), and (4) Short Tons (sTons).

Although, JDPAC would like to forecast Cubic Feet [22], none of the databases examined here can forecast cubic feet. Furthermore, none of these databases can forecast all of the metrics. Instead, each database is only able to forecast a subset of metrics. GATES and DCBS are the only databases that can contribute to forecasting Ton Miles (TM) and short tons (sTons). While REMIS and the hybrid database are the only databases that can contribute to forecasting tails and flying hours. As an extension of the current JDPAC forecasting metrics, this research evaluates adding Number of Passengers as a metric, models for which can be supported by only DCBS and the Hybrid databases. A summary of what each database can forecast, based on the data sets provided, is summarized in the Table 3 and 4 for cargo and passenger data respectively.

Table 3. Cargo Data Comparison

DB/Purpose	Tails	Flying Hrs	TM	sTons	Cubic Ft
GATES			X	X	
DCBS			X	X	
REMIS	X	X			
Hybrid	X	X			

Table 4. Passenger Data Comparison

DB/Purpose	Tails	Flying Hrs	TM	sTons	Cubic Ft	# of Pax
GATES			X	X		
DCBS			X			X
REMIS	X	X				
Hybrid	X	X				X

Please note that nine passengers represents a short ton of passengers.

Table 3 and 4 present clear competition between selected databases on specific metrics. For example, GATES and DCBS are in competition as they are the databases that can forecast Ton Miles (TM) and short tons, whereas REMIS and Hybrid compete on forecasting tails and flying hours.

4.2.2 Training Data & Testing Data

To evaluate which databases are 'best' in their ability to forecast, we begin by splitting the data into training and testing sets. The intention for splitting the data was to create a 80/20 percent split. In reality, this was accomplished by using the first four fiscal years (FY14-17) as training data and the last year (FY18) as test data. Partitioning the data in this way is in line with Hyndman's [20] recommendation to keep the test set at about "20% of the total sample" as well as "at least as large as the maximum forecast horizon required." To implement this partition, the data for each database was first split by origin-destination pair. Then, the training and test data were split from each other.

4.2.3 Forecasting Models Examined

To generate workload forecasts, JDPAC forecasts a variety of models with different model types and the uses the best fitting one, subject to subject matter expertise as part of their five-step process detailed in Section 2.5. More specifically, JDPAC currently forecasts with ARIMA and exponential smoothing (with/without seasonality, trends, etc.) [23]. In this research, five forecasting model types were used to emulate this process. These methods are as ARIMA, SES, ETS, Last-Naive, and Mean-Naive.

Auto Regressive Integrated Moving Average (ARIMA) and exponential smoothing "are the two most widely-used approaches to time series forecasting, and provide complimentary approaches to the problem" [20]. ARIMA and exponential smoothing are also the methods currently used by USTRANSCOM [23].

This research uses two methods that fall under the umbrella of exponential smoothing. The first method is Simple Exponential Smoothing (SES) which is suitable for forecasting data with no clear trend or seasonal component. The second method is ETS (Error, Trend, Seasonality). ETS gives more flexibility in examining exponential smoothing models. Depending on the use of its parameters, ETS can represent more advanced exponential smoothing models that take into account additive and multiplicative errors.

The next forecasting method used in this research is Naive forecasting. In this method, the last period (i.e., month) of the training data is used as the projected workload for all future months. Despite its simplicity, this method works remarkably well for many time series applications [20]. Naive forecasts are also called *random walk* forecasts [20].

The final method is similar to the fourth method. In *mean Naive*, the average of the training months is used as the workload forecast for future months. This method represents an intuitive approach to forecasting.

4.2.4 Implementation and Coding

The forecasting for this project was accomplished in R and relied heavily on Hyndman and Athanasopoulous's "forecast" package. This package is available from the Comprehensive R Archive Network (CRAN).

Because forecasting requires time series data, the format of dates was important in this research. Instead of using the dates directly, each entry was assigned an integer value corresponding to the month within the training data. The integers ranged from 1 to 60, representing the 60 months within fiscal years 2014 through 2018. For the sake of simplicity and because the process was not repeated often, this data preparation task was accomplished in Microsoft Excel before the development of forecast models in R. However, if the code in this research were to be used as part of a repeated periodic process by JDPAC, the R code should be modified to accomplish this step.

This research treated the data within each database as if it were full and complete. Hence, missing data was not imputed. This assumption is reasonable because, if there was no data for an origin-destination pair in a given month, it is assumed that there simply was no throughput over that pair in the month.

Additionally, forecasts were not generated if there was no data for the last two years. This was implemented in an attempt to reduce the time it took to generate forecasts. This is a reasonable simplification of the problem because two years of inactivity is likely to yield a zero throughput and unnecessary forecast. Unfortunately, this had a relatively small effect and only skipped a few forecasts per database.

The R code to implement the testing begins by reading in a database. For each origin-destination pair, it calculates the sum of each prediction metric (e.g., short tons) for each month. The data is subsequently split into a training and testing set. If no data exists for a certain month, then these months are filled in with zeros. This step is necessary to ensure the time series is the right length and interpreted accurately. Next, the data for each variable is used for each of the five forecast modeling methods. These forecasts are compared, and the forecast with the lowest Mean Absolute Error (MAE) is saved as the best forecast. Example code for this process is provided in Appendix A.

The following functions found in Hyndman's Forecast package were used for forecasting: auto.arima(), ets(), ses(), naive(), and meanf().

The auto.arima() function uses a variation of the Hyndman-Khandakar algorithm, which combines unit root tests and minimization of the corrected Akaike Information Criterion (AICc) and maximum likelihood estimation (MLE) [20]. An overview of this method can be found in Figure 5. Auto.arima() uses the typical ARIMA parameters:

p = order of the autoregressive part, d = degree of first differencing involved, and q = order of the moving average part. As seen in Figure 5, the parameter d is restricted to $d \in \{0, 1, 2\}$, and the function's automated parametric search algorithm considers only selected p and q values over the same parameter space.

1. The number of differences $0 \leq d \leq 2$ is determined using repeated KPSS tests.
2. The values of p and q are then chosen by minimising the AICc after differencing the data d times. Rather than considering every possible combination of p and q, the algorithm uses a stepwise search to traverse the model space.
   a. Four initial models are fitted:
      - ARIMA$(0, d, 0)$,
      - ARIMA$(2, d, 2)$,
      - ARIMA$(1, d, 0)$,
      - ARIMA$(0, d, 1)$.

      A constant is included unless $d = 2$. If $d \leq 1$, an additional model is also fitted:
      - ARIMA$(0, d, 0)$ without a constant.

   b. The best model (with the smallest AICc value) fitted in step (a) is set to be the "current model".
   c. Variations on the current model are considered:
      - vary p and/or q from the current model by ± 1;
      - include/exclude c from the current model.

      The best model considered so far (either the current model or one of these variations) becomes the new current model.
   d. Repeat Step 2(c) until no lower AICc can be found.

Figure 5. Hyndman-Khandakar Algorithm for Automatic ARIMA modelling

The ses() function has two parameters: a smoothing parameter ($\alpha \in [0, 1]$) and the first fitted value at time 1 (ℓ_0). For the ses() function, the unknown parameters and values are calculated by minimizing the sum of squared errors (SSE) as described by Hyndman [20].

Though the same method could be used for the ets() function, Hyndman chooses to estimate the unknown parameters and initial values by maximizing the likelihood instead. After the initial values are selected, the best model is selected using AICc. The ets() function has the following smoothing parameters: α, β, γ, and ϕ. These parmeters have the following restrictions: $\alpha \in (0, 1)$, $\beta \in (0, \alpha)$, $\gamma \in (0, 1 - \alpha)$, and

$\phi \in (0.8, 0.98)$ (specific to R function).

Additionally, ets() has the following initial states: ℓ_0, b_0, s_0, s_{-1},...,s_{-m+1}.

The naive() and meanf() functions have no model parameters, apart from the data, because they simply forecast future observations using the last period and mean period values, respectively.

Two of these five functions select the best model using AICc. AICc is a small-sample (second-order) correction of the Akaike Information Criterion, and it is the current selection criteria used by JDPAC to compare accuracy across model types. Based on the limited amount of data for each origin-destination pair, it is the most appropriate criteria for comparing models of the same type. AICc is appropriate for comparing models of the same type and using the data, however, it is not appropriate for comparing models of different types, as will be discussed in Section 4.2.5.

4.2.5 Best Model Identification

Currently, JDPAC uses AICc to evaluate and compare different forecast types [23]. Originally, this research had the intention on emulating this process to best replicate how JDPAC forecasts. However, in researching Akaike's Information Criterion (AIC) it was discovered that the process used by JDPAC, as of late 2018, may have model accuracy hindered by a common and major forecasting fallacy. Because of the way in which AIC and AICc are calculated, comparing either the AIC or AICc from one model type with another model type is inappropriate. Depending on the specifics of the model, one may not be able to accurately compare models using AIC. Hyndman gives the simple example that "you cannot compare the AIC from an ETS model with the AIC from an ARIMA model" [24]. Additionally, "you cannot [use AIC to] compare an ARIMA model with differencing to an ARIMA model without differencing" [24]. To explain the problem simply, models treat initial values differently, which results in likelihood functions being calculated differently. More detail about this complication is given in "Facts and Fallacies of the AIC" [24]. The consequence of this revelation is that there was and may still exist an issue with the process utilized by JDPAC to select forecasting models for subject matter expert validation.

This revelation inadvertently answers the question: does the current JDPAC use of the best fitting model portend quality forecasts? The answer is 'no'. Since their current procedure incorrectly uses AICc when selecting the best performance, this method will not always select the best fitting model. It may still select a good model, but it is less likely to be the best model. Furthermore, this process will likely select the same type of model (e.g., ARIMA) for each comparison because of a consistent model-based bias for AICc computations.

As a result of not being able to use AICc to compare different types of models, mean absolute error (MAE) is used to compare models in this research. MAE uses the same scale its data. Hence, it is reported in the same units as the data which is desirable. Errors were calculated using the "mae" function from Hammer and Frasco's "Metrics" R package. Of note, mean absolute scaled error (MASE) was considered but not used because it provides errors that are undefined or infinite when actual or predicted data is zero-valued.

After the MAE of each forecast for an origin-destination pair is calculated, the errors were compared. The model with the lowest MAE was recorded as the best model.

4.3 Comparison of Database-Specific Forecasts

4.3.1 Comparison of Forecast Errors

We still seek to answer the second research question: which data set yields the best performing forecasts over the set of forecast needs within USTRANSCOM? To answer this question, we need to compare the forecasts for each databases against the forecasts of the other databases.

After forecasting each origin-destination pair, these forecasts were compared against their test data sets. The training and test MAEs were collected and organized by database and metric resulting in 26 distributions of errors, wherein each error (i.e., measure of MAE) corresponds to an origin-destination pair within the given database. Since we make no assumptions about the distribution of the errors, we use Kolmogorov-Smirnov (K-S) Tests to compare error distributions. K-S Tests let us examine whether distributions of errors likely arise from the same distributions without making assumptions about the family of distribution. K-S tests do not depend on the underlying cumulative distribution function being tested and do not depend on an adequate sample size to be valid.

This research performed twelve K-S Tests comparing both the Training and Test MAEs across data sets and metrics. To clarify, pairwise comparisons of error distributions were performed for the training data, across databases and metrics, and then the same was done for the error distributions for the test data. These twelve K-S tests and graphical displays of forecast error distributions can be found in Appendix B. Using $\alpha = 0.05$, only one test (Test 2: REMIS v Hybrid Tails Test MAE) signified that the MAEs were from the same distribution. However the p-value is 0.095, which is still low and would not be significant if α were increased to a commonly-used value of 0.10. All other K-S Tests found that the distributions of errors were not from the same distribution.

Additionally, two-tailed Welch's t-tests were performed to compare the means of these distributions. These tests are also found in Appendix B. Welch's t-tests were used instead of Student's t-test since the former is more reliable when samples have unequal sample sizes or variances. These t-tests showed that the means of the error distributions were significantly different with $\alpha = 0.05$. Since the sets of error distributions, with one exception, and their means were found to be different, recommendations could be provided on which databases to use for which metrics. These recommendations are made based on minimizing mean error. (Minimizing standard deviation would yield the same results.) These recommendations, displayed in Table 7, show that all data sets are needed to forecast as well as possible, since each database is the best database at forecasting a specific metric.

Tables 5 and 6 show the mean, standard deviation, 95% confidence interval around the mean, and sample size for the training and testing distributions, respectively. Table 5 displays how well the respective databases enable the forecasting procedure to obtain good fitting models, while Table 6 displays how well the respective databases enable the forecasting procedure to generate accurate forecasts.

It is important to note that their was one outlier in the MAE data. This outlier occurred in the Gates Cargo data for both short tons and ton miles. This outlier was identified to represent the Ramstein Air Force Base to McGuire Air Force Base channel. This outlier was removed before the K-S tests were performed.

4.4 Summary

The initial data quality assessment was useful but insufficient to answer the both research questions. This chapter reviewed which metrics were forecast, how the available data was prepared, which models were used, how these models were implemented, and how the 'best' model was selected. This chapter then discussed comparing errors across databases to inform recommendations on which databases should be used for forecasting specific metrics. Additionally, this chapter identified a possible shortcoming to in JDPAC's current forecasting procedure.

Table 5. Training MAE Statistics

Test	Database	Metric	Mean	St Dev	95% CI	Sample Size
1	REMIS	Tails	0.223	0.387	(0.2, 0.24)	1662
1	Hybrid	Tails	0.351	0.839	(0.32, 0.38)	2869
3	REMIS	Hours	0.872	2.097	(0.77, 0.97)	1662
3	Hybrid	Hours	12.236	38.908	(10.81, 13.66)	2869
5	DCBS Cargo	sTons	1.645	12.59	(1.11, 2.18)	2153
5	GATES Cargo	sTons	11.086	20.205	(7.73, 14.45)	139
7	DCBS Cargo	Ton Miles	6347	51667	(4157, 8537)	2138
7	GATES Cargo	Ton Miles	45640	114674	(26577, 64704)	139
9	DCBS Pax	Ton Miles	2755	12092	(1810, 3700)	629
9	GATES Pax	Ton Miles	77484	199120	(47374, 107595)	168
11	DCBS Pax	# Pax	7.124	26.605	(5.05, 9.20)	631
11	Hybrid Pax	# Pax	1.395	3.373	(1.27, 1.52)	2863

Table 6. Testing MAE Statistics

Test	Database	Metric	Mean	St Dev	95% CI	Sample Size
2	REMIS	Tails	0.24	0.568	(0.21, 0.27)	1662
2	Hybrid	Tails	0.478	1.608	(0.42, 0.54)	2869
4	REMIS	Hours	1.019	4.592	(0.80, 1.24)	1662
4	Hybrid	Hours	21.977	144.522	(16.69, 27.27)	2869
6	DCBS Cargo	sTons	2.605	23.8	(1.60, 3.61)	2153
6	GATES Cargo	sTons	14.765	33.081	(9.27, 20.26)	139
8	DCBS Cargo	Ton Miles	10005	86262	(6349, 13662)	2138
8	GATES Cargo	Ton Miles	61085	175963	(31832, 90337)	139
10	DCBS Pax	Ton Miles	2704	12305	(1742, 3665)	629
10	GATES Pax	Ton Miles	88051	220618	(54689, 121412)	168
12	DCBS Pax	# Pax	6.797	24.884	(4.86, 8.74)	631
12	Hybrid Pax	# Pax	1.963	10.179	(1.59, 2.34)	2863

Table 7. Best Databases for Forecasting A Specific Metric

Type/Metric	Tails & Flying Hrs	TM	sTons	# of Passengers
Cargo	REMIS	DCBS	DCBS	-
Passenger	REMIS	DCBS	GATES	Hybrid

V. Conclusions and Future Research

Sometimes a scream is better than a thesis.

–Ralph Waldo Emerson

5.1 Conclusions

USTRANSCOM is the single manager of the United States military's transportation system. To calculate rates for its customers, USTRANSCOM forecasts future demand for their resources.

In this study, databases from across USTRANSCOM's multiple agencies are evaluated to determine validity of their use in USTRANSCOM forecasting. This study includes an evaluation of the quality of four databases as well as insight into how these databases affect USTRANSCOM's current forecasting process. More specifically, this research sought to answer the following questions:

1. Is any single data set used by different agencies to generate channel (cargo and passenger) air transport workload forecasts superior in *data quality*?

2. Which data set yields the best performing forecasts over the set of forecast needs within USTRANSCOM?

With regard to the first research question, no data set is clearly superior in data quality. In both a qualitative and quantitative sense, each data set has strengths and weaknesses, preventing a preliminary recommendation for any database to be set aside for workload forecasting use. With regard to the second research question, no single data set yields the best performing set of forecasts. Instead, each database is best at forecasting some metric, and all should be retained. Additional insights were also garnered, as detailed in the following recommendations.

5.2 Recommendations for JDPAC

The most important contribution of this work is the insight that JDPAC is likely comparing models incorrectly. As discussed, AICc can easily be misused and, as a result, fail to accurately compare models. Thus, this research recommends that JDPAC no longer use AICc when comparing models across different model types. Instead, it is suggested that JDPAC use an alternative metric such as Mean Absolute Error (MAE). It is still acceptable, however, to use AICc when selecting a model within a specific model type.

Also, Chapter 4 of this research compares forecasting errors across data sets. Based on the information available to this research, Table 7 summarizes the databases recommended for use by JDPAC when forecasting certain metrics.

Table 7. Best Databases for Forecasting A Specific Metric

Type/Metric	Tails & Flying Hrs	TM	sTons	# of Passengers
Cargo	REMIS	DCBS	DCBS	-
Passenger	REMIS	DCBS	GATES	Hybrid

5.3 Recommendations for Future Research

Forecasting at USTRANSCOM is a problem that has been studied many times, as the literature review of this research reveals. Yet, this problem of producing accurate forecasts still remains exceedingly difficult. It is, after all, a problem of predicting the future. Despite this difficultly and the great research that has been done, much can still be done to improve these forecasting efforts. Most research in the past has focused on mathematical techniques for improving forecasting. With this in mind, the best of approach for future research on this problem may lie with the structure and processes of USTRANSCOM's forecasting efforts.

As discussed in the literature review, a significant amount of work has been done by the United States Government Accountability Office (GOA) to identify the need for improved forecasting structure at USTRANSCOM. GOA states that "[USTRANSCOM] lacks an effective process to gather workload projections from its customers," "no longer uses forecasting accuracy metrics," "has not established forecast accuracy goals to monitor its performance," and does not have a plan to improve the increasing inaccuracy of its forecasts [1].

GAO appears to believe that the issue with USTRANSCOM forecasting lies with their processes and procedures. Their recommendations include improving the timing of forecasting and budget requests as well as implementing accuracy goals and a corrective action plan [1]. Additionally, they conjecture the possible benefit of implementing Sales and Operations Planning processes that resulted in a 50 percent reduction in forecast error in specific Army forecasting efforts.

With regard to future USTRANSCOM forecasting efforts, GAO believes "Until TRANSCOM establishes a process to collect projected workload information from its customers, uses forecast accuracy metrics and goals to monitor its performance, and implements a corrective action plan, forecast accuracy and [Airlift Readiness Account] estimates are not likely to improve" [1].

Appendix A. R Code Example

Appendix usually means "small outgrowth from large intestine," but in this case it means "additional information accompanying main text." Or are those really the same things?

–Pseudonymous Bosch, *The Name of This Book Is Secret*

The following pages contain example code showing how this research generated forecasts using GATES Cargo data.

```
########## Set Up ##########
rm(list = ls()) #remove variable
cat("\014") #clear command window
start_time=proc.time() #start timer
library(Metrics) #load packages
library(forecast)
library(grDevices)
f_months=12
max_months=60
train_cut_off=48
j=0
My_Forecasts=""
channel_names=""

########## Read Data ##########
Gates_Cargo = readRDS("//fsv-afit-617/common/Lunday/Advising/MS22.
    Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R
    Files/Gates_Cargo_FY14_18.rds")
Origin=unique(Gates_Cargo$PoeApc)
Destination=unique(Gates_Cargo$PodApc)

```

```
for (cur_Origin in Origin) {
  for (cur_Dest in Destination) {

    Channel_Subset = subset(Gates_Cargo,Gates_Cargo$PodApc == cur_
   Origin & Gates_Cargo$PoeApc == cur_Dest)

    if (nrow(subset(Channel_Subset,Channel_Subset$FY<2018)) > 0 ) {

        j=j+1

        data_train=aggregate(Channel_Subset$TStons, by=list(Category
   =Channel_Subset$Month), FUN=sum)
        names(data_train)=c("Month","SUM_sTons")
        data_train_tm=aggregate(Channel_Subset$TonMiles, by=list(
   Category=Channel_Subset$Month), FUN=sum)
        names(data_train_tm)=c("Month","Ton_Miles")

        ########## Filling in zeros / fixing time series ##########
        data_train_zeros = data.frame(1:max_months,rep(0,max_months)
   )
        data_train_tm_zeros = data_train_zeros
        names(data_train_zeros)=c("Month","SUM_sTons")
        names(data_train_tm_zeros)=c("Month","Ton_Miles")

        for (i in 1 : max(nrow(data_train),nrow(data_train_tm)) ) {
          if (is.na(data_train[i,2])==FALSE) {
            data_train_zeros[data_train[i,1],2]=data_train[i,2]
          }
          if (is.na(data_train_tm[i,2])==FALSE) {
            data_train_tm_zeros[data_train_tm[i,1],2]=data_train_tm[
   i,2]
          }
```

```
    }

    data_train=data_train_zeros[1:train_cut_off,]
    data_train_tm=data_train_tm_zeros[1:train_cut_off,]
    data_test=data_train_zeros[(train_cut_off+1):max_months,]
    data_test_tm=data_train_tm_zeros[(train_cut_off+1):max_months,]

    rm(data_train_zeros,data_train_tm_zeros)

    ########## Auto Arima ##########
    aa_fc_st = auto.arima(data_train$SUM_sTons,ic = "aicc")
    aa_fc_tm = auto.arima(data_train_tm$Ton_Miles,ic = "aicc")

    ########## SES ##########
    ses_fc_st=ses(as.ts(data_train$SUM_sTons),h=f_months) #h is time periods
    ses_fc_tm=ses(as.ts(data_train_tm$Ton_Miles),h=f_months)

    ########## ETS ##########
    ets_fc_st=ets(data_train$SUM_sTons)
    ets_fc_tm=ets(data_train_tm$Ton_Miles)

    ########## Naive ##########
    n_fc_st=naive(data_train$SUM_sTons,h=f_months) #Real Naive
    n_fc_tm=naive(data_train_tm$Ton_Miles,h=f_months) #Real Naive

    ########## Mean Naive ##########
    mean_fc_st=meanf(ts(data_train$SUM_sTons),h=f_months) #Mean
    mean_fc_tm=meanf(ts(data_train_tm$Ton_Miles),h=f_months) #Mean
```

```

    #Compare short tons
    aa_st_mae=mae(aa_fc_st$x,aa_fc_st$fitted)
    aa_tm_mae=mae(aa_fc_tm$x,aa_fc_tm$fitted)
    ses_st_mae=mae(ses_fc_st$x,ses_fc_st$fitted)
    ses_tm_mae=mae(ses_fc_tm$x,ses_fc_tm$fitted)
    ets_st_mae=mae(ets_fc_st$x,ets_fc_st$fitted)
    ets_tm_mae=mae(ets_fc_tm$x,ets_fc_tm$fitted)
    n_st_mae=mae(n_fc_st$x,rep(n_fc_st$x[train_cut_off],train_cut_off))
    n_tm_mae=mae(n_fc_tm$x,rep(n_fc_tm$x[train_cut_off],train_cut_off))
    mean_st_mae=mae(mean_fc_st$x,mean_fc_st$fitted)
    mean_tm_mae=mae(mean_fc_tm$x,mean_fc_tm$fitted)

    model_MAE_st=aa_st_mae
    best_name_st="aa"
    best_forecast_st=aa_fc_st

    if(ses_st_mae< model_MAE_st){
      model_MAE_st=ses_st_mae
      best_name_st="ses"
      best_forecast_st=ses_fc_st
    }
    if(ets_st_mae< model_MAE_st){
      model_MAE_st=ets_st_mae
      best_name_st="ets"
      best_forecast_st=ets_fc_st
    }
    if(n_st_mae< model_MAE_st){
      model_MAE_st=n_st_mae
      best_name_st="n_last"
```

```
      best_forecast_st=n_fc_st
    }
    if(mean_st_mae< model_MAE_st){
      model_MAE_st=mean_st_mae
      best_name_st="mean"
      best_forecast_st=mean_fc_st
    }

    #Compare ton miles
    model_MAE_tm=aa_tm_mae
    best_name_tm="aa"
    best_forecast_tm=aa_fc_tm

    if(ses_tm_mae< model_MAE_tm){
      model_MAE_tm=ses_tm_mae
      best_name_tm="ses"
      best_forecast_tm=ses_fc_tm
    }
    if(ets_tm_mae< model_MAE_tm){
      model_MAE_tm=ets_tm_mae
      best_name_tm="ets"
      best_forecast_tm=ets_fc_tm
    }
    if(n_tm_mae< model_MAE_tm){
      model_MAE_tm=n_tm_mae
      best_name_tm="n_last"
      best_forecast_tm=n_fc_tm
    }
    if(mean_tm_mae< model_MAE_tm){
      model_MAE_tm=mean_tm_mae
      best_name_tm="mean"
      best_forecast_tm=mean_fc_tm
```

```
        }

        #Mean Absolute Error Calculations
        st_fc=forecast::forecast(best_forecast_st,h=f_months)
        st_fc=as.data.frame(cbind((train_cut_off+1):max_months,st_fc$mean[1:12],data_test[,2]))
        names(st_fc) <- c("Month","Forecast","Actual")
        MAE_st=Metrics::mae(actual = st_fc$Actual,predicted = st_fc$Forecast)

        tm_fc=forecast::forecast(best_forecast_tm,h=f_months)
        tm_fc=as.data.frame(cbind((train_cut_off+1):max_months,tm_fc$mean[1:12],data_test_tm[,2]))
        names(tm_fc) <- c("Month","Forecast","Actual")
        MAE_tm=Metrics::mae(actual = tm_fc$Actual,predicted = tm_fc$Forecast)

        #Creating vectors to store data
        vector1 = c(aa_st_mae,ses_st_mae,ets_st_mae,n_st_mae,mean_st_mae)
        names(vector1) <- c("aa_fc_st","ses_fc_st","ets_fc_st","n_fc_st","mean_st_mae")

        vector2 = c(aa_tm_mae,ses_tm_mae,ets_tm_mae,n_tm_mae,mean_tm_mae)
        names(vector2) <- c("aa_fc_tm","ses_fc_tm","ets_fc_tm","n_fc_tm","mean_tm_mae")

        list1<-list(cur_Origin,cur_Dest,best_name_st,model_MAE_st,best_forecast_st,best_name_tm,model_MAE_tm,best_forecast_tm,MAE_st,MAE_tm,vector1,vector2)
        names(list1) <- c('Origin','Destination','Best_st_FC_Type','
```

```
Best_st_FC_Value','Best_st_FC','Best_tm_FC_Type','Best_tm_FC_Value','Best_tm_FC',"MAE_st","MAE_tm","st_FCs","tm_FCs")

      My_Forecasts[j]=list(list1)
      channel_names[j]<-paste(cur_Origin,"->",cur_Dest,sep = "")

      if (j==1) {
        graph_vector_st=data.frame(cbind(model_MAE_st,MAE_st))
        graph_vector_tm=data.frame(cbind(model_MAE_tm,MAE_tm))
      }else {
        graph_vector_st=rbind(graph_vector_st,c(model_MAE_st,MAE_st))
        graph_vector_tm=rbind(graph_vector_tm,c(model_MAE_tm,MAE_tm))
      }

      #Removing variables
      rm(aa_fc_st,ses_fc_st,ets_fc_st,n_fc_st,mean_fc_st)
      rm(aa_fc_tm,ses_fc_tm,ets_fc_tm,n_fc_tm,mean_fc_tm)
      rm(list1,vector1,vector2,MAE_st,MAE_tm)
      rm(best_forecast_st,best_name_st,model_MAE_st,data_train)
      rm(best_forecast_tm,best_name_tm,model_MAE_tm,data_train_tm)
      rm(data_test,data_test_tm)
      rm(aa_st_mae,ses_st_mae,ets_st_mae,n_st_mae,mean_st_mae)
      rm(aa_tm_mae,ses_tm_mae,ets_tm_mae,n_tm_mae,mean_tm_mae)
      rm(st_fc,tm_fc)

  } #end if nrow(subset(Channel_Subset,Channel_Subset$FY<2018)) > 0

} #end dest for
print(proc.time()-start_time) #print processing time wind window
```

```
} #end origin for

names(My_Forecasts)<-channel_names #labels to "My_Forecasts" saved data

#Graphing results
  #Graphing setup
    names(graph_vector_st)<-c("x","y")
    names(graph_vector_tm)<-c("x","y")

    graph_vector_st$x=as.numeric(graph_vector_st$x) #forcing data to be numeric
    graph_vector_tm$x=as.numeric(graph_vector_tm$x)
    graph_vector_st$y=as.numeric(graph_vector_st$y)
    graph_vector_tm$y=as.numeric(graph_vector_tm$y)

    graph_vector_st=graph_vector_st[order(graph_vector_st$x),] #ordering data for graphing
    graph_vector_tm=graph_vector_tm[order(graph_vector_tm$x),]
    graph_vector_st=graph_vector_st[1:(nrow(graph_vector_st)-1),] #deleting last row because it was on outlier
    graph_vector_tm=graph_vector_tm[1:(nrow(graph_vector_tm)-1),] #deleting last row because it was on outlier

    mod_st <- lm(y ~ x -1, data = graph_vector_st) #creating a linear model to plot line on graph
    mod_tm <- lm(y ~ x -1, data = graph_vector_tm)

  #Graphs
    names(graph_vector_st)<-c("model_MAE","MAE") #adding data labels
    names(graph_vector_tm)<-c("model_MAE","MAE")
```

```
 png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22.Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/Gates_Cargo_st_plot.png")
 plot(x=graph_vector_st$model_MAE,y=graph_vector_st$MAE,xlab = "model_MAE",ylab = "MAE",main = "Gates Cargo: sTons")
 lines(x=graph_vector_st$model_MAE,y=predict(mod_st, list(x = graph_vector_st$model_MAE)))
 dev.off()

 png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22.Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/Gates_Cargo_tm_plot.png")
 plot(x=graph_vector_tm$model_MAE,y=graph_vector_tm$MAE,xlab = "model_MAE",ylab = "MAE",main = "Gates Cargo: Ton Miles")
 lines(x=graph_vector_tm$model_MAE,y=predict(mod_tm, list(x = graph_vector_tm$model_MAE)))
 dev.off()

#Histograms
 #Test Mean Absolute Error
 png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22.Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/Forecast_MAE_Hist/Gates_Cargo_st_hist.png")
 hist(graph_vector_st$MAE,main="Histogram of Test MAE",xlab="Test MAE",xlim = c(0,15),nclass = 250)
 dev.off()
 png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22.Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/Forecast_MAE_Hist/Gates_Cargo_tm_hist.png")
```

```
    hist(graph_vector_tm$MAE,main="Histogram of Test MAE",xlab="Test MAE",xlim = c(0,10000),nclass=1000)
    dev.off()

    #Test/Training MAE Ratio
    png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22. Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/MAE_Ratio_Hist/Gates_Cargo_st_ratio_hist.png")
    hist(graph_vector_st$MAE/graph_vector_st$model_MAE,main="Histogram of Test/Training MAE Ratio",xlab = "Test/Training MAE Ratio",xlim = c(0,4))
    dev.off()
    png(filename ="//fsv-afit-617/common/Lunday/Advising/MS22. Thompson (Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Plots_3/MAE_Ratio_Hist/Gates_Cargo_tm_ratio_hist.png")
    hist(graph_vector_tm$MAE/graph_vector_tm$model_MAE,main="Histogram of Test/Training MAE Ratio",xlab = "Test/Training MAE Ratio",xlim = c(0,4))
    dev.off()

#Clean Up R
print(proc.time()-start_time) #printing total run time

rm(Channel_Subset,j,Origin,cur_Origin,i,max_months,train_cut_off)
rm(f_months,Gates_Cargo,cur_Dest,Destination,start_time)

mySave=list(My_Forecasts,graph_vector_st,graph_vector_tm,mod_st,mod_tm)
names(mySave)<-c('My_Forecasts','graph_vector_st','graph_vector_tm','mod_st','mod_tm')
```

```
saveRDS(mySave,
        file = "//fsv-afit-617/common/Lunday/Advising/MS22.Thompson(Database Anal for Forec)/Data/Current Data (FY14-18)/R_Code_Thesis/Final/Forecasts/Gates_Cargo.rds")

warnings()
summary(mod_st)
summary(mod_tm)
```

Appendix B. Kolmogorov-Smirnov Tests

The following pages contain twelve Kolmogorov-Smirnov Tests and Welch's t-tests which compare the distributions and means of test and training errors between data sets.

K-S Tests and Welch's t-tests

2d Lt Maxwell Thompson

19 Feb 2020

Abstract

KS Tests comparing the Training and Test MAEs across datasets and metrics. Alpha = 0.05. Only one test (Test 2: Remis v Hybrid Tails Test MAE) signified that the MAEs were from the same distribution. But the p-value is 0.9, which is still pretty low. Additionally, recommendations are provided on which database to use for forecasting each metric. These recommendations are made based on minimizing mean and st dev. No datasets that differ have overlapping confidence intervals. These results show that all datasets are needed.

R Markdown

```
rm(list = ls()) #remove variables
cat("\014") #clear command window
```



```
Remis <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/Remis.rds")
Hybrid <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/Hybrid.rds")
Gates_Pax <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/Gates_Pax.rds")
DCBS_Cargo <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/DCBS_Cargo_v2.rds
")
DCBS_Pax <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/DCBS_Pax.rds")
Gates_Cargo <- readRDS("C:/Users/Max's USAFA PC/Desktop/Final/Final/Forecasts/Gates_Cargo.rds"
)

alpha=0.05
```

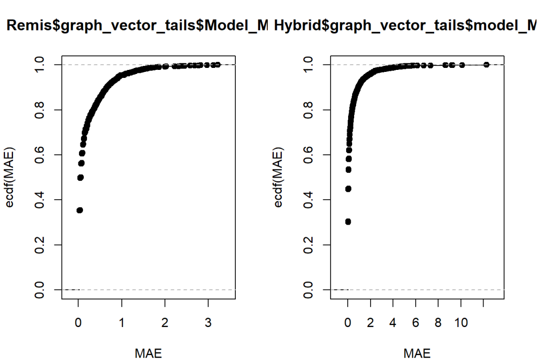
```
my_KS_funct <- function(vect1,vect2) {
  par(mfrow = c(1,2));plot(ecdf(vect1),xlab="MAE",ylab="ecdf(MAE)",main=deparse(substitute(vec
t1)));plot(ecdf(vect2),xlab="MAE",ylab="ecdf(MAE)",main=deparse(substitute(vect2)))
  test=suppressWarnings(ks.test(vect1,vect2))

  if (test$p.value>=alpha) {
    print("Not different")
    print(paste("p-value: ",round(test$p.value,3)))
  }else{
    print("Different")
    print(paste("p-value: ",round(test$p.value,3)))
  }

  print(paste("mean: ",round(mean(vect1),3), " & sd: ",round(sd(vect1),3)," & n: ",length(vect
1)))
  print(paste("mean: ",round(mean(vect2),3), " & sd: ",round(sd(vect2),3)," & n: ",length(vect
2)))

  #95% confidence intervals of the mean 1
  x=vect1
  sem<-sd(x)/sqrt(length(x));    #computation of the standard error of the mean
  print(paste("95% Conf Int: (",round(mean(x)-1.96*sem,2),",", round(mean(x)+1.96*sem,2),")"))


  #95% confidence intervals of the mean 2
  a=mean(x)-1.96*sem;b=mean(x)+1.96*sem; #from vect1
  x=vect2
  sem<-sd(x)/sqrt(length(x));    #computation of the standard error of the mean
  print(paste("95% Conf Int: (",round(mean(x)-1.96*sem,2),",", round(mean(x)+1.96*sem,2),")"))


  #print
  if ((a > mean(x)-1.96*sem & b > mean(x)+1.96*sem) | (a < mean(x)-1.96*sem & b < mean(x)+1.96
*sem)) {
    print("No overlap")
  }else{
    print("OVERLAP")
  }

  if ( (mean(vect1) < mean(vect2)) & (sd(vect1) < sd(vect2)) ){
    print("Recommendation: 1")
```



```
  }else if ( (mean(vect1) > mean(vect2)) & (sd(vect1) > sd(vect2)) ){
    print("Recommendation: 2")
  }else{
    print("No recommendation")
  }

  if (length(vect1) > length(vect2)) {
    print("1 is longer")
  }else{
    print("2 is longer")
  }

  t.test(vect1,vect2)

}
```

Test 1: Remis v Hybrid Tails Training MAE

```
my_KS_funct(Remis$graph_vector_tails$Model_MAE,
            Hybrid$graph_vector_tails$model_MAE)
```




```
## [1] "Different"
## [1] "p-value:  0.008"
## [1] "mean:  0.223  & sd:  0.387  & n:  1662"
## [1] "mean:  0.351  & sd:  0.839  & n:  2869"
```

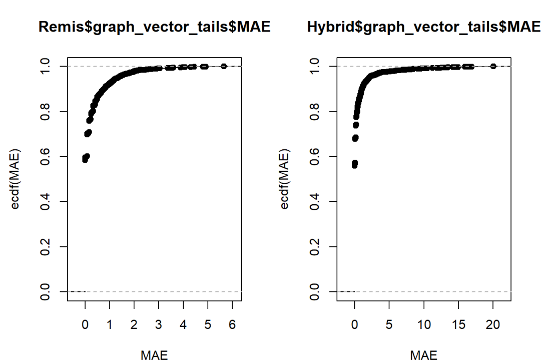


```
## [1] "95% Conf Int: ( 0.2 , 0.24 )"
## [1] "95% Conf Int: ( 0.32 , 0.38 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -7.0034, df = 4348.9, p-value = 2.882e-12
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.16421471 -0.09238326
## sample estimates:
## mean of x mean of y
## 0.2229436 0.3512425
```

Test 2: Remis v Hybrid Tails Test MAE

```
my_KS_funct(Remis$graph_vector_tails$MAE,
            Hybrid$graph_vector_tails$MAE)
```



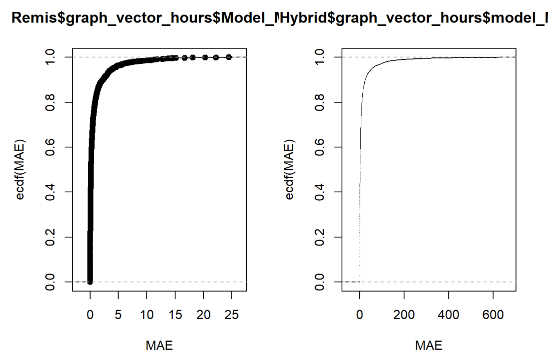

```
## [1] "Not different"
```

```
## [1] "p-value:  0.095"
## [1] "mean:  0.24  & sd:  0.568  & n:  1662"
## [1] "mean:  0.478  & sd:  1.608  & n:  2869"
## [1] "95% Conf Int: ( 0.21 , 0.27 )"
## [1] "95% Conf Int: ( 0.42 , 0.54 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -7.2041, df = 3922.9, p-value = 6.973e-13
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.3033627 -0.1735675
## sample estimates:
## mean of x mean of y
## 0.2397205 0.4781856
```

Test 3: Remis v Hybrid Hours Training MAE

```
my_KS_funct(Remis$graph_vector_hours$Model_MAE,
            Hybrid$graph_vector_hours$model_MAE)
```

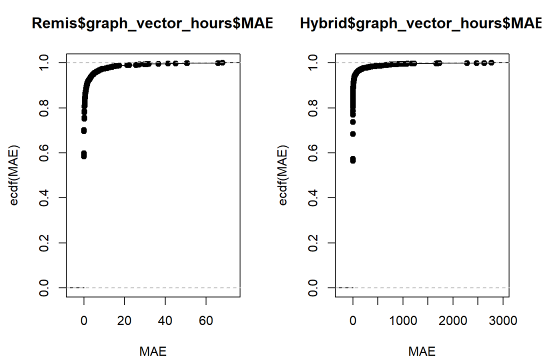




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  0.872  & sd:  2.097  & n:  1662"
## [1] "mean:  12.236  & sd:  38.908  & n:  2869"
## [1] "95% Conf Int: ( 0.77 , 0.97 )"
## [1] "95% Conf Int: ( 10.81 , 13.66 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -15.606, df = 2896.7, p-value < 2.2e-16
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -12.792148  -9.936429
## sample estimates:
##  mean of x  mean of y
##  0.8718051 12.2360934
```

Test 4: Remis v Hybrid Hours Test MAE


```
my_KS_funct(Remis$graph_vector_hours$MAE,
            Hybrid$graph_vector_hours$MAE)
```



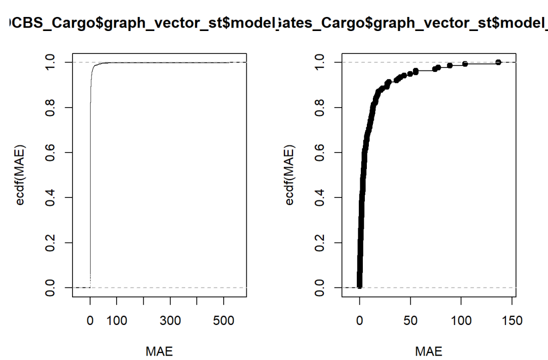

```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  1.019  & sd:  4.592  & n:  1662"
## [1] "mean:  21.977  & sd:  144.522  & n:  2869"
## [1] "95% Conf Int: ( 0.8 , 1.24 )"
## [1] "95% Conf Int: ( 16.69 , 27.27 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -7.7605, df = 2878, p-value = 1.166e-14
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -26.25252 -15.66224
## sample estimates:
## mean of x mean of y
##  1.019225 21.976603
```


Remis is better than Hybrid for hours. Remis is better than hybrid for Tails Training MAE.

Test 5: DCBS Cargo v Gates Cargo sTons Training MAE

```
my_KS_funct(DCBS_Cargo$graph_vector_st$model_MAE,
            Gates_Cargo$graph_vector_st$model_MAE)
```



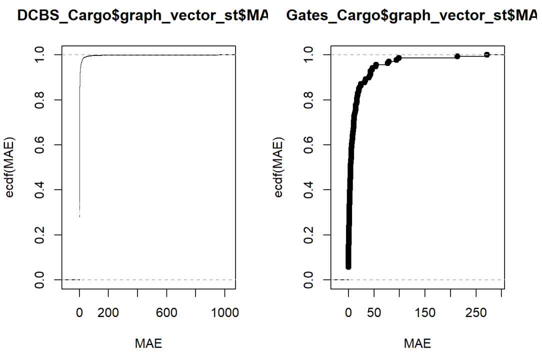

```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  1.645  & sd:  12.59  & n:  2153"
## [1] "mean:  11.086  & sd:  20.205  & n:  139"
## [1] "95% Conf Int: ( 1.11 , 2.18 )"
## [1] "95% Conf Int: ( 7.73 , 14.45 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -5.4415, df = 145, p-value = 2.195e-07
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -12.871273  -6.012386
## sample estimates:
## mean of x mean of y
##  1.644598 11.086427
```

Test 6: DCBS Cargo v Gates Cargo sTons Test MAE

```
my_KS_funct(DCBS_Cargo$graph_vector_st$MAE,
          Gates_Cargo$graph_vector_st$MAE)
```




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  2.605  & sd:  23.8  & n:  2153"
## [1] "mean:  14.765  & sd:  33.081  & n:  139"
## [1] "95% Conf Int: ( 1.6 , 3.61 )"
## [1] "95% Conf Int: ( 9.27 , 20.26 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

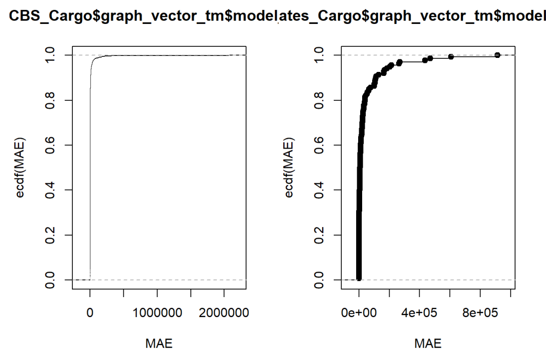
```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
```



```
## t = -4.2632, df = 147.37, p-value = 3.582e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -17.796898  -6.523288
## sample estimates:
## mean of x mean of y
##  2.604679 14.764772
```

Test 7: DCBS Cargo v Gates Cargo Ton Miles Training MAE

```
my_KS_funct(DCBS_Cargo$graph_vector_tm$model_MAE,
          Gates_Cargo$graph_vector_tm$model_MAE)
```




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  6347.15  & sd:  51667.057  & n:  2138"
## [1] "mean:  45640.288  & sd:  114674.236  & n:  139"
## [1] "95% Conf Int: ( 4157.04 , 8537.26 )"
## [1] "95% Conf Int: ( 26576.27 , 64704.31 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

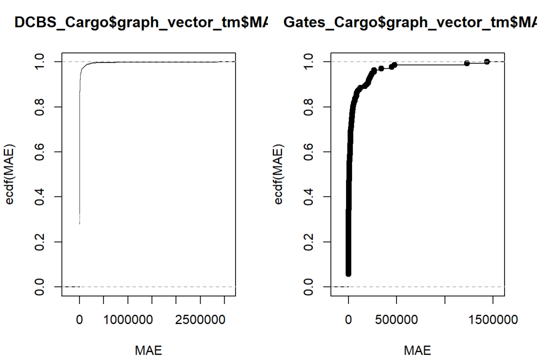
```
##
```



```
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -4.0134, df = 141.67, p-value = 9.659e-05
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -58647.53 -19938.75
## sample estimates:
## mean of x mean of y
##   6347.15  45640.29
```

Test 8: DCBS Cargo v Gates Cargo Ton Miles Test MAE

```
my_KS_funct(DCBS_Cargo$graph_vector_tm$MAE,
          Gates_Cargo$graph_vector_tm$MAE)
```




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  10005.194  & sd:  86261.663  & n:  2138"
## [1] "mean:  61084.52  & sd:  175962.945  & n:  139"
## [1] "95% Conf Int: ( 6348.66 , 13661.73 )"
## [1] "95% Conf Int: ( 31831.55 , 90337.48 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

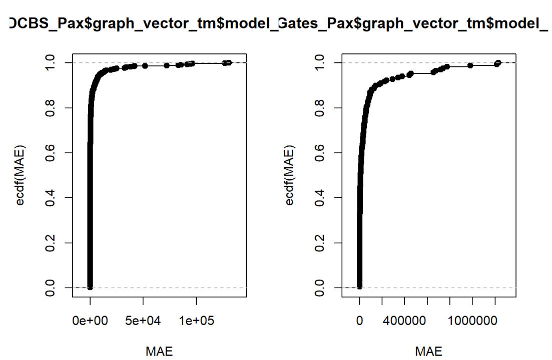


```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -3.396, df = 142.34, p-value = 0.0008866
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -80812.17 -21346.48
## sample estimates:
## mean of x mean of y
##  10005.19  61084.52
```

DCBS Cargo is better than Gates Cargo for sTons and ton miles.

Test 9: DCBS Pax v Gates Pax Ton Miles Training MAE

```
my_KS_funct(DCBS_Pax$graph_vector_tm$model_MAE,
          Gates_Pax$graph_vector_tm$model_MAE)
```



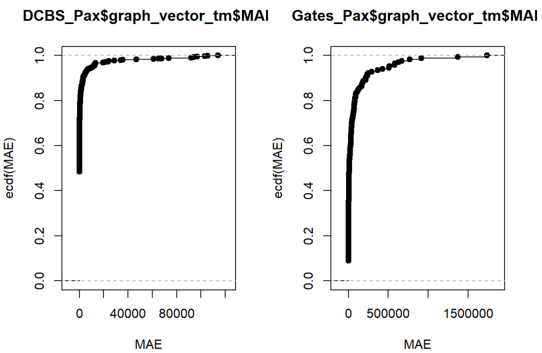

```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  2754.622  & sd:  12092.385  & n:  629"
## [1] "mean:  77484.396  & sd:  199120.334  & n:  168"
```

```
## [1] "95% Conf Int: ( 1809.6 , 3699.65 )"
## [1] "95% Conf Int: ( 47373.96 , 107594.83 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -4.862, df = 167.33, p-value = 2.661e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -105073.9  -44385.6
## sample estimates:
## mean of x mean of y
##  2754.622 77484.396
```

Test 10: DCBS Pax v Gates Pax Ton Miles Test MAE

```
my_KS_funct(DCBS_Pax$graph_vector_tm$MAE,
            Gates_Pax$graph_vector_tm$MAE)
```




```
## [1] "Different"
```



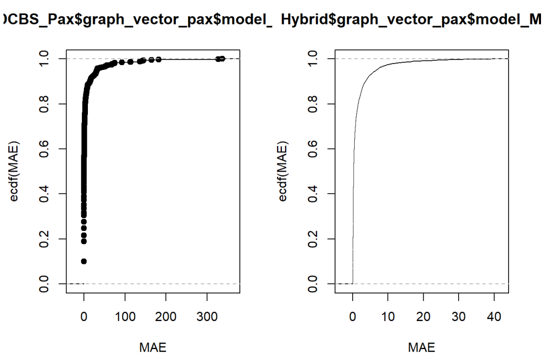
```
## [1] "p-value:  0"
## [1] "mean:  2703.712  & sd:  12304.515  & n:  629"
## [1] "mean:  88050.589  & sd:  220618.452  & n:  168"
## [1] "95% Conf Int: ( 1742.11 , 3665.31 )"
## [1] "95% Conf Int: ( 54689.27 , 121411.91 )"
## [1] "No overlap"
## [1] "Recommendation: 1"
## [1] "1 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = -5.0121, df = 167.28, p-value = 1.361e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -118964.7  -51729.1
## sample estimates:
## mean of x mean of y
##  2703.712 88050.589
```

DCBS Pax is better than Gates Paxfor ton miles.

Test 11: DCBS Pax v Hybrid Pax, Pax Training MAE

```
my_KS_funct(DCBS_Pax$graph_vector_pax$model_MAE,
            Hybrid$graph_vector_pax$model_MAE)
```

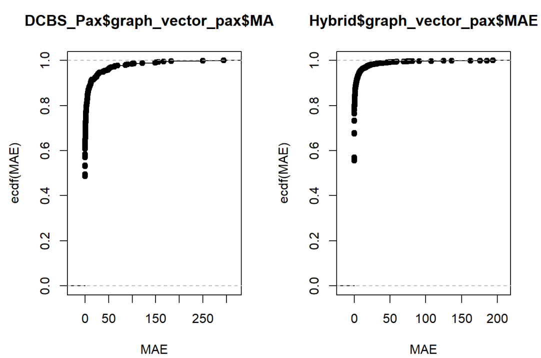




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  7.124  & sd:  26.605  & n:  631"
## [1] "mean:  1.395  & sd:  3.373  & n:  2863"
## [1] "95% Conf Int: ( 5.05 , 9.2 )"
## [1] "95% Conf Int: ( 1.27 , 1.52 )"
## [1] "No overlap"
## [1] "Recommendation: 2"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = 5.3999, df = 634.47, p-value = 9.438e-08
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  3.645804 7.812855
## sample estimates:
## mean of x mean of y
##  7.124143  1.394814
```

Test 12: DCBS Pax v Hybrid Pax, Pax Test MAE


```
my_KS_funct(DCBS_Pax$graph_vector_pax$MAE,
            Hybrid$graph_vector_pax$MAE)
```




```
## [1] "Different"
## [1] "p-value:  0"
## [1] "mean:  6.797  & sd:  24.884  & n:  631"
## [1] "mean:  1.963  & sd:  10.179  & n:  2863"
## [1] "95% Conf Int: ( 4.86 , 8.74 )"
## [1] "95% Conf Int: ( 1.59 , 2.34 )"
## [1] "No overlap"
## [1] "Recommendation: 2"
## [1] "2 is longer"
```

```
##
##  Welch Two Sample t-test
##
## data:  vect1 and vect2
## t = 4.7923, df = 677.12, p-value = 2.028e-06
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  2.853487 6.814691
## sample estimates:
## mean of x mean of y
##  6.796671  1.962583
```

Hybrid Pax is better than Gates Pax for pax.

Summary:

All databases are needed. All the means of the distributions are significantly different by two-tailed Welch's t-tests.

Remis is better than Hybrid for hours. Remis is better than hybrid for Tails Training MAE. DCBS Cargo is better than Gates Cargo for sTons and cargo ton miles. DCBS Pax is better than Gates Pax for pax ton miles. Hybrid Pax is better than Gates Pax for # pax. Only Gates Pax can do pax short tons.

Bibliography

1. United States Government Accountability Office, "Defense Logistics: DOD Needs to Improve Budget Estimates and Accuracy of Forecasts for Airlift Services," Tech. Rep. GAO-18-557, U.S. Government Accountability Office, 2018.

2. D. L. Hunter, "United States Transportation Command: 10 Years of Excellence, 1987 TO 1997," tech. rep., United States Transportation Command, 1997.

3. M. T. Small, *Predicting Global Disposition of U.S. Military Personnel via Open-Source, Unclassified Means*. Master's thesis, Department of Operational Sciences, 2018.

4. D. A. Hyder, *Liner Sustainment Workload Forecasting using Exogenous Data*. Master's thesis, Air Force Institute of Technology, 2019.

5. C. J. Bradshaw, *Contingency Workload Demand Forecast Techniques for Cargo and Flying Hours*. Master's thesis, Air Force Institute of Technology, 2016.

6. USTRANSCOM, "USTRANSCOM Contact Information," 2020, Retrieved February 3, 2020, from https://www.ustranscom.mil/cmd/contactus.cfm.

7. F. E. James Jr., *Forecasting Transportation Tonnage by an Adaptive Smoothing Model*. Master's thesis, Air Force Institute of Technology, 1973.

8. J. M. Park, *Optimizing Forecasting Methods for USTRANSCOM Railcar Demands*. Master's thesis, Air Force Institute of Technology, 2016.

9. D. Ardagna, C. Cappiello, W. Samá, and M. Vitali, "Context-aware data quality assessment for big data," *Future Generation Computer Systems*, vol. 89, pp. 548–562, 2018.

10. M. Abdallah, "Big Data Quality Challenges," in *Proceedings of the 2019 International Conference on Big Data and Computational Intelligence, ICBDCI 2019*, IEEE, 2019.

11. A. Panimalar, V. Shree, and V. Kathrine, "The 17 V's Of Big Data," *International Research Journal of Engineering and Technology*, vol. 4, no. 9, pp. 329–333, 2017.

12. IBM, "The Four V's of Big Data," 2020, https://www.ibmbigdatahub.com/infographic/four-vs-big-data.

13. L. L. Pipino, Y. W. Lee, and R. Y. Young, "Data Quality Assessment," *Communications of the ACM*, vol. 45, no. 4, pp. 211–218, 2002.

14. B. H. Archer, "Forecasting Demand and Intuitive Techniques," *International Journal of Tourism Management*, vol. March, pp. 5–12, 1980.

15. S. S. Jones, R. S. Evans, T. L. Allen, A. Thomas, P. J. Haug, S. J. Welch, and G. L. Snow, "A multivariate time series approach to modeling and forecasting demand in the emergency department," *Journal of Biomedical Informatics*, vol. 42, pp. 123–139, feb 2009.

16. A. Sfetsos and A. H. Coonick, "Univariate and Multivariate Forecasting of Hourly Solar Radiation with Artificial Intelligence Techniques," Tech. Rep. 2, 2000.

17. S. M. Phyoe, R. Guo, and Z. W. Zhong, "An Air Traffic Forecasting Study and Simulation," *International Journal of Science and Technology*, vol. 2, no. 3, pp. 55–69, 2016.

18. B. L. Bowerman, R. T. O'Connell, and A. B. Koehler, *Forecasting, Time Series, and Regression*. Belmont, Ca: Brooks/Cole, 4 ed., 2005.

19. G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. Hoboken, New Jersey: John Wiley & Sons, Inc., 5 ed., 2016.

20. R. J. Hyndman and G. Athanasopoulos, *Forecasting: Principles and Practice*. OTexts, Middletown, DE, second ed., 2018.

21. W. Y. Peng and C. W. Chu, "A comparison of univariate methods for forecasting container throughput volumes," *Mathematical and Computer Modelling*, vol. 50, no. 7-8, pp. 1045–1057, 2009.

22. M. M. U. Thompson, "JDPAC Forecasting Sync," tech. rep., USTRANSCOM/TCAC-O, 2019.

23. J. M. U. Schofield, "USTRANSCOM Workload Forecasting," tech. rep., USTRANSCOM/TCAC-O, 2016.

24. R. J. Hyndman, "Facts and Fallacies of the AIC," 2013, Retrieved February 3, 2020, from https://robjhyndman.com/hyndsight/aic/.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704–0188

1. REPORT DATE *(DD–MM–YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From — To)*
26–03–2020	Master's Thesis	August 2018 — March 2020

4. TITLE AND SUBTITLE

Database Analysis to Improve U.S. Transportation Command Forecasting Processes

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Thompson, Maxwell C., 2nd Lt, USA

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
WPAFB OH 45433-7765

8. PERFORMING ORGANIZATION REPORT NUMBER

AFIT-ENS-MS-20-M-176

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

USTRANSCOM/JDPAC
Mr. Reed Jorgenson, JDPAC Futures Div.
508 Scott Drive
Scott Air Force Base, IL 62225-5357
(618) 220-5139 (reed.w.jorgenson.civ@mail.mil)

10. SPONSOR/MONITOR'S ACRONYM(S)

USTRANSCOM, JDPAC

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 United States Transportation Command (USTRANSCOM) facilitates air, land, and sea transportation for the DOD. On a periodic basis, a myriad of different agencies within USTRANSCOM project future workload to facilitate resource planning, budgeting, and reimbursable rate identification. Within USTRANSCOM, there are a variety of databases and metrics utilized for workload forecasts; neither a standard nor a preferred technique is prescribed. Currently, USTRANSCOM faces challenges in producing accurate workload forecasts. These challenges can lead to unreliable budget requests and, ultimately, hinder the effectiveness and efficiency of USTRANSCOM. For the purpose of routine aircraft movements of cargo and personnel, this research seeks to answer (1) whether any data sets are dominated with respect to data quality, allowing for their removal from consideration and (2) the degree to which any data set is superlative with respect to informing high quality air workload forecasts. Furthermore, this research identifies a possible major problem with USTRANSCOM's current forecasting procedure and provides recommendations on how to best utilize the data sets readily available for use.

15. SUBJECT TERMS

Database analysis, forecasting

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Brian J. Lunday, AFIT/ENS
U	U	U	UU	65	19b. TELEPHONE NUMBER *(include area code)* (937) 255-3636, x4444; brian.lunday@afit.edu