



**ANALYSIS OF BENCHMARK FORECASTING MODELS VERSUS A RECURRENT
NEURAL NETWORK FORECAST OF CHANNEL CARGO DEMAND**

Graduate Research Paper

Abraham N. Umanah, Major, USAF

AFIT-ENS-MS-22-J-056

**DEPARTMENT OF THE AIR FORCE
AIR UNIVERSITY**

AIR FORCE INSTITUTE OF TECHNOLOGY

Wright-Patterson Air Force Base, Ohio

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

The views expressed in this Graduate Research Paper are those of the author and do not reflect the official policy or position of the United States Air Force, Department of Defense, or the United States Government. This material is declared a work of the U.S. Government and is not subject to copyright protection in the United States

AFIT-ENS-MS-22-J-056

ANALYSIS OF BENCHMARK FORECASTING MODELS VERSUS A RECURRENT
NEURAL NETWORK FORECAST OF CHANNEL CARGO DEMAND

Graduate Research Paper

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 Management

Abraham N. Umanah, BS, MA,

Major, USAF

June 2022

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

AFIT-ENS-MS-22-J-056

ANALYSIS OF BENCHMARK FORECASTING MODELS VERSUS A RECURRENT
NEURAL NETWORK FORECAST OF CHANNEL CARGO DEMAND

Abraham N. Umanah, BS, MA,

Major, USAF

Committee Membership:

Bruce A. Cox, PhD
(Primary Research Advisor)

Abstract

Given the longstanding history of inaccurate cargo demand forecast and underutilization of contracted commercial airlift augmentation, the primary objective of this paper is to present a simple forecasting model developed using Python. This study compares benchmark time series forecasting models against a Recurrent Neural Network model to determine which model is best for predicting monthly channel cargo airlift demand. To determine which model produces the best forecast, univariate time series analysis is conducted on Integrated Development Environment/Global Transportation Network Convergence data using readily available statistical modules and Machine Learning algorithms within the Python ecosystem. Of the models examined, the Seasonal Auto Regression Integrated Moving Average model performed the best and was recommended for use at the Global Cargo Management level. The research concludes with recommendations for future research that includes aircraft type on specific routes using a multivariable time series analysis.

AFIT-ENS-MS-22-J-056

To my son

Acknowledgements

I would like to extend my deepest gratitude to my advisor, Dr. Cox, for investing countless hours into my academic journey and ensuring my on-time completion of this research endeavor. I am indebted to Lt Col Thompson for providing the data used in this research and for making time to answer all my data related questions. I am also grateful for Dr. Reiman's initial feedback on my literature review and for Mr. Don Anderson's guidance on the direction of this paper.

Abraham Umanah

Table of Contents

Abstract	iv
Acknowledgements	vi
List of Figure	ix
List of Equations	x
I. Introduction	1
Background	1
Problem Statement	6
Objective	7
Assumptions	7
Limitations	8
II. Literature Review	9
Chapter Overview	9
Forecasting in the Air Force	9
Civil Reserve Air fleet	10
Forecasting Research Over Time	10
Current Forecasting Methods	16
III. Methodology	19
Chapter Overview	19
Data Processing	19
Univariate Time Series Forecasting - ARIMA	20
Differencing	20
Autoregression Model	21
Moving Average Model	22
ARMA Model	23
ARIMA Model	23

SARIMA Model.....	24
Long Short-Term Memory (LSTM) Recurrent Neural Network.....	25
Measures of Forecast Performance	28
IV. Results and Analysis	30
Chapter Overview	30
Python	30
Processing	31
ARIMA	33
SARIMA	36
LSTM.....	38
V. Conclusion	40
Chapter Overview	40
Conclusion from Research.....	40
Recommendations.....	41
Areas for Further Research	41
Appendix A: Abbreviations	42
Bibliography	44

List of Figures

Figure

1: Commercial Airlift Process Flow	5
2: Sample Long Short-Term Memory Unit	26
3: Raw Data.....	31
4: Cleaned and Aggregated Data	31
5: Seasonal Decomposition.....	32
6: Train Dataset Partial Autocorrelation Plot.....	33
7: Train Dataset Autocorrelation Plot	33
8: ARIMA Model Results	34
9: ARIMA Model Predicted vs Test Data.....	34
10: ARIMA Model Diagnostics.....	35
11: SARIMA Model Results	36
12: SARIMA Model Predicted vs Test Data	37
13: SARIMA Model Diagnostics	37
14: LSTM Model	38
15: Loss Per Epoch	38
16: LSTM Model Predicted vs Test Data	39

List of Equations

Equation

1: Autoregression Model.....	21
2: Moving Average Model.....	22
3: ARMA Model	23
4: ARIMA Model	23
5: SARIMA Model	24
6: Mean Absolute Error	28
7: Root Mean Square Logarithmic Error	28
8: Mean Absolute Percent Error	29

ANALYSIS OF BENCHMARK FORECASTING MODELS VERSUS A RECURRENT NEURAL NETWORK FORECAST OF CHANNEL CARGO DEMAND

I. Introduction

Background

In today's competitive global environment with increasing multidomain threats from near peer adversaries, advantages must be sought within all facets that constitute national power. In a future fight, the margin of error across all military operations will become very thin. One way to edge out competition is to become more efficient in some operations to allow for less efficiency in operations where speed of delivery is of higher priority. This stance aligns with the 2018 non-classified Summary of the National Defense Strategy (NDS) which highlights the need to "deliver performance at the speed of relevance" by eliminating the "wasteful application of resources." (DoD, 2018). The NDS also concedes that success is no longer assured to the nation with the newest technology rather, success tilts towards the nation that can rapidly and seamlessly integrate new technologies for the changing character of warfare (DoD, 2018).

Furthermore, when assessing national power indicators, excessive spending overestimates military capability. As an example, the National Intelligence Council's use of military spending, GDP, etc., as part of an index to evaluate "global power trends" would result in an overestimation of military power if there is excessive spending within the military (Beckley, 2018). An insidious consequence of this overestimation of power indicators appears in strategic plans that are not in sync with available resources.

Beyond the strategic implications of real power estimation, military spending is of great interest to congress. On 6 July 2017, House Report 115-200 was published to accompany the National Defense Authorization Act for Fiscal Year 2018. This House Report was direct

communication from the Committee on Armed Services to the Comptroller General of the United States requesting the submission of “a report to the congressional defense committees by August 1, 2018, regarding allocation of the Airlift Readiness Account (ARA) in conjunction with the Transportation Working Capital Fund (TWCF) during fiscal years 2007 to 2016, inclusive” (H. Rept. 115-200). This led the Government Accountability Office (GAO) to publish its findings in a Report to Congressional Committees titled “DOD Needs to Improve Budget Estimates and Accuracy of Forecasts for Airlift Services.” The report highlighted Department of Defense (DoD) forecasting shortfalls and key relationships between the United States Transportation Command (USTRANSCOM), U.S. Air Force’s Air Mobility Command (AMC) and commercial airlines. Specifically, the report highlighted the relationship between USTRANSCOM’s management of the TWCF and AMC’s execution of the ARA.

USTRANSCOM manages the TWCF, a revolving fund, and uses it to cover the cost of providing a suite of transportation services to include airlift (Government Accountability Office [GAO], 2018). To stay competitive, USTRANSCOM sets its rates, those covered by the TWCF, below cost and the rest of the expenses are covered by the Airlift Readiness Account (GAO 18-557).

The proportion of transportation costs covered by the TWCF and ARA is a direct function of USTRANSCOM’s workload or consumer demand forecast (GAO-18-557). USTRANSCOM’s workload forecasts informs the agency’s determination of rates. More pertinent to this paper, consumer demand forecast dictates how much ARA funds USTRANSCOM must request from the Air Force for the next fiscal year. The ARA is a DoD mandate codified to ensure that USTRANSCOM always has seasoned military aircrew and proper equipment to adequately provide airlift capacity that will meet national security interests around the globe (DoDFMR Vol 2B, 9-17). The Air Force manages the ARA as part of its

Operations and Maintenance (O&M) annual budget request under the Airlift Operations subactivity group and must rely on an accurate and timely ARA forecast from USTRANSCOM (DAF FY21 Budget Estimates, 331). Understandably, workload forecasts include quantitative analysis, direct interaction with the consumer and other operations-based adjustments (GAO -18-557).

This paper focuses on quantitative analysis of channel cargo demand as a viable option to reduce historical inaccuracies in forecasting. Though some forecast variance is expected due to emerging requirements, grossly inaccurate forecasts of consumer demand will lead to inefficient execution of airlift, unreliable budget request and in some cases, defunding other programs (GAO-18-557).

In full transparency, GAO's website reflects that DoD has implemented all the recommendations of GAO-18-557. Specifically, the United States Air Force (USAF) started reporting specifically requested ARA information on its O&M budget requests; USTRANSCOM and the Air Force Financial Management Budget Office (SAF/FMB) formed an official partnership aimed at meeting budget submission requirements in a timely fashion; USTRANSCOM overhauled its workload forecast instructions to create a more robust conduit for operational input from services and Geographic Combatant Commands and, developed metrics to evaluate forecast accuracy. USTRANSCOM also established a Transportation Workload Forecast Working Group (TWWFG) that convenes monthly with one of its main charters being to continuously finetune the workload forecast (GAO).

Study of USTRANSCOM's organizational structure reveals that the Joint Distribution Process Analysis Center (JDPAC) is the analysis and engineering center at the core of implementing workload forecast improvements. In response to GAO's findings, JDPAC

employed powerful new technologies that leverage cutting edge industry practices that include analysis using artificial intelligence (AI) and machine learning (ML). One example of JDPAC's tools is a software called Demand Guru by LLamasoft, a subsidiary of Coupa. According to Coupa, Demand Gurus is a "visually interactive tool for causative and life cycle demand modeling and prediction" (Coupa). According to JDPAC's Senior Mobility Analyst, the software is a commercial off the shelf (COTS) program utilizing ensemble forecasting methodology to power statical methodologies and produce a forecast. Specifically, the software uses ML to optimally determine if autoregression integrated moving average (ARIMA), autoregression integrated moving average with exogenous input (ARIMAX), exponential smoothing, model averaging, naive forecasting, quantile random forest, simple moving average, stochastic gradient boosting, or linear support vector machine (SVM) is the best fit for any given time series dataset.

Although USTRANSCOM has access to this very powerful tool, the software itself is for the most part limited to personnel within the combatant command. Additionally, although USTRANSCOM has written instructions centralizing the workload forecast process, along with JDPAC's Workload Forecasting Overview document, both documents are still drafts not meant for implementation nor compliance. Further, commercial airlift augmentation purchases are accomplished by AMC's Global Cargo Management team, and it is not clear that they have Demand Guru usage rights nor that they are completely plugged in to USTRANSCOM's new forecasting policies. The Global Cargo Management team views USTRANSCOM's Command Forecast as a strategic guidance within which to operate rather than a product to employ operationally during commercial airlift augmentation purchases. Phone calls with some Global Cargo Management personnel highlight that commercial airlift augmentation forecasts are

created internally from analysis of historical data, demand signals from combatant commanders through Distribution Lane Validation, and inputs from USTRANSCOM, AMC/A3 and TACC leadership. A forecast is then pushed through the Commercial Operations Integrated System (COINS) to the Contracting Division. Figure 1 provides a high-level overview of the commercial augmentation purchase process. The analysis that occurs immediately before purchase submission in COINS is where the analysis presented in this paper could pay the most dividends.

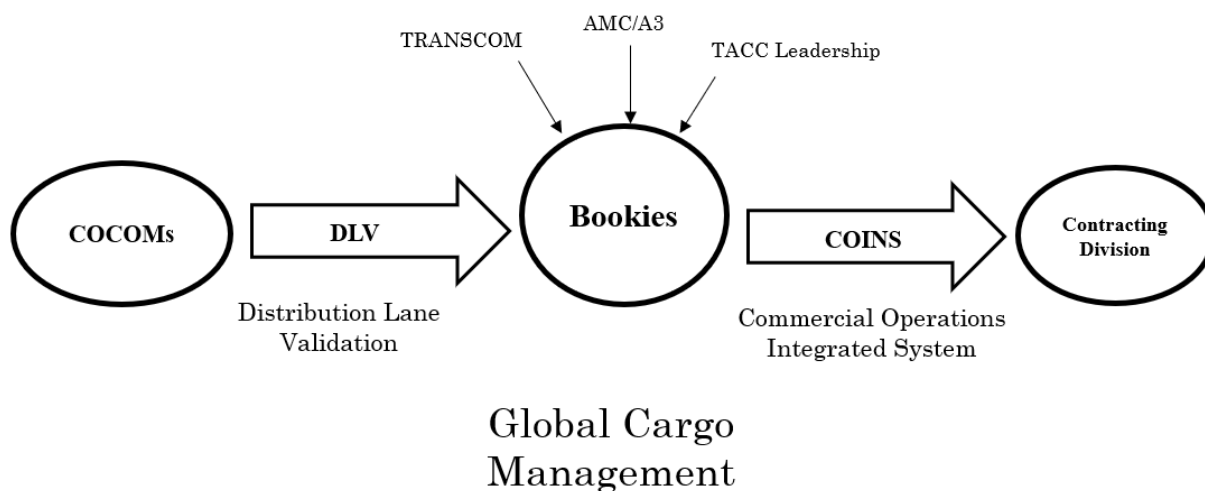


Figure 1: Commercial Airlift Process Flow

GAO documents show a long history of inefficient utilization of airlift assets. In 1983, the GAO published NSIAD-83-19 citing unused cargo capacity on returning legs of Military Airlift Command (MAC) missions. The report estimated that if the empty space was fully maximized, the DoD would save “as much as \$3.4 million from October 1, 1980 to March 31, 1982” (NSIAD-83-19). To mitigate this issue, GAO recommended effective management of cargo movement. In 1992 GAO/NSIAD-92-263 highlighted “unfilled capacity” on gray tails flying overseas missions even when the DoD was simultaneously paying for commercial

movement of lower priority cargo. If the problem was fixed, GAO claimed DoD would save \$21 million per year. Though previous analyses were on utilization of military aircraft, GAO-18-557 took a more robust and wholistic approach in determining a flawed workload forecasting process which had ramifications on budget requests and commercial airlift augmentation contracts. NSIAD-83-19, NSIAD-92-263 and GAO-18-557 clearly establish an enduring problem with inefficient utilization of airlift assets. While there are certain mission sets where success far outweighs efficiency, peacetime commercial airlift augmentation of channel cargo movement is not one of them. Though USTRANSCOM has a powerful piece of technology, in the Demand Guru Software, to address the shortfalls highlighted in GAO-18-557, it is not available to all the key players who purchase commercial airlift.

Problem Statement

The continued lack of robust consumer demand forecasting models and ensuing projection inaccuracies exacerbate the problem of underutilization. Specifically, these inaccurate forecasts result in over committing to contracts for commercial augmentation airlift through Fixed Award contracts. A quick analysis of AMC/A9's 2021 Boeing 747 channel cargo data shows a 40.77% utilization rate. The utilization rate for each B747 contracted for 100 short tons of cargo is, on average, 40.77 short tons of cargo. Though USTRANSCOM is motivated to provide incentives for commercial partners to participate in the Civil Reserve Air Fleet (CRAF) program, flying around partially loaded aircraft is simply a waste of resources. Additionally, in those situations where the allotted ARA is exceeded, extra funds to cover TWCF shortfalls are taken out of the O&M funds. While there are benefits to a generous estimation of annual fix buy contracts, increasing global challenges within a resource constrained budget environment highlight the need for savings where possible.

Objective

Given the longstanding history of inaccurate consumer demand forecast, the primary objective of this paper is to present a simple forecasting model developed using a free programming language (Python). Through Python's data analysis libraries, technicians have access to both traditional statistical techniques and ML models. The resulting forecasting model could assist by feeding into the overall analysis at the Global Cargo Management level. To determine which model produces the most suitable forecast for the given dataset, this paper compares the forecasting accuracy of traditional statistical models against a ML-recurrent neural network (RNN) model to predict future customer demand for channel cargo missions using historical demands.

Assumptions

Several assumptions were required while developing the models presented in this paper. The first assumption is that the aggregated and sanitized data received from JDPAC accurately accounts for the entirety of actual channel cargo moved by both military and commercial airlift. The implication is that there are no significant input and collection errors that could affect the results of the model.

The second assumption is that the dataset has monthly seasonality. This assumption allows for the aggregation of multiple channel cargo missions into monthly totals enabling monthly predictions. This assumption implies that the dataset is not stationary where the cumulative distribution function of a stochastic process is independent of time. As such, distribution parameters such as the variance and mean are not constant over time.

Despite the assumption of seasonality, the next assumption is that monthly channel cargo demand is stable with no significant deviations or outliers and, actual cargo airlift demand is known for at least three months prior to the prediction. This assumption is necessary for the recurrent neural network model. The fourth assumption is that city pairs in the current dataset will not change and is representative of city pairs to be used in the future. Lastly, this model assumes peacetime steady state operations.

Limitations

The research presented in the paper is not without limitations. The first limitation is that this research and the resulting models focus solely on channel missions. Additionally, this research uses univariate time series analysis with no consideration for customer signals, additional causative factors or any other exogenous inputs. The models use historical and aggregated monthly tonnage to predict future cargo movement in tons per month. All models produced from this research do not account for macroeconomic factors and geopolitical events that could affect cargo movement. During data processing, categorical independent variables were not used to represent unique events like holidays, natural disasters, or significant geopolitical events. Further this research does not delineate between cargo carried by military airlift versus cargo carried by commercial airlift. Monthly cargo prediction represents total channel cargo moved across the system by both military and civilian aircraft. Lastly, cargo weight is not tied to a specific airframe nor an exact origination and destination.

II. Literature Review

Chapter Overview

To get a full grasp of any problem set and, in turn, understand how to make meaningful contributions, a thorough review of the existing body of literature is required. This section starts off with a brief history of strategic forecasting in the Air Force before scoping down to specific forecasting methodologies. Since the focus of this paper is to fine tune channel cargo forecasts to right size commercial airlift augmentation contracts, a brief history of the CRAF and commercial airlift augmentation is also presented. What follows is a chronological review of techniques that could be useful or, that have been employed as solutions to the forecasting problem set. This type of literature review is necessary to avoid duplicating any previously completed works and to help define a path for this paper. Additionally, the review helps to illuminate hidden relationships and trends between variables that may seem unrelated. Finally, an overview of current airlift forecasting methods is presented.

Forecasting in the Air Force

The value of forecasting in the Air Force goes all the way back to the “Toward New Horizons” and “Wood Hole Summer” studies in the 1950s. These forecasts were very broad and strategic in scope and sought to chart a path for the service’s technological focus (Gorn, 1988). Like the forecasts discussed in this research, these historical technology forecasts evolved over time charting newer paths for the Air Force. After the drawdown in Vietnam and the Middle East Oil Embargo of 1973, the Air Force further morphed its forecasting requirements to mitigate severe scarcity in resources (Gorn, 1988). Today, the Air Force continues to operate in a resource constrained environment at a time when the nation is strategically pivoting to compete against peer adversaries (Mills et al., 2017).

Civil Reserve Air Fleet

On August 22, 2021, Secretary of Defense Lloyd J. Austin III ordered USTRANSCOM to activate the CRAF (DoD PR, 2021). Though the activation was for a small fraction of the entire CRAF fleet involving only Delta, American, Atlas, Omni, Hawaiian and United airlines, it marked only the third activation in the program's history (DoD PR, 2021). The CRAF derived from the Defense Production Act of 1950 and served as a wartime ready augmentation program as defined by 32 CFR, Section 243.3 (Cornell Law School). Those U.S. air carriers that volunteered to participate in the CRAF entered a contract with the DoD (DoT). The contractual agreement to fly peacetime cargo, however, is based on forecasted demand by DoD planners (GAO). It is this demand forecast that is the focus of this research.

Forecasting Research Over Time

The earliest airlift research reviewed was a document titled "A Methodology for Accurately Predicting Demand for Airlift of Military Cargo to Overseas Destinations" published in 1966 by William E. Billion and Lawrence G. Regan of the Research Analysis Corporation (RAC) (Billion & Regan, 1966). During the time of the research, the C-5 was being brought on to the Air Force's fleet as a means of executing rapid wartime contingency deployment of troops. However, the DoD also recognized the aircraft's capacity for peacetime movement and as a result, hired the RAC to measure the economic benefit of using organic airlift. In the paper, the authors reject time series extrapolation for an economic utilization of airlift model where the elasticity of demand at various air ton-mile rate was calculated. Essentially, the authors built a demand curve based on total distribution cost of delivering cargo overseas versus movement of Air Force priority cargo. The model was developed by correlating what they determined to be "economically air-eligible commodities" with known volume and traffic flow data to find a

“break-even air rate” (Billion & Regan, 1966). Given that the number of aircraft flying and cargo airlifted today is exponentially greater and more varied than in 1966, replicating this type of analysis would quickly exceed the resources and scope of this research.

In the next paper, Morton et al. construct an objective function based on relevant decision variables and constraints to optimize strategic airlift asset utilization. More precisely, the authors expanded on the Mobility Optimization Model (MOM), which came from the 1991 Mobility Requirement Study (MRS), and THRUPUT Model to develop an optimization model formulated as a “multi-period, multi-commodity network-based linear programming model” (Morton et al., 1996). The researchers’ intent was to answer specific scenario-based questions that give planners the necessary insight to optimally employ strategic airlift assets. They analyzed increases in known level of demand and sought to identify how resources must be employed to meet those increases in demand. Ultimately the research argues that optimization models have the advantage over simulations in answering “what’s-best” questions (Morton et al., 1996).

Surdyk (2010) highlighted the inability of a major transload port to forecast follow-on destinations for arriving pallets beyond 2-3 days. Historical pallet data from the Global Air Transportation Execution System (GATES) along with aircraft type were used to create weekly forecasts. The research sought to extend this forecast window for follow-on sustainment cargo beyond 3 days and possibly create a tool available for broad application to other Time Phased Force Deployment Data (TPFDD) movements. The research utilized Microsoft Excel to develop several statistical forecasting models including Naïve Forecast, Moving Averages, Simple Exponential Smoothing, Holt’s Method, Damped Trend Forecast, Winter’s Method, Multiple Regression, Adaptive Rate Exponential Smoothing and three ‘Novel’ models developed specifically for the research. The research concluded that all models performed better than time

series forecasts after aircraft type was incorporated into forecasting transloading operations through major nodes (Surdyk, 2010).

In “A Practical Guide to Volatility Forecasting Through Calm and Storm” the researchers reviewed forecasting in the field of risk analysis and provided a richer understanding of forecasting techniques across various domains. The research compared five volatility models within the autoregressive conditional heteroskedasticity (ARCH) class of models to determine if the performance of the models is affected by “choices for estimation window length, innovation distribution, and frequency of parameter reestimation” (Brownlee et al., 2012, p. 3). The research concludes that model performance is not sensitive to forecasting horizons and, short term prediction windows hold up during periods of turmoil or volatility. Volatility forecasting could prove valuable if properly adapted in efforts to forecast airlift demand during contingencies or tumultuous periods around the globe.

DeYoung (2012) researched the use of time series forecasting to develop long-range outlooks that CRAF partners could use, in the context of a contracted economy, for their decision-making processes. The scope of the research was limited to missions with an aerial port of debarkation (APOD) in Iraq, Afghanistan or a select few locations within the United States Central Command (USCENTCOM). DeYoung found that seasonal exponential smoothing and sometimes, Autoregressive Integrated Moving Average (ARIMA) models yielded high R squared values and, when applied to disaggregated AOR data, produced results with very low comparative sum of squared error values (DeYoung, 2012). The analysis was conducted on data from the GATES database.

For a more complete approach, Leonard (2013) integrated Monte Carlo simulation, time series forecasting techniques and heuristic optimization with operational planning to yield short

term (1 month) and long term (1 year) forecasts. The research proposes that more accurate forecasts allow for greater fixed buy contracts which both incentivizes CRAF partners and allows flexibility with monthly and quarterly commercial airlift augmentation purchases. Since it is cheaper to cancel quarterly and monthly buys than to purchase extra airlift capacity during the year, the research argues there is an overall cost savings. The research used excel based TACC Strategic Airlift Simulation (TACCSAS) and JMP to analyze data from GATES (Leonard 2013).

In “The Myth of Strategic and Tactical Airlift” Maywald et al. (2017) offered their take on optimal airlift planning. Using a mathematical model called Aircraft Selection Model (ASM), a given amount of cargo is transported via selected city pairs using the optimal selection of available aircraft. While fuel savings is the primary objective of the model, the research highlights variables that influence overall system cargo airlift capacity. The research does not seek to forecast cargo demand, instead it provides an optimal mix of aircraft for a given level of cargo requirement with fuel savings as the objective. Along the same lines of broad system wide analysis, Small (2018) in “Predicting Global Disposition of U.S. Military Personnel via Open-Source Unclassified Means,” analyzed U.S. military personnel total end strength in support of USTRANSCOM across the globe. The research developed exponential smoothing, ARIMA and ARIMAX models in the programming language of R to forecast personnel end strength (Small, 2018).

Beyond modeling, “Aligning Incentives in the Transportation Working Capital Fund” highlights the ramifications of inaccurate workload estimates that affect rate setting and in turn, influence airlift demand (Connor et al., 2019). Admittedly, the focus of this research was not on demand forecasts, instead the research recommended pricing strategies that influence customer

transportation demand. The value, however, is that the research contextualized the complexity of airlift demand. Specifically, it highlighted variables that may not be readily obvious and how those variables affect cost which ultimately affects demand. For example, there are USTRANSCOM missions where cost minimization or cost efficiency are not desired, and charges should be below cost because those missions are intrinsically beneficial to wartime readiness. The research begins with defining USTRANCOM's scope of responsibilities consisting of airlift, sealift, and surface transportation alongside the management of multi-modal ports across the globe. The authors then illustrate the continuum of USTRANSCOM's operations from peacetime to wartime and how some peace time operations are crucial to wartime readiness. Analysis is then conducted on available data including literature on industry standards, budget, and cost data along with stakeholder interviews. Ultimately, the lessons gleaned are assessed through a series of case studies leading to the development of recommendations that help USTRANSCOM favorably influence customers' choices about their transportation needs. The first recommendation was recovery of fixed rate costs through means other than charging customers and the second recommendation was to charge the lowest rate possible for those missions that USTRANSCOM prioritizes for wartime readiness. Both recommendations tie back to workload forecast since deliberate and specific rate setting rely on accurate forecasts (Connor et al., 2019).

The next paper, like others previously reviewed, directly addresses the problem of forecasting sustainment of channel cargo airlift demand. In "Channel Mission: Forecasting Sustainment Cargo Requirements", Hemken (2019) developed four models to forecast weekly, monthly, and quarterly cargo demand in the pacific region in an attempt to improve forecasted workload demand and increase savings. Specifically, the research develops an ordinary

regression, ARIMA, Seasonal Exponential Smoothing and Winters' Method models in JMP using GATES data. For scope, the research uses specific commodity codes to isolate the precise cargo of interest transported on commercial airlift. The research finds that quarterly model performed best but recommends monthly models due to high data fidelity, practical application of forecast, and prediction consistency.

A more granular take on airlift allocation and demand is found in "Applying the Multiple Multidimensional Knapsack Assignment Problem to a Cargo Allocation and Transportation Problem with Stochastic Demand" (Maus, 2019). In this paper, the researcher adapts the Multiple Multidimensional Knapsack Assignment Problem (MMKAP) to a notional airlift network with supply and demand nodes complete with categorization of ferried items and aircraft types based on historical cargo movements. Once the optimal solution was found, the research introduced exogenous and stochastic demand factors to determine what demand expectations or forecasts can be planned to minimize fixed and variable cost penalties. This research highlighted applications of a combinatorial optimization model, with MMKAP, to solve cargo allocation and transportation with calculated fuel and distance estimations. Ultimately the research identified that there is a level of expected demand that, when planned, can minimize cost (Maus, 2019).

While Maus' research involved complex applied mathematical models beyond the scope of this research, the next reviewed work involved the use of multiple linear regressions against 5 years' worth of data to forecast augmented cargo and passenger services on channel missions (Callahan, 20). Data for the model was downloaded from the Global Decision Support System (GDSS) and COINS. The research utilized regression models to forecast sorties flown per year, specific city pair routes flown, and short tons airlifted. The research concluded with at least one

recommendation for AMC to adopt multiple linear regression in workload forecast efforts (Callahan, 20).

Data analysis plays a key role in ascertaining results or key insights that may not be readily evident. To achieve reliable results, the quality of data should be assessed. This is the focus of “Database Analysis to Improve U.S. Transportation Command Forecasting” (Thompson, 2020). The research explored four databases: GATES, Defense Enterprise Accounting and Management System’s Component Billing System (DCBS), Reliability and Maintainability Information System (REMIS) and a hybrid of GDSS and GATES. These databases were assessed for volume, completeness, representation, ease of use and relevancy. Further, the research emulated JDPAC’s forecasting processes as one way to test whether any of the databases was superior in data quality for forecasting channel airlift workload. The research revealed that no one dataset was significantly superior. The research also highlighted that use of the Akaike Information Criterion (AIC) as a discriminator when comparing different models is illogical.

Current Forecasting Methods

Most of the literature review so far has covered traditional statistical methods of forecasting including exponential smoothing, and ARIMA. Exponential smoothing assigns exponential decreasing weights such that as the data gets older it has less weight. Specifically, the Winter’s Three Parameter Linear and Seasonal Exponential Smoothing models account for increasing and decreasing trends in addition to seasonality in data. To account for intermittent and sporadic demand signal, a modification of the Exponential smoothing model known as the Croston model was reviewed. The Croston model is used to predict average demand volume and the average time between demand known as intermittent demand. On the other hand, ARIMA

forecasting method relies on the autoregressive, integrated and moving average parts of the data being analyzed.

Since this research also employs ML, literature on ML forecasting was also reviewed. In “A Comprehensive Literature Review of the Demand Forecasting Methods of Emergency Resources From the Perspective of Artificial Intelligence” the authors combed through 1,235 academic papers dealing with forecasting methodologies and applications, between 1980 and 2018 (Zhu et al., 2019). The research into existing papers led to the categorization of existing papers into time series analysis, case-based reasoning, mathematical models, information technology and the conclusion that there is a need for more dynamic demand prediction adaptable to emergency situations (Zhu et al., 2019). "Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward" takes a more direct approach in its evaluation and comparison of ML forecasting methods against traditional statistical forecasting models (Makridakis et al., 2018). The authors analyzed several ML models including Multi-Layer Perceptron (MLP), Bayesian Neural Network (BNN), Radial Basis Functions (RBF), Generalized Regression Neural Network (GRNN)/Kernel regression, K-Nearest Neighbor regression (KNN), CART regression trees (CART), Support Vector Regression (SVR), Gaussian Processes (GP), RNN and Long Short Term Memory network (LSTM) using symmetric Mean Absolute Percentage Error (sMAPE) and Mean Absolute Scaled Error (MASE) as accuracy measures (Makridakis et al., 2018). The authors found MLPs to be the best ML model for one step ahead forecasts followed by Bayesian Neural Networks (BNNs) but, also demonstrated that ML models performed worse than simpler statistical models.

The most impactful literature review on ML and RNNs was “Recurrent Neural Network for Time Series Forecasting: Current Status and Future Directions” (Hewamalage et al., 2021).

The authors began by illustrating the historical perspective of neural networks (NNs) and speculated on reasons why NNs may have been perceived as underperforming established benchmarks such as ARIMA. Specifically, the research provided an overview and categorization of relevant literature, empirical analysis of popular RNN architectures, and best practices and guidelines for RNN modeling. The research concludes that when the series in a “dataset follow homogenous seasonal pattern, with all of them covering the same duration in time at sufficient lengths, RNNs are capable of capturing the seasonality without prior deseasonalization” (Hewalage et al., 2021). Overall, the research demonstrates, empirically, that RNNs “outperform statistical benchmarks that are currently the state-of-the-art” in many forecasting scenarios (Hewalage et al., 2021).

III. Methodology

Chapter Overview

This chapter outlines key concepts used to complete this research. What follows is a description of the origin, scope, and steps taken to process the data. An explanation and summary of each model used for univariate time series forecasting is also presented. Lastly, the chapter describes measures of performance used to determine which model is most suitable.

Data Processing

The data for this research is a hybrid of information from GATES and GDSS and was downloaded from the Integrated Development Environment/Global Transportation Network Convergence (IGC) database. JDPAC provided the data in a comma-separated-values (CSV) format. The dataset contained 11 years of information spanning from 2011 to 2022 and consisted of 38 columns by 16.8 million rows of data. The `Air_Mission_Type` column contained text identifiers that described what type of mission was in each row.

Using the `Air_Mission_Type` column, the researcher was able to filter any non-channel missions out of the dataset. This approach differs from past research where focus was placed on commodity codes for further scoping. Rather, this research focused solely on the `Air_Mission_Type` to determine if the mission was a channel mission. Next, the researcher created a new dataset containing only the `TCN_Weight` and `Latest_Lift_Date` columns. `TCN_Weight` represents the total cargo airlifted by a specific mission and `Latest_Lift_Date` represents the last date each Transportation Control Number (TCN) shows up in the system and thus, arrival at destination. Together these two columns form the time series dataset used for the univariate time series analysis. Next, the `TCN_Weight` data was resampled to aggregate tonnage info into monthly sums. Further, the `Latest_Lift_Date` was parsed as a date-time object and

indexed as the independent variable. Finally, the dataset was split into a training and test-set at 80% and 20% respectively as recommended by Hyndman and Athanasopoulos and, in line with JDPAC's forecasting process (2018). This means that the first 104 months were categorized as Training data while the remaining months were categorized as Test data.

Univariate Time Series Forecasting - ARIMA

ARIMA is one of the most widely used models in forecasting. However, to fully understand an ARIMA model it is necessary to address the individual components of an ARIMA model. An ARIMA model can be decomposed into an autoregression component, integration element and a moving average.

Differencing

The implication of Wold's Decomposition Theorem is that any stationary series can be modeled using an ARMA process (Williams & Hoel, 2003). As discussed in the literature review chapter, a data set is stationary when the cumulative distribution function of a stochastic process is independent of time. Meaning, distribution parameters such as the variance and mean are approximately constant no matter what time in the series these parameters are measured. The implication is that there is no seasonality or trends in the series.

Just as the time plot of a dataset reveals stationarity or lack thereof, an Autocorrelation Function (ACF) plot is useful for identifying non stationarity (Hyndman & Athanasopoulos, 2018). For stationary time series, the ACF plot drops to zero quickly while the ACF plot of a non-stationary data decreases gradually.

Non-stationary datasets can be transformed into a stationary series through a process known as differencing (Hyndman & Athanasopoulos, 2018). It is a process where the difference

between consecutive observations is calculated. Differencing helps to stabilize the mean of a time series by removing changes in the level of a time series and in turn mitigating trend and seasonality. The I in ARIMA stands for integration and represents the fact that differenced values have been integrated into the original series transforming the non-stationary (series with trend or seasonality) series to a stationary series (Hyndman & Athanasopoulos, 2018).

Autoregression Model

As in multivariable regressions where multiple regressors are used to predict the variable of interest, in the autoregression (AR) model, the variable of interest is predicted using a linear combination of its lagged or past values (Hyndman & Athanasopoulos, 2018). For example, Lag 1 would mean that at the current time, the value of the dependent variable is equal to the previous value. In this research the variable of interest or dependent variable is the monthly sum of TCN_WEIGHT because the intent is to forecast monthly sums of total channel cargo weight airlifted based on its past values.

Mathematically the AR part of the model is shown below:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (1)$$

Where,

y_t dependent variable – regression of its lagged values

c intercept term

ϕ_i autoregressive parameter

y_{t-p} p th lagged value of y

The value of p is equal to the number of lags affecting the current value of y_t

ε_t error term and considered white noise

If there is any kind of pattern in the plot of error terms over time, then the dependent variable is not fully explained by the lag regressors. To find the value of the autoregressive parameter, we use the Partial Autocorrelation Function (PACF) plot to find how many lags are impacting the current value of the dependent variable. The PACF plot shows the pure contribution of each lag without the effects of preceding lags. If the error term in the AR model is not observed to be white noise after $p = 2$, proceed to the MA model (Hyndman & Athanasopoulos, 2018).

Moving Average Model

In a moving average (MA) model, prediction of the dependent variable relies on lagged forecast error terms. Each error term corresponds to a specific time point in the autoregressive model. Put differently, it is the moving average of the error terms represented as a linear combination of the errors.

Mathematically, the MA part of the model is shown below:

$$y_t = c + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q} \quad (2)$$

y_t dependent variable – a weighted MA of past forecast errors

c intercept term

θ_i moving average parameter

ε_{t-q} q th lagged value of the error term

ε_t error term

q can be determined by using the ACF chart. If the error term in the MA model is not white noise after $q = 3$, then proceed to using ARMA model (Hyndman & Athanasopoulos, 2018).

ARMA Model

The ARMA model is a combination of the AR and MA models. Mathematically, the ARMA model is represented as shown below:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} + \varepsilon \quad (3)$$

If the series was differenced to achieve stationarity, then the integration term is introduced into the ARMA model making it an ARIMA model. The ARMA models can be represented using ARIMA by simply setting $d = 0$. ARIMA is sufficient for those time series where the historical values of the series have sufficient pattern and information to predict its future values. For example, a stock market price chart may not be the best sole predictor of future prices since there are several other factors that affect the value of future prices (Hyndman & Athanasopoulos, 2018). The auto regressive, integrated, and moving average components of an ARIMA model have corresponding parameter terms p , d and q respectively.

ARIMA Model

The ARIMA model is often depicted using a backshift notation. The backshift notation for the ARIMA is

$$(1 - \phi_1 B - \cdots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \cdots + \theta_q B^q) \varepsilon_t \quad (4)$$

Where,

$$(1 - \phi_1 B - \cdots - \phi_p B^p) \rightarrow AR(p)$$

$$(1 - B)^d y_t \rightarrow \text{differences } (d)$$

$$c + (1 + \theta_1 B + \cdots + \theta_q B^q) \varepsilon_t \rightarrow MA(q)$$

And the backshift notation is such that:

$$By_t = y_{t-1}, \text{ to annotate one 1 lag}$$

$$B(By_t) = B^2y_t = y_{t-2}, \text{ to annotate two lags}$$

And for differencing,

$$y'_t = y_t - y_{t-1} = y_t - By_t = (1 - B)y_t, \text{ for first differencing}$$

$$y''_t = y_t - 2y_{t-1} + y_{t-2} = (1 - 2B + B^2)y_t = (1 - B)^2y_t$$

Generally, the d th order difference is written as:

$$(1 - B)^d y_t$$

Of note, ARIMA is not appropriate for time series with seasonality. If a time series has seasonality, the more appropriate model is Seasonal ARIMA (Hyndman & Athanasopoulos, 2018).

SARIMA Model

When seasonality must be accounted for in a time series, seasonal terms $(P, D, Q)m$ are added to the ARIMA model already developed. The resulting model for $SARIMA(p, d, q)(P, D, Q)m$ is given by:

$$\Phi_p(B^m)\phi(B)\nabla_m^D\nabla^d y_t = \theta_Q(B^m)\theta(B)\epsilon_t \quad (5)$$

Where,

$\phi(B)$ is the ordinary autoregressive component (p)

$\theta(B)$ is the ordinary moving average component (q)

∇^d is the ordinary difference component (d)

$\Phi_p(B^m)$ is the seasonal autoregressive component (P)

$\theta_Q(B^m)$ is the season moving average component (Q)

∇_m^D is the seasonal differencing component (D)

B is the backshift operator

To expand,

$$\phi(B) = 1 - \phi_1 B - \dots - \phi_p B^p$$

$$\Phi_p(B^m) = 1 - \phi_1 B^m - \dots - \phi_p B^{pm}$$

$$\theta(B) = 1 + \theta_1 B + \dots + \theta_q B^q$$

$$\theta_Q(B^m) = 1 + \theta_1 B^m + \dots + \theta_Q B^{Qm}$$

$$\nabla^d = (1 - B)^d$$

$$\nabla_m^D = (1 - B^m)^D \text{ (Chang et al. 2012)}$$

Long Short-Term Memory (LSTM) Recurrent Neural Network

Traditional neural networks are unable to use information from previous events to inform later events and therefore have an issue with memory persistence. RNNs have loops that allow information persistence. However, recurrent neural networks suffer from a lack of long-term memory due to a vanishing gradient problem during back propagation. LSTMs are a specialized variant of an RNN which add a cell state. Through this cell state, LSTMs have the capacity to learn long-term dependencies. Information passing through the cell state is regulated by structures called gates. The cell state can be thought of as an information highway that serves as the long-term memory for the system and, the gates serve the function of determining which long-term information is relevant. Same as traditional RNNs, LSTMs also possess a hidden state; compared to the cell state, the hidden state represents the short-term memory component of an LSTM. The gates typically have a sigmoid neural net layer and a pointwise multiplication

operation that allow each gate to determine how much of the past and current data is to be forgotten, kept, and propagated (Hewamalage, et al, 2020).

With the major components of a LSTM identified, the flow of a sequence of data is relatively straight forward. Past hidden state cell data and the current input are concatenated prior to reaching the gating mechanism. The hidden state and current input are processed through the forget gate and the result is pointwise multiplied with the past cell state. The past hidden state and current input are also put through the input gate and hyperbolic tangent (tanh) activation function for regulation. The resulting data from the input gate and tanh activation function (candidate cell state) are pointwise multiplied and then pointwise added to the cell state. The resulting data become the new cell state which feeds into the next cell and also into a tanh activation function where it is pointwise multiplied with the product of the output gate to make a new hidden state (Hewamalage, et al, 2020). Figure 2 is an illustration of a LSTM unit.

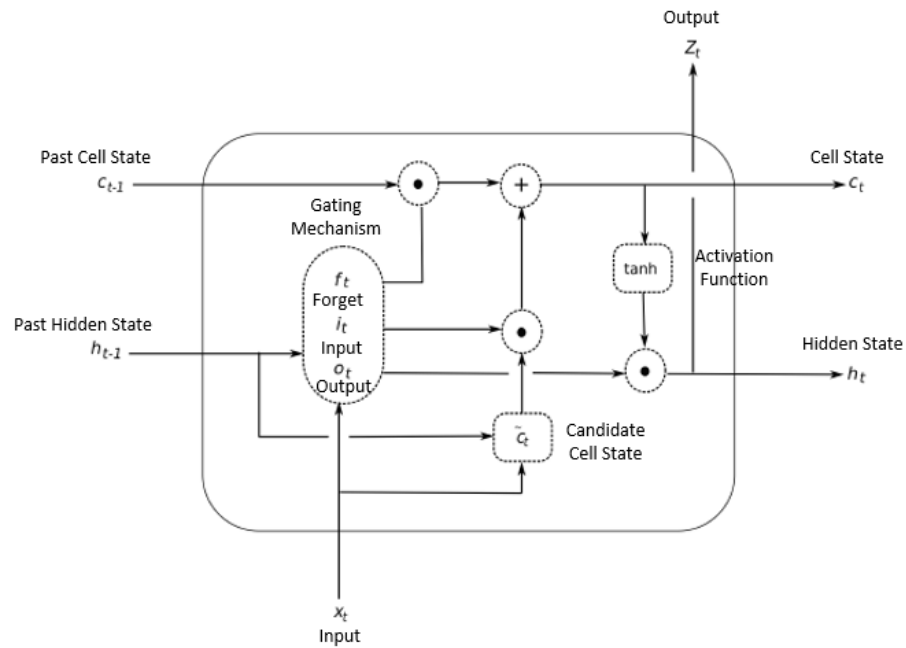


Figure 2: Sample Long Short-Term Memory Unit (Hewamalage, et al, 2020)

Mathematically, the operations within a LSTM are shown as:

$$i_t = \sigma(W_i \cdot h_{t-1} + V_i \cdot x_t + b_i)$$

$$o_t = \sigma(W_o \cdot h_{t-1} + V_o \cdot x_t + b_o)$$

$$f_t = \sigma(W_f \cdot h_{t-1} + V_f \cdot x_t + b_f)$$

$$\tilde{C}_t = \tanh(W_c \cdot h_{t-1} + V_c \cdot x_t + b_c)$$

$$C_t = i_t \odot \tilde{C}_t + f_t \odot C_{t-1}$$

$$h_t = o_t \odot \tanh(C_t)$$

$$Z_t = h_t$$

Where,

$\odot$, pointwise multiplication

$h_t \in \mathbb{R}^d$, hidden state of the cell

$C_t \in \mathbb{R}^d$, cell state

$\tilde{C}_t \in \mathbb{R}^d$, candidate cell state

d , cell dimensions or size of the input

$x_t \in \mathbb{R}^d$, input at time step t

$z_t \in \mathbb{R}^d$, output at time step t

$i_t \in \mathbb{R}^d$, input gate vector

$o_t \in \mathbb{R}^d$, output gate vector

$f_t \in \mathbb{R}^d$, forget gate vector

σ , sigmoid function with output range $[0, 1]$

$W_i, W_o, W_f, W_c, \in \mathbb{R}^{d \times d}$, weight matrices for the input gate (i), output gate(o), forget gate(f), and cell state(c).

$V_i, V_o, V_f, V_c, \in \mathbb{R}^{d \times d}$, weight matrices for the current input

$b_i, b_o, b_f, b_c, \in \mathbb{R}^d$, bias vector (Hewamalage, et al, 2020)

Of note, though the preceding discussion references tanh activation function, this research uses a rectified linear activation function (ReLU) for simplicity and better performance.

Measures of Forecast Performance

The Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE) and Root Mean Squared Logarithmic Error were determined for each model and used to determine which model was most accurate. RMSLE was chosen to help scale down any errors from outliers that would otherwise explode the value of the Root Mean Square Error (RMSE). An equation for each measure of performance is given below.

$$MAE = \frac{\sum_{i=1}^n |y_i - \hat{y}_i|}{n} \quad (6)$$

Where:

n = number of observations

y_i = independent variable (actual value)

$\hat{y}_i$ = forecasted value

and,

$$RMSLE = \sqrt{\frac{\sum_{i=1}^n (\log(y_i+1) - \log(\hat{y}_i+1))^2}{n}} \quad (7)$$

$$MAPE = \frac{100}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (8)$$

IV. Results and Analysis

Chapter Overview

This chapter will evaluate and then compare the performance of the ARIMA, SARIMA and LSTM models using IGC data for the year 2013-2022. The chapter begins with a short description of the software used for analysis and model development. What follows is a discussion of the general characteristics of the dataset. Each model is then presented along with goodness of fit measures for comparison. Lastly, the model predictions are plotted against Test data for graphical presentation of model performance.

Python

The main purpose of this discussion is to showcase the versatility of python and generate discussion on possible application. Secondly, this section gives a brief overview of the software used for analysis. Python is the world's fastest growing programming language and according to the Tiobe index, the most popular in 2022 (Tiobe, 2022). The Tiobe Index is useful for making "a strategic decision about what programming language should be adopted" (Tiobe). Part of what makes Python popular is its beginner friendly syntax, versatile application in the field data analysis, and visualization and cross platform utilization (Python). Python's versatility serves as a force multiplier and is line with command's priority for digital adept airmen as it gives airmen the tools to solve a myriad of problems. The researcher used Python to clean the data and develop models used to forecast monthly cargo figures. Overall, the SARIMA model performed the best in forecasting test values.

Processing

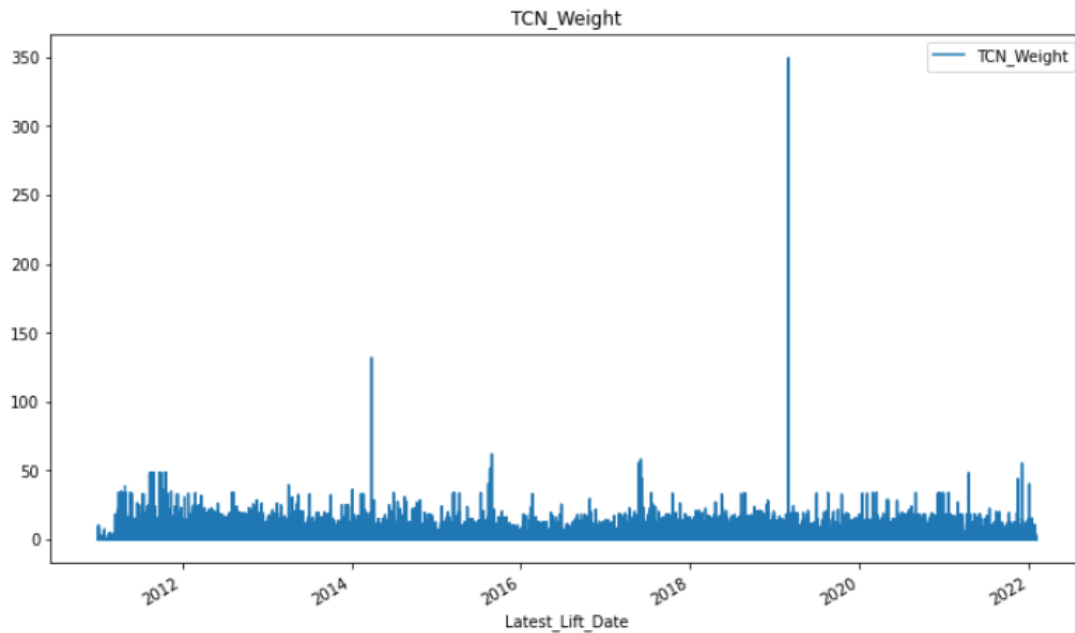


Figure 3: Raw Data

Figure 3 is plot of the raw TCN_Weight data for only those missions designated as a CHANNEL in the Air_Mission_Type column. After monthly aggregation and further cleaning, the dataset was replotted. Figure 4 is a plot of the cleaned data.

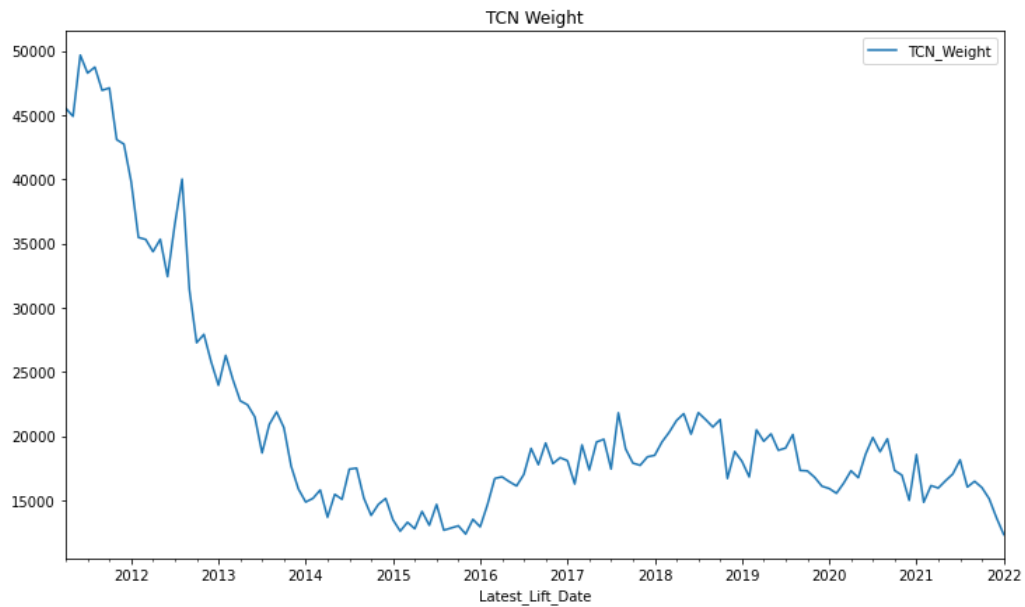


Figure 4: Cleaned and Aggregated Data

Information about the data's trend, seasonality, and residuals was obtained using a seasonal decomposition tool in Python. Figure show 5 shows an overall decreasing trend over time, seasonality, and a residual plot that has the approximate pattern of the seasonal plot. Visual inspection of these plots confirm non-stationarity.

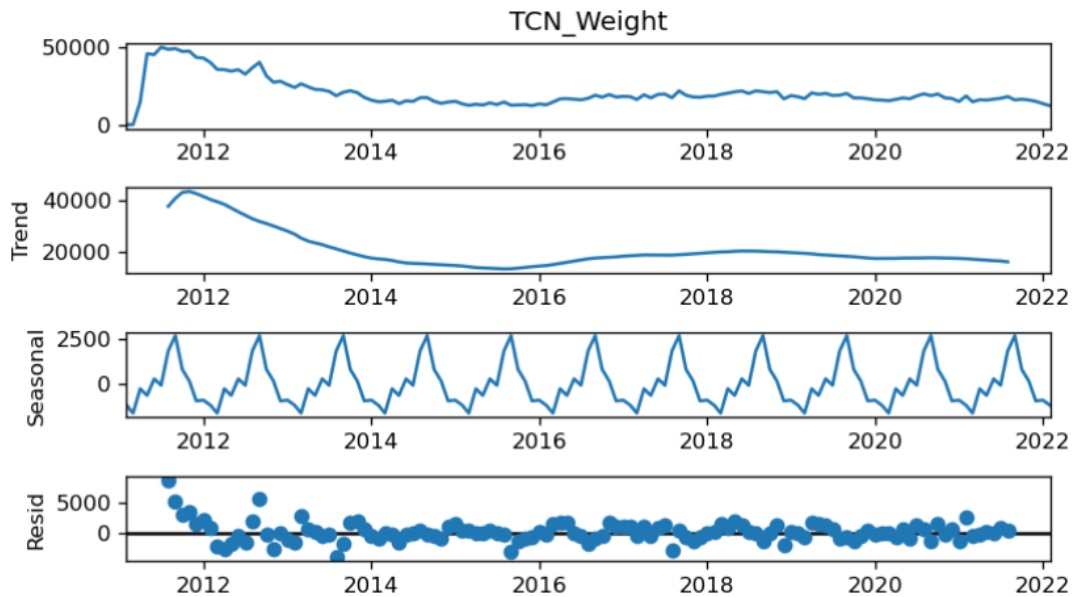


Figure 5: Seasonal Decomposition

The data was split at the 2019-11-13 point for an approximate 80% - 20% train - test ratio. Train data are used to train the model and the predictions are evaluated against the Test dataset. The autocorrelation and partial autocorrelation for the Train dataset were plotted (Figures 6 & 7) to glean a better understanding of the lags. For example, the plots reveal the number of lags that are statistically significant and in turn, help identify the order of the ARIMA model.

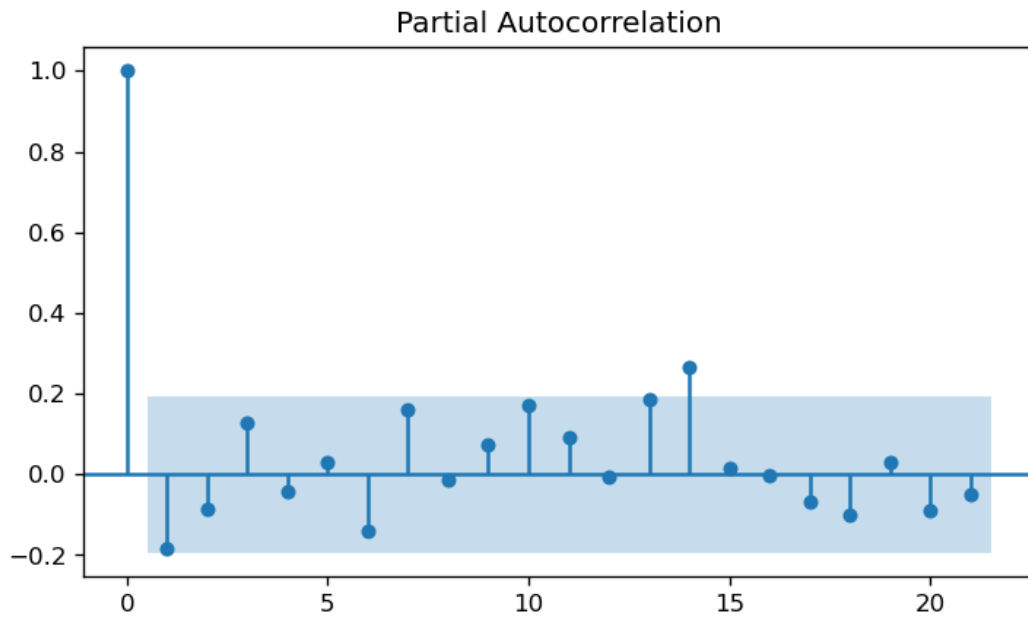


Figure 6: Train Dataset Partial Autocorrelation Plot

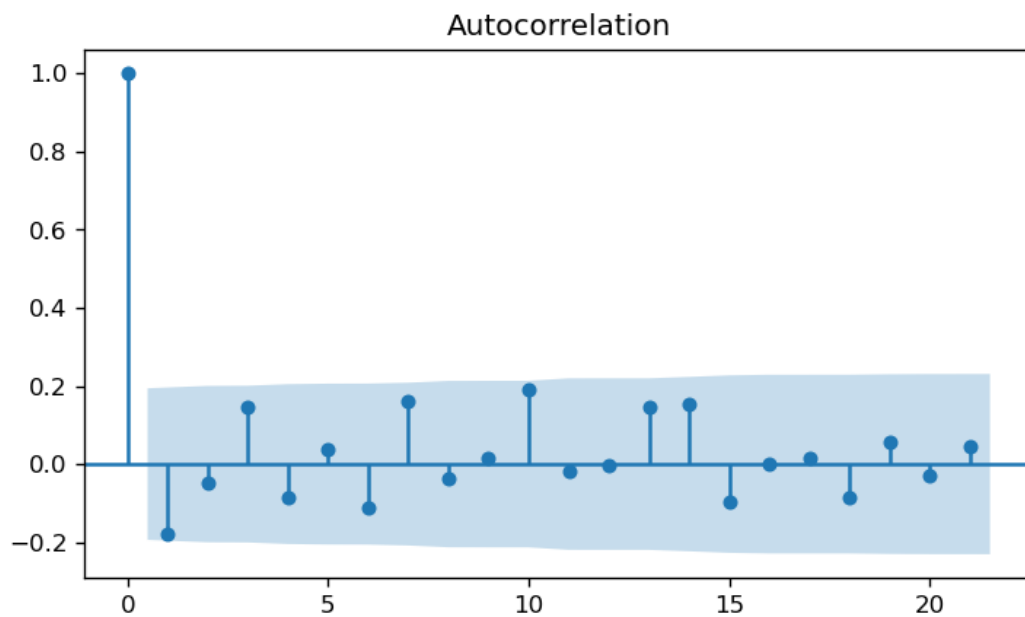


Figure 7: Train Dataset Autocorrelation Plot

ARIMA

With use of the *pm.auto_arima* function in Python, the best ARIMA model was determined to have p, d, q values of 0,2,1 with characteristics summarized in Figure 8.

Dep. Variable:	y	No. Observations:	103			
Model:	SARIMAX(0, 2, 1)	Log Likelihood	-939.811			
Date:	Thu, 19 May 2022	AIC	1883.622			
Time:	06:11:15	BIC	1888.852			
Sample:	0	HQIC	1885.739			
	- 103					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.4554	0.043	-10.528	0.000	-0.540	-0.371
sigma2	6.816e+06	6.18e+05	11.032	0.000	5.61e+06	8.03e+06
Ljung-Box (L1) (Q):	4.53	Jarque-Bera (JB):	40.10			
Prob(Q):	0.03	Prob(JB):	0.00			
Heteroskedasticity (H):	0.47	Skew:	-0.79			
Prob(H) (two-sided):	0.03	Kurtosis:	5.65			

Figure 8: ARIMA Model Results

The model was trained using the Train dataset and its predictions were plotted against the values of Test dataset. Figure 9 is a graph showing the overlay of model predictions and Test values.

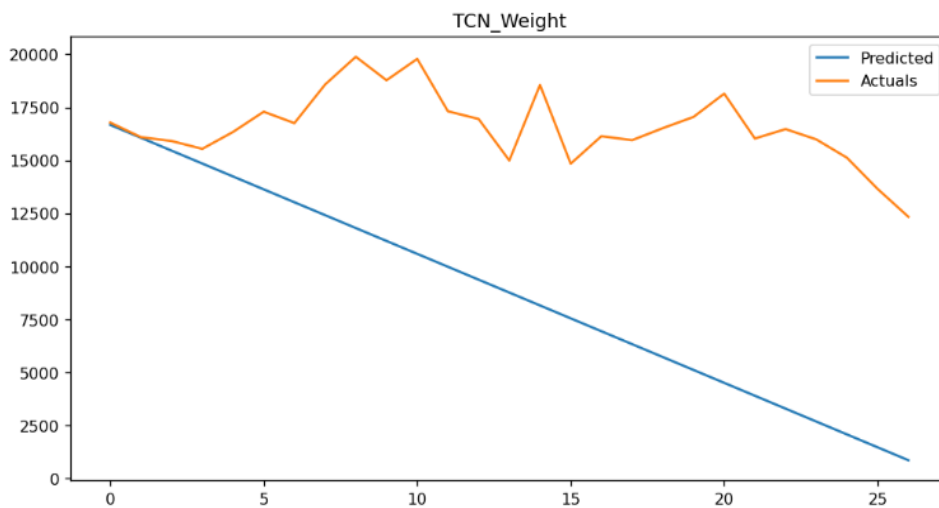


Figure 9: ARIMA Model Predicted vs Test Data

The model did not perform very well in predicting the values of the Test dataset. In effect, the model produced a downward sloping linear approximation of the future values due to the lack of seasonal parameters in the ARIMA model. The behavior of the predicted values makes sense as the ARIMA model is not accounting for seasonal effects. The model had a MAE of 7820.67, a RMSLE of 1.11 and a MAPE of 242.25. Figure 10 is a summarization of the ARIMA model's performance consisting of standardized residual, density distribution, normal Q-Q and correlogram plots.

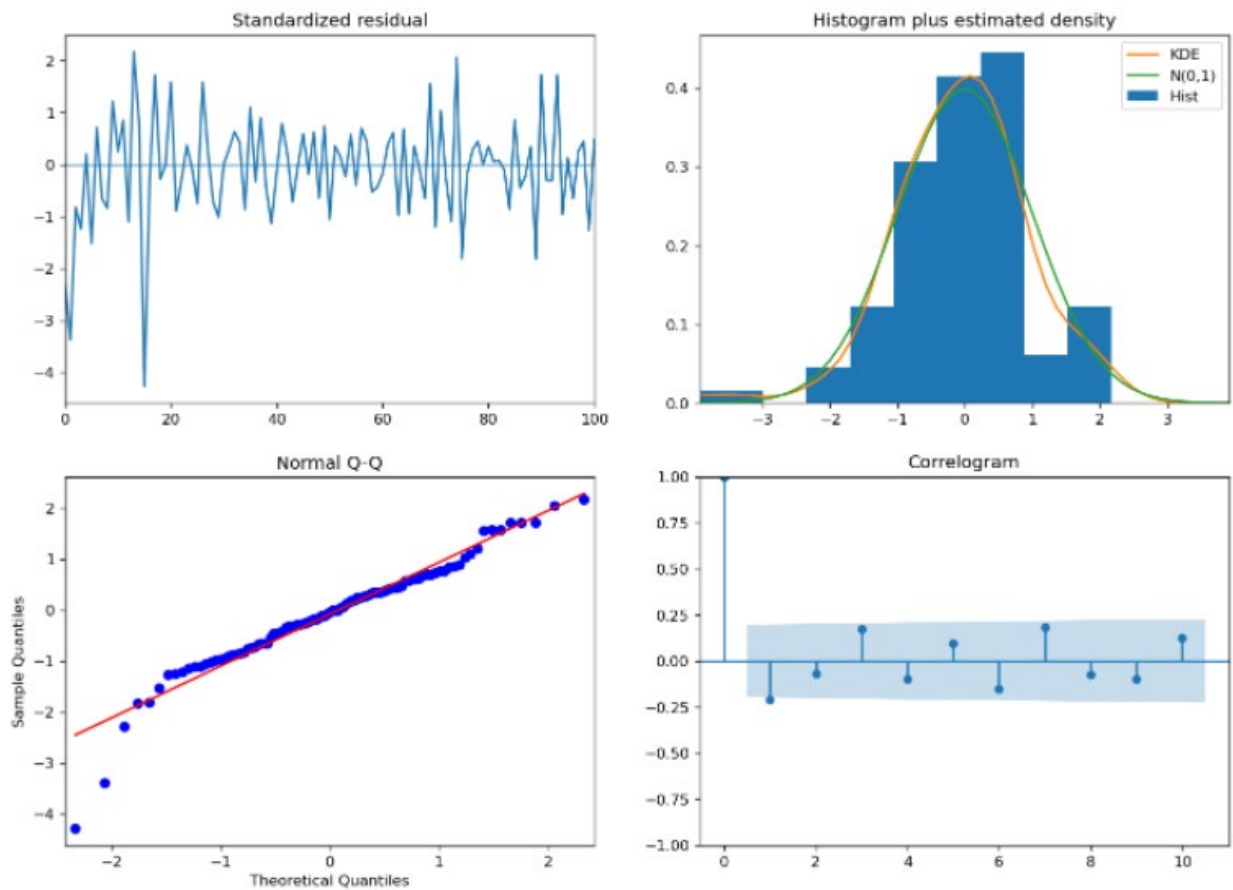


Figure 10: ARIMA Model Diagnostics

SARIMA

The process for developing the SARIMA model was very similar to the ARIMA process except that it included seasonal parameters. Using the *pm.auto_arima* function in Python, the best SARIMA model was determined to have p, d, q values of 0,1,1 and P, D, Q, m values of 1,1,1, 12 with characteristic summarized in Figure 11.

Dep. Variable:	y	No. Observations:	103			
Model:	SARIMAX(0, 1, 1)x(1, 1, 1, 12)	Log Likelihood	-831.451			
Date:	Thu, 19 May 2022	AIC	1670.902			
Time:	06:20:16	BIC	1680.901			
Sample:	0	HQIC	1674.934			
	- 103					
Covariance Type:	opg					
	coef	std err	z	P> z	[0.025	0.975]
ma.L1	-0.1131	0.072	-1.576	0.115	-0.254	0.028
ar.S.L12	0.4546	0.109	4.179	0.000	0.241	0.668
ma.S.L12	-0.8376	0.217	-3.854	0.000	-1.264	-0.412
sigma2	6.694e+06	1.28e+06	5.225	0.000	4.18e+06	9.21e+06
Ljung-Box (L1) (Q):	1.40	Jarque-Bera (JB):	1.88			
Prob(Q):	0.24	Prob(JB):	0.39			
Heteroskedasticity (H):	0.27	Skew:	-0.17			
Prob(H) (two-sided):	0.00	Kurtosis:	3.62			

Figure 11: SARIMA Model Results

With the inclusion of seasonal parameters the model performed much better than the ARIMA model. Figure 12 is a plot of the model's predictions overlayed with the Test data.

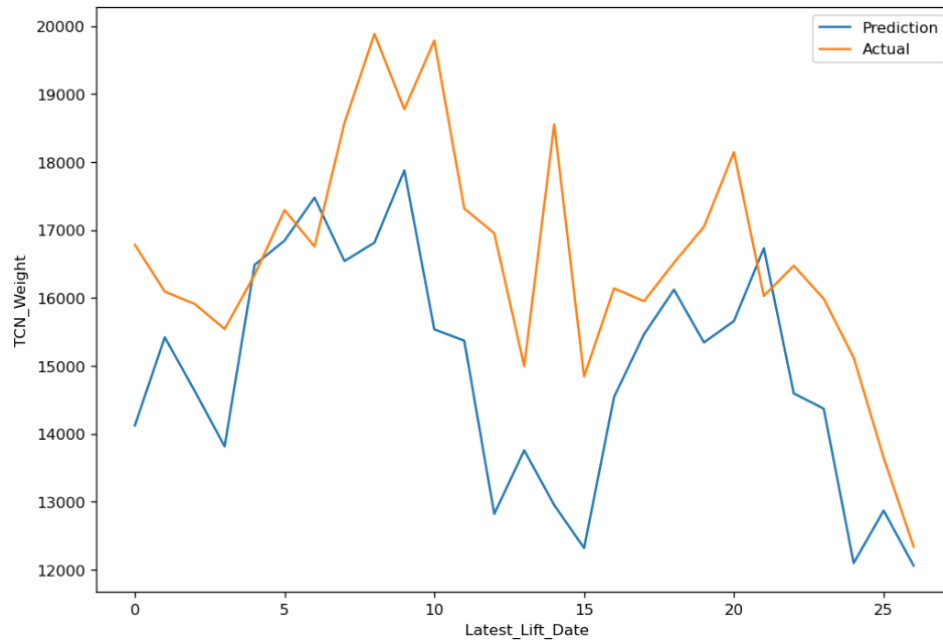


Figure 12: SARIMA Model Predicted vs Test Data

The model had a MAE of 1789.58, a RMSLE of 0.14 and a MAPE of 12.52. Figure 13 is a summarization of model's performance consisting of standardized residual, density distribution, normal Q-Q and correlogram plots.

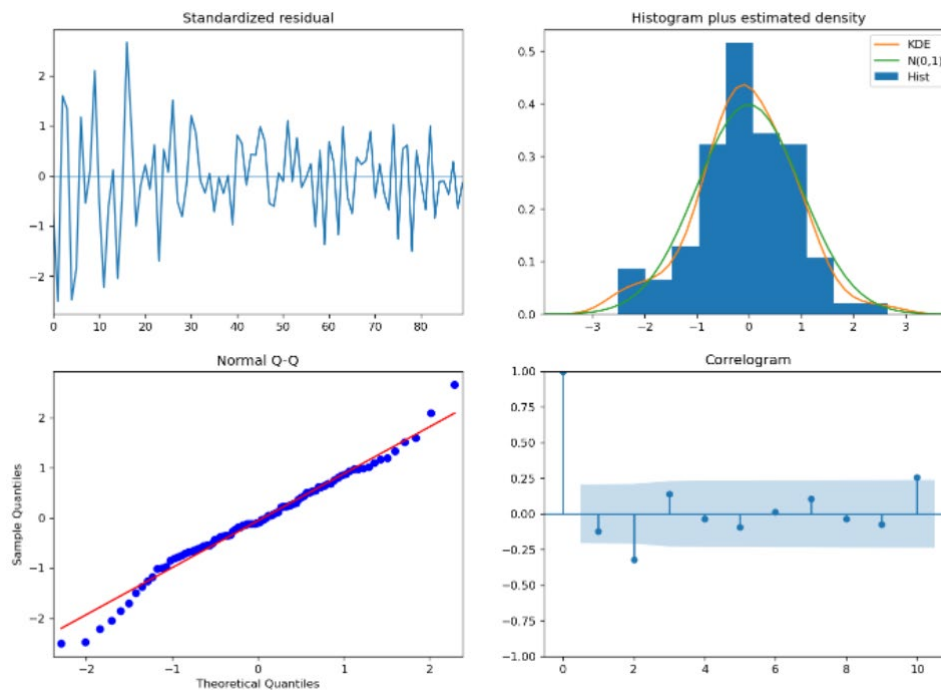


Figure 13: SARIMA Model Diagnostics

LSTM

The LSTM model was developed by importing the required modules, instantiation, adding layers and compiling the model using an Adam Optimizer. Finally, the model was fit to the Train dataset. A summary of the the LSTM is shown below in Figure 14.

Model: "sequential"

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 250)	252000
dense (Dense)	(None, 1)	251

=====
Total params: 252,251
Trainable params: 252,251
Non-trainable params: 0
=====

Figure 14: LSTM Model

Figure 15 illustrates the reduction in loss rate as the LSTM iterated through each epoch indicating that the model is sound.

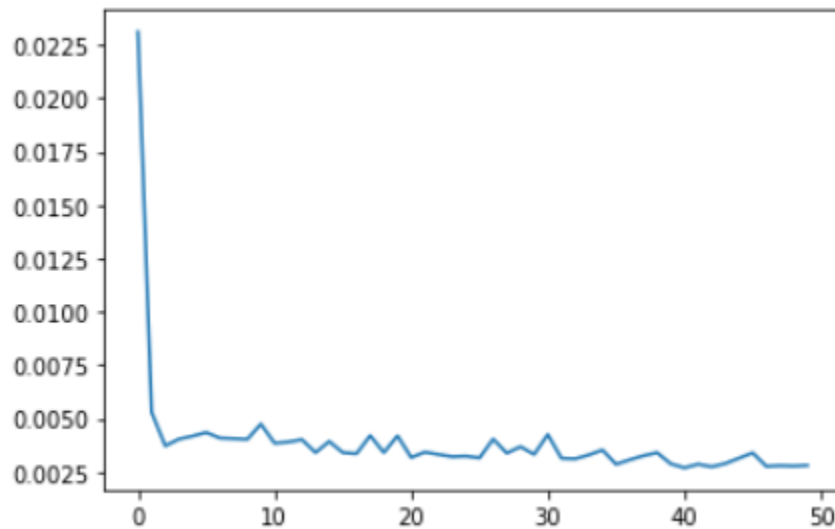


Figure 15: Loss Per Epoch

As seen in Figure 16, the LSTM model performed marginally better than the ARIMA model and clearly not as good as the SARIMA model. The results from the graph are backed up by an MAE of 4821.58, a RMSLE of 0.56 and a MAPE of 78.64.

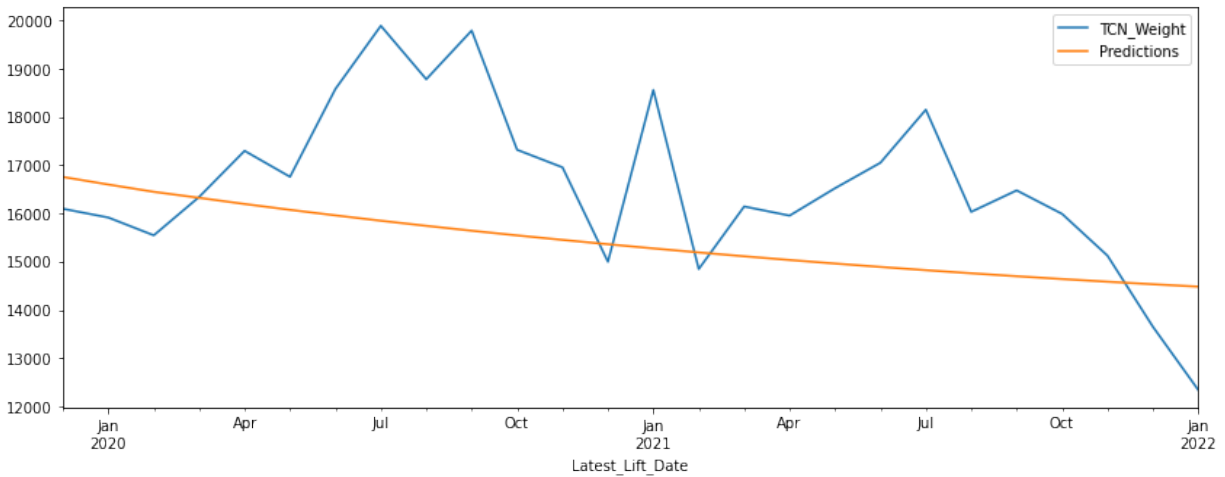


Figure 16: LSTM Model Predicted vs Test Data

V. Conclusion

Chapter Overview

The purpose of this research was to create a cheap yet accurate tool for forecasting channel cargo demand. With this objective as the guide, the researcher analyzed benchmark forecasting models like ARIMA and SARIMA and compared their performance to an LSTM model. What follows is a discussion of the overall findings of the research along with recommendations for implementation.

Conclusion from Research

In today's environment of increasing global military threats from near peer adversaries, inefficient use of resources can be an Achilles heel. Underutilization of contracted commercial augmentation airlift through fixed buy is an enduring problem and a prime example of inefficient use of resources. To be clear, there are cases where a commander's priority should not be efficiency instead the commander must deploy resources as necessary to ensure success. Channel cargo airlift for peacetime sustainment, however, is not one of those cases. Further, though USTRANSCOM is motivated to provide commercial partners with incentives to participate in CRAF, mechanisms for precise monetary incentives are preferred over ad hoc and opportune circumstances. This research attempts to mitigate this problem by analyzing different models for best performance using open-source code.

Data from IGC consisting of channel cargo airlifted between 2011 and 2022 were processed and aggregated into monthly sums to align with the target output for the forecasting models. The dataset was then split into a Train and Test dataset. Using pre-built modules within Python, ARIMA, SARIMA, and LSTM models were built. The models were fit to the Train dataset and their respective performance was evaluated using the Test dataset. The predictions

from each model were plotted against the Test dataset for a graphical representation of model performance. Additionally, the MAE, RMSLE and MAPE were used to compare model performance. Overall, the SARIMA had the best performance followed by the LSTM model and lastly, the ARIMA model. It is worth noting that the performance of the LSTM model could improve with the addition of other independent variables.

Recommendations

Though an LSTM can be incredibly powerful and precise when applied to forecasting, it underperformed the SARIMA model in analysis of monthly forecast of channel cargo demand across the globe. Based on the results of this research it is recommended that the SARIMA model be employed when using historical tonnage data to forecast monthly channel cargo demand at the Global Cargo Management level.

Python is a free and multifaceted tool that's available to all airmen who wish to become more digitally adept. Even better, this free tool can be used to conduct analysis that leads to better decision-making and ultimately help reduce unnecessary costs.

Areas for Further Research

This research did not account for specific city pairs and aircraft type. Additionally, the models developed in this research were all univariate and depended only on historically airlifted channel cargo. Though these were stated limitations of this research, areas of future research include LSTM forecasts that incorporate cargo transported on specific routes and by specific aircraft for added realism. Additionally, incorporating additional causative factors and independent variables takes full advantage of statistical algorithms that are available at no cost.

Appendix A

Abbreviations

ACF Autocorrelation Function

AIC Akaike Information Criterion

AMC Air Mobility Command

APOD Aerial Port of Debarkation

AR Autoregression

ARA Airlift Readiness Account

ARCH Autoregressive Conditional Heteroskedasticity

ARIMA Autoregressive Integrated Moving Average

ASM Aircraft Selection Model

BNN Bayesian Neural Network

CART Classification and Regression Trees

COINS Commercial Operations Integrated System

COTS Commercial Off the Shelf

CRAF Civil Reserve Air Fleet

CSV Comma Separated Values

DCBS Defense Enterprise Accounting and Management System's Component Billing System

DoT Department of Transportation

GAO United States Government Accountability Office

GATES Global Air Transportation Execution System

GDSS Global Decision Support System

GP Gaussian Processes

GRNN Generalized Regression Neural Network

IGC Integrated Development Environment/Global Transportation Network Convergence

JDPAC Joint Distribution Process Analysis Center

KNN Kernel regression, K-Nearest Neighbor regression

LSTM Long Short-Term Memory network

MA Moving Average
MAE Mean Absolute Error
MAPE Mean Absolute Percentage Error
MASE Mean Absolute Scaled Error
ML Machine Learning
MLP Multi-Layer Perceptron
MOM Mobility Optimization Model
MRS Mobility Requirement Study
MMKAP Multiple Multidimensional Knapsack Assignment Problem
NDS National Defense Strategy
NN Neural Network
PACF Partial Autocorrelation Function
RAC Research Analysis Corporation
RBF Radial Basis Functions
REMIS Reliability and Maintainability Information System
RMSLE Root Mean Squared Logarithmic Error
RNN Recurrent Neural Network
SAF/FMB Air Force Financial Management Budget Office
SARIMA Seasonal Autoregressive Integrated Moving Average
sMAPE symmetric Mean Absolute Percentage Error
SVM Support Vector Machine
SVR Support Vector Regression
TCN Transportation Control Number
TPFDD Time Phased Force Deployment Data
TWCF Transportation Working Capital Fund
TWFWG Transportation Workload Forecast Working Group
USCENTCOM United States Central Command
USTRANSCOM United States Transportation Command

Bibliography

About Python, Python, www.python.org/about/.

Aman, Saimn, et al. "Prediction Models for Dynamic Demand Response: Requirements, Challenges, and Insights." IEEE International Conference on Smart Grid Communications, 5 Nov. 2015. IEEE Xplore, ieeexplore.ieee.org/document/7436323.

Ariton, Lleyton. "A Thorough Introduction to Holt-Winters Forecasting." Medium, 22 Feb. 2021, medium.com/analytics-vidhya/a-thorough-introduction-to-holt-winters-forecasting-c21810b8c0e6.

Billion, William E., and Lawrence G. Regan. "A Methodology for Accurately Predicting Demand for Airlift of Military Cargo to Overseas Destinations." RAC PAPER RAC-P-25, Dec., doi:<https://apps.dtic.mil/sti/pdfs/AD0686767.pdf>.

Beckley, Michael. The Power of Nations: Measuring What Matters. International Security 2018; 43 (2): 7–44. doi: https://doi.org/10.1162/isec_a_00328.

Brownlees, Christian, et al. "A Practical Guide to Volatility Forecasting Through Calm and Storm." *The Journal of Risk*, vol. 14, no. 2, 2012, pp. 3-22, www.sas.upenn.edu/~fdiebold/papers/misc/Brownlees.pdf.

Callahan, Keith D. "Commercial Augmentation and the Need for Demand Forecasting." 1 June 2020. Defense Technical Information Center, apps.dtic.mil/sti/pdfs/AD1106530.pdf.

Chang, Xinghua, et al. "Seasonal Autoregressive Integrated Moving Average Model for Precipitation Time Series." Journal of Mathematics and Statistics, 20 Feb. 2013, doi:[10.3844/jmssp.2012.500.505](https://doi.org/10.3844/jmssp.2012.500.505), thescipub.com/pdf/jmssp.2012.500.505.pdf.

Connor, Kathryn, Michael Vasseur, and Laura H. Baldwin, Aligning Incentives in the Transportation Working Capital Fund: Cost Recovery While Retaining Readiness in Military Transportation. Santa Monica, CA: RAND Corporation, 2019. https://www.rand.org/pubs/research_reports/RR2438.html.

Cornell Law School, Legal Information Institute (LII), "32 CFR § 243.3 – Definitions" <https://www.law.cornell.edu/cfr/text/32/243.3#:~:text=The%20Civil%20Reserve%20Air%20Fleet%20%28CRAF%29%20is%20a,any%20level%20of%20national%20emergency%20or%20defense-orientated%20situations>.

Coupa Introduction to Demand Guru <https://help.llama.ai/release/native/demand/Home.htm>.

DeYoung, Daniel S. "TIME SERIES FORECASTING OF AIRLIFT SUSTAINMENT CARGO DEMAND." 1 June 2012. *Defense Technical Information Center*, apps.dtic.mil/sti/pdfs/ADA566289.pdf.

DoDFMR Vol 2B. https://comptroller.defense.gov/Portals/45/documents/fmr/Volume_02b.pdf.

DOD Needs to Improve Budget Estimates and Accuracy of Forecast for Airlift Services, U.S. Government Accountability Office, Sept. 2018, www.gao.gov/assets/700/694350.pdf.

DoD Release, "Department of Defense Activates Civil Reserve Air Fleet to Assist with Afghanistan Efforts", <https://www.defense.gov/Newsroom/Releases/Release/Article/2741564/departments-of-defense-activates-civil-reserve-air-fleet-to-assist-with-afghanistan/>.

"Fiscal Year (FY) 2021 Budget Estimates, Operation and Maintenance, Air Force Volume 1." *Financial Management and Comptroller*, U.S. Government Accountability Office, Feb. 2020, www.saffm.hq.af.mil/Portals/84/documents/FY21/OM_/ACTIVE/FY21%20Air%20Force%20Operation%20and%20Maintenance%20Vol%20I.pdf?ver=2020-02-10-154101-027.

Department of Transportation, "Civil Reserve Airfleet" <https://www.transportation.gov/mission/administrations/intelligence-security-emergency-response/civil-reserve-airfleet-allocations>.

Greater Use of Peacetime Airlift Cargo Capacity Would Reduce Cost, U.S. Government Accountability Office, Sept. 1992, www.gao.gov/assets/nsiad-92-263.pdf.

Gorn, Michael H., *Harnessing the Genie: Science and Technology Forecasting for the Air Force, 1944-1986*. https://books.googleusercontent.com/books/content?req=AKW5Qad-6EG06zUN3n8xydWU8oOhFu7aVoPYuUjU2v8bVRAJMX3TT23MQCXgNI14n-57N17FbFUU8naR-VSkAe7wAWvfn52KkMqbhZEcoA9uk2jsFR7FfUO5xJtmsikST71MGpCUHWIv582dOfWkI4q7NgN1sARbC-1BMIUXyUIEygV1LMrfN9iQdCO6nWyLdDaFqsCCQsiAU29bRuY9bfFIpD8JrZssS119kkVKDu2BQeNP_Tzdx_8SnaKzJLvMjHiCQSzZlBDb.

Hemken, Katherine B. "Channel Mission: Forecasting Sustainment Cargo Requirements." 1 June 2019. *Defense Technical Information Center*, apps.dtic.mil/sti/pdfs/ADA582032.pdf.

Hewamalage, Hansika, et al. "Recurrent Neural Networks for Time Series Forecasting: Current Status and Future Directions." *International Journal of Forecasting*, vol. 37, no. 1, 2021, pp. 388-427, doi:<https://doi.org/10.1016/j.ijforecast.2020.06.008>, arxiv.org/pdf/1909.00590.pdf.

- "H. Rept. 115-200 - NATIONAL DEFENSE AUTHORIZATION ACT FOR FISCAL YEAR 2018." Congress.gov, Library of Congress, 30 April 2022, <https://www.congress.gov/congressional-report/115th-congress/house-report/200/1>.
- Hyndman, Rob J., and George Athanasopoulos. *Forecasting: Principles and Practice*. 2ndnd ed., OTexts, 2018, otexts.com/fpp2/.
- Leonard, Taylor J. "Operational Planning of Channel Airlift Missions Using Forecasted Demand." 1 Mar. 2013. Defense Technical Information Center, apps.dtic.mil/sti/pdfs/ADA582032.pdf.
- Makridakis, Spyros, et al. "Statistical and Machine Learning forecasting methods: Concerns and Ways Forward." PLOS ONE, 27 Mar. 2018, journals.plos.org/plosone/article%3Fid%3D10.1371/journal.pone.0194889.
- Mills, Patrick, et al. "Support to the Air Force Installation and Mission Support Center Enabling AFIMSC's Role in Agile Combat Support Planning, Execution, Monitoring, and Control." *RAND Project Air Force*, RAND Corporation, Jan. 2017, www.rand.org/content/dam/rand/pubs/research_reports/RR1500/RR1555/RAND_RR1555.pdf.
- Morton, David P., et al. "Optimization modeling for Airlift Mobility." *Military Operations Research*, 1996, pp. 49-67, www.usna.edu/Users/math/uhan/sa475a/guidelines/articles/morton.pdf.
- Rajkumar, Nitarshan. "Weight Uncertainty in Neural Networks." Nitarshan Rajkumar, 4 June 2018, www.nitarshan.com/bayes-by-backprop/.
- Small, Matthew T., "Predicting Global Disposition of U.S. Military Personnel via Open-Source, Unclassified Means" (2018). *Theses and Dissertations*. 1862. <https://scholar.afit.edu/etd/1862>.
- Summary of the National Defense Strategy. (DoD, 2018) <https://dod.defense.gov/Portals/1/Documents/pubs/2018-National-Defense-Strategy-Summary.pdf>.
- Surdyk, Brian A. "AN ANALYSIS OF TIME SERIES FORECASTING METHODS FOR THE AIRLIFT OF PALLETIZED SUSTAINMENT CARGO." June 2010. *Defense Technical Information Center*, doi:<https://apps.dtic.mil/sti/pdfs/ADA526862.pdf>, apps.dtic.mil/sti/pdfs/ADA526862.pdf.

Thompson, Maxwell C., "Database Analysis to Improve U.S. Transportation Command Forecasting. Processes" (2020). Theses and Dissertations. 3202.
<https://scholar.ait.edu/etd/3202>.

TIOBE Index: TIOBE Index for May 2022, Tiobe, www.tiobe.com/tiobe-index/.

Unused Cargo Space on Military Aircraft Returning to the United States, U.S. Government Accountability Office, 9 Aug. 1983, www.gao.gov/assets/nsiad-83-19.pdf.

Williams, Billy M., and Lester A. Hoel. "Modeling and Forecasting Vehicular Traffic Flow as a Seasonal ARIMA Process: Theoretical Basis and Empirical Results." *Journal of Transportation Engineering*, vol. 129, no. 6, Nov. 2003, pp. 664-72, doi:[https://doi.org/10.1061/\(ASCE\)0733-947X\(2003\)129:6\(664\)](https://doi.org/10.1061/(ASCE)0733-947X(2003)129:6(664)).

Zhu, X., Zhang, G. & Sun, B. A comprehensive literature review of the demand forecasting methods of emergency resources from the perspective of artificial intelligence. *Nat Hazards* 97, 65–82 (2019). <https://doi.org/10.1007/s11069-019-03626-z>.

REPORT DOCUMENTATION PAGE					<i>Form Approved</i> OMB No. 0704-0188	
The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY) 03/06/2022		2. REPORT TYPE Graduate Research Paper			3. DATES COVERED (From - To) July 2021 - June 2022	
4. TITLE AND SUBTITLE ANALYSIS OF BENCHMARK FORECASTING MODELS VERSUS A RECURRENT NEURAL NETWORK FORECAST OF CHANNEL CARGO DEMAND				5a. CONTRACT NUMBER		
				5b. GRANT NUMBER		
				5c. PROGRAM ELEMENT NUMBER		
6. AUTHOR(S) Umanah, Abraham N. Maj, USAF				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Institute of Technology Graduate School of Engineering and Management (AFIT/EN) 2950 Hobson Way Wright-Patterson AFB OH 45433-7765					8. PERFORMING ORGANIZATION REPORT NUMBER AFIT-ENS-MS-22-J-056	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Intentionally Left Blank					10. SPONSOR/MONITOR'S ACRONYM(S)	
					11. SPONSOR/MONITOR'S REPORT NUMBER(S)	
12. DISTRIBUTION/AVAILABILITY STATEMENT Distribution Statement A. Approved for public release; distribution is unlimited.						
13. SUPPLEMENTARY NOTES This work is declared a work of the U.S. Government and is not subject to copyright protection in the United States.						
14. ABSTRACT Given the longstanding history of inaccurate cargo demand forecast and underutilization of contracted commercial airlift augmentation, the primary objective of this paper is to present a simple forecasting model developed using Python. This study compares benchmark time series forecasting models against a Recurrent Neural Network model to determine which model is best for predicting monthly channel cargo airlift demand. To determine which model produces the best forecast, univariate time series analysis is conducted on Integrated Development Environment/Global Transportation Network Convergence data using readily available statistical modules and Machine Learning algorithms within the Python ecosystem. Of the models examined, the Seasonal Auto Regression Integrated Moving Average model performed the best and was recommended for use at the Global Cargo Management level. The research concludes with recommendations for future research that includes aircraft type on specific routes using a multivariable time series analysis.						
15. SUBJECT TERMS Models, Forecast						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT UU	18. NUMBER OF PAGES 59	19a. NAME OF RESPONSIBLE PERSON Dr. Bruce A. Cox	
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U			19b. TELEPHONE NUMBER (Include area code) (937) 255-3636 ext 4676	