Time series forecasting with high stakes: A field study of the air cargo industry

Anonymous Author(s)

ABSTRACT

Time series forecasting in the air cargo industry presents unique challenges due to volatile market dynamics and the significant impact of accurate forecasts on generated revenue. This paper explores a comprehensive approach to demand forecasting at the origin-destination (O&D) level, focusing on the development and implementation of machine learning models in decision-making for the air cargo industry. We leverage a mixture of experts framework, combining statistical and advanced deep learning models to provide reliable forecasts for cargo demand over a six-month horizon. The results demonstrate that our approach outperforms industry benchmarks, offering actionable insights for cargo capacity allocation and strategic decision-making in the air cargo industry. While this work is applied in the airline industry, the methodology is broadly applicable to any field where forecast-based decision-making in a volatile environment is crucial.

KEYWORDS

Time Series Forecasting, Mixture of Experts Framework, Revenue Management, Seasonal Trends, Deep Learning, Meta-Learning, Case Study

1 INTRODUCTION

The increasing availability and complexity of time series data have amplified the need for robust forecasting techniques across various industries, including finance, healthcare, and transportation [27, 34, 37]. In the air cargo sector, accurate forecasting is critical for maximizing revenue generated through each unit of capacity used by cargo bookings [5, 21]. Unlike passenger sales, air cargo bookings are notoriously volatile, are prone to be cancelled, and are influenced by many factors such as economic conditions, seasonal events, and disruptions affecting global trade [24]. Additionally, cargo bookings are often part of longer-term customer commitments, which help airlines plan their network schedules and budgeting efforts. Finally, since there are significantly fewer

cargo bookings per flight as compared to passenger bookings, the importance of making the right decision (informed by an accurate forecast) is increased. This paper presents a field study on time series forecasting in the air cargo industry, focusing on the challenges and solutions associated with mid-to-long-term demand prediction.

Recent advancements in sensing technologies and data analytics have paved the way for more sophisticated forecasting models [3, 17]. However, the air cargo industry faces unique challenges, including irregular booking patterns, varying shipment types, and the need for high accuracy in forecasts. Traditional forecasting methods often fail to address these complexities, necessitating the development of specialized machine-learning approaches.

In this study, we introduce a comprehensive forecasting framework that combines multiple statistical and machine learning models to predict cargo demand at the origin and destination level. Our approach involves a detailed data preparation process, integrating various data sources to capture the full spectrum of factors affecting cargo demand. We employ a mixture of experts framework, which allows us to select the best-performing model for each sample based on historical performance, thereby enhancing the overall prediction accuracy.

Our research is conducted within the context of a major air cargo carrier, providing a real-world application of the proposed method, and validated across diverse market segments. We evaluate the performance of our models using key metrics such as root mean square error (RMSE) and weighted normalised root mean square (WnRMSE), demonstrating significant improvements over existing industry benchmark forecasting techniques. The findings of this study offer valuable insights for applied teams in various industries, enabling more informed decision-making leading towards better adoption of forecasting models.

Our contributions are as follows:

- We present an applied real-world case study where time-series forecasting is crucial in decision-making. Additionally, we describe the modeling process to make machine learning solution adoption smoother by the end user.
- We propose a practically usable framework of a mixture of experts. Also, our training methodology of meta-learning on machine learning models outperformed industry benchmarks.
- We share results from our implementation backtested on real data from a major airline carrier, aiming to make high-stakes decisions that significantly impact the majority of the annual revenue.

1.1 Background

In the context of airline cargo, mid-term refers to a forecasting period of up to six months ahead, effectively covering everything beyond the typical cargo booking window of up to 30 days before

the flight departure. Mid-term forecasting is crucial because it allows stakeholders to have a proper idea of upcoming trends and demands, enabling them to make informed decisions on mid-term allotment requests. Mid-term allotment contracts in air cargo are agreements between airlines and shippers or freight forwarders to allocate a certain amount of cargo space on flights over a defined period, typically ranging from a few months to a couple of years [16, 21, 33]. These contracts facilitate more accurate demand forecasting and reduce risks associated with fluctuating cargo volumes and market rates [2, 31]. For shippers and freight forwarders, these contracts provide guaranteed access to cargo space, enhancing service reliability and customer satisfaction, especially during peak seasons or unexpected demand surges [30, 41]. The contracts typically detail the volume of cargo space, routes, duration, and pricing terms, with negotiations often considering historical cargo volumes, projected future needs, and current market conditions [2]. Despite the advantages, challenges arise from the dynamic nature of global trade, economic fluctuations, and unexpected events like the COVID-19 pandemic, which impact cargo volumes and market rates. Flexibility and adaptive strategies are essential in managing these contracts amidst such uncertainties [4, 12]. Overall, mid-term allotment contracts are pivotal in ensuring stability and predictability in the air cargo industry, benefiting both airlines and their customers by providing a structured approach to capacity allocation and revenue management [13, 39].

1.2 Related Work

Forecasting techniques have significantly evolved, with traditional methods like ARIMA and exponential smoothing remaining popular due to their simplicity and effectiveness in stable environments [11]. Recent advancements in computational capabilities have led to the development of more sophisticated models, particularly neural network-based approaches. These models, including convolutional neural networks (CNNs) and recurrent neural networks (RNNs), have demonstrated superior performance in capturing complex, non-linear relationships in data, thereby enhancing forecasting accuracy [22, 36]. Additionally, attention-based neural network architectures have also shown some promise in time series forecasting [26, 35].

Applications of such techniques in airline revenue management have picked up recently. For instance, application of machine learning-based solutions for dynamic pricing and passenger flow estimation [23, 40] in passenger business. Recent research on gravity models, dynamic linear model, fuzzy regression forecasting models and simplified variants of neural networks have shown promising results in air cargo industry [1, 8, 9, 32]. Historically, there have been fewer publications on air cargo demand forecasting, despite its considerable importance, especially when compared to the extensive research on air passenger demand forecasting [3].

2 DECISION FACTORS

In this section, we explain the two primary factors that are responsible for mid-term allotment decision-making: demand estimation and, airline network and capacity.

2.1 Demand estimation

Estimating the upcoming mid-term demand is crucial for airline cargo carriers as it allows them to allocate capacity efficiently between long-term stable contracts and short-term volatile bookings, which typically arrive within 30 days. Demand for air cargo is influenced by various factors including the origin and destination, product category, global events (i.e., COVID-19 and Suez Canal blockade), and macro trends [24]. For instance, countries like Columbia and Ethiopia are significant exporters of flowers, resulting in a peak in cargo volumes before Valentine's Day and Mother's Day. Similarly, there is a notable increase in import volumes of electronics and technology components from Asia to Europe and North America during Black Friday and Cyber Monday. Moreover, macroeconomic policies and trends play a critical role in influencing both the demand and supply of products.

2.2 Network and capacity

Passenger airlines with significant revenue from air cargo typically operate a hub-and-spoke network model [6], where flights are routed through a central hub to various destinations. This model facilitates efficient passenger transfers and operational management. In contrast, air cargo logistics often prioritize the origin and destination points without significant regard to intermediary stops, focusing instead on cost-effectiveness and efficient routes for their shipments. To optimize the overall logistics, a comprehensive network-wide consideration of capacity is essential. This includes evaluating forward-looking schedules and sellable capacities, encompassing both passenger and cargo flights for each flight leg [25]. Such a holistic approach ensures that both passenger and cargo needs are met efficiently, maximizing the use of available resources and improving overall service delivery.

3 FORECASTING FRAMEWORK

We want to predict the total demand in terms of expected shipped weight, volume, and revenue for each week for each origin and destination (O&D) for the next 6 months. Airline carriers may have a lot of cargo destinations that they fly to and therefore, the possible O&D combinations for forecasting could be in the order of tens of thousands. To make the problem simpler, we only focus on O&D that contribute to the top 90% of cargo revenue for the airline. We call this group of O&Ds the significant cluster. This reduces the problem space to a few 1000s of O&Ds. Through exploratory data analysis, we found that most of the variability and important patterns in the data are captured by the high-frequency signals at the departure date level. Therefore in our approach, we make predictions at the O&D - departure date level, and then aggregate them up to the O&D - departure week level.

3.1 Models

We experimented with eight different statistical models for baseline analysis. Additionally, the industry-wide benchmark, used by the majority of airlines, relies on year-over-year predictions. We used Croston Classic, Historic Average, Auto ETS, Window Average, Dynamic Optimized Theta (DOT), Simple Exponential Smoothing (SES), Holt-Winters, and Seasonal Naive for statistical baseline [7, 10, 15, 19, 20]. With their strengths and weaknesses, these

methods collectively aim to capture a broad range of trend and seasonality patterns in the time series that help set a baseline for building advanced models. Furthermore, our study focused on the implementation of three machine learning model architectures.

3.1.1 DNN-LADD. We implemented a simplified deep neural network (DNN) based primarily on fully connected layers assuming independent and identically distributed data points as inputs. Each timestamp point is considered independent of others without considering the sequential nature of the time series. The only time series considered is a small period that "Looks Around the Departure Dates" (LADD). These features help capture the static information in the time series and any seasonal and calendar events that the airline cares about. The LADD features are currently set to be equal to harmonic features [29] and are encoded using a bidirectional LSTM [18] as illustrated in the figure 1. All categorical features go through their embedding layers. These embeddings and encodings along with the dense features are concatenated together before being fed to the fully connected layers. The primary hypothesis for implementing this architecture is to capture patterns based on context rather than driven by traditional time series concepts, leveraging the flexibility of neural networks to model complex relationships within the data.
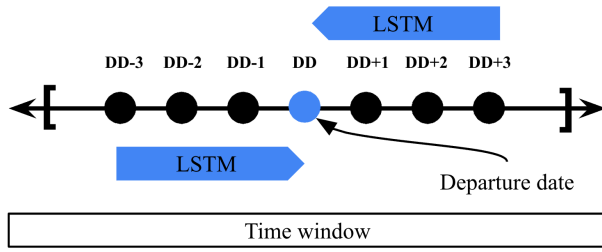

Figure 1: Illustration of look around departure date (LADD) method using the bidirectional LSTM

3.1.2 NBEATS. This is a state-of-the-art deep-learning-based time-series forecasting model proven to work well on short-horizon forecasting [28]. It takes a time-series input to predict the entire forecasting horizon in one shot. The model tries to sequentially learn the residual from the inputs with the use of stacks. By default, NBEATS is a univariate time-series model, that is, it only takes time-series being forecasted as input. To incorporate additional input features, we modified its architecture to add context in the form of encodings concatenated together to the input vector. We observed that this model performs better when patterns are seasonal and struggles when data is sparse or highly context-driven such as capacity or events. LADD and embedding features are learned the same way as the DNN-LADD model. Dense features are embedded using an attention block [38].

3.1.3 TFT. Temporal Fusion Transformer (TFT) is one of the transformational methods in time-series forecasting utilizing the transformer architecture [26]. Unlike NBEATS, TFT is designed to support input features and can learn from static metadata, time-varying

past inputs, and time-varying a priori known future inputs. It incorporates LSTM layers and a multi-head self-attention transformer block to enable efficient information flow, enhanced by skip connections. However, being a transformer-based architecture, this model is computationally expensive and requires significant memory to train, especially for long-horizon forecasting.

3.2 Meta learning

Meta-learning is an advanced approach in artificial intelligence aimed at training systems to effectively learn new tasks or adapt to novel situations based on past experiences, rather than merely mastering specific tasks. This paradigm, often described as "learning to learn," enables a system to acquire new concepts or tasks with minimal additional training quickly. Given the variability in data sizes across different Origin-Destination (O&D) pairs, traditional machine learning models can easily overfit to O&Ds with larger datasets. Meta-learning addresses this issue by implementing a 2-step optimization process rather than standard gradient descent, as detailed in the Model-Agnostic Meta-Learning (MAML) framework [14].

In this approach, each O&D is treated as a distinct task, and gradients for model updates are derived by moving toward the local minima for a set of O&Ds. This is achieved through a stochastic gradient descent update for a particular O&D on a model copy in the inner loop using the support set. Concurrently, the gradients for the main model updates are gathered using a query set from the updated model copy in the outer loop. During inference, the model undergoes fine-tuning using a subset of each O&D's training data for a few steps before making predictions. A significant advantage of this meta-learning approach is its applicability to O&Ds with limited historical data, such as the bottom 10% of O&Ds. By fine-tuning the model with minimal steps, accurate predictions can be made without extensive retraining, provided the data supports the creation of such O&Ds.

Algorithm 1 O&D Meta-Learning

Require: $p(\mathcal{D})$: distribution over O&D
Require: α, β: step size hyperparameters
1: Randomly initialize θ
2: **while** not done **do**
3:     Sample batch of O&D: $\mathcal{D}_i \sim p(\mathcal{D})$
4:     **for all** $\mathcal{D}_i$ **do**
5:         Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{D}_i}(f_\theta)$ using support set
6:         Compute adapted parameters with gradient descent: $\theta'_i = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{D}_i}(f_\theta)$
7:     **end for**
8:     Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{D}_i \sim p(\mathcal{D})} \mathcal{L}_{\mathcal{D}_i}(f_{\theta'_i})$ using query set
9: **end while**

In the context of the meta-learning algorithm, let $p(\mathcal{D})$ represent the distribution over tasks, where each O&D (Origin-Destination pair) is considered a separate task. The step size hyperparameters, α and β, are employed in the inner and outer loop updates respectively, governing the magnitude of parameter adjustments during gradient descent. The model parameters are denoted by θ, while θ'_i refers to the adapted parameters for a specific task $\mathcal{D}_i$ after the inner

loop update. The loss function for a task $\mathcal{D}_i$ is represented by $\mathcal{L}_{\mathcal{D}_i}$, which evaluates the model's performance on that task. Finally, f_θ signifies the model parameterized by θ, encapsulating the function learned by the meta-learning algorithm. These variables collectively enable the systematic optimization and fine-tuning of the model across diverse tasks as mentioned in algorithm 1.

3.3 Mixture of experts

For each O&D pair $\mathcal{D}_i$, the model f_θ with the lowest loss value $\mathcal{L}$ on the validation set is selected, as denoted in equation 1. The selected model $f_{\theta^*_{\mathcal{D}}}$ is then used to make predictions on the test set for the corresponding O&D pair. By selecting the best-performing model for each O&D pair based on the validation loss, the overall prediction error across all O&D pairs is minimized. This approach leverages the strengths of different models, acknowledging that no single model consistently performs best across all O&D pairs. The implementation of this framework involves training each model on the training set, evaluating their performance on the validation set to compute the loss for each O&D pair, and then selecting and using the model with the lowest loss for making predictions on the test set. This process ensures that the forecasting approach is tailored to the specific characteristics of each O&D pair, ultimately aiming to minimize the overall prediction error through a data-driven model selection process.

$$f_{\theta^*_{\mathcal{D}_i}} = \arg\min_{f_\theta \in \mathcal{F}} \mathcal{L}_{\mathcal{D}_i}(f_\theta) \qquad (1)$$

We chose the mixture of experts approach to provide the most accurate and reliable forecasts for air cargo analysts, thereby enabling them to make well-informed decisions. This method allows us to leverage baseline forecasts for inference, particularly in scenarios where the existing modeling approaches fall short, ensuring comprehensive coverage across various origin and destination pairs. Furthermore, this approach is apt for applications with low-frequency inference due to high computational complexity and hence affordable to mid-term forecasts as it needs to be updated only a few times a month. Therefore, the mixture of experts framework is a practically viable solution, aligning with the operational needs.

4 RESULTS

The target variable chosen for all experiments is the total shipped cargo weight in kilograms on a given date for a given O&D pair. All models are trained from April 2019 to March 2023, and validated on a period starting from April 2023 to September 2023 [1].

The primary metrics chosen for evaluation are root mean squared error (RMSE), unweighted normalized root mean squared error (nRMSE), and weighted normalized root mean squared error (WnRMSE). The WnRMSE formulation is described in equation 2.

$$\text{WnRMSE} = \sum_{OD} w_{OD} \times \text{nRMSE}_{OD} \qquad (2)$$

where $w_{OD} = \frac{sum(\text{weight})_{OD}}{\sum_{OD} sum(\text{weight})_{OD}}$, $\text{nRMSE}_{OD} = \frac{\text{RMSE}_{OD}}{avg(\text{weight})_{OD}}$

---

[1]The exact dates and revenue figures cannot be included due to proprietary, privacy, and sensitivity restrictions

Table 1: Model performance and win ratio on significant cluster

Model Name	Model Type	Win ratio	nRMSE	WnRMSE
TFT	ML	0.245	1.493	1.042
Auto ETS	Stat	0.160	1.459	1.016
NBEATS	ML	0.151	1.440	1.000
DOT	Stat	0.087	1.447	1.006
DNN-LADD	ML	0.084	1.618	1.092
Holt-Winters	Stat	0.070	1.698	1.178
SES	Stat	0.068	1.001	1.037
Window Average	Stat	0.051	1.521	1.044
Historic Average	Stat	0.044	1.626	1.094
Croston Classic	Stat	0.027	1.489	1.030
Seasonal Naive	Stat	0.006	1.898	1.210

RMSE, by definition, is biased towards high-value data points, leading to higher magnitude values for bigger O&Ds in the airline network and potentially skewing network-wide RMSE values. Normalized RMSE (nRMSE) addresses this bias by normalizing RMSE values by the average of the target variable. Therefore, we calculate nRMSE for each O&D to ensure fair performance comparison across the network. We use a weighted mean of the individual nRMSE values to obtain the network-wide error, where weights are the total shipped weight of each O&D pair over the evaluation period. This approach provides a performance assessment of the overall network with proportionate relevance of O&Ds.

Individual model performance is presented in Table 1. Here, average nRMSE and WnRMSE are recorded for each model across the entire validation period. The win ratios are calculated for the significant cluster due to resource constraints. Advanced techniques like TFT and NBEATS are 2 of the top 3 best performing models within the significant cluster. Also, statistical methods like Auto ETS effectively capture seasonal trends, resulting in improved performance.

Table 2: Aggregated metrics for clustered O&Ds

O&D's	Sample Size	Share of Revenue	nRMSE	WnRMSE
Significant cluster	>2000	0.90	1.04	0.52
Top 100	100	0.37	0.34	0.30
101-500	400	0.30	0.56	0.51
501-1000	500	0.12	0.75	0.69
Above 1001	>1000	0.11	1.32	0.99

To do further analysis, we ranked the entire dataset based on weight contribution. A significant O&D cluster is a subset of the dataset that contributes to 90% of the total revenue across the airline network in the year before the validation period. This significant cluster contains 2000+ O&D pairs. Aggregated metrics are reported in Table 2. Both RMSE and WnRMSE indicate performance degrades towards the tail end of the ranked O&D. Additionally, the significant cluster has low WnRMSE. This implies that the forecast can capture

trends where variability averages out due to volume as compared to a more stochastic tail end. In table 3, we use the same dataset slices and evaluate the winning ratios of the machine learning methods as compared to (a) statistical baseline and (b) industry benchmark i.e. year-over-year predictions. Although, the winning percentage for machine learning models is about 50% as compared to the baseline, we outperformed the industry benchmark overall. Hence, the utility of a mixture of experts is relevant as it provides a fallback option where machine learning models are not performing better than the statistical baselines.

Table 3: Win ratios of ML models for clustered O&Ds

O&D's	Statistical baseline	Industry benchmark
Significant cluster	0.49	0.94
Top 100	0.42	0.96
101-500	0.49	0.96
501-1000	0.46	0.93
Above 1001	0.50	0.94

5 FUTURE WORK

In the future, we intend to extend this work to other target variables that are as crucial as shipped weight for the airline in its decision-making. These include shipped volume (in m^3), chargeable weight (normalized metric between gross weight and volumetric weight), yield, and/or revenue.

Furthermore, the proposed mixture of experts framework currently misses out on integrating information from models other than the best performing one on a particular O&D pair, limiting its predictive power. Therefore, we propose extending the framework to a two-step forecasting process. In the first step, we train simpler models such as statistical benchmarks, decision stumps, and simple DNN networks. The predictions from these models would serve as additional features for advanced deep learning models like NBEATS and TFT, enhancing overall forecast accuracy. Integrating information from these simpler models allows the deep learning model to focus on complex, context-driven scenarios without being burdened with learning the simple trend and seasonality patterns already captured by the statistical models in the first step.

Lastly, we would like to build an attribution system to measure the revenue uplift from decisions made using improved forecasting techniques. Thereby, validating the overall approach for time series forecast in the air cargo industry.

6 CONCLUSION

The field study presented in this paper demonstrates the efficacy of advanced machine learning models in improving mid-term demand forecasting for the air cargo industry. Our experiments show that machine learning models perform better than statistical baseline models. Additionally, mixture of experts which combines both statistical and machine learning models, significantly outperforms machine learning only approaches. The real-world application of these models within a major air cargo carrier underscores their practical utility and potential for broader adoption in other volatile market environments. Future research will focus on extending this framework to additional target variables and enhancing the predictive power by incorporating a two-step forecasting process. This work lays a strong foundation for further advancements in the adoption of machine learning techniques in air cargo demand forecasting, ultimately contributing to more efficient and informed decision-making processes in the industry.

REFERENCES

[1] David W Alexander and Rico Merkert. 2021. Applications of gravity models to evaluate and forecast US international air freight markets post-GFC. *Transport Policy* 104 (2021), 52–62.

[2] Kannapha Amaruchkul, William L Cooper, and Diwakar Gupta. 2011. A note on air-cargo capacity contracts. *Production and Operations Management* 20, 1 (2011), 152–162.

[3] Juan Gerardo Muros Anguita and Oscar Díaz Olariaga. 2023. Air cargo transport demand forecasting using ConvLSTM2D, an artificial neural network architecture approach. *Case Studies on Transport Policy* 12 (2023), 101009.

[4] Bjoern Becker and Nadja Dill. 2007. Managing the complexity of air cargo revenue management. *Journal of Revenue and Pricing Management* 6 (2007), 175–187.

[5] John S Billings, Adriana G Diener, and Benson B Yuen. 2003. Cargo revenue optimisation. *Journal of Revenue and Pricing Management* 2 (2003), 69–79.

[6] Jan K Brueckner, Nichola J Dyer, and Pablo T Spiller. 1992. Fare determination in airline hub-and-spoke networks. *The RAND Journal of Economics* (1992), 309–333.

[7] Chris Chatfield. 1978. The Holt-winters forecasting procedure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 27, 3 (1978), 264–279.

[8] Shu-Chuan Chen, Shih-Yao Kuo, Kuo-Wei Chang, and Yi-Ting Wang. 2012. Improving the forecasting accuracy of air passenger and air cargo demand: the application of back-propagation neural networks. *Transportation Planning and Technology* 35, 3 (2012), 373–392.

[9] Tsung-Yu Chou, Gin-Shuh Liang, and Tzeu-Chen Han. 2011. Application of fuzzy regression on air cargo volume forecast. *Quality & Quantity* 45, 6 (2011), 1539–1550.

[10] John D Croston. 1972. Forecasting and stock control for intermittent demands. *Journal of the Operational Research Society* 23, 3 (1972), 289–303.

[11] Jan G De Gooijer and Rob J Hyndman. 2006. 25 years of time series forecasting. *International journal of forecasting* 22, 3 (2006), 443–473.

[12] Wouter Dewulf. 2014. *The strategy of air cargo operators: about carpet sellers and cargo stars*. Ph. D. Dissertation. University of Antwerp.

[13] Bo Feng, Yanzhi Li, and Huaxiao Shen. 2015. Tying mechanism for airlines' air cargo capacity allocation. *European Journal of Operational Research* 244, 1 (2015), 322–330.

[14] Chelsea Finn, Pieter Abbeel, and Sergey Levine. 2017. Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *Proceedings of the 34th International Conference on Machine Learning (Proceedings of Machine Learning Research, Vol. 70)*, Doina Precup and Yee Whye Teh (Eds.). PMLR, 1126–1135. https://proceedings.mlr.press/v70/finn17a.html

[15] José Augusto Fioruci, Tiago Ribeiro Pellegrini, Francisco Louzada, and Fotios Petropoulos. 2015. The optimised theta method. *arXiv preprint arXiv:1503.03529* (2015).

[16] Rolf Hellermann. 2006. *Capacity options for revenue management: theory and applications in the air cargo industry*. Vol. 575. Springer Science & Business Media.

[17] Hong-jun Heng, Bing-zhong Zheng, and Ya-jing Li. 2009. Study of SVM-based air-cargo demand forecast model. In *2009 International Conference on Computational Intelligence and Security*, Vol. 2. IEEE, 53–55.

[18] Sepp Hochreiter and Jürgen Schmidhuber. 1997. Long short-term memory. *Neural computation* 9, 8 (1997), 1735–1780.

[19] Rob J Hyndman and George Athanasopoulos. 2018. *Forecasting: principles and practice*. OTexts.

[20] Garima Jain and Bhawna Mallick. 2017. A study of time series models ARIMA and ETS. *Available at SSRN 2898968* (2017).

[21] Raja G Kasilingam. 1997. Air cargo revenue management: Characteristics and complexities. *European Journal of Operational Research* 96, 1 (1997), 36–44.

[22] Taewook Kim and Ha Young Kim. 2019. Forecasting stock prices with a feature fusion LSTM-CNN model using different representations of the same data. *PloS one* 14, 2 (2019), e0212320.

[23] Ismail Koc and Emel Arslan. 2021. Dynamic ticket pricing of airlines using variant batch size interpretable multi-variable long short-term memory. *Expert Systems with Applications* 175 (2021), 114794.

[24] Franziska Kupfer, Hilde Meersman, Evy Onghena, and Eddy Van de Voorde. 2017. The underlying drivers and future development of air cargo. *Journal of*

Air Transport Management 61 (2017), 6–14.

[25] Tatsiana Levina, Yuri Levin, Jeff McGill, and Mikhail Nediak. 2011. Network cargo capacity management. *Operations Research* 59, 4 (2011), 1008–1023.

[26] Bryan Lim, Sercan Ö Arık, Nicolas Loeff, and Tomas Pfister. 2021. Temporal fusion transformers for interpretable multi-horizon time series forecasting. *International Journal of Forecasting* 37, 4 (2021), 1748–1764.

[27] Bryan Lim and Stefan Zohren. 2021. Time-series forecasting with deep learning: a survey. *Philosophical Transactions of the Royal Society A* 379, 2194 (2021), 20200209.

[28] Boris N Oreshkin, Dmitri Carpov, Nicolas Chapados, and Yoshua Bengio. 2019. N-BEATS: Neural basis expansion analysis for interpretable time series forecasting. *arXiv preprint arXiv:1905.10437* (2019).

[29] H.A. Panofsky and G.W. Brier. 1958. *Some Applications of Statistics to Meteorology.* Mineral Industries Extension Services, College of Mineral Industries, Pennsylvania State University. https://books.google.com/books?id=UyRRAAAAMAAJ

[30] Andrea Popescu, Pinar Keskinocak, and I al Mutawaly. 2010. *Intermodal transportation: Moving freight in a global economy* (2010), 209–237.

[31] Chan Seng Pun and Diego Klabjan. 2010. Air Cargo Allotment Planning. (2010).

[32] Yesid Rodriguez, Wilmer Pineda, and Oscar Diaz Olariaga. 2020. Air traffic forecast in post-liberalization context: a Dynamic Linear Models approach. *Aviation* 24, 1 (2020), 10–19.

[33] Michael Sales and Sebastiaan Scholte. 2023. *Air cargo management: Air freight and the global supply chain.* Routledge.

[34] Omer Berat Sezer, Mehmet Ugur Gudelek, and Ahmet Murat Ozbayoglu. 2020. Financial time series forecasting with deep learning: A systematic literature review: 2005–2019. *Applied soft computing* 90 (2020), 106181.

[35] Shun-Yao Shih, Fan-Keng Sun, and Hung-yi Lee. 2019. Temporal pattern attention for multivariate time series forecasting. *Machine Learning* 108 (2019), 1421–1441.

[36] Sima Siami-Namini, Neda Tavakoli, and Akbar Siami Namin. 2018. A comparison of ARIMA and LSTM in forecasting time series. In *2018 17th IEEE international conference on machine learning and applications (ICMLA)*. IEEE, 1394–1401.

[37] José F Torres, Dalil Hadjout, Abderrazak Sebaa, Francisco Martínez-Álvarez, and Alicia Troncoso. 2021. Deep learning for time series forecasting: a survey. *Big Data* 9, 1 (2021), 3–21.

[38] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. *Advances in neural information processing systems* 30 (2017).

[39] Masato Wada, Felipe Delgado, and Bernardo K Pagnoncelli. 2017. A risk averse approach to the capacity allocation problem in the airline cargo industry. *Journal of the Operational Research Society* 68 (2017), 643–651.

[40] Jiangni Yu. 2022. Short-term airline passenger flow prediction based on the attention mechanism and gated recurrent unit model. *Cognitive Computation* 14, 2 (2022), 693–701.

[41] Rudolf Zivcic. 2022. Applicability of Dynamic Pricing Concept for Air Cargo. (2022).