DATA-DRIVEN APPROACHES FOR JET FUEL DEMAND FORECASTING

ABSTRACT

Jet fuel demand forecasting plays an important role in optimizing supply chain operations in the aviation market. In particular, fuel distributors need to accurately estimate the demand in order to avoid under- and over-stocking. However, many industry practitioners tackle this task using deterministic or expertise-based models, and there are not many studies that analyze the jet fuel demand forecasting problem with machine learning models. In this work, we assess the performance of data-driven approaches on a relevant amount of data related to the Danish market, gathered from DCC & Shell Aviation Denmark A/S. In particular, we compare the predictive performance obtained with traditional time series based models (SARIMAX), LSTM sequence-to-sequence neural networks, and hybrid models. While developing such models, the main challenges lie in the needed forecasting horizon and in the novelty of the approach. Indeed, to support the company's sourcing strategy, the fuel demand needs to be predicted for the upcoming month. To assess the reliability of the data-driven approaches, three different case studies are proposed and analyzed.

***K*eywords** Aviation, Demand forecasting, time series models, LSTM, hybrid models.

1 Introduction

Fuel consumption is a crucial research area in transportation business and management. In particular, in the aviation industry, jet fuel is one of the largest operational expenses of airline companies [1]. To seek protection against sudden price fluctuations, many airlines resort to the use of fuel hedging strategies [2]. One of the most basic hedging strategies is referred to as a jet fuel swap, which is an agreement where a good with a volatile price is exchanged at a fixed price over a specified time frame. To optimize this strategy, airlines take into account their tentative flight schedule, which is usually made publicly available for a long time horizon. Fuel providers also tend to use these flight schedules to support the decision-making process related to fuel sourcing. However, in many cases, the schedule differs from the actual number of flights eventually planned, resulting in the risk of under- or over-stocking. This highlights how being able to accurately estimate the fuel demand represents a crucial step in achieving an efficient sourcing strategy. However, while many researchers focused on fuel demand forecasting on a macroeconomic level, very few case studies have been published related to short-term jet fuel demand forecasting from a supply chain point of view. Moreover, many industry practitioners do not use advanced models for prediction and mostly rely on the tentative schedule provided by the airlines.

In this work, we evaluate the effectiveness of machine learning models in providing a reliable forecast of the jet fuel demand in the Danish market, using data from DCC & Shell Aviation Denmark A/S. The project includes three case studies. The first two cases are related to the volume of fuels purchased by Ryanair and Fly emirates, and the final case study is related to the total volume of fuel ordered by all the airlines flying through Copenhagen Airport (CPH). It should be noted that, in the third case study only airlines that are customers of DCC & Shell Aviation have been considered. However, it still includes a significant number of airlines as DCC & Shell Aviation Denmark A/S covers approximately 35% of the CPH market (August 2022) and is the largest company operating in the market.

The remainder of this paper is organized as follows. Section 2 analyzes the current literature and similar problems to motivate the chosen methods together with the challenges in demand forecasting that this project is facing. In section 3, we provide a summary of the machine learning models used in the experiments. In section 4, three case studies are used to evaluate the effectiveness of the data-driven strategies. Finally, section 5 provides some conclusions.

2 Related work

Despite the economical and managerial impacts of being able to correctly forecast jet fuel demand, there are not many works showing the potential of data-driven approaches to tackle this task. Indeed, to the best of our knowledge, there is currently no publicly available research on aviation fuel demand forecasting using deep learning or time series models. Previous studies in this area have primarily focused on modeling single aircraft consumption rather than market demands. For example, Li et al. [3] and Bauman et al. [4] presented studies on modeling aircraft and flight-specific statistics, such as wind and time, to forecast consumption based on aircraft type. Another topic researched in the field is flight time prediction, with studies such as Zhu et al. [5] using deep learning to predict flight time for more efficient fuel loading scheduling. While these studies are relevant, our research aims to assess jet fuel demands from a wider perspective and with a purely data-driven approach.

Time series analysis is the most natural approach to the problem as it is stated. Many techniques, such as regression (Božidar [6], Su et al. [7]) or simple time series models can be applied as starting points to forecast future values given previous time steps. For instance, Wu et al. [8] and Parasyris et al.[9] applied $ARIMA$ and $SARIMA$ models to forecast future tourist arrivals or meteorological variables. Even considering that the 2 papers achieve good results, more advanced models can be deployed. Among the traditional machine learning models, anyway, SARIMAX is known to perform very well given its ability to include history and many variables at the same time. Plenty of examples of SARIMAX usage can be found in energy demand predictions: Taşpınar et al. [10], and Yucesan et al.[7] predicted natural gas demands in regions of Turkey, Su et al. predicted hourly natural gas demands [11].

As mentioned before, more complex models capable of taking into consideration nonlinear relationships or performing internal feature selection can be deployed. In the most recent years, artificial neural networks (ANN) became of primary interest among many machine learning techniques. Most of the research presented above tends to compare SARIMAX with neural networks [10][11][7] [12] obtaining comparable or even better results. LSTM especially are known to perform very well, sometimes outperforming simpler neural networks or time series approaches.

Even more complex are hybrid methods. By hybrid it is intended a prediction model based on 2 or more methods somehow merged together. Since most machine learning algorithms are designed for a particular dataset or task, combining multiple ML algorithms can greatly improve the overall result by either helping tune one another, generalize, or adapt to unknown tasks. [13] In literature it can be found several hybrid and ensemble methods applied to time series: Petropoulos et al. used Bagging to predict time series [14] and Pan et al. created a new model by merging together Artificial Neural Networks and autoregression in a so-called "extreme learning machine" [15]. Anyway, most of the hybrid methods used for time series forecasting in literature are based on the stacking of linear methods and non-linear methods through the predictions of residuals. Galadino et al.[16] improved the accuracy of the models that compose the hybrid system from successive corrections using the residual modeling. So, a set of models can be generated to forecast each pattern (linear and nonlinear), where the posterior model corrects the last one through the modeling of their residuals. Parasyris et al.[9] used a $SARIMA$ model residuals to fit a univariate $LSTM$ neural network to catch the nonlinearity and improved the forecast scores. Similarly, Wu et al. [8] used the same technique, deploying an $ARIMA$ model followed by an $LSTM$ fit on residuals.

The first difficulty to overcome is data collection. To have a reliable prediction long historical data is required. Not only that but also data must be meaningful. This leads to the second common issue in the field: unpredictable factors that may affect demand or offer. The companies on either side of the supply chain impact the company's ability to keep promises. Examples can be: the recent Covid-19 pandemic, resulting in a drop in flights due to restrictions, the sudden acquisition of new contracts, and acquiring new routes or markets can change a lot one airline flights volumes. Improper planning and stocking can produce a massive loss even in large and organized companies. It is important to have a plan for handling extreme variations and edge cases. A model to be successful and resilient to changes must include all the possible relevant factors influencing the target variable. It is therefore a primary importance challenge to identify and embed the correct features.

3 Methods

3.1 SARIMA and SARIMAX

Time series models such as Seasonal Autoregressive Integrated Moving Average (SARIMA) and Seasonal Autoregressive Integrated Moving Average with Exogenous Variables (SARIMAX) are commonly utilized in demand forecasting due to their efficacy in capturing and modeling the inherent patterns and structures present in time series data. SARIMA, as the acronym suggests, considers both the autoregressive (AR) and moving average (MA) components of a time series in addition to any seasonality or periodicity exhibited by the data. The "I" in SARIMA signifies that the model involves differencing of the data to attain stationarity. The model is defined by three key parameters, namely p, d, and q, which indicate the orders of the autoregressive, integrated, and moving average components, respectively. The seasonal aspect of the model is represented by the parameters P, D, and Q, which correspond to the orders of the seasonal autoregressive, seasonal integrated, and seasonal moving average components, respectively.

SARIMA models are comprised of two fundamental components, AR and MA, and do not require any external variable. The model also incorporates an option for integration (I) and the embedding of seasonality, where the seasonal period, denoted by s, defines the number of observations that constitute a seasonal cycle. For instance, daily data may exhibit weekly seasonality with $s = 7$, monthly seasonality with $s = 30$, or yearly seasonality with $s = 365$. The value of s is assumed to be fixed in the series. SARIMA$(p, d, q)(P, D, Q, s)$ introduces three additional parameters that adjust the autoregressive, integration, and moving average values of the seasonal aspect of the model. The seasonal component of the model consists of terms that are analogous to the non-seasonal components but involve back shifts of the seasonal period. The operator Δ_s^D applies differentiation to the seasonal component of the model. A $SARIMA(p, d, q)(P, D, Q, s)$ can be percieved as the product of 2 processes: $ARMA(P, Q)$, utilizing seasonal time lags and the L^s operator instead of the conventional lag operator, P, and Q are again seasonal time lags; and a normal $ARIMA(p, d, q)$ applied to $\Delta_s^D yX_t$ [17]:

$$ARMA(P, Q) : \quad \Delta_s^D X_t = \theta(L^s)^P \Delta_s^D X_t + \phi(L^s)^Q \Delta_s^D \varepsilon_t + \Delta_s^D \varepsilon_t$$

Absorbing the constants into the polynomials we have the concise form:

$$\Theta(L^s)^p \Delta_s^D X_t = \Phi(L^s)^Q \Delta_s^D \varepsilon_t$$

Same process for $ARIMA$:

$$ARIMA(p, d, q) : \quad \Delta^d X_t = \Theta(L)^p \Delta^d X_t + \Phi(L)^q \Delta^d \varepsilon_t + \Delta^d \varepsilon_t$$

Absorbing the constants into the polynomials we have the concise form:

$$\Theta(L)^p \Delta^d X_t = \Phi(L)^q \Delta^d \varepsilon_t$$

This last model is now applied to $\Delta_s^D X_t$ by multiplying the seasonal model with it, giving the general equation of a $SARIMA(p, d, q)(P, D, Q, s)$ process:

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D X_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \varepsilon_t$$

SARIMAX is an extension of SARIMA that allows for the inclusion of exogenous variables in the model. Exogenous variables are external variables that may have an impact on the time series being modeled, such as the price of oil or the number of flights being operated by an airline in the case of jet fuel demand forecasting. The "X" in SARIMAX stands for "Exogenous", and the model is specified by additional parameters that capture the relationship between the exogenous variables and the time series being modeled. In practice, the process of fitting a SARIMA or SARIMAX model involves selecting the appropriate values of the p, d, q, P, D, and Q parameters based on an analysis of the time series data. This can be done using statistical techniques such as maximum likelihood estimation or Bayesian methods. Once the model has been fit to the data, it can be used to generate forecasts of future values of the time series, as well as to identify and analyze any patterns or structures in the data.

In the case where the observed time series is driven by some "forcing" time series, which may or may not have a causal effect on the observed series, a multivariate time series model is required. A multivariate time series consists of more than one time-dependent variable, and each variable depends not only on its past values but also has some dependency on other variables. The target variable in this case depends on the extra variables, which will be included as exogenous variables in the models. In Singh et al. [18], the (S)ARMA model has been developed by additional exogenous variables such as temperature, wind, etc., giving rise to the models (S)ARIMAX or (S)ARMAX. In order to enrich the predictive power of the model, the feature engineering process is revealed to be fundamental in producing useful variables.

A $SARIMAX(p, d, q)(P, D, Q, s)$ process is defined by adding the exogenous terms to the $SARIMA$ equation:

$$\Theta(L)^p \theta(L^s)^P \Delta^d \Delta_s^D y_t = \Phi(L)^q \phi(L^s)^Q \Delta^d \Delta_s^D \varepsilon_t + \sum_{i=1}^{n} \beta_i x_t^i$$

Where y_t is the target variable, there are n exogenous variables at each time step t, denoted by x_t^i for $i \leq n$ with the coefficients β_i.

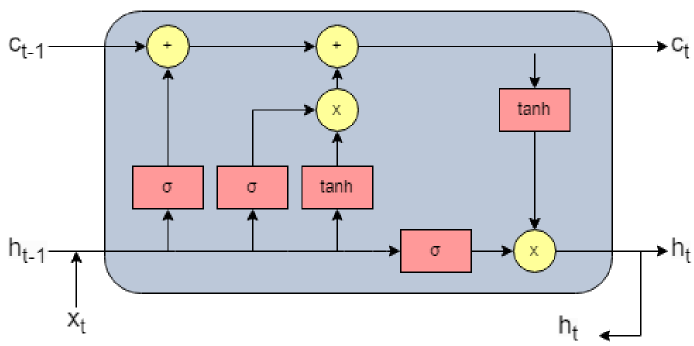
The SARIMA and SARIMAX models are robust tools for demand forecasting that enable the capture and modeling of intricate structures and patterns present in time series data, including trend and seasonality, and the influence of exogenous variables. These models provide analysts and forecasters with valuable insights into the behavior of the time series under analysis and facilitate the generation of accurate and reliable forecasts of future values.

3.2 LSTM Neural Networks

Recurrent Neural Network (RNN) models have been widely employed in real-world forecasting problems, with most proposed models being based on shallow structures comprising a single hidden layer. [19] Long Short-Term Memory (LSTM) neural networks represent a specific type of RNNs. Unlike standard feed-forward neural networks, LSTMs possess feedback connections and can process not only individual data points, such as numbers, but also entire sequences of data, such as speech or time series. The LSTM architecture is designed to provide an RNN with short-term memory that can persist for thousands of timesteps, hence the name "long short-term memory."[20] A typical LSTM cell is depicted in Figure 1. An LSTM neural network is comprised of LSTM layers, with each layer composed of multiple units, each unit consisting of a cell, an input gate, an output gate, and a forget gate. The cell retains values over arbitrary time intervals, while the three gates control the inflow and outflow of information into and out of the cell.[20] At a given point in time t, the output of an LSTM is determined by three elements:

- The current long-term memory of the network — known as the cell state (c_{t-1})
- The output at the previous point in time — known as the previous hidden state (h_{t-1})
- The input data at the current time step (x_t)

Figure 1: An LSTM cell and its components. h_t is the hidden state, c_t is the cell state and x_t is the input value, all at time t.



The equations below show in detail how the h_t and c_t parameters are obtained following the flowcharts presented in 1. The lowercase variables represent vectors. Matrices W_q and U_q contain, respectively, the weights of the input and recurrent connections, where the subscript $_q$ can either be the input gate i, output gate $_o$, the forget gate f or the memory cell c, depending on the activation being calculated. In this section, we are thus using a "vector notation". So, for example, $c_t \in \mathbb{R}^h$ is not just one unit of one LSTM cell but contains h LSTM cell's units. The compact forms of the equations for the forward pass of an LSTM cell with a forget gate are:[20]

$$
\begin{aligned}
\mathrm{f}_t &= \sigma_g \cdot (W_f \cdot x_t + U_f \cdot h_{t-1} + b_f) \\
\mathrm{i}_t &= \sigma_g \cdot (W_i \cdot x_t + U_i \cdot h_{t-1} + b_i) \\
\mathrm{o}_t &= \sigma_g \cdot (W_o \cdot x_t + U_o \cdot h_{t-1} + b_o) \\
\widetilde{c}_t &= \sigma_c \cdot (W_c \cdot x_t + U_c \cdot h_{t-1} + b_c) \\
\mathrm{c}_t &= f_t \odot c_{t-1} + i_t \odot \widetilde{c}_t \\
\mathrm{h}_t &= o_t \odot \sigma_h(c_t)
\end{aligned}
$$

where the initial values are $c_0 = 0$ and $h_0 = 0$ and the operator $\odot$ denotes the Hadamard product (element-wise product). The subscript $_t$ indexes the time step. Variables:

- $x_t \in \mathbb{R}^d$: input vector of the LSTM unit.
- $f_t \in (0,1)^h$: forget gate's activation vector
- $i_t \in (0,1)^h$: input/update gate's activation vector
- $o_t \in (0,1)^h$: output gate's activation vector
- $h_t \in (-1,1)^h$: hidden state vector also known as output vector of the LSTM unit
- $\widetilde{c}_t \in (-1,1)^h$: cell input activation vector

- $c_t \in \mathbb{R}^h$: cell state vector
- $W \in \mathbb{R}^{hxd}$, $U \in \mathbb{R}^{hxh}$ and $b \in \mathbb{R}^h$: weight matrices and bias vector parameters which need to be learned during training where the superscripts $_d$ and $_h$ refer to the number of input features and a number of hidden units, respectively.

Activation functions:

- σ_g: sigmoid function, defined as[21]:
$$f(x) = \frac{1}{1 + e^{-x}}$$
where x is the input value.
- σ_c,σ_h: hyperbolic function, defined as[21]:
$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$
where x is the input value.

In this project, the input vector x_t consists of a sequence of fuel volumes and possibly some exogenous support variables. The LSTM networks are well-suited for processing, classifying, and making predictions based on time series data, as they can capture lags of unknown duration between significant events in a time series.

Given the problem setup with only four years of data and a monthly prediction window, it is not feasible to train the network using a monthly window. Data is available from 2018 or 2019 up to the end of August, considering one data point picked for each month we obtain a limited-size dataset. The limited data points would not suffice for a deep learning model. Therefore, the standard LSTM architecture has been modified to leverage as many training points as possible.

The network must predict the following k using n days, with the input array x_t and output array y_t comprising n and k elements, respectively. This type of network is known as a seq-to-seq LSTM neural network, which uses a sequence of input data points to predict an output sequence.

The network consists of two primary components: the training and validation stage, followed by the prediction stage. In this case, the training network error and validation phases are conducted with a daily increasing step, while the prediction phase is performed using a time window moving 30 days at a time. Various activation functions, including linear and nonlinear ones, are tested with this neural network.

This solution can utilize daily data points for training and predict the next months one month at a time, starting from every day.

The networks used in all the case studies are shallow, comprising only one or two LSTM layers and a single output layer. Occasionally, a single dense hidden layer is added to improve the exploitation of activation functions. As demonstrated by Bourennani et al. [19], flat networks often perform exceptionally well and are widely used. Additionally, $n = 60$ and $k = 30$ are fixed due to architectural limitations and the necessity to exploit weekly seasonality. This approach solely relies on learning previous data's future behavior, implying that any external aspect that affects the data will most likely result in the model's failure to predict the behavior without additional information. Examples may include changes in schedules due to business reasons or incorrect weather forecasts.

Considering the limitations of the univariate approach in handling external factors, supportive exogenous variables are included in the previous model. The network learns the relationship between the present confirmed volume and the future ones, including external factors directly related to it. Thus, the network must use both present and future exogenous factors. In this scenario, this means that we must encode future schedules, customer magnitude, and so on in the input.

Most available implementations use the last n time lags as predictors for the future target. However, this method has proven insufficient in effectively predicting future data since sudden changes in values were not learned, and the network was slow to adapt to seasonal or trend changes. The LSTM fails to exploit the most relevant information in our problem, i.e., the future values of exogenous variables, which are the best predictors for future volumes. Fortunately, in this project, future exogenous data is available. The most natural approach to incorporate future exogenous data is to train the network using past examples and providing the future as input to generate the output, similar to what is done in SARIMAX models. However, it is not feasible to incorporate exogenous data at a later point than the input in neural networks, as they require a fixed-size tensor as input and one or more outputs. Therefore, to include this information, the input tensor must be modified to include future exogenous data, which will assist in learning future behavior based on the present confirmed quantity and the relationship between present and past exogenous variables.

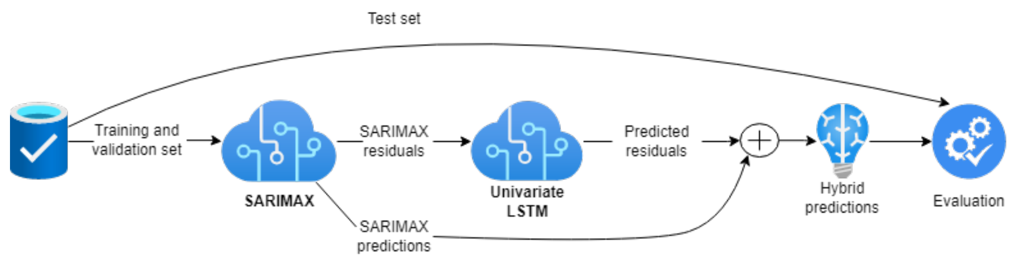
3.3 Residuals hybrid

The hybrid model is designed to address the issue of linearity and nonlinearity in real-world time series. Such time series typically exhibit both linear and nonlinear patterns, and a suite of models can be developed to forecast each pattern (i.e., linear and nonlinear), where the posterior model corrects the previous one through the modeling of their residuals. This approach has been explored in the literature [16, 22], and in this project, we aim to obtain the most performant linear method, SARIMAX, followed by an LSTM neural network to generate the final predictions. The architecture of our model is similar to that reported in Galdino et al. [16], with the key difference that our model does not rely on an iterative approach. Specifically, our model can be described as a stacking ensemble of two models: SARIMAX and LSTM.

The first model, SARIMAX, is trained to learn the linear pattern from the training data, and its performance is evaluated using a validation set. Subsequently, residuals are computed as the mathematical difference between real values and the predicted values, where the predicted values correspond to all the training sets. The difference between the training set and the predictions for the same period is treated as the new training set, and the same process is applied to the validation and test sets. Thus, the residuals become the dataset for the next stage of the model.

In the second stage, a univariate LSTM model is fitted to the residuals to predict future residuals. These predicted residuals are then added back to the predictions generated by the SARIMAX model, resulting in a final output: the hybrid predictions. The parameters of the network are estimated by assessing the scores of the new validation set, which consists of the residuals from the original validation set. Finally, the performance of the model is evaluated by comparing the hybrid predictions with the original test set.

Figure 2: Schema of the SARIMAX-LSTM hybrid model.



The employment of a hybrid method presents a key advantage, aside from the anticipated improvement in theoretical accuracy, namely the ability to leverage the predictive capacity of neural networks in conjunction with the interpretability of time series. This advantage is rooted in the core of the hybrid method, namely SARIMAX, which can furnish explanations pertaining to essential features and internal weights of time lags. Neural networks, in turn, serve to address errors generated by the first model, with the aim of identifying a means to model and predict such errors.

4 Case studies

4.1 Data

The data used in this study is composed of discrete data points collected over time intervals, enabling the tracking of changes over time. Each point represents a day, with data collected from 2019 (or 2018 for Fly Emirates) up to August 31, 2022. However, the data is limited as the actual SAP system was adopted between 2018 and 2019. The target variable in this study is the "Confirmed Quantity," which refers to the amount of fuel the company has sold to customers on a given day, measured in metric tonnes. The data is collected automatically during fueling operations at the airport and sent to the company's SAP database, then accessed through online software and exported as an Excel spreadsheet.

Given the nature of the problem, correctly identifying features is crucial. In this study, the approach followed was to replicate the decision process used by the company and replicate the features deemed important by trading experts. The goal is to enable the machine learning approach to better utilize the data and identify hidden patterns. It is worth noting that the feature identification process was not structured, and forward feature engineering was performed. Due to time constraints, only a subset of the relevant features were coded and included in the project, one after the other. The set of features was identified by observing the experts making predictions multiple times and validated together with them during the project development. Given the business-oriented nature of this study, a reduced number of exogenous features were preferred to focus on the model's interpretability and delve deeper into the aviation business.

The identified variables across the three case studies were: number of customers, magnitude of customers, number of daily scheduled flights, Weekend, and Market Share. Further details about the features and data used will be provided in the case studies section.

It is also essential to note that all the volumes reported in this study are re-scaled with different measures to ensure the protection of DCC Shell Aviation A/S interests.

4.2 Training and prediction process

All the case studies were divided into train, validation, and test sets. The model parameters were assessed using the validation set, with the Akaike Information Criteria used for the time series model. All other validation and performance assessments used the SMAPE error metric, with a lower error being preferred. The best methods identified in the validation set were then trained on the training and validation sets together to predict daily volumes. These daily predictions were then aggregated on a monthly basis and compared against actual volumes.

4.3 Performance metrics

Regrettably, due to the market-oriented nature of this project and the sensitive nature of the presented data, the utilization of an absolute error measure is not feasible, as it would entail the risk of competitors exploiting the information. As a result, SMAPE, a relative error metric, is employed as an alternative.

The Symmetric Mean Absolute Percentage Error (SMAPE), also referred to as the "Adjusted MAPE", was originally introduced by Armstrong [23] and is defined as follows:

$$SMAPE = \frac{1}{n}\sum_{t=1}^{n} \frac{|F_t - A_t|}{(|F_t| + |A_t|)/2}$$

SMAPE was chosen because being a relative percentage error metric is easy to interpret and confront also for not technical people.

4.4 Baseline: Company Approach

The project's fundamental premise is predicated on evaluating the performance of the company. The forecasts generated by DCC & Shell Aviation are limited to Copenhagen Airport as the company adopts a singular customer approach. Thus, this initial baseline shall serve as a direct reference point for comparing against the third case study and as a general indicator of acceptable performance. Additionally, the forecasts are expressed as monthly figures, warranting comparison solely against the monthly predictions. Performance is quantified by a score calculated using selected metrics. These scores represent the disparity between the company's forecast and the actual sales figures. Historical scores covering the period from March to August 2022 shall be presented, which corresponds to the Test set of both the Copenhagen Airport and Ryanair case studies. The company's performance data shall be detailed in Table 1.

Pred. interval	SMAPE
Historical	23.4
Test set	17.5

Table 1: DCC Energi & Shell Aviation A/S performances in predicting volumes for Copenhagen Airport. In Metric Tonnes for the absolute values.

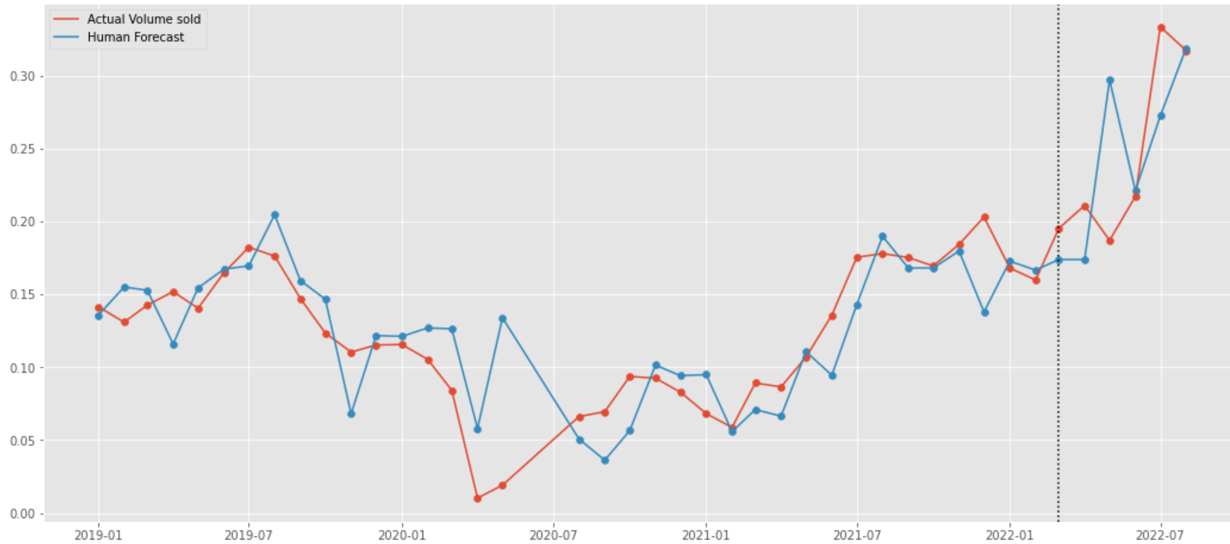
Finally, in figure 3 the performance of the company is shown as a graph.

4.5 Emirates

Emirates is one of two official carriers of the United Arab Emirates, operating from Garhoud, Dubai. It functions as a subsidiary of The Emirates Group, which is owned by the Investment Corporation of Dubai, a governmental entity.

The airline operates daily flights from Dubai to Copenhagen, as well as a weekly departure from Copenhagen to Chicago. Although the airline's schedule is considered dependable, the frequency of flights is relatively low, with only one per day and two on weekends. This is largely due to the considerable amount of fuel required for each flight, given the lengthy distances involved in direct routes to and from far-flung destinations.

Figure 3: DCC Energi & Shell Aviation A/S performances in predicting volumes for Copenhagen Airport.The test set is separated with a vertical line.



4.5.1 Company overview

The acquisition of the company as a customer occurred in 2018, providing access to a long historical dataset. However, this dataset is highly susceptible to cancellations or variations due to the low frequency of flights per day. Figure 4 depicts a stable period before the COVID-19 pandemic up to March 2020, which constitutes the dataset used for the present prediction.

Following this period, there was a considerable increase in volume that did not correspond to the recovery rate of other companies. As of June 2022, the airline has returned to its normal schedule. It is worth noting that Emirates tends to exceed the schedule published in OAG, as evidenced in Figure4 Considering the regularity of the flights and the expected regular behavior of the customer following the completion of this project, only the period prior to the COVID-19 outbreak until March 2020 is taken into account for this initial case study. Therefore, the training data consists of 2018 and 2019, while the validation is based on the first month of 2020. Additionally, the test set covers the period up to March 15th.

4.5.2 Models

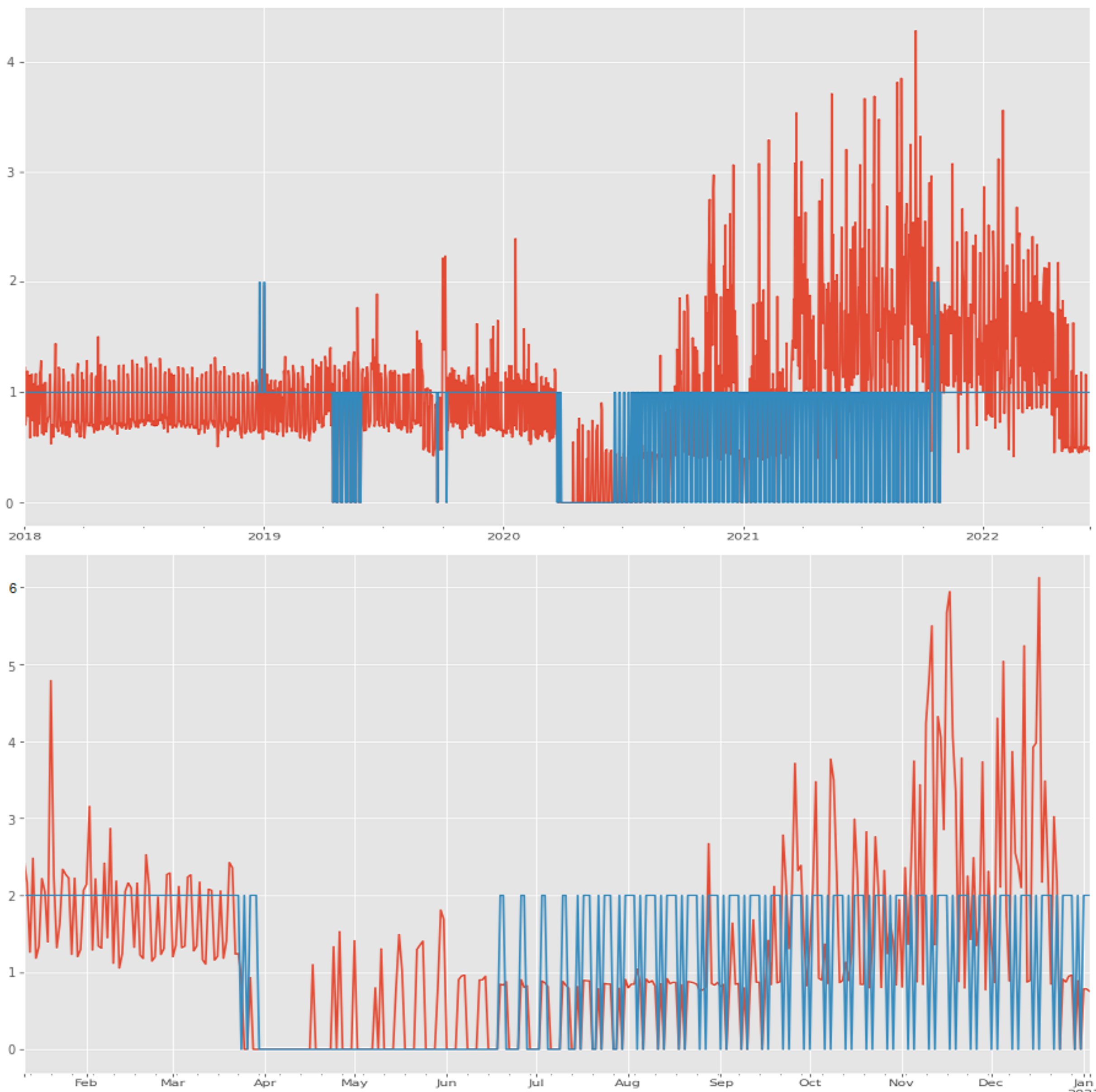
The linear model with the best performance, selected based on the AIC on the validation set, is chosen for this study. The search process led to the identification of $SARIMA(4,1,2) \times (1,1,1,7)$ and $SARIMAX(4,1,1) \times (1,1,1,7)$ as the most suitable models.

Table 2 indicates that the univariate SARIMA outperformed the SARIMAX model, contrary to the initial hypothesis. This outcome is possibly attributed to the high stability of Emirates' schedule and its unreliability. Notably, the test set included three major outliers in terms of demand, resulting in the imperfection of the predictions as some days overestimated the demand, while others underestimated it. However, this model behavior is beneficial for this particular study as the volume suffers from a few significant outliers. Therefore, the model serves as a reliable baseline for volume predictions. Nevertheless, the DCC experts need to anticipate outliers by analyzing real-time news and communicating with airlines.

To further explore the prediction models, both normal and exogenous variables LSTM were utilized, incorporating similar architectures. The network architecture comprised of one LSTM layer, one Dense layer, and minimal regularization. It was observed that without any regularization, the network tended to overfit the training data and fail in predicting the outcome. Thus, Dropout and Early stopping techniques were introduced. In general, the network architecture was as follows, with minor modifications depending on the activation function used:

- 1 input LSTM layer with 30 $\times n$ ($n =$ number of input days) hidden units and 5/10% dropout rate. Followed by the activation function.

Figure 4: Daily volumes of Emirates (MT) and the scheduled flights for that day (rescaled).



- 1 hidden layer, to better exploit the non-linearity, using 25 hidden neurons. Followed by the relu activation function.
- A learning rate of 0.01 or 0.03.
- Early stopping with the patience of 15 and delta = 0 and an upper limit of 2000 epochs.
- Adam optimizer.

First of all the neural networks are observed to not be overfitting. This is thanks to the careful application of regularization techniques. The error percentages can be observed in Table 2. It can be observed that contrary wise to the previous technique errors are sensibly lower and the method using exogenous variables is performing better. The predicted values are shown in Figure 6.

It can be observed that both networks follow reasonably the data. First, observing the training data it can be observed how both the networks are fitting the data well, showing good learning, and also resilient to outliers. The predictions

Figure 5: Train, Validation and Test splits for the emirates dataset.

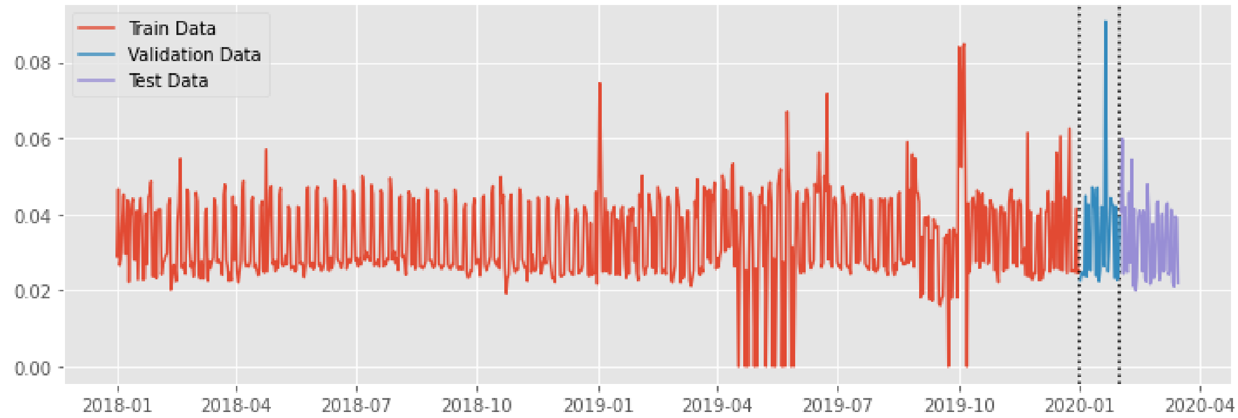
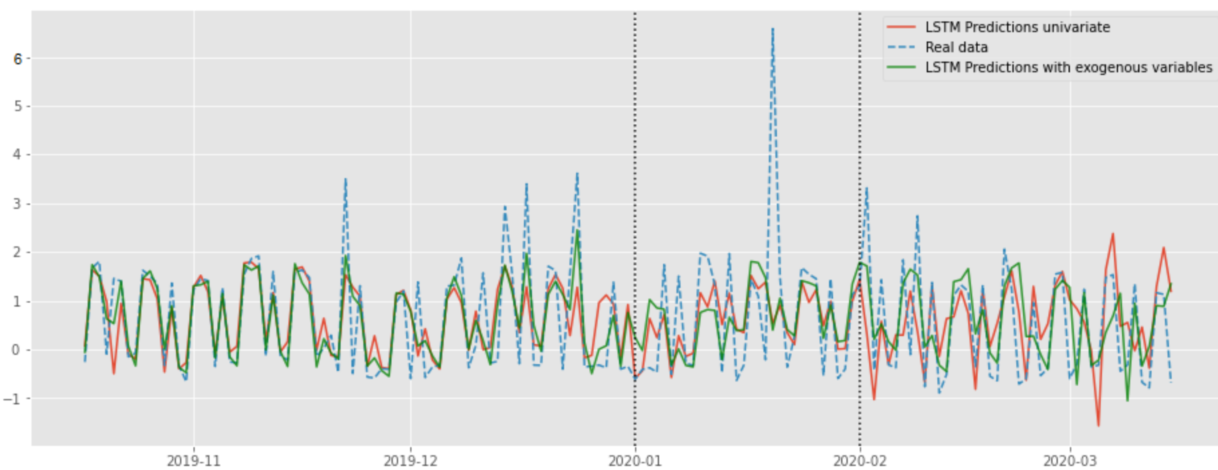


Figure 6: Fly Emirates. Predictions of the LSTM networks. Validation and Test set are separated with vertical lines. Note: training set is truncated for visualization purposes



turned slightly worse in the validation set. Finally, the test set can be observed. The green line, the LSTM embedding exogenous variables is clearly predicting better the future data. Furthermore, it can be observed how different the predictions are from the SARIMAX in the previous section, they appear to be more chaotic, probably due to the non-linearities of neural networks. Nonetheless, performances are sensibly better with this technique.

Finally, the hybrid model is presented. The results obtained by the hybrid must be observed in 2 ways. The first one is to analyze the scores of the neural network to assess whether there is some meaningful learning or if the residuals are just random noise. In the first case, it means that the neural network is somehow capturing some non-linearities in data that the time series model didn't catch. The second way to evaluate it is to observe the score of the hybrid predictions together with the relative plots. The first fact meaningful to report is that the network necessary to have a proper fit of the residual was slightly deeper than the one used for the original dataset and took a much larger number of epochs. The network is composed in this way:

- 2 input LSTM layer with 20 $\times n$ (n = number of input days) hidden units and 10% dropout rate.
- 1 hidden layer (only for Relu and TanH networks) to better exploit the non-linear activation function of size and 15 hidden neurons.
- A learning rate of 0.03
- Early stopping with patience of 25 and delta = 0 and an upper limit of 6000 epochs.
- Adam optimizer.

When exclusively considering the LSTM component of the hybrid model, it becomes challenging to determine the quality of the output. Thus, to obtain a comprehensive evaluation of the hybrid model's performance, the final scores are presented in Table 2.

Upon examination of the final symmetric mean absolute percentage error (SMAPE) and a comparison with prior techniques, it is apparent that the linear method's performance has improved significantly, with the error rate dropping to nearly 5%. In light of this finding and the impressive results produced by the LSTM networks, it is likely that the SARIMAX model is incapable of capturing certain non-linearities attributed to external factors.

Model	SMAPE
SARIMA	7.1
SARIMAX	10.1
Relu LSTM multiv.	3.8
Relu LSTM univ.	6.2
Hybrid	5.0

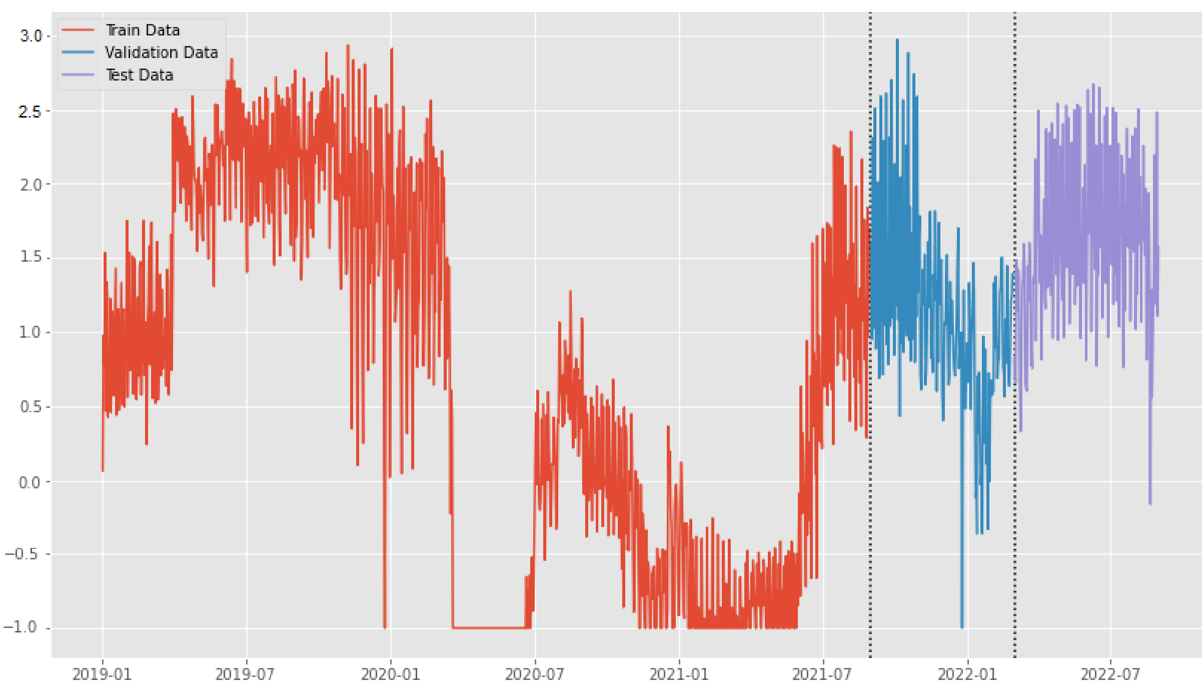
Table 2: Fly Emirates. Error score of the various algorithms, monthly predictions for the test set (February to mid-March 2020)

4.6 Ryanair

4.6.1 Company overview

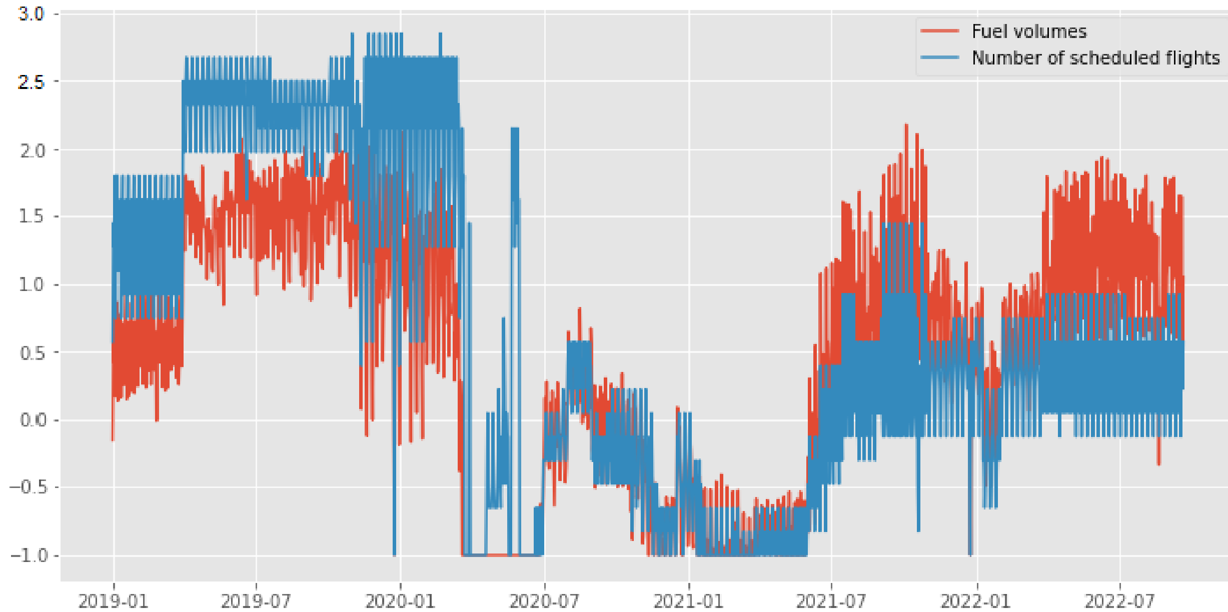
Ryanair DAC is an Irish ultra-low-cost carrier established in 1984, with its headquarters situated in Swords, Dublin, Ireland, and primary operational bases at Dublin and London Stansted airports. The airline is a significant component of the Ryanair Holdings group of airlines, and it has several sister airlines, including Ryanair UK, Buzz, Lauda Europe, and Malta Air. The company became a customer in 2019, and its behavior is noted for its unpredictability due to

Figure 7: Ryanair. Daily fueling volumes, in metric tonnes



the numerous short-haul flights it operates within Europe. Ryanair optimizes fuel costs by purchasing fuel in the cheapest countries, and even a small difference in fuel prices can result in considerable cost savings when considering airplane consumption and flight volume. Additionally, the company is presently expanding its business, leveraging the post-COVID-19 uncertainty affecting other low-cost carriers. Its current strategy aims to further expand and internationalize, with the objective of increasing its passenger numbers from 130 million annually to over 200 million by 2024 [24]. These two factors make it challenging to predict Ryanair's volume in advance, as evidenced by the high variance and unpredictable trends in the volumes shown in Figure 7. While it is commonly known in the aviation industry that the summer period is characterized by high demand, Ryanair did not follow this trend in the last year, as evident in the final section of the Figure 7. The observed increase in passenger volume for Ryanair can be attributed

Figure 8: Ryanair. Daily volumes and the number of scheduled flights, scaled.

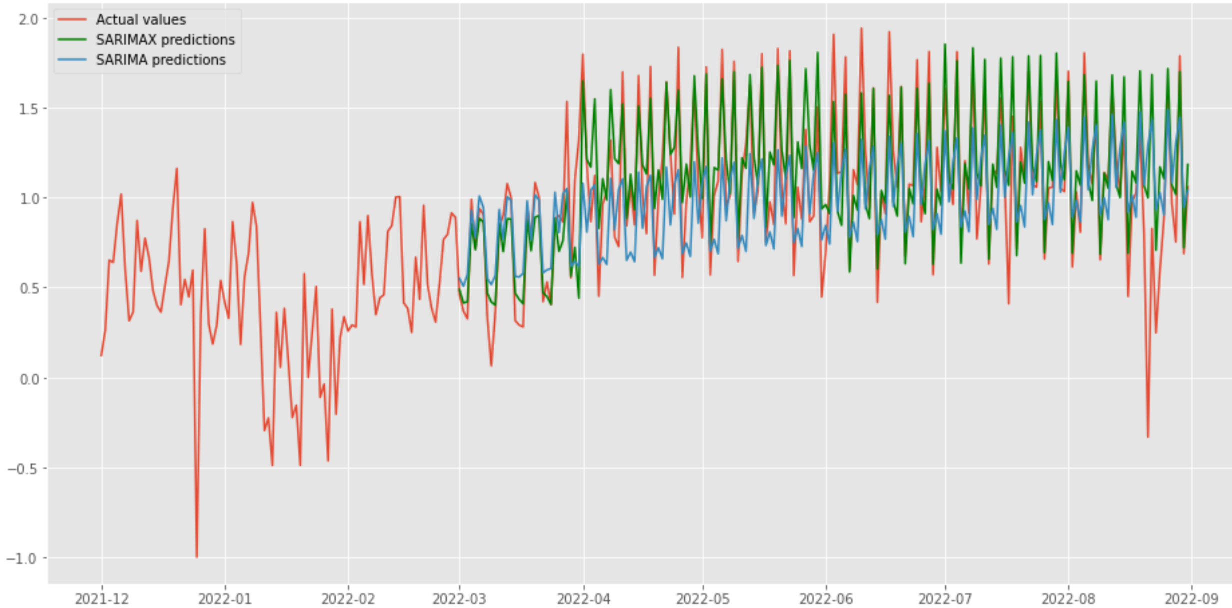


to the airline's strategic efforts to expand its route offerings and surpass competitors. It is worth noting that Ryanair maintains a high level of precision in publishing and updating flight schedules. Furthermore, the airline's frequent small-volume purchases serve to mitigate any discrepancies between scheduled and actual flight times. As shown in Figure 8, the scaled daily schedules exhibit a degree of correspondence with the observed volumes.

4.6.2 Models

In the context of the SARIMA(X) model, a parameter search pipeline is implemented to identify the optimal values for p, q and sp, sq parameters with the lowest Akaike Information Criterion (AIC) score on the validation set. The best model is found to be a $SARIMA(6, 1, 6) \times (4, 1, 4, 7)$ based on this approach. A similar model selection loop is applied to the SARIMAX model that includes exogenous variables, while considering that the AIC metric tends to favor models with fewer parameters. As a result, a $SARIMAX(0, 1, 6) \times (1, 1, 1, 7)$ model is determined to be the most appropriate. The performance of these two models is presented in Table 4.6.2. It is worth noting that although the SARIMA model can yield satisfactory results, it is slow in adapting to changes, as depicted in Figure 9. In contrast, the SARIMAX model can quickly detect changes in volumes caused by variations in schedules and accurately forecast the test data, including its peaks. Regarding the SARIMAX model, the following variables are incorporated: Confirmed Quantity, Scheduled Ryanair, Weekend, and Scheduled CPH. Notably, the number of flights in Denmark, the number of customers, and the size of the airlines are found to be ineffective in improving the model's performance, leading to a decrease in the validation score. The forecasted outcomes are illustrated in Figure 9. Moreover, the significance attributed by the model to its parameters can be observed in Table 4.6.2. Notably, the most important variable is the schedule, which has a significant impact on the accuracy of the time series model. This is due, in part, to the punctuality of Ryanair's flights, which contributes to the model's ability to achieve a high degree of precision. Following an examination of the variable importance summary, attempts were made to predict the data using a reduced number of exogenous variables. Unfortunately, this did not result in an improved score, even when the model was very close to doing so, unless the schedule feature was removed.

Figure 9: Ryanair. Predictions of the SARIMA and SARIMAX model. Note: training set is truncated to give a more clear visualization.



	coef	std err	z	P> |z|	[0.025	0.975]
Scheduled Ryanair	0.4982	0.012	42.977	0.000	0.476	0.521
Scheduled CPH	0.1959	0.030	6.483	0.000	0.137	0.255
Weekend	-7.965e-09	3.62e-11	-220.048	0.000	-8.04e-09	-7.89e-09
ma.L1	-0.6212	0.018	-33.773	0.000	-0.657	-0.585
ma.L2	-0.0234	0.023	-1.035	0.301	-0.068	0.021
ma.L3	-0.1318	0.026	-5.073	0.000	-0.183	-0.081
ma.L4	-0.0477	0.025	-1.876	0.061	-0.098	0.002
ma.L5	0.0211	0.030	0.695	0.487	-0.038	0.081
ma.L6	-0.0660	0.029	-2.303	0.021	-0.122	-0.010
ar.S.L7	0.1763	0.035	5.020	0.000	0.107	0.245
ma.S.L7	-0.8008	0.023	-35.009	0.000	-0.846	-0.756
sigma2	0.0445	0.001	53.282	0.000	0.043	0.046

Table 3: Simpler $SARIMAX(0,1,6)(1,1,7)$ parameters importance

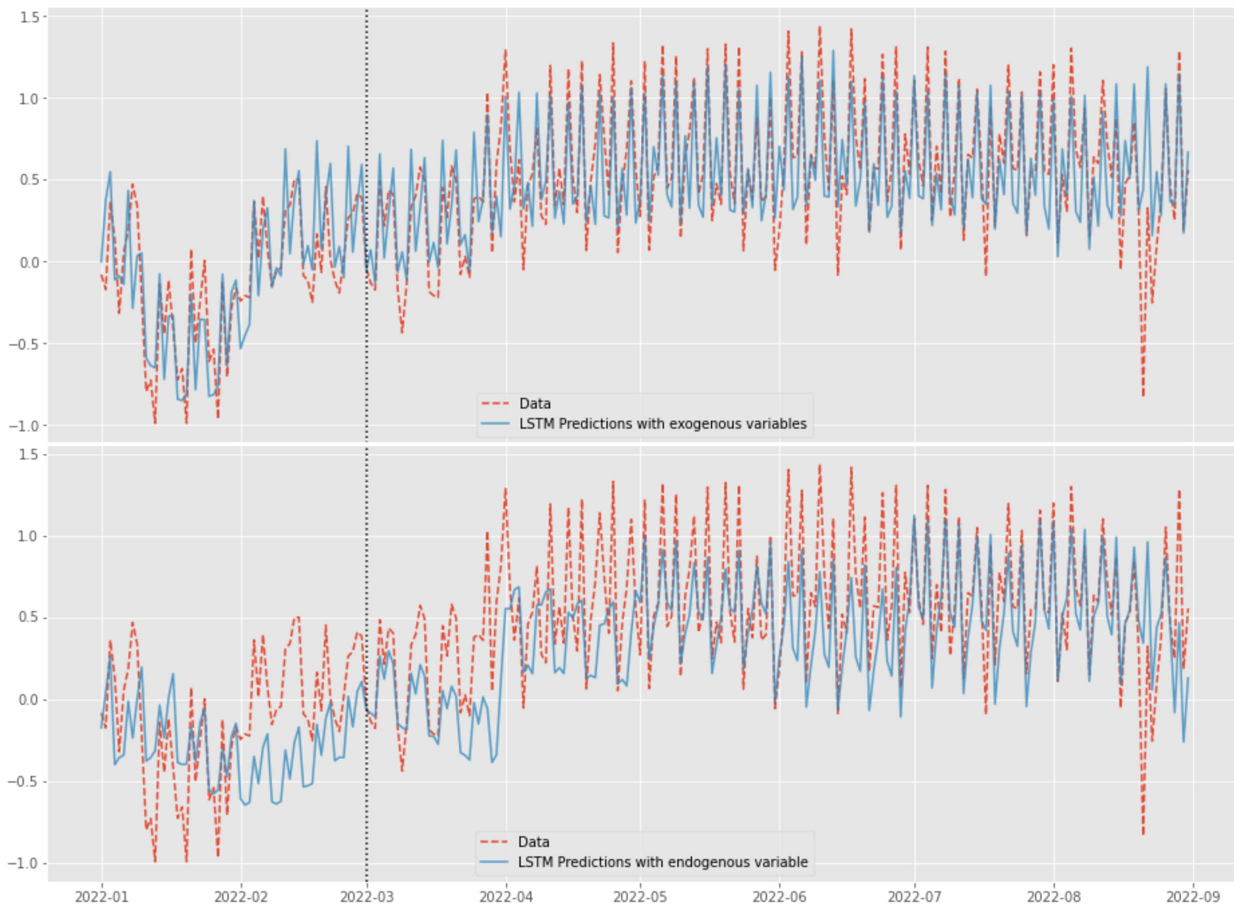
Similar architectures, utilizing both normal and exogenous variables LSTM, were employed. A shallow network comprising of one LSTM, one Dense layer, and minimal regularization was used. Without regularization, the network was observed to overfit the training data, leading to poor predictive performance. Dropout and Early stopping techniques were subsequently incorporated. Generally, the network architecture is as follows (with slight modifications depending on the chosen activation function):

- 1 input LSTM layer with 20 (30 univ.) $\times n$ (n = number of input days) hidden units and 50 (10 univ.)% dropout rate.
- 1 hidden layer to better exploit the non-linear activation function of size and 40 hidden neurons. Relu activation function is used in both networks.
- A learning rate of 0.01 for the univariate and 0.0005 for the multivariate.
- Early stopping with patience of 15 and delta = 0 and an upper limit of 2000 epochs.
- Adam optimizer.

In this study, all features used in the time series, including Confirmed Quantity, Scheduled Ryanair, Weekend, and Scheduled CPH, were found to contribute to higher scores. It should be noted that while neural networks are capable of performing feature selection internally, this approach has its limitations. Despite this, the results suggest that the neural networks utilized in this study did not overfit, thanks to the application of careful regularization techniques. Interestingly, the two networks differed significantly in terms of architecture, with the LSTM network with exogenous variables exhibiting a tendency to overfit rapidly due to its smallness, thus requiring a heavy dropout.

Table 4.6.2 presents the error percentages obtained in this study, revealing significantly lower errors compared to previous techniques. Notably, the method utilizing exogenous variables demonstrated superior performance. Figure 10

Figure 10: Fly Emirates. Predictions of the LSTM networks. Validation and Test set are separated with vertical lines. Note: training set is truncated for visualization purposes



displays the predictions of the two selected networks. Both LSTM networks were observed to perform well, with the first network utilizing exogenous variables achieving almost perfect prediction accuracy on the test set, suggesting that Ryanair's fueling behavior is predictable.

The hybrid model is similar to the one in the previous case. To begin, the used LSTM network is presented. The architecture is similar to the one used in the previous case study:

- 2 input LSTM layer with 20 $\times n$ ($n =$ number of input days) hidden units and 10% dropout rate.
- 1 hidden layer and relu activation function.
- A learning rate of 0.02
- Early stopping with patience of 25 and delta = 0 and an upper limit of 6000 epochs.
- Adam optimizer.

A linear and interpretable predictor in the form of $SARIMAX(0, 1, 6) \times (1, 1, 1, 7)$ will be utilized in the analysis. The performance of the hybrid model and the residuals Long Short-Term Memory (LSTM) model are presented in Table 4.6.2. The close agreement between the predicted values and the actual data suggests that the SARIMAX model is well-constructed. Additionally, an examination of the hybrid model scores alongside those of the SARIMAX model in Table 4.6.2 indicates that the hybrid approach improves upon the scores obtained by SARIMAX alone. However, despite this improvement, the monthly predicted SARIMAX and the multivariate LSTM models still outperform the hybrid method in this particular case study. Ensemble methods are recognized to perform well in cases of overfitting or underfitting[13]. Given the low error scores attained in this analysis, it can be surmised that the SARIMAX model is of high quality, thereby explaining why the ensemble technique does not significantly enhance the performance metrics.

Technique	SMAPE
SARIMA	6.7
SARIMAX	5.7
Relu LSTM multiv	4.1
Relu LSTM univ	8.6
Hybrid	6.0

Table 4: Ryanair. Error score of the various algorithms, monthly predictions for the test set (March to end of August 2022)

4.7 Copenhagen Airport

4.7.1 Airport overview

Copenhagen Airport, Kastrup is the largest international airport serving Copenhagen, Denmark, the rest of Zealand, the Øresund Region, and a significant part of southern Sweden, including Scania. As of 2019, the airport was the largest in the Nordic countries, handling close to 30.3 million passengers. It is important to note that the volumes analyzed in this section refer exclusively to the customers of DCC Shell Aviation A/S, estimated to represent approximately 35% of the total airport market (August 2022). The company's market share has increased significantly in recent years, contributing to the strong upward trend seen in volumes. The airport quantities are characterized by diverse customers with varying fuelling behavior. The data has been collected since 2019, and customers at CPH (IATA reference name for Copenhagen airport) differ greatly in terms of fleet size, frequency, single volumes, seasonality, and reliability. This suggests that small companies with less predictable behavior might be overshadowed in predictions by larger, more regular companies. Airport data has been collected since the beginning of 2019. For this study, a validation set of 6 months, from September 2021 to the end of February 2022, was chosen, similar to the Ryanair case. The test set spans from March 2022 to the end of August, comprising 5 months. This choice was made to leverage almost 3 entire years of training data and have a fresh year to test the previously fitted model, using an intuitive and business-friendly approach. The volumes and data split can be observed in Figure 11. In line with the significance of previous case studies, the schedule is analyzed and compared again. It is apparent from Figure 12 that the schedule following the Covid-19 outbreak behaves differently from before. There may be two reasons for this: either airlines tend to refuel more, or the increased market share of DCC Shell Aviation A/S is interfering with the proportions. Possibly both. Despite the pandemic, there are still fewer flight schedules, and this can be attributed to two main reasons. Firstly, companies are less diligent in publishing schedules, or secondly, there are fewer flights with higher volumes. In reality, it is the second reason, as companies such as Fly Emirates or Thai Air operate direct routes from Asia with a few flights having extremely high volumes, which are often not published in schedules. Moreover, it is observed that the latest peak during the summer shows no change in terms of schedules. To predict this, more features are required. Therefore, the number of customers and their magnitude might be very relevant features in this case. This assumption has led to the creation of the two corresponding features.

4.7.2 Models

The AIC loop has yielded the best SARIMA(X) models, namely, $SARIMA(6,1,6) \times (2,1,4,7)$ and $SARIMAX(1,1,1) \times (2,1,1,7)$. These results are consistent with the Ryanair case study, as the ACF and PACF functions exhibited similar patterns. All available features, including CPH schedule, Weekend, Number of customers, customer magnitude, and market share, were incorporated in the models. The performance scores of the two models are presented in Table **??**, indicating that, once again, the model incorporating exogenous variables has a lower error score. Detailed predictions are presented in Figure .13. The SARIMA model failed to capture the trend change in the data when the summer schedule began and a sudden increase in flights occurred. Without exogenous variables, the model was unable to detect this change in the data. As anticipated, the inclusion of exogenous variables enhanced the performance of the models significantly. Table 5 reveals that almost all the exogenous variables, except the "Magnitude of Customers," were found to be relevant. The p-values of the variables were found to be very low, indicating that the

Figure 11: Copenhagen Airport. Daily fuelling volumes, in metric tonnes.

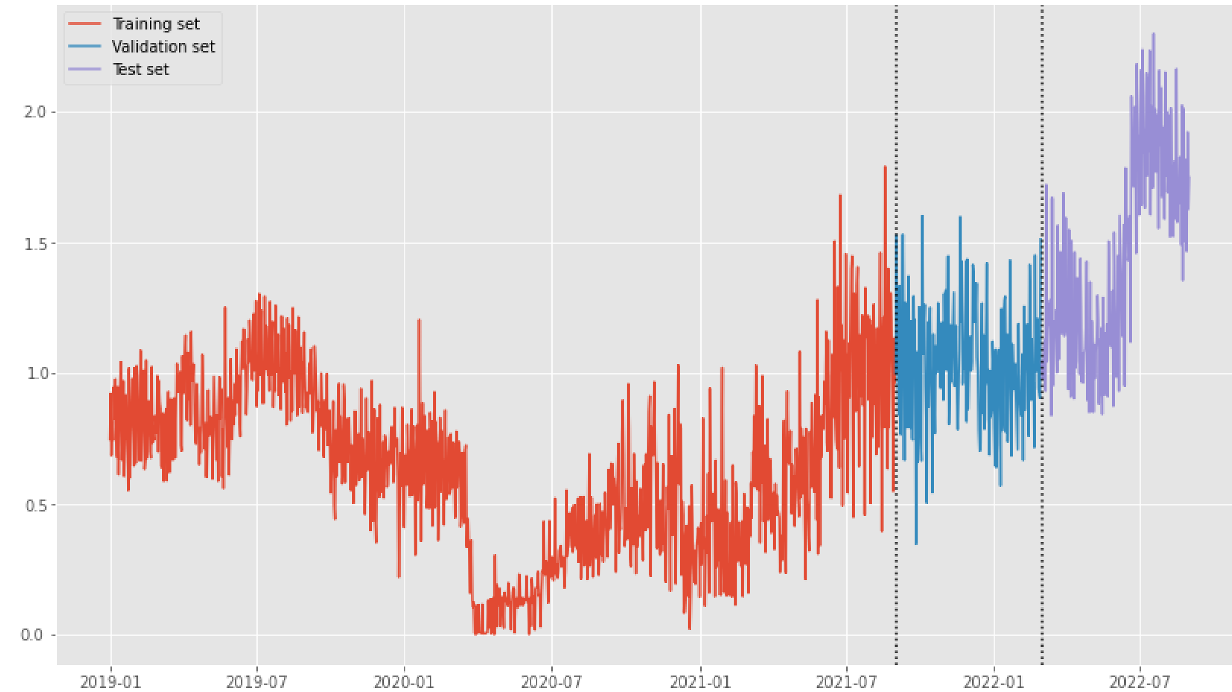
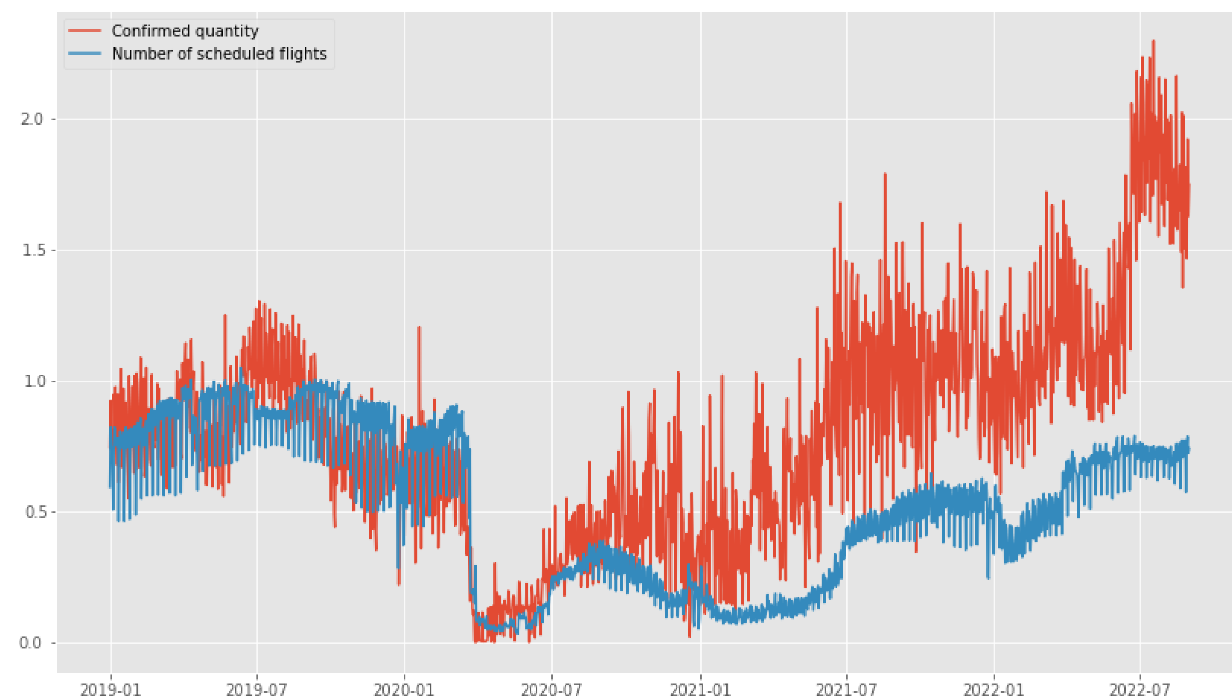
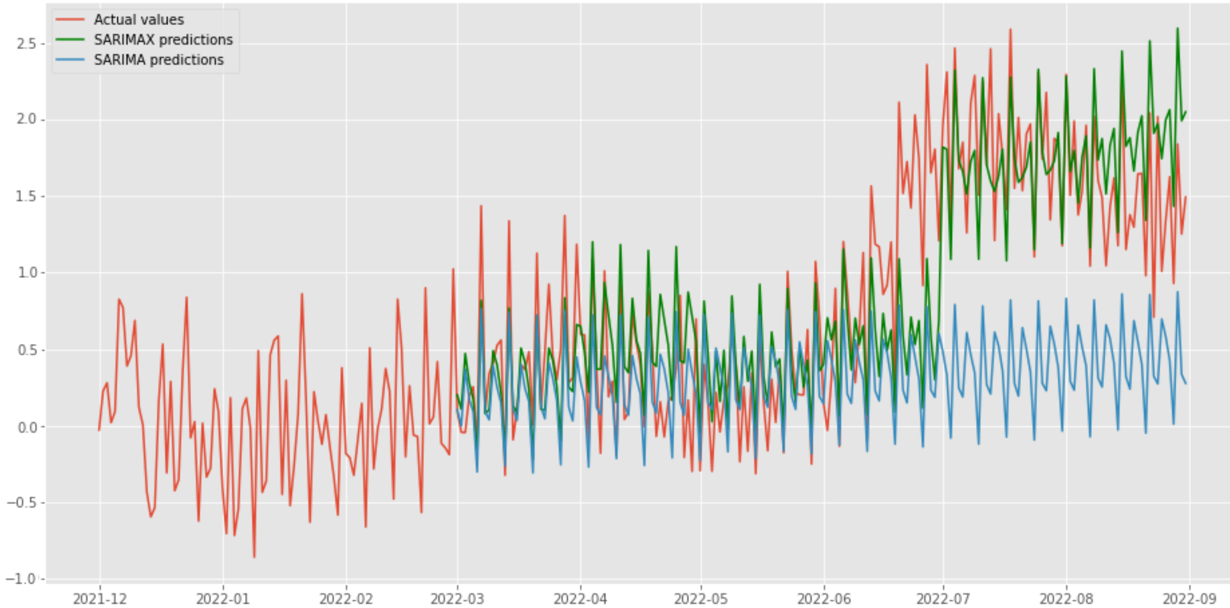


Figure 12: Copenhagen Airport. Daily fuelling volumes (MT) compared with the daily scheduled planes.



model heavily relies on them to produce accurate predictions. The magnitude of customers was still included in the model, despite a lower validation score without it. Another noteworthy observation from the p-values is that the 14th seasonal moving average lag and the 7th autoregressive were deemed crucial by the model, confirming the presence of strong weekly seasonality in the data.

Figure 13: Ryanair. Predictions of the SARIMA and SARIMAX model. Note: training set is truncated to give a more clear visualization.



	coef	std err	z	P> |z|	[0.025	0.975]
Scheduled CPH	0.4134	0.100	4.140	0.000	0.218	0.609
Number of Customers	0.0279	0.050	0.557	0.578	-0.070	0.126
Weekend	9.149e-08	3.53e-10	259.542	0.000	9.08e-08	9.22e-08
Market Share	0.1345	0.101	1.329	0.184	-0.064	0.333
ar.L1	0.0256	0.027	0.949	0.342	-0.027	0.078
ma.L1	-0.8724	0.014	-60.791	0.000	-0.901	-0.844
ar.S.L7	0.2469	0.026	9.490	0.000	0.196	0.298
ar.S.L14	0.1053	0.028	3.728	0.000	0.050	0.161
ma.S.L7	-0.9185	0.014	-63.738	0.000	-0.947	-0.890
sigma2	0.1982	0.005	36.179	0.000	0.187	0.209

Table 5: $SARIMAX(1, 1, 1)(2, 1, 1, 7)$ parameters

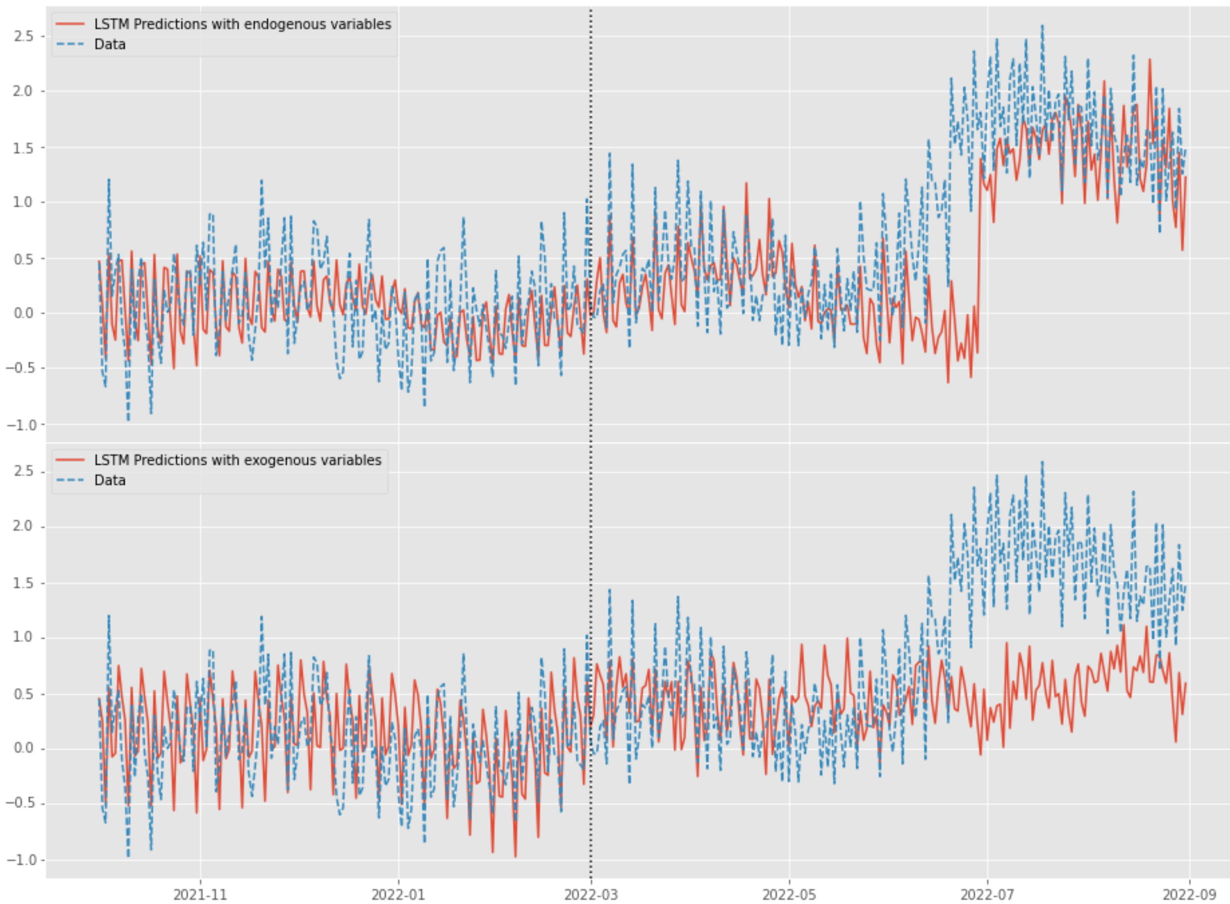
Subsequently, an LSTM network was fitted using all the features used in the time series, along with the univariate approach. Any other or fewer features resulted in lower scores. The features included were Confirmed Quantity, Weekend, Scheduled CPH, Number of customers, and Market Share. While neural networks are capable of performing feature selection internally, there are limitations due to the curse of dimensionality, and the addition of infinite variables is not possible. The network architecture was derived from the univariate approach, with some adjustments to achieve optimal performance:

- 2 input LSTM layer with 40 $\times n$ ($n =$ number of input days, fixed to 30 in this case given the architecture constraints) hidden units and 10% dropout rate. Followed by the activation function.
- 1 hidden layer (only for Relu and TanH networks) to better exploit the non-linear activation function using 40 or 30 hidden neurons. Followed by the activation function.
- A learning rate of 0.01. For tanH only the rate was 0.005.
- Early stopping with patience of 25 and delta = 0 and an upper limit of 2000 epochs.
- Adam optimizer.

To begin with, it is important to note that none of the networks are suspected of overfitting the validation set, and their behavior is consistent with previous cases, which has been omitted here for the sake of simplicity.

When considering the LSTM scores presented in Table 6 in light of previous case studies, it becomes apparent that the model's performance is exhibiting some peculiar characteristics. Evidence suggests that, in this case, the exogenous variables are leading the network to make accurate predictions on the test set. However, the network using only the target variable is performing better. Despite more complex architecture and extensive parameter tuning, the error could not be reduced. Therefore, in this particular case study, the univariate model outperforms the multivariate model. The predictions for the LSTM model, which is the best-performing one, are displayed in Figure 14.

Figure 14: Predictions of the univariate (relu) and multivariate (tanh) LSTM. The summer peak is very difficult for the second network to predict.



Lastly, as was done in previous cases, the hybrid results will be analyzed in two ways: first the LSTM model, and then the entire hybrid model. The LSTM network used is presented first, and its architecture is similar to the one used in previous case studies:

- 2 input LSTM layer with 20 $\times n$ ($n =$ number of input days) hidden units and 10% dropout rate.
- 1 hidden layer (only for Relu and TanH networks) to better exploit the non-linear activation function of size and 15 hidden neurons.
- A learning rate of 0.02
- Early stopping with patience of 25 and delta = 0 and an upper limit of 6000 epochs.
- Adam optimizer.

A $SARIMAX(1,1,1)(2,1,2,7)$ will be used as the linear interpretable predictor.

The scores obtained can be compared to those of the single SARIMAX and the LSTM (Table 6). It is observed that the hybrid model improves upon the SARIMAX, but it does not outperform the LSTM, with both models having a SMAPE of 8.9. It can be argued that the hybrid model captures some nonlinearity that was not detected by the SARIMAX model.

Technique	SMAPE
SARIMA	13.9
SARIMAX	8.9
Relu LSTM multiv	5.9
Relu LSTM univ	5.5
Hybrid	8.9
Company current	17.5

Table 6: Copenhagen Airport. Error score of the various algorithms, monthly predictions for the test set (March to the end of August 2022)

The present study results are reported in Table 6, where the final row discloses the current error of the company under examination. The data indicate that, during the testing time frame of March through August of 2022, each data-driven model proposed has demonstrated a notably diminished margin of error.

5 Conclusion

The present paper demonstrates the effectiveness of data-driven demand forecasting in the aviation industry, with a particular emphasis on predicting the future volume demand for jet fuel through the examination of three distinct case studies. Our findings suggest that machine learning techniques, namely SARIMAX, Hybrid, and LSTM, can produce reliable forecasts with low prediction errors on the testing set. The incorporation of exogenous variables in the models can enhance their performance, particularly for linear models such as SARIMAX. However, our results indicate that LSTM neural networks are the most effective models for mid-term jet fuel forecasting.

Furthermore, our study highlights that shallow neural networks with no more than three layers can yield satisfactory performance in forecasting jet fuel demand, which is consistent with prior research in this area. This is a pertinent practical consideration for implementation, as shallow networks are simpler and faster to train and possess fewer parameters to optimize.

To conclude, the results of our case studies demonstrate the potential of novel, data-driven approaches to bolster forecasting accuracy and inform decision-making in the aviation industry. By leveraging machine learning models to analyze past data and incorporate exogenous variables, fuel providers can gain a better understanding of and make more precise predictions about future jet fuel demand. This can enable them to allocate resources more efficiently, budget and plan more accurately, and ultimately achieve cost savings and increased profitability.

References

[1] Peter Morrell and William Swan. Airline jet fuel hedging: Theory and practice. *Transport Reviews*, 26(6):713–730, 2006.

[2] Siew Hoon Lim and Yongtao Hong. Fuel hedging and airline operating costs. *Journal of Air Transport Management*, 36:33–40, 2014.

[3] Lisha Li, Shuming Yuan, Yue Teng, and Jing Shao. A study on sustainable consumption of fuelmdash;an estimation method of aircraft. *Energies*, 14(22), 2021.

[4] Sebastian Baumann and Uwe Klingauf. Modeling of aircraft fuel consumption using machine learning algorithms. *CEAS Aeronautical Journal*, 11:1–11, 09 2019.

[5] Xinting Zhu and Lishuai Li. Flight time prediction for fuel loading decisions with a deep learning approach. *Transportation Research Part C: Emerging Technologies*, 128:103179, 2021.

[6] Božidar Soldo. Forecasting natural gas consumption. *Applied Energy*, 92(C):26–37, 2012.

[7] Melih Yucesan, Engin Pekel, Erkan Celik, Muhammet Gul, and Faruk Serin. Forecasting daily natural gas consumption with regression, time series and machine learning based methods. *Energy Sources, Part A: Recovery, Utilization, and Environmental Effects*, 0(0):1–16, 2021.

[8] Don Chi Wai Wu, Lei Ji, Kaijian He, and Kwok Fai Geoffrey Tso. Forecasting tourist daily arrivals with a hybrid sarima–lstm approach. *Journal of Hospitality & Tourism Research*, 45(1):52–67, 2021.

[9] Antonios Parasyris, George Alexandrakis, Georgios V. Kozyrakis, Katerina Spanoudaki, and Nikolaos A. Kampanis. Predicting meteorological variables on local level with sarima, lstm and hybrid techniques. *Atmosphere*, 13(6):878, May 2022.

[10] Fatih Taşpınar, Numan Çelebi, and Nedim Tutkun. Forecasting of daily natural gas consumption on regional basis in turkey using various computational methods. *Energy and Buildings*, 56:23–31, 2013.

[11] Huai Su, Enrico Zio, Jinjun Zhang, Mingjing Xu, Xueyi Li, and Zongjie Zhang. A hybrid hourly natural gas demand forecasting method based on the integration of wavelet transform and enhanced Deep-RNN model. *Energy*, 178:585–597, July 2019.

[12] Zhihui Hu, Yongxing Jin, Qinyou Hu, Sukanta Sen, Tianrui Zhou, and Mohd Tarmizi Osman. Prediction of fuel consumption for enroute ship based on machine learning. *IEEE Access*, 7:119497–119505, 2019.

[13] Mohammed Abdelrahim, Carlos Merlos, and Taehyung (George") Wang. Hybrid machine learning approaches: A method to improve expected output of semi-structured sequential data. In *2016 IEEE Tenth International Conference on Semantic Computing (ICSC)*, pages 342–345, 2016.

[14] Fotios Petropoulos, Rob J. Hyndman, and Christoph Bergmeir. Exploring the sources of uncertainty: Why does bagging for time series forecasting work? *European Journal of Operational Research*, 268(2):545–554, 2018.

[15] F. Pan, H. Zhang, and M. Xia. A hybrid time-series forecasting model using extreme learning machines. In *2009 Second International Conference on Intelligent Computation Technology and Automation*, volume 1, pages 933–936, 2009.

[16] Eraylson Galdino, Paulo De Mattos Neto, and Joao Oliveira. Hybrid system for time series using iterative residual forecasting models. pages 872–877, 10 2019.

[17] C Chatfield. The analysis of time series: An introduction, sixth edition (6th ed.). *Chapman and Hall/CRC.*

[18] Rayman Preet Singh, Peter Xiang Gao, and Daniel J. Lizotte. On hourly home peak load prediction. *2012 IEEE Third International Conference on Smart Grid Communications (SmartGridComm)*, pages 163–168, 2012.

[19] Farid Bourennani, Raazi Rizvi, and Shahryar Rahnamayan. Optimal photovoltaic solar power farm design using the differential evolution algorithms. 09 2010.

[20] Sepp Hochreiter and Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 9(8):1735–1780, 11 1997.

[21] Tomasz Szandala. Review and comparison of commonly used activation functions for deep neural networks. *CoRR*, abs/2010.09458, 2020.

[22] Winita Sulandari, Subanar, Muhammad Hisyam Lee, and Paulo Canas Rodrigues. Indonesian electricity load forecasting using singular spectrum analysis, fuzzy systems and neural networks. *Energy*, 190:116408, 2020.

[23] Steven P. Schnaars. Long-range forecasting: From crystal ball to computer: J. scott armstrong, 2nd ed. (wiley, new york, 1985) [uk pound]22.95 (paper), pp. 689. *International Journal of Forecasting*, 2(3):387–390, 1986.

[24] George-Cristian Prichinet. Strategic analysis of ryanair, 2020.