Prediction of Traffic Take-Off Times at Out-stations

A Case Study at Schiphol Airport

T.M. Ruskamp

Delft University of Technology

TU Delft

Prediction of Traffic Take-Off Times at Out-stations: A Case Study at Schiphol Airport

by

T.M. Ruskamp

to obtain the degree of Master of Science
at the Delft University of Technology,

Student number:	4696875	
Project duration:	February 10, 2023 – January 16, 2024	
Thesis committee:	Dr. M.J. Ribeiro,	TU Delft, supervisor
	Dr. A. Bombelli,	TU Delft
	Dr. A. Jamshidnejad,	TU Delft
	F. Dijkstra	KDC Mainport Schiphol

Cover: a tall white tower with a white top and a sky background by Niels Baars

An electronic version of this thesis is available at http://repository.tudelft.nl/.

Preface

This thesis marks the culmination of my master's program in Aerospace Engineering at Delft University of Technology, conducted at LVNL on behalf of KDC. It encompasses both the preliminary report and the final academic paper and has been written under the supervision of Dr. M.J. Ribeiro.

I am deeply grateful to my supervisors, Marta Ribeiro and Ferdinand Dijkstra, for their invaluable guidance and continuous encouragement throughout this journey. Your passion and enthusiasm have made this journey enjoyable, and without you, not a letter of this report would've been written.

To my parents and sister, thank you for always supporting and motivating me throughout my studies, and beyond. To my girlfriend Merlijne, thank you for always being there and listening to me during this challenging journey. It would not have been possible without you.

Contents

List of Figures

Acronyms

4D Four Dimensional. 36

A-CDM Airport Collaborative Decision Making. 4, 30–33, 59

ADS-B Automatic Dependent Surveillance-Broadcast. 59, 61

AI Artificial Intelligence. 47, 53

ALDT Actual Landing Time. 32

ANN Artificial Neural Network. 4, 52–54

ANSP Air Traffic Service Provider. 28, 30, 39, 44, 59

AO Aircraft Operator. 32, 33

AOBT Actual Off-Block Time. 31, 32

ATC Air Traffic Control. 28, 31, 32, 34, 38, 42, 44, 45, 61

ATCos Air Traffic Controllers. 28, 29, 36, 48

ATFCM Air Traffic Flow Capacity Management. 59, 60

ATFM Air Traffic Flow Management. 28–30, 33, 34, 40, 42, 44–46, 58–60, 62

ATM Air Traffic Management. 27, 30, 36–39, 46, 48–50, 62

ATOT Actual Take Off Time. 30, 32, 40, 59

ATS Air Traffic Service Provider. 33

CFMU Central Flow Management Unit. 44, 45

CNN Convolutional Neural Network. 4, 53–57

CTOT Calculated Take Off Time. 30–33, 46

DCB Demand Capacity Balancing. 27, 28, 33, 38

DPI Departure Planning Information. 59

DST Decision Support System. 28–30, 48, 51, 63

EOBT Estimated Off-Block Time. 31, 32

ETOT Estimated Take Off Time. 30–32

FIR Flight Information Region. 32

FMPC Flow Management Position Controler. 29, 30, 34, 37, 48

FUM Flight Update Messages. 59

GH Ground Handler. 32, 33

GNN Graph Neural Network. 53, 54

GRU Gated Recurrent Unit. 56, 58

ICAO International Civil Aviation Organization. 44

KDC Knowledge & Development Centre. 37

LSTM Long-Short Term Memory Cell. 4, 56–58

LVNL Luchtverkeersleiding Nederland. 27, 28, 36, 37, 63

METAR Meteorological Aerodrome Reports. 60

ML Machine Learning. 27, 48–50, 62

MTTT Minimum Turn Round Times. 32

NM Network Manager. 28–30, 36, 37, 59, 60

NMOC Network Manager Operations Centre. 46

OCC Operations Control Centre. 33

RF Random Forest. 51, 52, 57

RNN Recurrent Neural Network. 53, 56

SESAR Single European Sky Aviation Research. 37, 47, 59

TACT Tactical Load Factor Calculation and Distribution System. 44

TBO Trajectory Based Operations. 29, 30, 37

TOBT Target Off-Block Time. 31–33, 47

TP Trajectory Prediction. 38

TSAT Target Start-Up Approval Time. 31–33

TTOT Target Take Off Time. 30–33

VTT Variable Taxi-Times. 33

Scientific Article

Prediction of Traffic Take-Off Times at Out-stations: A Case Study at Schiphol Airport

T.M. Ruskamp
Under the supervision of Dr. M.J. Ribeiro (TU Delft) and F. Dijkstra (LVNL)

Sustainable Air Transport, Faculty of Aerospace Engineering
Delft University of Technology Delft, The Netherlands

***Abstract*—Reducing uncertainty in air traffic flow management is crucial for maintaining safety and efficiency in modern aviation. Additionally, forecasting Actual Take-Off Times (ATOT) for flights across Europe is particularly challenging due to the diverse flight-specific variables and operational conditions. This study focuses on enhancing ATOT prediction for flights arriving at Amsterdam Schiphol Airport from European out-stations by leveraging machine learning techniques, specifically a Long Short-Term Memory (LSTM) neural network, augmented with a Multihead Attention mechanism. A model capable of capturing complex temporal dependencies and operational factors influencing the ATOT is developed utilizing data from Electronic Flight Data (EFD) messages, weather reports and a EUROCONTROL dataset. The model's performance is evaluated against traditional ensemble methods and the current Decision Support Tool (DST) system used by Luchtverkeersleiding Nederland (LVNL). Results indicate that the LSTM model outperforms existing models including a reproduction of the DST, achieving a Mean Absolute Error (MAE) of 12.05 minutes at a forecast horizon of 4 hours, demonstrating significant improvements. This assessment underscores the importance of factors such as the knock-on effect in delay prediction and suggests that integrating advanced machine learning models can significantly enhance demand forecasting, leading to more efficient air traffic management and reduced delays at Schiphol Airport.**

***Index Terms*—Air Traffic Management, Demand Forecasting, Departure Time Prediction, Long-Short-Term Memory Neural Network (LSTM)**

I. INTRODUCTION

With the steady increase in global air travel, Air Navigation Service Providers (ANSPs) faces the dual challenges of managing the growing complexity of airspace while maintaining safety and efficiency. Despite not yet returning to pre-COVID traffic levels, the demand for air travel is steadily rising, with 2023 seeing a significant recovery compared to 2022 (EUROCONTROL [1]). However, this recovery has been accompanied by an increase in delays, as only 70.6% of flights arrived within 15 minutes of their scheduled time, and only 65.3% of flights departed within 15 minutes of scheduled time (EUROCONTROL [1]). This alarming trend highlights the urgency for innovative solutions to mitigate delays and enhance the overall reliability of air traffic flow management systems.

Central to this challenge is balancing the demand for airspace with its available capacity. Demand, defined as the total number of aircraft seeking to land within a specific timeframe, must be continuously assessed against capacity, which is dynamically influenced by factors such as weather conditions, runway configurations, and staffing shortages. To manage overloads, ANSPs can increase capacity—for example, by opening additional runways—or reduce demand by delaying arrivals. In the case of Schiphol, the optimal timeframe to issue such regulations is 3-4 hours before take-off, as sufficient flights are still in the pre-departure phase. However, this differs per use case, depending on airport size and geographical location. Extensive studies have focused on forecasting demand by predicting individual departures and arrivals. Accurate forecasting is essential for adaptability, enabling ANSPs to implement actions such as real-time data monitoring, predictive analytics, and integrated decision support systems. Additionally, enhancing communication and collaboration with stakeholders, adopting flexible operational procedures, and investing in advanced technologies like machine learning and automation are crucial. These measures empower ANSPs to make informed, real-time decisions that optimize airspace management, reduce delays, and improve overall efficiency.

The effectiveness of decision-support tools like the Decision Support Tool (DST) is intrinsically linked to the accuracy and quality of their input data. Precise forecasting of traffic load and demand is essential to align operations with capacity constraints—the maximum number of aircraft that can safely occupy the airspace and runways. Inaccurate predictions can lead to suboptimal decisions, causing unnecessary delays and safety risks. Therefore, enhancing demand forecasting accuracy is paramount for the DST to function effectively and for Air Traffic Controllers (ATCos) to manage air traffic efficiently.

To address these challenges, Luchtverkeersleiding Nederland (LVNL) has introduced a DST designed to enhance its demand capacity balancing processes. By leveraging real-time data, advanced analytics, and machine learning algorithms, the DST provides air traffic controllers with actionable insights to better predict and manage traffic flows. However, the accuracy of the demand prediction by the DST has proven to be limited. According to Vos [2] this is largely due to the uncertainty in departure time prediction. Therefore, the objective of the model developed in this study is to improve the Actual Take Off Time (ATOT) prediction for all flights inbound Schiphol that depart within Europe from the 50 out-stations with most flight inbound Schiphol.

This paper is organized as follows. Section II reviews related studies, focussing on demand prediction and various flight prediction methodologies. Section III presents the case study of Schiphol Airport, detailing its operational context and the challenges it faces in air traffic flow management. Section IV outlines the methodology, including data preparation, model architecture, and training procedures using machine learning techniques, specifically Long-Short Term Memory Cell (LSTM) networks. The results are presented in Section V, where the predictive accuracy of the developed model is compared with baseline approaches, and the performance of the DST is evaluated. Section VI discusses the implications of the findings, addressing limitations and opportunities for future research. Finally, Section VII concludes the paper by summarizing its contributions and significance.

II. Related work

This section reviews the extensive research conducted on demand prediction, focusing on methodologies for forecasting departure times. It is organized into three main parts: subsection II-A discusses traditional trajectory-based and aggregate demand prediction methods, evaluating their strengths and limitations; subsection II-B explores conventional forecasting techniques and the advancements introduced by machine learning algorithms. Finally, subsection II-C identifies the research gap.

A. Demand Prediction

This subsection explores the primary methodologies employed in predicting air traffic demand, evaluates their strengths and limitations, and examines the key factors influencing demand in air transportation.

Predicting air traffic demand involves estimating the number and flow of aircraft within various airspace sectors over time. These predictions are essential for ensuring safety, strategic planning, and optimizing operational efficiency. Various methodologies have been developed to forecast air traffic demand, each with its own advantages and limitations. These methods can be broadly categorized into trajectory-based approaches and aggregate models. The following subsections explore these traditional methods in detail, laying the groundwork for understanding the advancements brought by machine learning techniques:

1) Trajectory-based Demand Prediction: Traditionally, demand forecasting in Air Traffic Management (ATM) has relied on trajectory-based methods, which predict aircraft positions along planned flight paths. These methods utilize Four Dimensional (4D) trajectories, latitude, longitude, altitude, and time for each aircraft to predict sector occupancy as reported by de Leege et al. [3]. The process involves two critical stages:

1) **Pre-Departure Prediction**: Before departure, models estimate departure times and predict flight trajectories using historical data and statistical techniques. This stage incorporates variables such as scheduled departure times, historical flight durations, traffic flow patterns, and weather conditions to estimate when and where an aircraft will be during its flight (Ye et al. [4]).
2) **Post-Departure Updates**: Once airborne, real-time radar data refines trajectory predictions, ensuring alignment with actual flight paths and enhancing forecast reliability (Wu and Pan [5]).

While trajectory-based methods provide detailed and accurate short-term predictions, they face significant challenges. As Pérez Moreno et al. [6] highlight, segmenting trajectories for precision increases computational demands and data requirements. Additionally, these models struggle with unplanned deviations, such as rerouting due to congestion or adverse weather, which can degrade prediction accuracy (Vos [2]; SESAR Joint Undertaking [7]).

2) Aggregate Demand Prediction: To address the limitations of trajectory-based approaches, researchers have developed aggregate models that simplify airspace management by focusing on traffic flow between defined airspace blocks rather than individual flight paths. A study by Sridhar et al. [8] introduced an aggregate model for the U.S. National Airspace System, partitioned into 22 airspace blocks and one international block. The study showed that combining multiple aggregate models with hypothesis testing improved demand forecasting accuracy, achieving root-mean-square errors between 1.79 and 2.64 aircraft.

Aggregate models offer computational efficiency and scalability, making them suitable for large-scale demand forecasting. However, they lack the granularity of trajectory-based methods, potentially overlooking individual flight behaviors and dynamic changes in airspace utilization (Delahaye et al. [9]; Bubalo and Daduna [10]). Additionally, their predictive accuracy diminishes over longer forecasting horizons due to the static nature of the transition matrices and their inability to account for real-time disruptions (Könnemann [11]).

To overcome the limitations of traditional demand prediction methods, researchers like Vos [2] and Li et al. [12] have turned to machine learning techniques. These advanced computational methods offer the ability to handle complex, non-linear relationships and adapt to real-time changes, addressing the challenges of scalability, computational demands, and prediction accuracy inherent in traditional models.

B. Forecasting methods

Traditional methods, such as stochastic modelling and Poisson distributions, have been foundational in predicting flight delays and demand. However, they often overlook the complex dependencies and variables influencing airport operations [13, 14]. Time series methods, like the Clustered Airport Modeling (CAM), incorporate network-based information of airports to enhance prediction accuracy by leveraging structural features and clustering airports with similar delay patterns [15].

Advancements in computational technologies and the need for real-time responsiveness have driven a shift toward more adaptable and sophisticated forecasting methods. Machine learning techniques have emerged as essential tools for predicting and managing the complex, multi-faceted nature of air traffic. They excel at modelling non-linear relationships and processing large datasets from diverse sources—such as weather conditions, flight schedules, and air traffic flow patterns—to enhance prediction accuracy and reliability.

Within the field of machine learning, supervised learning stands out as particularly well-suited to these applications. Supervised learning leverages historical data labelled with actual outcomes—such as recorded departure times or airspace demand metrics—to train models. By learning the mapping between inputs and known outputs, supervised learning algorithms develop predictive capabilities that can be applied to new, unseen data. This approach not only captures the inherent complexities of air traffic patterns but also allows for continuous improvement as new data becomes available, making it a robust and scalable solution for dynamic air traffic management challenges.

1) Supervised Learning in Demand Prediction: Supervised learning techniques, including Random Forest (RF), Graph Neural Network (GNN), Convolutional Neural Network (CNN) and Recurrent Neural Network (RNN), LSTM networks have been applied to predict various outcomes. These outcomes include individual and aggregate arrival and departure delays, as well as airspace complexity.

- **Random Forest (RF):** An ensemble supervised learning technique suitable for handling complex, non-linear interdependencies. Guo et al. [16] proposed a hybrid method leveraging Random Forest Regression and a Maximal Information Coefficient to effectively predict flight departure delays, outperforming conventional models.
- **Graph Neural Network (GNN):** Extend neural networks to data represented as graphs, capturing dependencies in both feature space and data structure. Guo et al. [17] developed a Spatio-temporal Graph Dual-Attention Neural Network to predict departure delays at long horizons.
- **Convolutional Neural Network (CNN):** Effective in processing data with grid-like topology, capturing complex structures and patterns. Qu et al. [18] presented methods using CNN models to predict flight delays by integrating flight data and meteorological information.
- **Recurrent Neural Network (RNN):** Designed to recognize patterns in sequences of data, making them ideal for tasks where the order of events is important. Li et al. [12] introduces a two-stage CNN-LSTM-Random Forest model integrating spatial and temporal data for flight delay predictions. Sun et al. [19] employed an LSTM model to predict arrival and departure delays using the EUROCONTROL R&D dataset, achieving high precision in their predictions. Finally, Wang et al. [20] developed a convolutional LSTM with a multi-head attention mechanism for predicting civil traffic flow, which outperformed several other models. Their multi-head attention LSTM framework has served as a foundation for this research.

Finally, in research with similar data available and with an emphasis on the knock-on effect, Ramon Dalmau [21] estimated Estimated Time of Arrival (ETA) for flights crossing the MUAC area with a 30% accuracy increase compared to baseline predictions.

In summary, supervised learning techniques provide significant advancements in predicting flight delays and air traffic outcomes. Ensemble methods, like RF excel at capturing non-linear relationships, while GNN effectively model spatial dependencies in networked data. CNN are well-suited for identifying patterns in grid-like data, and RNN, particularly LSTM-based models, leverage temporal dependencies for sequential predictions. Hybrid approaches, such as CNN-LSTM and attention-based frameworks, combine spatial and temporal features for improved accuracy. These methods complement each other, with the choice depending on data structure and prediction needs, enabling more precise forecasts and better operational efficiency in air traffic management.

C. Research Gap

Despite the advancements in predictive modelling techniques, several gaps remain in the literature. Existing models that forecast the ATOT for flights often make predictions on shorter horizons. Traditional trajectory-based and aggregate demand forecasting methods provide valuable insights but cannot incorporate non-linear, dynamic interactions and long-term temporal dependencies that significantly influence departure delays.

Although promising, current machine learning approaches, including ensemble models and conventional neural networks, have shown limitations in accounting for factors such as the knock-on effect, where delays in one flight propagate to subsequent flights sharing the same aircraft. Additionally, the integration of diverse and real-time as Electronic Flight Data (EFD) messages, weather reports and filed flight data across a larger area—has not been fully explored, leaving substantial room for improvement.

To be noted that the DST currently employed by LVNL relies on a Random Forest with limited predictive accuracy, largely due to the uncertainty surrounding departure times. Enhancing ATOT prediction models to accommodate a broader scope of flight data, incorporating network effects as the knock-on effect and other congestion factors, and exploit advanced neural architectures like an LSTM network is therefore potentially improving. Such improvements are expected to result in more reliable demand forecasts, enabling better demand-capacity balancing, and leading to reduced delays and enhanced operational efficiency at airports like Amsterdam Schiphol.

III. Case study: Schiphol Airport

Amsterdam Schiphol Airport is one of Europe's busiest hubs, handling 441,969 flights in 2023 (Royal Schiphol Group [22]). At such high traffic volumes, forecasting departure times accurately is crucial to maintain efficient flow management, avoid congestion, and minimize holding patterns. Although LVNL has already deployed a DST to assist ATCos and Flow Management Position Controler (FMPC) in balancing capacity with predicted demand, the tool's predictive accuracy remains limited, especially for longer horizons.

A key focus of this forecasting challenge is the ATOT, which has proven to be the largest uncertainty in determining the ETA. Accurate ATOT predictions have proven to be difficult because they are dependent on operational inefficiencies (e.g., runway capacity, gate availability, turnaround delays), environmental conditions (e.g., weather disruptions), and systemic pressures (e.g., congestion, inter-airport dependencies). By capturing these interrelated factors, refined ATOT forecasts can enhance both the predictability and the overall efficiency of air traffic operations.

For Schiphol, extending the forecast horizon to four hours is particularly valuable, as it ensures significant flights are still in their pre-departure phase—thereby enabling proactive measures to manage impending demand peaks. However, forecasting at this timescale is challenging due to weather variability, potential knock-on effects from upstream airports, and the inherent complexity of ground operations. The Airport Collaborative Decision Making (A-CDM) system, which coordinates real-time ground-operation data among airlines, airport authorities, and ground handlers, provides critical inputs for improving DST accuracy. Fully leveraging these data streams, however, calls for more advanced forecasting methods that can accommodate Schiphol's complex operational dynamics.

According to Dijkstra and Calis [23], the DST must deliver precise, real-time traffic analyses, identify runway and airspace capacity constraints, and offer features such as Air Traffic Flow Management (ATFM) regulation, scenario comparison, and delay balancing. Additionally, the system should generate early overload alerts to facilitate timely interventions. Collectively, these requirements aim to keep Schiphol's operations safe and efficient, even under fluctuating traffic demands and evolving operational conditions. In the next chapter, the machine learning approach that integrates operational, weather, and systemic factors to yield more accurate ATOT predictions is described.

IV. Methodology

The methodology utilized to develop a time series prediction model employs a LSTM network augmented with an attention mechanism (LSTM-MHA). This approach includes data preparation, the design of a neural network architecture, and a training procedure aimed at improving predictive performance during inference. This section provides a comprehensive overview of the methodology employed to develop a time series prediction model, structured as follows: Subsection IV-A outlines the data sources used, emphasizing their significance and preparation processes; Subsection IV-B details the preprocessing steps, including feature engineering, standardization, and sequence construction; Subsection IV-C describes the architecture of the proposed LSTM-MHA model, explaining its integration of LSTM layers and attention mechanisms; Subsection IV-D elaborates on the training procedure, including loss functions, optimization, and evaluation metrics; Subsection IV-E introduces the Recursive LSTM-MHA forecaster for real-time predictions; and Subsection IV-F compares alternative machine learning models to validate the effectiveness of the proposed approach.

A. Data

The success of the predictive methods outlined in the previous chapter is fundamentally dependent on the availability and quality of relevant data. Accurate forecasting and modelling, especially for predicting flight delays, require extensive datasets that encompass a wide range of factors, including weather conditions, ATFM regulations, and A-CDM data. However, accessing these diverse data sources can be challenging due to availability constraints and real-time accessibility. This subsection discusses the data sources utilized in this research, detailing the types of data collected and their significance in enhancing prediction models.

A critical consideration in selecting these data sources is the desired prediction horizon of four hours. The reliability of data often diminishes as the time gap between the prediction and the actual event increases. At this horizon, updates to flight plans have proven to be insufficiently accurate. Consequently, supplementary data sources are required to enhance the predictive capabilities of the models. While flight plan data from other flights can provide valuable insights into factors like knock-on delays, this research focuses on data sources that are available in real-time to align with the operational needs of the DST, even though the actual data used is historical.

For training, the dataset is constructed such that the first timestep begins 5 hours before the ATOT. In real-time scenarios, where the ATOT is unavailable, the data points in the testing dataset start 5 hours before the filed Estimated Take Off Time (ETOT) and continue up to the actual ATOT. Both datasets contain the latest update on a 5 minute resolution.

1) EFD Messages: The primary data source for this project are EFD messages, provided by LVNL. These messages are the primary stream of information between individual flights and LVNL with EUROCONTROL serving as an intermediary. The messages are sent regularly, in addition to whenever a significant update can be given. The data extracted out of the EFD messages can be found in Table I, and more thorough explanation in Appendix A.

Among other features, however, not all features impact the ATOT and thus are not used in this research.

2) Weather Data: Weather conditions significantly impact airport capacity and flight operations. Despite inherent uncertainties in weather forecasts, they are essential for making informed decisions regarding flight scheduling and departures, particularly within the four-hour prediction horizon. Terminal Aerodrome Forecast (TAF) data, which provides routine weather predictions at airport weather stations, is particularly useful for forecasting and predicting the impact of weather on departure times.

The features from TAF reports which are used are:

TABLE I: EFD Message Fields and Data Types

Field	Data Type
Timestamp	Datetime
Flightstate	Categorical
Flightplan ID	Identifier
Aircraft ID	Identifier
Aircraft Type	Categorical
EOBT	Datetime
TSAT	Datetime
TOBT	Datetime
CTOT	Datetime
ETA	Datetime
ADEP	Categorical (Airport Code)
ADES	Categorical (Airport Code)
Flightplan	Text/String
Message Type	Categorical

- **Wind:** Speed, direction, and gusts, which are crucial for runway utilization and influence takeoff and landing operations.
- **Visibility:** Low visibility can trigger air traffic flow restrictions, potentially causing delays. Maximum visibility is $10km$ or CAVOK.

More features could be extracted, however, for this research, these were selected as the most crucial as the focus is on developing a model that works on days without significant weather. Predictions for weather conditions are available up to 18 hours in advance, which is within the scope of the 4-hour horizon desired for this research. TAF data thus provides critical insight into the weather conditions that could affect airport operations and flight schedules. The TAF data was extracted from Ogimet [24].

3) EUROCONTROL DDR dataset: The EUROCONTROL Demand and Capacity Data Reporting (DDR) dataset provides crucial insights into air traffic flow across Europe, with a specific focus on filed flight plans submitted by airlines. This dataset is instrumental in understanding how air traffic demand interacts with available capacity in European airspace, particularly for forecasting delays and optimizing traffic management. The primary component of the DDR data used in this research are the Filed Flight Messages. These messages contain essential information about flight plans filed by airlines, which include details such as the aircraft ID, departure and destination airports, flight routing, and the planned departure and arrival times. They are continuously updated and provide real-time flight information that can be used to track the progress of individual flights. Filed flight messages allow for the analysis of flow disruptions, as delays or routing changes can be traced, and their effects on airport and airspace capacity can be assessed.

B. Data Preparation

The dataset consists of both fixed and time-varying features. Fixed features represent the filed flight data from both the EFD and DDR and weather data, whereas time-varying features capture the updates which are given within the window.
The set of fixed features is denoted as $\mathbf{X}_{\text{fixed}} \in \mathbb{R}^{N\times F}$, where N represents the number of samples and F the number of fixed features. Time-varying features are represented as $\mathbf{X}_{\text{time}} \in \mathbb{R}^{N\times T\times V}$, with T indicating the number of time steps and V the number of time-varying features. The preparation process involves several key steps:

1) Knock on Effect: The knock-on effect arises when delays propagate from one flight to subsequent flights sharing the same aircraft. The knock-on delay is calculated by considering the dependencies between successive flights. The method relies on both historical and real-time data to estimate the readiness of the aircraft for its next departure.

The process is started by firstly retrieving the most recent Actual Arrival Time (ATA) or ETA of the preceding flight. If no arrival information is available, an estimated arrival time is calculated using filed schedule data, such as the Estimated Off-Block Time (EOBT) and typical flight durations. The expected ready time for the aircraft is then computed by adding the minimum turnaround time to the arrival time. This turnaround time varies per aircraft type and accounts for activities such as passenger deboarding, refuelling, and cleaning.

The knock-on delay is determined by comparing the computed ready time of the aircraft to the scheduled departure time of the subsequent flight. If the ready time exceeds the scheduled departure time, the difference represents the knock-on delay. This value is constrained to be non-negative, as early readiness does not contribute to delays. Mathematically, the knock-on delay δ for flight n is computed as:

$$\delta_n = \max(0, t_{\text{ready,n}} - t_{\text{departure,n}}) \tag{1}$$

where t_{ready} represents the computed aircraft ready time, and $t_{\text{departure}}$ is the scheduled departure time of the current flight.

2) Feature Preparation: Several features receive special attention due to their inherent meaning. Time of day-related features are transformed into circular representations to more accurately capture the cyclical nature of time—for example, representing the time difference between 23:00 and 01:00 as two hours. Categorical features, such as departure airport and day of the week, as well as string-based features like model type, flight state, and flight offblock status, are encoded using dummy variables. This encoding, as shown in equations (2) and (3), facilitates the model's ability to interpret and leverage these cyclical inputs effectively.

$$\text{Time}_{\sin} = \sin\left(2\pi\frac{\text{Time}}{T}\right) \tag{2}$$

$$\text{Time}_{\cos} = \cos\left(2\pi\frac{\text{Time}}{T}\right) \tag{3}$$

3) Standardization: To ensure numerical stability and improve convergence during training, the features are standardized using the a standard scaler. This transformation is applied separately to each of the fixed features and each of the time-varying features. Because of the recursive approach which will be explained later in this section, it is important that features are scaled irrespective of their timestep.

a) Fixed Features.: For the fixed features $\mathbf{X}_{\text{fixed}}$, the standardization is performed independently for each feature across all samples. That is, for feature j:

$$\mathbf{X}'_{\text{fixed}}[i,j] = \frac{\mathbf{X}_{\text{fixed}}[i,j] - \mu_{\text{fixed},j}}{\sigma_{\text{fixed},j}}, \tag{4}$$

where $\mu_{\text{fixed},j}$ and $\sigma_{\text{fixed},j}$ are the mean and standard deviation of feature j across all samples.

b) Time-varying Features.: For the time-varying features $\mathbf{X}_{\text{time}}$, the standardization is performed independently for each base feature across all timesteps and samples. Specifically, for feature j:

$$\mathbf{X}'_{\text{time}}[i,t,j] = \frac{\mathbf{X}_{\text{time}}[i,t,j] - \mu_{\text{time},j}}{\sigma_{\text{time},j}} \tag{5}$$

where $\mu_{\text{time},j}$ and $\sigma_{\text{time},j}$ are the mean and standard deviation of feature j, aggregated over all timesteps and samples.

c) Binary Features.: Binary features are excluded from standardization because scaling them can distort their categorical nature, leading to unintended interpretations by the model.

This preprocessing ensures that both fixed and time-varying features contribute comparably to the model, avoiding dominance by features with larger magnitudes or higher variance.

4) Sequence Construction: Time-varying features are reshaped to form sequences suitable for input into the LSTM:

$$\mathbf{X}_{\text{seq}} = [\mathbf{X}'_{\text{fixed}}, \mathbf{X}'_{\text{time}}] \tag{6}$$

where $\mathbf{X}_{\text{seq}} \in \mathbb{R}^{N \times T \times (F+V)}$ represents the combined feature set repeated across all time steps.

5) Target Variable Scaling: To improve convergence and the stability of the training process the target variable $y \in \mathbb{R}^N$ is also standardized:

$$y'_n = \frac{y_n - \mu_y}{\sigma_y} \tag{7}$$

After compiling all the relevant features, Figure 1 presents a correlation heatmap of the numerical variables at both the initial and latest timesteps.

C. LSTM-MHA Architecture

The model architecture integrates a stacked LSTM network with a Multi-Head Attention mechanism (LSTM-MHA) to effectively capture both local and global temporal dependencies while enhancing feature weighting. As visualized in Figure 2, the input sequence is first processed through multiple LSTM layers, extracting sequential patterns and retaining long-term memory. The output is then passed to the Multi-Head Attention mechanism, where attention heads compute content-based and relative positional scores to dynamically focus on the most relevant timesteps. The aggregated outputs from the attention heads are transformed by a Dense layer and regularized with a Dropout layer to prevent overfitting. Finally, the model's output layer generates predictions, ensuring accurate time series forecasting by combining temporal dependencies and dynamic feature weighting. While this architecture provides robust interpretability and scalability, it is computationally intensive and memory-demanding, which may limit its applicability in real-time or resource-constrained environments.

1) LSTM Architecture: The core part of the LSTM-MHA is the LSTM cells. A schematic of an LSTM cell can be found in Figure 3, showcasing the multiple gates:

- Forget Gate (FG): This gate decides what information from the cell state should be discarded or retained. It uses the current input and the output from the previous step to generate values between 0 and 1 through a sigmoid layer, with 0 indicating that the information should be forgotten, and 1 indicating that it should be retained and used to modify the hidden state.
- Input Gate (IG): Similar to the forget gate, the input gate decides which of the current input information should be used to update the cell state. It determines the relevance of the input information in the context of the new cell state.
- Cell-State Gate (CG): This gate calculates the candidate values for the new cell state by considering the previous output and the current input. It uses a hyperbolic tangent activation function, which outputs values in the range of -1 to 1, to generate these candidate values. The new cell state is then determined by combining the outputs of the forget gate, input gate, and the cell-state gate.
- Output Gate (OG): The output gate decides what the next output should be based on the cell state and the current input. This gate determines the next cell output, which is modified according to the cell state to produce the final output.

The LSTM network processes input sequences to model temporal dependencies. For each time step t, the LSTM computes:

$$\mathbf{i}_t = \sigma\left(\mathbf{W}_i \mathbf{x}_t + \mathbf{U}_i \mathbf{h}_{t-1} + \mathbf{b}_i\right) \tag{8}$$

$$\mathbf{f}_t = \sigma\left(\mathbf{W}_f \mathbf{x}_t + \mathbf{U}_f \mathbf{h}_{t-1} + \mathbf{b}_f\right) \tag{9}$$

$$\mathbf{o}_t = \sigma\left(\mathbf{W}_o \mathbf{x}_t + \mathbf{U}_o \mathbf{h}_{t-1} + \mathbf{b}_o\right) \tag{10}$$

$$\tilde{\mathbf{c}}_t = \tanh\left(\mathbf{W}_c \mathbf{x}_t + \mathbf{U}_c \mathbf{h}_{t-1} + \mathbf{b}_c\right) \tag{11}$$

$$\mathbf{c}_t = \mathbf{f}_t \odot \mathbf{c}_{t-1} + \mathbf{i}_t \odot \tilde{\mathbf{c}}_t \tag{12}$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh\left(\mathbf{c}_t\right) \tag{13}$$

In these equations, $\mathbf{x}_t$ represents the input at time t, $\mathbf{h}_t$ the hidden state, $\mathbf{c}_t$ the cell state, $\sigma(\cdot)$ the sigmoid activation function, and $\odot$ denotes element-wise multiplication. $\mathbf{h}_t$ represents the prediction for timestep t. The weight matrices $\mathbf{W}$ and $\mathbf{U}$, along with biases $\mathbf{b}$, are learnable parameters.

2) Multi-Head Attention Mechanism: To enhance the model's ability to focus on relevant features, a multi-head

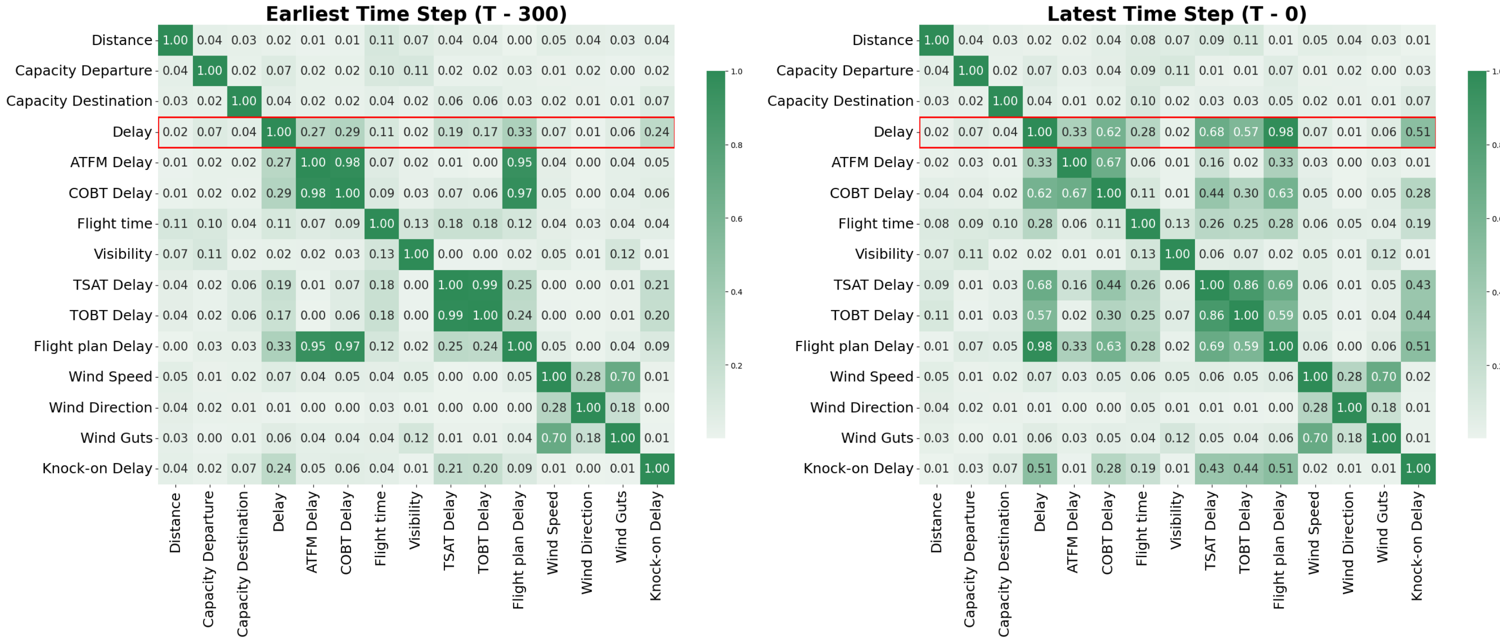


Fig. 1: Correlation heatmap for all numerical features for the first and latest timestep. Highlighted value is the departure delay.

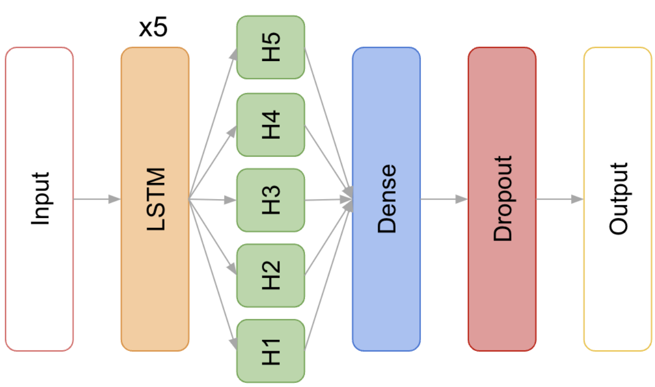


Fig. 2: Schematic of the LSTM-MHA

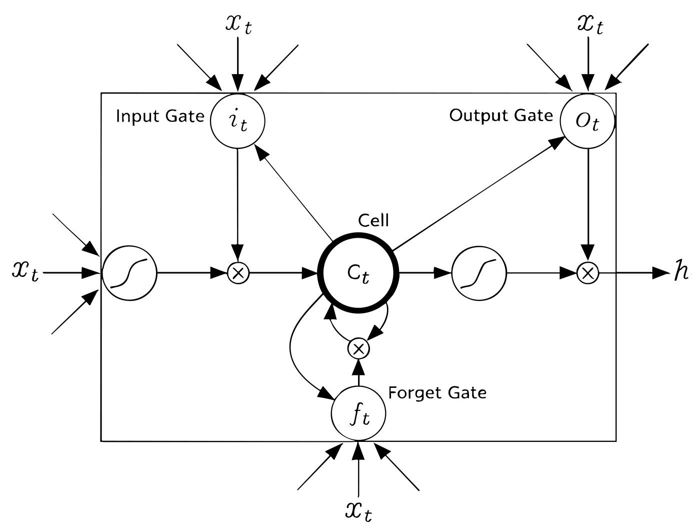


Fig. 3: Schematic of the Long Short-Term Memory Cell [25]

attention mechanism with relative positional encoding is employed. This mechanism allows the model to focus on different parts of the input sequence, capturing various dependencies and patterns. By computing attention scores, the model can assess the importance of each time step in the sequence when making predictions.

a) Scaled Dot-Product Attention: For each head, attention weights are computed using queries $\mathbf{Q}$, keys $\mathbf{K}$, and values $\mathbf{V}$ derived from the LSTM outputs:

$$\mathbf{Q} = \mathbf{H}\mathbf{W}_Q \tag{14}$$
$$\mathbf{K} = \mathbf{H}\mathbf{W}_K \tag{15}$$
$$\mathbf{V} = \mathbf{H}\mathbf{W}_V \tag{16}$$

where $\mathbf{H} \in \mathbb{R}^{T \times H}$ is the hidden state sequence, and $\mathbf{W}_Q$, $\mathbf{W}_K$, $\mathbf{W}_V$ are projection matrices.

Attention scores are calculated as:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\mathbf{Q}\mathbf{K}^\top}{\sqrt{d_k}} + \mathbf{R}\right)\mathbf{V} \tag{17}$$

where d_k is the dimensionality of the keys, and $\mathbf{R}$ represents the relative positional encoding.

b) Relative Positional Encoding: Relative positions between time steps are encoded using an embedding matrix:

$$\mathbf{R}_{i,j} = \mathbf{E}_{\text{pos}}(i - j + \Delta) \tag{18}$$

where Δ is a constant to shift the indices into a valid range, and $\mathbf{E}_{\text{pos}}$ is the positional embedding matrix.

c) Multi-Head Attention: Multiple attention heads capture information from different representation subspaces. The outputs from each head are concatenated along the feature dimension and then projected to the desired dimensionality:

$$\text{MH}(\mathbf{Q},\mathbf{K},\mathbf{V}) = [\mathbf{head}_1;\mathbf{head}_2;\ldots;\mathbf{head}_h]\,\mathbf{W}_O, \quad (19)$$

$$\mathbf{head}_i = \text{Attention}(\mathbf{Q}_i,\mathbf{K}_i,\mathbf{V}_i). \quad (20)$$

Here, $\mathbf{W}_O \in \mathbb{R}^{(h\cdot d_k)\times d_{\text{model}}}$ is the output projection matrix, and h is the number of attention heads.

3) Output Layer: Outputs from the attention mechanism are passed through a fully connected layer to produce final predictions:

$$\hat{y}_t = \mathbf{W}_{\text{FC}}\mathbf{z}_t + \mathbf{b}_{\text{FC}}, \quad (21)$$

where $\mathbf{z}_t$ is the attention output at time t, and $\mathbf{W}_{\text{FC}}$, $\mathbf{b}_{\text{FC}}$ are learnable parameters.

D. Training Procedure

1) Loss Function: The model is trained to minimize a composite loss function combining Mean Squared Error (MSE) and Mean Average Error (MAE)

$$\mathcal{L} = \lambda_{\text{MSE}} \cdot \text{MSE} + \lambda_{\text{MAE}} \cdot \text{MAE} \quad (22)$$

where:

$$\text{MSE} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left(\hat{y}_{n,t} - y_{n,t}\right)^2 \quad (23)$$

$$\text{MAE} = \frac{1}{NT}\sum_{n=1}^{N}\sum_{t=1}^{T}\left|\hat{y}_{n,t} - y_{n,t}\right| \quad (24)$$

The weights λ_{MSE} and λ_{MAE} control the contribution of each loss component, both taken as 0.5.

2) Optimization: The Adam optimizer is employed with a learning rate η and weight decay β for L2 regularization:

$$\theta^{(k+1)} = \theta^{(k)} - \eta\left(\nabla_\theta \mathcal{L} + \beta\theta^{(k)}\right), \quad (25)$$

where θ represents the model parameters.

3) Early Stopping: To prevent overfitting, early stopping is applied based on validation loss with a patience parameter p:

$$\text{Stop training if } \mathcal{L}_{\text{val}}^{(k)} > \mathcal{L}_{\text{val}}^{(k-p)} - \delta, \quad (26)$$

where δ is a minimal improvement threshold.

4) Data Augmentation with Sequence Shifting: To enhance robustness, training data is augmented by shifting sequences forwards and backwards in time. For a shift value s, input and target sequences are adjusted:

$$\mathbf{X}_{\text{shifted}}(t) = \mathbf{X}(t+s), \quad (27)$$

$$y_{\text{shifted}}(t) = y(t+s). \quad (28)$$

Padding is applied by repeating boundary values where necessary.

E. Recursive LSTM-MHA Based Forecaster

The Recursive LSTM-MHA Based Forecaster is a pivotal component of the predictive modelling framework, designed to iteratively generate forecasts by leveraging the LSTM-MHA's power. It is the last step to a fully functional model that can work in real-time. This subsection describes the operational mechanics of the Rolling Forecaster, integrating seamlessly with the previously established data preparation and LSTM architecture. The Recursive Forecaster methodology integrates the LSTM-MHA architecture with an iterative forecasting strategy to effectively model and predict time series data. The approach captures complex temporal dependencies and dynamically updates predictions, ensuring robust performance across varying forecast horizons.

1) Mathematical Framework: Taking a time series dataset comprising N samples as input, each consisting of a set of fixed features and time-varying features across T timesteps. The Recursive Forecaster aims to predict future values of the target variable by iteratively updating its input window with newly forecasted values.

a) Notation and Definitions:

- Let $\mathbf{X} \in \mathbb{R}^{N\times T\times F}$ denote the input feature tensor, where F represents the total number of features. The dataset used contains data ranging from 300 minutes before filed takeoff, untill actual takeoff with $\Delta t = 5$. This implies a variable time-series length per flight.
- $\mathbf{y} \in \mathbb{R}^N$ is the target variable matrix, where each entry y_n corresponds to the delay of flight n.
- The window size is fixed at $W = 61$, or 300 minutes.
- The prediction horizon is denoted by H, indicating the number of future steps ahead to forecast.
- t_{start} indicates the time 300 minutes before the predicted takeoff time, indicating the start of the forecasting window.

b) Forecasting Mechanism: The Recursive Forecaster operates through a recursive prediction strategy, encapsulated in the following mathematical formulation:

1) **Initialization:** For each sample n, initialize the input window with the first W timesteps because no information is available apart from the first timestep:

$$\mathcal{W}_n^{(0)} = \{\mathbf{X}_{n,0}, \mathbf{X}_{n,0}, \ldots, \mathbf{X}_{n,0}\} \quad (29)$$

2) **Iterative Prediction:** For each forecast step $h = 1, 2, \ldots, H$, perform the following:
   a) **Model Input:** Input the current window $\mathcal{W}_n^{(h)}$ into the trained LSTM-MHA network to obtain the prediction:

   $$\hat{y}_{n,W+h} = f(\mathcal{W}_n^{(h)}; \boldsymbol{\theta}) \quad (30)$$

   where f represents the predictive function parameterized by weights $\boldsymbol{\theta}$.

   b) Select the value at forecast horizon H from the output of length W.

c) **Window Update:** Incorporate the newly predicted y_n to derive t_{start} and create a new input window.

$$\mathcal{W}_n^{(h+1)} = \{\mathbf{X}_{n,t_{start}}, \mathbf{X}_{n,t_{start}+1}, \ldots, \mathbf{X}_{n,t_{start}+W}\} \quad (31)$$

Whenever data for a timestep t is unavailable, the input window is padded by repeating the last known feature vector until the window length reaches W.

3) **Recursive Forecasting:** Repeat the iterative prediction and window update steps until the flight has departed, which the model detects by the absence of data.

The forecasting mechanism is also given in the following pseudo-code:

Algorithm 1 Recursive LSTM-MHA Based Forecaster

for each sample n **do**
  Initialize:
  Set input window $\mathcal{W}_n^{(0)} = \{\mathbf{X}_{n,0}, \ldots, \mathbf{X}_{n,0}\}$ (repeated W times)
  Set predicted takeoff time $t_{\text{pred}} = 300$
  Initialize time indices $t_{\text{start}} = t_{\text{data}} = t_{\text{model}} = 0$
  while data is available **do**
    Update Features:
    Determine input window $\mathcal{W}_n^{(h)}$ based on indices from t_{start} to t_{data}, limited to length W
    Apply padding if necessary to maintain consistent input length
    Predict Next Value:
    Compute $\hat{y}_n = f(\mathcal{W}_n^{(h)}; \boldsymbol{\theta})$
    Inverse transform $\hat{y}_n$ to original scale
    Update Indices and Times:
    Update predicted takeoff time t_{pred} using $\hat{y}_n$
    Adjust t_{start}, t_{current}, t_{data}, and t_{model} based on new predictions
  end while
end for

2) Performance Evaluation: The performance of the Recursive Forecaster is assessed using standard evaluation metrics, including Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) scores. These metrics are computed both overall and on a per-timestep basis to provide granular insights into the model's predictive performance across different forecast horizons.

a) Error Analysis: To understand the distribution and propagation of prediction errors, the following analyses are conducted:

- **Overall Error Metrics:** Compute aggregated MAE, RMSE, and R^2 across all samples and timesteps.
- **Per-Timestep Metrics:** Calculate MAE, RMSE, and R^2 for each forecast horizon to identify patterns or inconsistencies in prediction accuracy over time.
- **Residual Analysis:** Analyze residuals to detect any systematic biases or anomalies in the forecasting process.

F. Comparison Models

To ensure the selection of the most suitable model, several state-of-the-art machine learning techniques were implemented and compared. These include ensemble methods such as RF, LightGBM, and CatBoost, which are widely used in literature but tailored to this specific application. Additionally, the Transformer model was evaluated for its extensive application in machine learning tasks and its inherent attention mechanism, which facilitates the modelling of complex dependencies.

1) Random Forest (RF): A RF model was implemented to assess the impact of the additional input features on predictive performance. RF is an ensemble method that constructs multiple decision trees using random subsets of features and data, combining their predictions to enhance robustness and mitigate overfitting. In this study, features were standardized and aggregated using statistical measures such as mean, maximum, and variance within each time-series window. Hyperparameters, including the number of trees and maximum depth, were fine-tuned using grid search and cross-validation.

2) LightGBM: LightGBM, a gradient-boosting framework, was utilized to evaluate the effectiveness of boosting methods in this context. Known for its speed and efficiency, LightGBM employs histogram-based feature binning. Features were processed similarly to RF, with categorical data label-encoded and time-varying features aggregated statistically. Key hyperparameters, including the number of leaves and the learning rate, were optimized through cross-validation to minimize the MSE.

3) CatBoost: CatBoost, another gradient-boosting algorithm, was chosen for its ability to natively handle categorical features without extensive preprocessing. This model was included to evaluate the benefits of this capability. Similar to LightGBM, time-varying features were aggregated using statistical measures. Hyperparameter tuning focused on parameters such as tree depth and the number of iterations, employing ordered boosting to improve accuracy and reduce overfitting.

4) Transformer: The Transformer model, leverages self-attention mechanisms to capture long-range dependencies in sequential data. The architecture comprises multiple encoder layers which allow the model to focus on different parts of the input sequence simultaneously, enhancing its ability to understand complex temporal patterns. The Transformer model processes the input data by first projecting it into a higher-dimensional space through a linear layer, applying positional encoding, and then passing it through the stacked Transformer encoder layers. The output from the final encoder layer is then projected back to the desired output dimension through another linear layer. The Transformer shares similarities with the LSTM in processing sequential data. However, while LSTM relies on recurrent connections to capture dependencies, the Transformer employs self-attention mechanisms to model relationships within the data, eliminating the need for traditional attention mechanisms.

5) LSTM: To evaluate the contribution of the Multihead Attention Mechanism, an LSTM model without an attention mechanism was included in the range of models for comparison.

To evaluate the models' performance during inference, algorithms were developed that allow for each model to predict in real-time. The ensemble methods, RF, LightGBM and CatBoost, which utilize a separate model for each timestep, largely follow a similar structure, where the subsequently used model is selected by the previous output. In contrast, the Transformer-based recursive forecaster has larger similarities to the Recursive LSTM-MHA based Forecaster, both in leveraging sequential dependencies and in their ability to maintain context over multiple timesteps.

V. Results

After experimenting with multiple models, this section presents the performance of the proposed LSTM-MHA architecture and compares it to other approaches and the existing DST. Subsection V-A describes the experimental setup, and Subsection V-B outlines the training process. Subsection V-C provides a direct comparison of models, while Subsection V-D measures performance against the baseline. Feature importance is discussed in Subsection V-E, and Subsection V-F explores results under different operational conditions. Finally, Subsection V-G examines how improved forecasts can enhance demand prediction.

A. Experimental Setup

A comparative analysis against the existing DST model was conducted to evaluate the performance of the proposed LSTM-MHA model and assess its suitability for operational deployment. The model was trained on data from March 24th, 2023, to August 31st, 2023, and evaluated on data spanning September 1st to October 28th, 2023. During this two-month evaluation period, a total of 14,142 flights were inbound to Schiphol Airport from European airports, providing a substantial dataset for analysis.

For the models discussed in previous sections, the primary metric of interest is the MAE with respect to the ATOT. However, from the perspective of LVNL, the most valuable metric is the MAE with respect to the time until the aircraft reaches Dutch airspace, referred to as the Cross Border Areas (CBAS) entry time.

Therefore, the selected LSTM-MHA model was compared to the flight plans, current DST model, and the RF with the same dataset. This comparison will highlight the potential improvements in prediction accuracy and operational efficiency that the LSTM-MHA model could offer over the existing model.

B. Training Process and Results

The training dataset was sorted on ETOT and split into training and validation sets, using an 80-20 split, to monitor the model's performance and prevent overfitting. The LSTM-MHA model was trained over multiple epochs, with each epoch representing a complete pass over the entire training dataset. The number of epochs was determined based on the model's convergence behaviour, monitored through the loss function.

Hyperparameter tuning was conducted to optimize the model's performance. Key hyperparameters adjusted included the number of LSTM layers, the hidden layer size, the learning rate, and batch size. The final model configuration was selected based on the best performance on the validation set. Table II lists the resulting parameters, which align with configurations found in the literature but introduce a relatively high level of complexity.

TABLE II: Selected parameters for the LSTM

Parameter	**Value(s)**
Hidden Layer Size	110
Number of Layers	5
Learning Rate	0.00005
Number of Epochs	20
Dropout Rate	0.3
L1 Penalty	0
L2 Penalty	0
Maximum Shift	5
Number of Attention Heads	4

C. Model Comparison

In order to thoroughly assess the performance of the LSTM-MHA model, it is crucial to compare it against other established models trained and evaluated on the same dataset. This comparative analysis ensures a fair assessment and highlights the relative strengths and weaknesses of each modelling approach. The models considered in this comparison include traditional ensemble methods: RF, LightGBM, and CatBoost, as well as neural network architectures like the Transformer and a LSTM without an attention mechanism.

Given the diverse architectures, learning methods, and training strategies of these models, a direct comparison of their computational characteristics and predictive performances can be challenging. The dataset, structured as a time series, naturally suits models like the Transformer and LSTM-based approaches. Ensemble models, by contrast, contain a separate model for each forecast horizon, a strategy that may lead to different resource requirements and error propagation over multiple timesteps. To ensure a balanced evaluation, this section examines several key factors: training speed, model complexity, total model size, simulation run time, and predictive accuracy.

Table III highlights each model's computational attributes, including training time, parameter counts, model size, and simulation runtime. All models were trained on the same NVIDIA 4090 GPU for a consistent performance baseline. The

RF model, with a training time of 85 minutes and 1.28×10^{10} parameters, stands out for its substantial computational footprint, ultimately producing a 14 GB model. This high resource demand can be attributed to its ensemble design, which aggregates numerous decision trees for improved predictive accuracy.

By contrast, gradient boosting methods such as LightGBM and CatBoost are considerably more resource-efficient. With training times of 36 and 31 minutes, respectively, and significantly fewer parameters, these models produce much more compact model files while maintaining competitive performance. Their leaner architectures thus offer a favourable compromise between accuracy and computational overhead.

The Transformer model stands out with the shortest training time (6 minutes) and the smallest size (4.1 KB), attributable to its efficient attention-based architecture. The standard LSTM model, while moderate in training time (17 minutes) and size (41.1 KB), has a relatively long simulation run time (97 minutes), reflecting the recurrent step-by-step processing of sequential data. Introducing multi-head attention to the LSTM resulting in the LSTM-MHA, increases both training time (20 minutes) and simulation run time (115 minutes), demonstrating that enhanced representational capacity comes at a computational cost.

These computational findings highlight the trade-offs inherent in model selection. While Transformer-based models and gradient boosting methods are computationally more efficient, LSTM-based models, particularly with multi-head attention, may yield better predictive accuracy at the expense of longer run times.

TABLE III: Training and Model Information for Various Models

Model	Training Time (min)	Trainable Parameters	Model Size	Simulation Run Time (min)
Random Forest	85	1.28×10^{10}	14 GB	51
LightGBM	36	9.3×10^{5}	84 MB	6
CatBoost	31	1.8×10^{7}	306 MB	12
Transformer	6	4.07×10^{5}	4.1 KB	20
LSTM	17	5.04×10^{5}	41.1 KB	97
LSTM-MHA	20	5.76×10^{5}	41.4 KB	115

In addition to computational factors, predictive performance remains a critical focus. Table IV compares the models based on average MAE, Root-Mean Square Error (RMSE), R^2, and the Standard Deviation (STD) of their errors. Figure 4 visualizes the evolution of MAE across forecasting horizons for all flights within the testing period. The ensemble methods—RF, LightGBM, and CatBoost—show similar performance trends, with limited improvement as the forecast horizon progresses, except for a slight enhancement in the final 100 minutes. While the Transformer model outperforms these ensemble methods, its accuracy is ultimately surpassed by the LSTM-based models. The incorporation of the MHA mechanism further enhances the predictive performance of the LSTM model, albeit with marginal gains. Figure 5, depicting the error distribution for the critical 4-hour horizon, highlights the tighter error margins achieved by the LSTM-MHA model. A key insight from Figure 4 is the tipping point observed at 140 minutes before the ATOT, driven by the introduction of A-CDM updates, such as slot issuance. The LSTM architectures excel at leveraging earlier data to handle this transition effectively, whereas other models struggle to adapt to these updates.

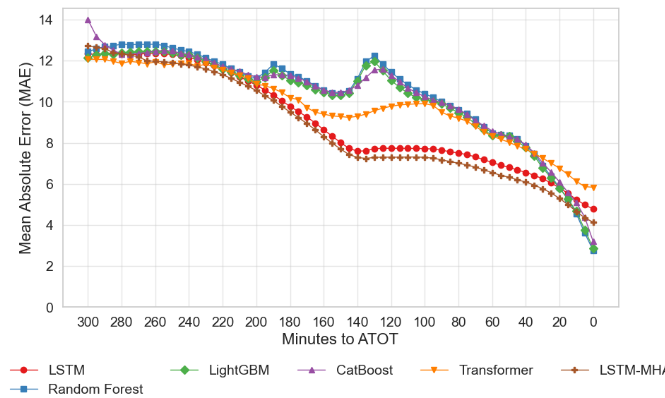


Fig. 4: Mean Absolute Error (MAE) comparison of models.

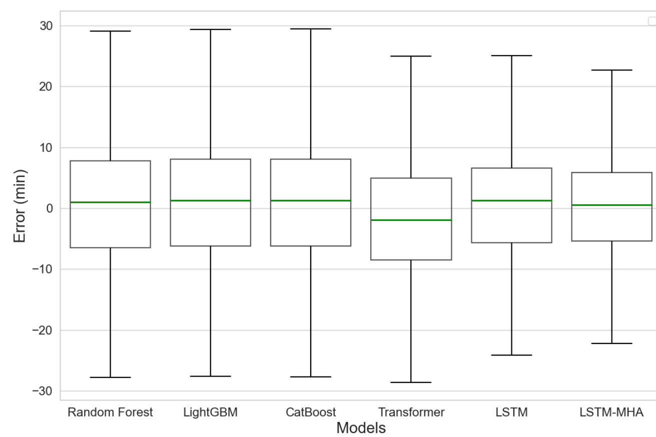


Fig. 5: Distribution of errors for ATOT prediction at a horizon of 4 hours.

Table IV presents the predictive metrics, offering a direct comparison of accuracy, error magnitude, and variability. The ensemble methods (RF, LightGBM, and CatBoost) cluster around a 10-minute average MAE and average R^2 values near 0.58–0.60, reflecting their similar architectures and timestep-by-timestep forecasting approach. Such methods, while straightforward, can suffer from error accumulation over multiple horizons. Neural network-based models, on the other hand, exhibit better performance. The Transformer reduces the average MAE to 9.62 minutes and achieves an average R^2 of 0.66. The LSTM model further improves to an average MAE of 7.65 minutes and R^2 of 0.68. The LSTM-MHA model attains an average MAE of 7.57 minutes and maintains an average R^2 of 0.69, indicating that attention-enhanced recurrent architectures effectively capture long-term dependencies and reduce prediction variability. Overall, the LSTM architecture proves to be the most accurate, and the MHA mechanism marginally improves performance.

TABLE IV: Average Simulation Results of Various Models for a 5-hour window

Model	Average MAE	Average RMSE	Average R^2	Average STD
Random Forest	10.41	15.36	0.58	11.29
LightGBM	10.22	15.02	0.60	11.01
CatBoost	10.33	15.27	0.59	11.26
Transformer	9.62	13.97	0.66	10.12
LSTM	7.65	10.66	0.68	7.26
LSTM-MHA	7.57	10.43	0.69	7.17

D. Comparison with Baseline

When comparing the best-performing model, the LSTM-MHA, to the current DST's RF, the RF with similar data and the flight plan updates, the first notable difference is, at longer horizons both the RF and LSTM-MHA achieve lower MAE, implying successful knock-on effect implementation. Moreover, it becomes evident that the LSTM-MHA more effectively captures the underlying trends and temporal dynamics present in the flight delay data. Figure 6 illustrates the differences in MAE relative to the time until CBAS entry, providing a visual comparison of the models' predictive accuracies. The time till CBAS entry represents the time till the aircraft reaches the Dutch airspace, which is a more important indicator for the demand at Schiphol than the ATOT.

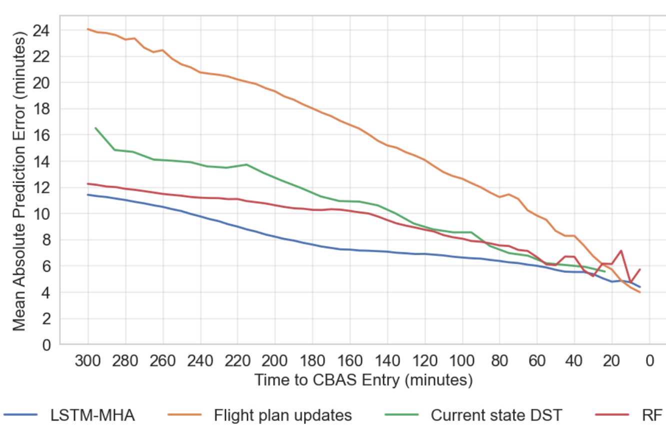


Fig. 6: MAE Comparison between LSTM-MHA, Flight Plan, DST and RF

The LSTM-MHA improves the critical 4-hour horizon MAE from approximately 13.8 minutes to 9.9 minutes, representing a roughly 28% improvement. Although the numerical data is not available, this visual comparison underscores the LSTM-MHA's superior performance compared to both the DST model and the RF with similar data.

Figure 7 further compares the mean prediction error and standard deviation between the LSTM-MHA and DST models. The LSTM-MHA demonstrates not only a lower standard deviation but also a more consistent mean error trajectory over time.

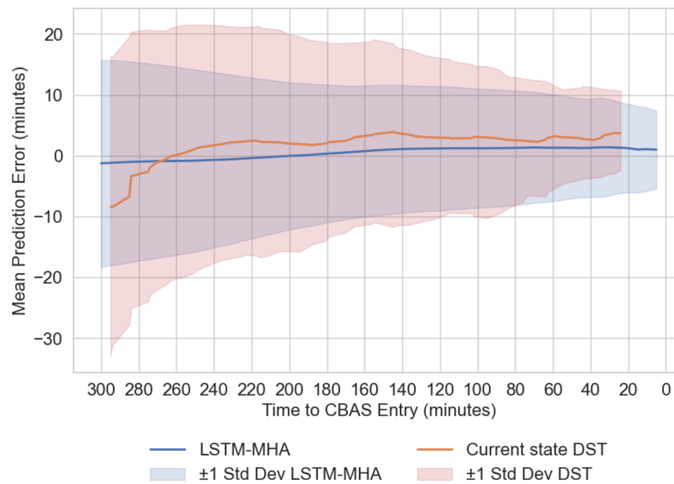


Fig. 7: Standard Deviation Comparison between LSTM-MHA and DST

E. Feature Importance Analysis

To understand the factors influencing the LSTM-MHA's performance, a feature importance analysis was conducted by permuting individual features and observing changes in MAE. Figure 8 shows the top six features impacting predictions. The numerical values for all features can be found in Appendix A

Flight Plan Delay is highly influential throughout the prediction window, peaking at 80% importance, increases as ATOT approaches, emphasizing the growing accuracy of the flight plan updates. *Knock-On Delay* starts with a strong influence but diminishes over time, suggesting that other features like *Flight Plan Delay* capture its effects. Binary feature *Flight State SI* (Slot Issued) which indicates whether a departure slot has been issued, becomes critical halfway through the prediction window, gaining around 150–200 minutes before ATOT. This aligns with the slot issuance window of 180 minutes. *TSAT Delay* and *TOBT Delay* remain consistently low, providing supplementary information, while binary feature *Model Type ACT* gains importance in the final 100 minutes, implying that actual updates are most meaningful in this stage.

The evolving feature importance underscores the need for dynamic modelling approaches to adjust feature weights over time. Early reliance on historical data and a shift toward real-time factors enhance prediction accuracy. These findings suggest operational practices should focus on addressing knock-on delays early and keep on improving on the flight plan updates.

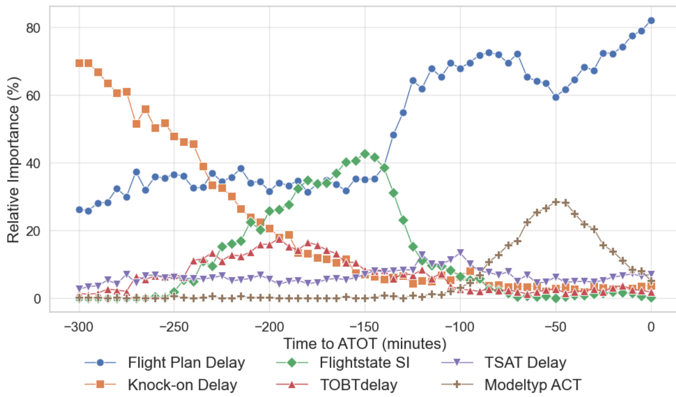


Fig. 8: Relative importance of top features over time.

TABLE V: MAE by Day of the Week at Different Horizons

Day of the Week	4-Hour Horizon	2-Hour Horizon	0-Hour Horizon
Monday	13.141	9.185	5.978
Tuesday	11.655	9.029	6.092
Wednesday	10.445	8.817	5.771
Thursday	13.477	9.002	5.883
Friday	13.839	9.350	6.149
Saturday	10.577	9.057	6.244
Sunday	11.752	9.365	6.283

F. Performance Analysis by Dimension

This section examines the predictive performance of the LSTM-MHA across various dimensions, including departure hour, day of the week, departure airport, and aircraft type. By exploring how the MAE changes under different grouping criteria, patterns and opportunities can be identified.

First, Figure 9 highlights the relationship between flight departure time (grouped by ETOT) and MAE. While differences are relatively small, a pattern emerges: early-morning and late-evening flights generally yield lower prediction errors, whereas midday periods—particularly morning and early afternoon—present greater challenges.

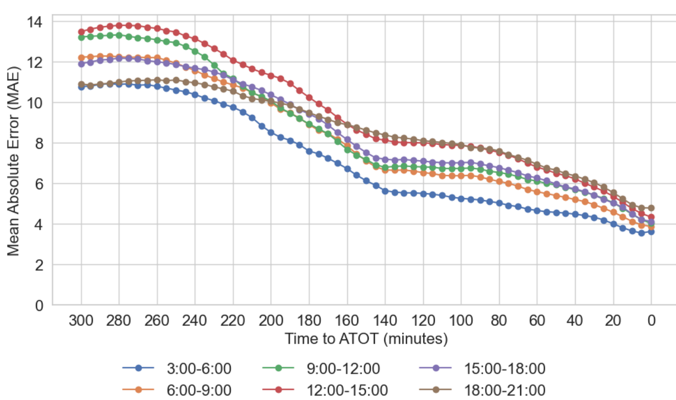


Fig. 9: MAE comparison of flights grouped by ETOT (hour of the day).

Beyond daily patterns, examining variations throughout the week can yield valuable insights. Table V presents the average MAE across different days at three forecasting horizons (4-hour, 2-hour, and 0-hour). The results suggest that on certain days predictions on longer horizons differ, but these effects diminish closer to takeoff. This potentially indicates differences in traffic intensity, scheduling regularity, or operational constraints that influence forecasting performance.

Geographic factors also play a role, as demonstrated in Figure 10, which shows the MAE distribution by departure airport. Although most airports cluster around a median MAE close to zero, some exhibit larger variability and numerous outliers. Airports such as EIDW (Dublin), EGKK (London Gatwick), EGGW (London Luton) and EGPH (Edinburgh) show wider error distributions, suggesting their operational complexity or unique local conditions make accurate forecasts more challenging. The occasional presence of extreme outliers highlights instances of significant prediction deviations, emphasizing the need for further refinement in the forecasting process.

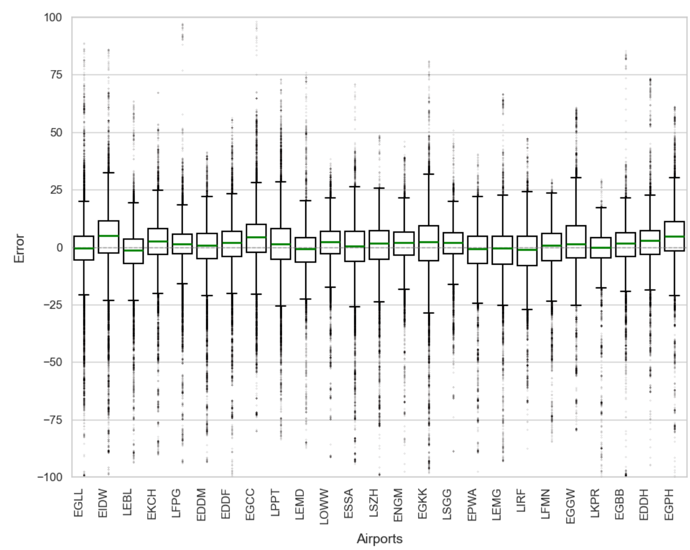


Fig. 10: MAE comparison of flights grouped by departure airport.

In addition to temporal and spatial considerations, the aircraft type also significantly influences forecasting accuracy. Figure 11 compares MAE values (red line and markers) against flight counts (grey bars) for different aircraft types. Commonly observed aircraft (e.g., B738 and A320) exhibit relatively low MAE, reflecting more stable and reliable predictions enabled by abundant training data. Less frequent aircraft types, with fewer than 50 flights during the testing period and grouped into the “Other” category, exhibit significantly higher MAE, highlighting the challenges of accurately predicting outcomes for underrepresented classes due to limited training data. The ’other’ category also included small aircraft, which generally are less schedule adherent. Finally, the A333 type similarly stands out with a higher MAE, reinforcing the need for additional modelling strategies or enhanced data sources to handle these rare cases effectively.

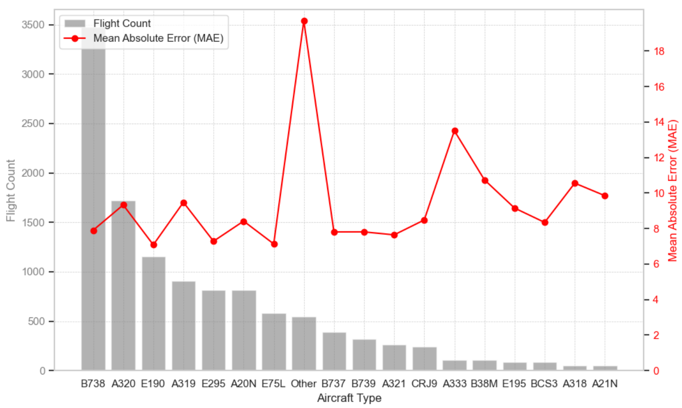


Fig. 11: MAE comparison of flights grouped by aircraft type.

In summary, the analysis reveals how predictive performance varies depending on the time of day, day of the week, departure airport, and aircraft type. Identifying where and when prediction errors are most pronounced provides a roadmap for improving forecasting models, whether by incorporating additional data, refining methodological approaches, or developing specialized strategies for atypical flight categories.

G. Demand Prediction Performance

Given the observed superior performance of the LSTM-MHA model in predicting flight delays, it is reasonable to anticipate that the demand forecast for airspace usage and airport resources would significantly benefit from these more accurate predictions. The LSTM-MHA model’s predictions are compared against the flight plan data, representing the scheduled flight times submitted by airlines.

To simulate the DST, the model is run for an entire day. Figure 12 gives the predicted demand at Schiphol on the critical 4-hour horizon. Significant deviations between the LSTM-MHA and flight plans in aircraft count, and thus demand can be seen, however, still large errors occur. Figure 13 underscores this but highlights the improvement over the flight plan data. It is important to note that, in this research, only ground-related delays and effects are considered. The flight time en route is assumed to be constant and is taken as the flight time provided in the flight plans.

Figure 12 visualizes the demand predictions for an entire day of traffic. While a significant shift in peak times is observed, Figure 13 indicates that the number of incorrect counts remains substantial. An interesting detail is the fact that both flight plans and the LSTM-MHA produce more one-off prediction errors than perfect predictions.

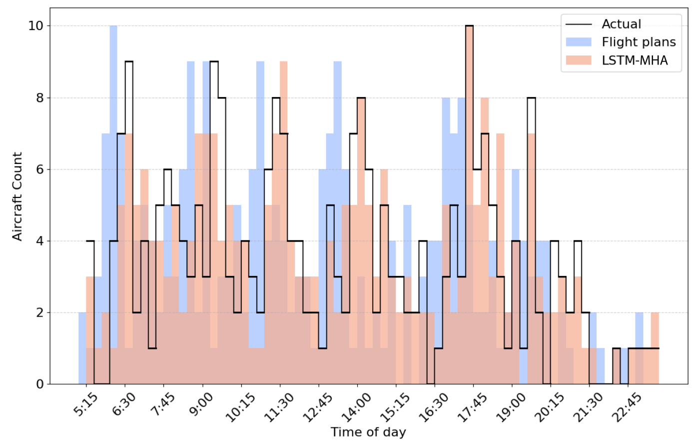


Fig. 12: Example of predicted demand of the LSTM-MHA and flight plans at a 4-hour horizon on September 3rd 2023.

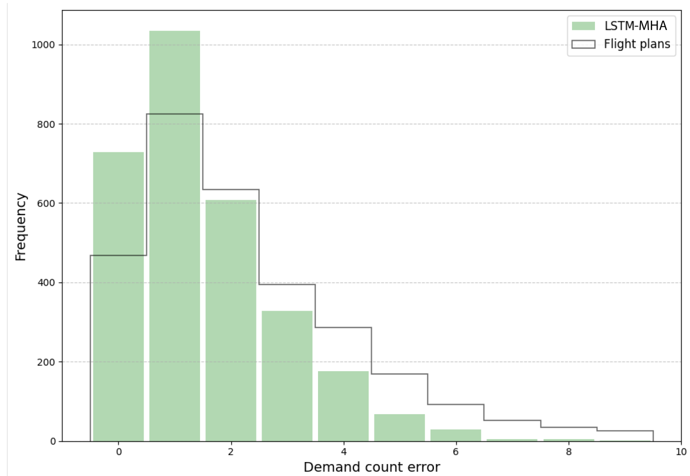


Fig. 13: Demand count error distribution of LSTM-MHA and flight plans

VI. Discussion

The analysis presented demonstrates that the LSTM-MHA model outperforms traditional ensemble methods, the baseline DST system, and simpler neural network architectures in predicting flight departure delays. By leveraging its recurrent and attention-based structure, the LSTM-MHA captures temporal dependencies and subtle variations in operational conditions, leading to more accurate and stable predictions.

Subsection VI-A begins by evaluating the performance gains of the LSTM-MHA relative to other models, noting the trade-off between accuracy and computational overhead. Next, in subsection VI-B the the knock-on effect and its implications for forecasting in dynamic conditions are examined. Subsection VI-C compares the model’s outputs with both the baseline DST and raw flight plans, underlining the

LSTM-MHA's operational advantages. Subsequently, subsection VI-D highlight how the tool can be integrated into existing workflows and generalized to other airport contexts. Finally, subsection VI-E addresses its current limitations, delineate future enhancements—such as periodic retraining and broader data integration—and recommend strategies that could further refine and extend the model's applicability.

A. Model Performance

Comparing multiple modelling approaches, including Random Forest, LightGBM, CatBoost, Transformer, and LSTM variants, reveals that neural network-based models consistently deliver more precise predictions. Among them, the LSTM-MHA stands out, achieving the lowest average MAE (7.26 minutes) between 300 and 0 minutes before departure, and exhibiting strong generalization. Although the Transformer model also shows promise, the LSTM, especially when augmented with a multi-head attention mechanism, yield incremental improvements. The ability to leverage all information available, and focus on the most important parts of this data, allows the LSTM-MHA to outperform the other models. However, these performance gains come at the cost of increased training times and computational complexity. This trade-off, while acceptable in a high-stakes environment like ATM, should be considered when scaling or implementing the model operationally.

B. Incorporating the Knock-on Effect

An essential aspect of accurate flight delay prediction is the knock-on effect—cascading delays stemming from earlier disruptions. This phenomenon has been receiving increasing attention, and accurately modelling it has significant potential to enhance prediction accuracy (EUROCONTROL [26]). Despite limited knock-on data in the current dataset, the feature importance analysis highlights that even a small amount of relevant information substantially improves forecasts. Expanding the data scope to include more European flights and leveraging real-time updates could amplify this effect. Such enrichment would be particularly beneficial for longer forecast horizons and periods closer to takeoff, where operational conditions rapidly evolve.

C. Comparison with Schiphol's Baseline and Flight Plan Data

The LSTM-MHA's performance improvement over the baseline DST (Random Forest-based) and raw flight plan data underscores its operational value. By reducing the 4-hour horizon MAE from approximately 13.8 minutes to 9.9 minutes, the LSTM-MHA demonstrates a roughly 28% improvement. Furthermore, while this research focuses on the longer horizons, the absence of the data after 25 minutes till CBAS entry, raises questions. Although exact DST metrics remain undisclosed, the substantial gains observed against equivalent datasets and the flight plans highlight the model's potential for enhancing prediction accuracy. Lower prediction variability and reduced standard deviations further indicate more reliable forecasts over time.

The success of the LSTM-MHA highlights key improvements that could inform enhancements to the baseline model. Incorporating temporal dependencies and multi-head attention mechanisms into the existing Random Forest-based DST could allow it to better capture sequential relationships and critical patterns in the data. Additionally, integrating spatial and temporal features, as demonstrated in the LSTM-MHA, may reduce prediction gaps and improve accuracy across various forecast horizons. These insights suggest that hybrid approaches, combining the strengths of ensemble methods like Random Forests with the sequential modelling capabilities of LSTM networks, could further optimize prediction performance and operational reliability.

D. Operational Use and Generalizability

From an operational standpoint, the LSTM-MHA forecaster can serve as both a strategic planning resource and a real-time decision aid. Updating predictions frequently (e.g., every five minutes) and starting as early as five hours before filed takeoff can improve long-horizon forecasts. While early-stage predictions offer valuable initial assessments of delay risks, their reliability strengthens considerably closer to departure, when real-time conditions stabilize. Thus, the model's outputs can guide various decision-making layers—from initial resource allocations to last-minute runway and gate adjustments.

Interpreting the model's forecasts as indicative rather than absolute can help air traffic managers and controllers integrate this information with other tools, expert judgment, and scenario analyses. Such a holistic approach encourages proactive regulation and mitigation strategies, minimizing cascading delays and congestion while optimizing resource usage.

Additionally, generalizing this tool to other airports depends only on data availability and local operational patterns. Airports with similar data availability and complexity should achieve comparable improvements, though site-specific configurations, traffic densities, and procedures will require model retraining and calibration.

The LSTM-MHA's complexity and computational demands may pose challenges in real-time or resource-constrained environments. High-end computing infrastructure is necessary to realize the model's full potential, which is feasible at large ATM organizations but may be restrictive elsewhere. Additionally, the model's reliance on historical data from a specific period means it may struggle to adapt to unforeseen events or structural changes in operations. Yearly retraining and adaptive modelling strategies are essential to ensure long-term relevance and resilience.

E. Limitations, Future Work, and Recommendations

This study operates within a constrained temporal window and is largely limited to data concerning Dutch airspace. Such a scope limits the model's ability to capture network-wide effects and long-term patterns fully. As highlighted in the results, the knock-on effect has significant importance on longer horizons. Broadening the dataset to encompass multiple regions, thus allowing for a more extensive knock-on effect, can significantly improve predictions. Additionally, exploring alternative architectures, including Transformers or hybrid configurations that combine LSTM and attention-based models, may simplify the modelling process while maintaining high predictive performance.

Furthermore, the model is trained on data from March 28th 2023 to August 31st 2023, implying that it only captures the dynamics during the summer period. Retraining the model for winter traffic is advised. Furthermore, significant disruptions, such as the closure of certain airspace regions or disruptive weather, could have a substantial impact on the model's effectiveness. Other unforeseen events might also disrupt the system, potentially rendering the model less useful or even disposable. This limitation highlights the importance of continuously updating the model to adapt to new conditions and ensure its long-term reliability.

Future work should focus on several key areas to further enhance the model's performance and applicability:

- **Enhance the Knock-on Effect Data:** Integrate comprehensive, real-time information on interconnected flights to improve long-horizon predictions.
- **Regular Model Updates:** Implement periodic retraining to keep the model aligned with seasonal variations and evolving operational conditions.
- **European-Wide Data Integration:** Broaden the geographic scope by incorporating A-CDM data from multiple European airspace regions, thereby capturing a wider range of disruptions and dependencies.
- **Optimize Data and Feature Engineering:** Adjust data refresh intervals and refine feature sets to include all critical updates, fostering more nuanced and reliable forecasts.

By pursuing these recommendations, advanced forecasting models like the LSTM-MHA can achieve even greater accuracy and operational utility, ultimately contributing to more efficient, reliable, and sustainable ATM.

VII. Conclusion

Reducing uncertainty in flow management for arriving traffic at Schiphol Airport is achieved by enhancing the prediction of Actual Take-Off Time (ATOT) from out-stations. The development of a Long Short-Term Memory with Multi-Head attention (LSTM-MHA) neural network model successfully captures the intricate temporal dependencies and operational complexities associated with flight departure delays. Demonstrating superior performance, the LSTM-MHA model outperforms traditional ensemble methods and the existing Decision Support Tool (DST), attaining a Mean Absolute Error (MAE) of 7.57 minutes at Schiphol Airport. Despite the increased computational complexity introduced by the LSTM-MHA model, the resulting enhancements in operational efficiency and safety are considerable. This underscores the potential of integrating advanced machine learning techniques into existing systems, leading to significant advancements in air traffic management.

Incorporating factors such as the knock-on effect into predictive models plays a pivotal role in enhancing accuracy. Improved ATOT predictions facilitate more precise demand forecasting, enabling Air Navigation Service Providers (ANSPs) to optimize demand-capacity balancing, allocate resources more efficiently, and proactively manage potential traffic overloads. This proactive approach not only mitigates delays but also contributes to operational efficiency and safety, underscoring the value of advanced predictive models in air traffic management.

Replacing the Random Forest of the DST with the LSTM-MHA model has the potential to significantly enhance air traffic flow management at Schiphol Airport, reducing delays and improving overall efficiency. This advancement not only benefits Schiphol Airport but is also readily transferable to other airports, establishing a benchmark for similar applications in other airports and regions, and promoting more reliable and efficient air travel operations. By leveraging neural network architectures, air traffic management systems can achieve higher levels of precision and responsiveness, ultimately contributing to a more sustainable and effective aviation ecosystem.

Future work should focus on expanding datasets to encompass a broader range of flight information and integrating real-time data streams to enhance model adaptability. Additionally, refining feature engineering to better represent critical factors like the knock-on effect and optimizing data update frequencies will likely enhance predictive accuracy and model robustness.

References

[1] EUROCONTROL, "Performance review report: An assessment of air traffic management in europe," May 2024. [Online]. Available: http://ansperformance.eu

[2] R. Vos, "Machine learning based trajectory prediction to support demand forecasting," Delft, Netherlands, October 2023, master's thesis.

[3] A. de Leege, M. M. Van Paassen, and M. Mulder, "A machine learning approach to trajectory prediction," in *A Machine Learning Approach to Trajectory Prediction*, 08 2013.

[4] B. Ye, B. Liu, Y. Tian, and L. Wan, "A methodology for predicting aggregate flight departure delays in airports based on supervised learning," *Sustainability*, vol. 12, no. 7, p. 2749, 2020. [Online]. Available: https://www.mdpi.com/2071-1050/12/7/2749

[5] K. Wu and W. Pan, "A 4-d trajectory prediction model based on radar data," in *Proceedings of the 27th Chinese Control Conference*. Kunming, Yunnan, China: IEEE, July 2008.

[6] F. Pérez Moreno, V. F. Gómez Comendador, R. Delgado-Aguilera Jurado, M. Zamarreño Suárez, D. Janisch, and R. M. Arnaldo Valdés, "Methodology of air traffic flow clustering and 3-d prediction of air traffic density in atc sectors based on machine learning models," *Expert Systems with Applications*, vol. 223, p. 119897, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0957417423003986

[7] SESAR Joint Undertaking, *Digitalising Europe's Aviation Infrastructure: MASTER PLAN Executive View 2020 Edition*. Luxembourg: Publications Office of the European Union, 2020. [Online]. Available: https://www.sesarju.eu/

[8] B. Sridhar, N. Y. Chen, and H. K. Ng, "An aggregate sector flow model for air traffic demand forecasting," in *9th AIAA*

Aviation Technology, Integration, and Operations Conference (ATIO). Hilton Head, South Carolina: American Institute of Aeronautics and Astronautics, 21–23 September 2009, pp. AIAA 2009–7129.
[9] D. Delahaye, P. Paimblanc, S. Puechmorel, J. Histon, and R. Hansman, "A new air traffic complexity metric based on dynamical system modelization," *Digital Avionics Systems Conference*, vol. 1, pp. 4A2–4A2, 2002.
[10] B. Bubalo and J. Daduna, "Airport capacity and demand calculations by simulation—the case of Berlin-Brandenburg International Airport," *NETNOMICS Economic Research and Electronic Networking*, vol. 12, pp. 161–181, Oct. 2011.
[11] F. Könnemann, "Performance impact of improved departure time prediction relative to sector demand & arrival time predictability," Master's thesis, Delft University of Technology, Delft, Netherlands, 12 2014.
[12] Q. Li, X. Guan, and J. Liu, "A cnn-lstm framework for flight delay prediction," *Journal of Intelligent Transportation Systems*, 2024.
[13] R. Shone, K. Glazebrook, and K. G. Zografos, "Applications of stochastic modeling in air traffic management: Methods, challenges and opportunities for solving air traffic problems under uncertainty," *European Journal of Operational Research*, vol. 292, pp. 1–26, 2021.
[14] P. Brooker, "Simple models for airport delays during transition to a trajectory-based air traffic system," *The Journal of Navigation*, vol. 62, no. 4, pp. 555–570, 2009. [Online]. Available: https://doi.org/10.1017/S0373463309990105
[15] M. Güvercin, N. Ferhatosmanoglu, and B. Gedik, "Forecasting flight delays using clustered models based on airport networks," *IEEE Transactions on Intelligent Transportation Systems*, vol. 22, no. 5, pp. 3179–3189, May 2021.
[16] Z. Guo, B. Y. M. Hao, W. Wang, Y. Jiang, and F. Zong, "A novel hybrid method for flight departure delay prediction using random forest regression and maximal information coefficient," *Aerospace Science and Technology*, vol. 116, p. 106822, 2021.
[17] f. Z. Guo, G. Mei, S. Liu, L. Pan, L. Bian, H. Tang, and W. Diansheng, "Sgdan—a spatio-temporal graph dual-attention neural network for quantified flight delay prediction," *Sensors*, vol. 20, no. 6433, 2020.
[18] J. Qu, T. Zhao, M. Ye, J. Li, and C. Liu, "Flight delay prediction using deep convolutional neural network based on fusion of meteorological data," *Neural Processing Letters*, vol. 52, pp. 1461–1484, 2020.
[19] J. Sun, T. Dijkstra, K. Aristodemou, V. Buzeţelu, T. Falat, T. Hogenelst, N. Prins, and B. Slijper, "Designing recurrent and graph neural networks to predict airport and air traffic network delays," in *10th International Conference for Research in Air Transportation*. FAA & Eurocontrol, 2022.
[20] C. Wang, F. Zhu, Y. Chen, and J. Chen, "A LSTM-Based Time Series Model for Delay Prediction in Air Traffic Flow Management," *Sensors*, vol. 22, no. 21, p. 8106, 2022. [Online]. Available: https://doi.org/10.3390/s22218106
[21] A. T. . S. K. Ramon Dalmau, "Peta: Combining machine learning models to improve estimated time of arrival predictions," SESAR INNOVATION Days, accessed: 2024-12-09.
[22] Royal Schiphol Group, "Movements — traffic review 2023," 2023, accessed: 2024-12-24. [Online]. Available: https://www.schipholannualtrafficreview.nl/2023/movements
[23] F. Dijkstra and G. Calis, "System Subsystem Specification (SSS) Decision Support Tool," LVNL INTERN, DRAFT, 4 2023, document number: DST-SSS, Classification: LVNL INTERN.
[24] Ogimet, "Ogimet: Meteorological data and weather information," 2024, accessed: 2024-04-20. [Online]. Available: https://www.ogimet.com/credits.phtml
[25] A. Graves, N. Jaitly, and A. rahman Mohamed, "Hybrid speech recognition with deep bidirectional lstm," in *Proceedings of the 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE. Vancouver, Canada: IEEE, 2013, pp. 273–277.
[26] EUROCONTROL, "Summer 2024 overview of network performance," 2024, accessed: 2024-06-18. [Online]. Available: https://www.eurocontrol.int/press-release/summer-2024-overview-network-performance

Appendix

TABLE VI: Description of EFD Message Fields

Field	Description
Timestamp	The precise time when the EFD message was sent or updated, crucial for tracking the progression of flight information over time.
Flightstate	The current flight status, such as airborne, on the ground, or in a holding pattern, determining the flight's operational status at any given time.
Flightplan ID	A unique identifier for the flight plan, distinguishing each flight and ensuring updates are correctly matched.
Aircraft ID	The unique identifier for the aircraft, typically a registration number or flight number, used to track progress and link other data.
Aircraft Type	The type of aircraft, which determines minimum turn-around time and operational category.
EOBT	Estimated Off-Block Time: The time the aircraft is expected to leave the gate or start taxiing, crucial for predicting takeoff delays.
TSAT	Target Start-Up Approval Time: The target time for receiving clearance to start engines.
TOBT	Target Off-Block Time: The target time for the aircraft to be ready to leave the stand.
CTOT	Calculated Take-Off Time: The computed time for the aircraft to take off, synchronized with air traffic flow.
ETA	Estimated Time of Arrival: The projected arrival time at the destination airport, is essential for traffic planning and ground service coordination.
ADEP	Aerodrome of Departure: The airport from which the aircraft is departing, used to track flight origin and relevant control zones.
ADES	Aerodrome of Destination: The destination airport, helping manage traffic flows, arrival slots, and coordination at the destination airport.
Flightplan	Detailed route information for the flight, including waypoints, airways, and altitude assignments, ensuring smooth coordination.
Message Type	The category of the message (ACT (Actual), CAL (Calculation), or EST (Estimate)), distinguishing the reliability of the data.

TABLE VII: Top Feature Importances per minutes before ATOT (Part 1)

	240	120	0
Flight Plan Delay	26.4	55.03	72.51
Knock-on Delay	36.98	4.53	3.28
Flightstate SI	3.99	9.9	0.18
TOBTdelay	9.08	7.45	1.61
TSAT Delay	4.77	11.46	6.27
Modeltyp ACT	0.03	0.45	4.59
fltstate other	3.58	0.02	0.36
modeltyp CAL	0.61	0.02	0.57
eflighttime	0.45	0.83	0.42
ADEP LPPT	0.56	0.4	0.84
ADEP EGLL	0.29	0.19	0.24
actype A20N	0.19	0.13	0.2
cos ETOT	0.05	0.15	0.08
cos ETA	0.06	0.02	0.0
atfmdelay	0.32	0.18	0.01
actype E195	0.27	0.24	0.03
cobt delay	0.1	0.02	1.0
ADEP EGGW	0.13	0.28	0.04
ADEPLat	0.03	0.27	0.32
ADEP LTFM	0.44	0.21	0.02
actype E295	0.04	0.14	0.04
actype BCS1	0.24	0.08	0.03
ADEP LFMN	0.03	0.06	0.14
actype A321	0.41	0.01	0.1
ADEP LSGG	0.05	0.01	0.05
ADEP EIDW	0.03	0.03	0.18
ADEP LEAL	0.04	0.06	0.3
ADEP LFPO	0.27	0.02	0.12
actype BCS3	0.09	0.03	0.03
ADEP LEBL	0.04	0.03	0.08
ADEP ESSA	0.05	0.2	0.05
ADEP EGSS	0.09	0.01	0.15
modeltyp EST	0.12	0.18	0.05
ADEP EDDH	0.01	0.3	0.06
distance	0.03	0.1	0.13
ADEP EBBR	0.1	0.2	0.04
ADEP LOWW	0.01	0.02	0.03
ADEP EGCC	0.15	0.02	0.03
ADEP LEMD	0.17	0.01	0.12
sin ETA	0.06	0.04	0.16
ADEP LEPA	0.0	0.06	0.23
actype E75L	0.01	0.12	0.04
ADEP EGBB	0.08	0.06	0.04
ADEP EDDS	0.06	0.03	0.05
actype CRJ9	0.05	0.09	0.02
ADEP LGAV	0.04	0.02	0.09
actype E290	0.32	0.06	0.05
ADEP EFHK	0.09	0.07	0.16
ADEP EKCH	0.13	0.1	0.03
ADEP LEMG	0.1	0.01	0.21
sin ETOT	0.11	0.03	0.08
actype B38M	0.01	0.11	0.11
flighttype S	0.14	0.07	0.07
ADEP LKPR	0.05	0.01	0.02
actype A333	0.07	0.03	0.14
actype E190	0.14	0.09	0.04
actype F2TH	0.23	0.04	0.01
actype A319	0.24	0.0	0.12
actype B788	0.11	0.0	0.01
ADEP LIMC	0.1	0.07	0.22
actype PC12	0.02	0.15	0.04

TABLE VIII: Top Feature Importances per minutes before ATOT (Part 2)

	240	120	0
ADEP EDDM	0.02	0.09	0.06
ADEP EDDL	0.05	0.08	0.04
ADEP EDDM	0.02	0.09	0.06
actype B77W	0.15	0.02	0.02
actype B748	0.06	0.06	0.0
actype GLF4	0.02	0.16	0.0
ADEP ENGM	0.21	0.04	0.01
actype A332	0.07	0.02	0.01
actype C68A	0.06	0.0	0.0
day of week 0	0.04	0.14	0.0
flighttype G	0.03	0.08	0.02
actype C25B	0.02	0.01	0.02
actype CRJX	0.05	0.0	0.01
actype LJ40	0.02	0.04	0.01
actype GA5C	0.0	0.0	0.0
actype C25C	0.01	0.09	0.01
actype CL30	0.02	0.11	0.0
actype BE40	0.06	0.03	0.0
actype C525	0.01	0.15	0.05
actype E550	0.04	0.04	0.02
actype A359	0.05	0.0	0.04
ADEP LIME	0.04	0.01	0.09
actype B39M	0.01	0.01	0.02
flighttype M	0.0	0.01	0.02
actype GLF5	0.04	0.01	0.0
actype GL5T	0.0	0.0	0.01
actype E55P	0.03	0.05	0.03
ADEP LSZH	0.01	0.17	0.01
ADEPLong	0.02	0.13	0.15
actype FA7X	0.02	0.0	0.06
ADEP capacity	0.03	0.03	0.03
actype LJ60	0.08	0.04	0.01
actype C560	0.01	0.01	0.0
actype E75S	0.04	0.02	0.0
actype B744	0.02	0.02	0.0
actype B752	0.03	0.04	0.03
actype LJ45	0.03	0.02	0.01
actype GLF6	0.01	0.0	0.01
actype CL60	0.23	0.02	0.03
actype HDJT	0.0	0.01	0.0
actype E145	0.0	0.0	0.01
actype C680	0.01	0.01	0.01
actype B350	0.02	0.0	0.0
actype FA8X	0.01	0.01	0.01
actype G280	0.01	0.04	0.0
actype G150	0.0	0.0	0.0
actype B772	0.0	0.0	0.0
actype DH8D	0.0	0.0	0.0
actype E135	0.0	0.0	0.0
cap DEP	0.0	0.0	0.0
actype H25C	0.0	0.0	0.0
actype H25B	0.0	0.0	0.0
actype BE20	0.0	0.0	0.0
actype B763	0.0	0.0	0.0
actype GALX	0.0	0.0	0.0
actype ASTR	0.0	0.0	0.0
actype E545	0.0	0.0	0.0
actype A343	0.0	0.0	0.0
actype C25M	0.0	0.0	0.0
actype AC80	0.0	0.0	0.0
actype EA50	0.0	0.0	0.0
actype F100	0.0	0.0	0.0

TABLE IX: Top Feature Importances per minutes before ATOT (Part 3)

	240	120	0
actype D328	0.0	0.0	0.0
actype LJ35	0.0	0.0	0.0
actype AT75	0.0	0.0	0.0
actype LJ35	0.0	0.0	0.0
ADEP ULLI	0.0	0.0	0.0
cap DES	0.0	0.0	0.0
actype E50P	0.0	0.01	0.01
actype GA6C	0.01	0.0	0.02
actype PRM1	0.01	0.01	0.01
actype BE9L	0.05	0.0	0.0
day of week 2	0.1	0.06	0.08
ADEP LTBA	0.01	0.0	0.02
actype B789	0.04	0.0	0.06
actype CL35	0.01	0.04	0.05
actype GLEX	0.01	0.02	0.03
actype C650	0.0	0.0	0.0
ADEP LPPR	0.0	0.0	0.05
actype B737	0.04	0.08	0.05
actype GL7T	0.01	0.01	0.0
actype B735	0.02	0.02	0.02
actype E170	0.01	0.02	0.01
actype PC24	0.01	0.08	0.02
actype C56X	0.05	0.02	0.04
ADEP EGPH	0.02	0.05	0.05
actype CRJ7	0.01	0.01	0.0
actype CRJ2	0.05	0.0	0.02
ADEP LFLL	0.15	0.06	0.04
actype A318	0.1	0.0	0.08
ADEP LHBP	0.12	0.05	0.0
actype E35L	0.0	0.0	0.04
actype A306	0.01	0.12	0.01
actype B739	0.02	0.07	0.1
ADEP EDDK	0.05	0.0	0.01
ADEP EGKK	0.07	0.03	0.16
flighttype X	0.07	0.12	0.01
actype B733	0.01	0.05	0.02
wdirec	0.24	0.1	0.04
flighttype N	0.05	0.06	0.02
ADEP EDDF	0.05	0.36	0.05
ADEP LIRF	0.13	0.09	0.08
actype A21N	0.18	0.1	0.02
actype C510	0.01	0.12	0.07
actype C295	0.03	0.03	0.0
actype B738	0.08	0.13	0.15
actype C55B	0.05	0.02	0.01
ADEP EPWA	0.07	0.09	0.12
actype B77L	0.14	0.08	0.0
ADEP LROP	0.12	0.01	0.03
day of week 4	0.17	0.04	0.0
actype F900	0.09	0.06	0.06
wguts	0.2	0.35	0.12
actype C25A	0.05	0.1	0.02
day of week 3	0.23	0.38	0.07
ADEP LFPG	0.18	0.05	0.04
visibility	0.13	0.05	0.11
fltstate FI	0.16	0.12	0.03
day of week 5	0.22	0.02	0.01
day of week 1	0.24	0.06	0.04
actype A320	0.26	0.01	0.11
day of week 6	0.27	0.05	0.03
wspeed	0.6	0.16	0.2

Chapter 1

Introduction

From the adventures of early pioneers to the web of global connectivity we see today, aviation has grown from a new innovation to a crucial, and indispensable pillar of modern society. A pillar that is expected to grow even further (SESAR Joint Undertaking [43]). This journey will not only continue to shrink the world but also create new challenges, among them the need to evolve Air Traffic Management (ATM) into a system that prioritizes sustainability while maintaining the highest standards of safety and efficiency that characterize global air travel. This thesis tries to contribute to a more sustainable ATM system, by improving on the standards found in operations of today.

At the core of modern ATM is the process of Demand Capacity Balancing (DCB), ensuring that the sky's capacity does not become a bottleneck for the flights passing through. In the Netherlands, Luchtverkeersleiding Nederland (LVNL) has developed a decision support tool designed to forecast traffic and manage the Dutch airspace up to five hours before aircraft arrival. This proactive approach reflects a broader commitment across the industry to enhance the predictive accuracy and operational efficiency of air traffic control systems.

Despite the advances, the existing systems are not without their limitations. Misalignments between predicted traffic flows and actual volumes can lead to inefficient and sometimes unnecessary traffic regulations, such as issuing flight delays both airborne and before departure. These inefficiencies not only disrupt airline operations but also contribute to increased fuel consumption and carbon emissions, underlining the urgent need for more advanced forecasting techniques.

This research positions itself at the intersection of these ongoing challenges and the emerging potentials of Machine Learning (ML) in trajectory prediction. By leveraging the capabilities of ML, this research aims to enhance the accuracy of demand forecasts, by focussing on the uncertainty during the pre-departure phase, thereby optimizing the flow within controlled airspaces and reducing unnecessary emissions. This aligns with broader environmental goals and supports the sustainability targets increasingly adopted by the aviation sector.

In this study, Chapter 2 delves deeply into defining the problem and stating baseline knowledge. Following this, Chapter 3 examines the application of demand prediction in current operations and investigates various alternative methodologies. Chapter 4 continues this exploration by delving into the intricacies that determine the departure time, setting the stage for incorporating machine learning techniques. Chapter 5 outlines the basic and more advanced algorithms that have been applied for departure time prediction, among other related metrics. This chapter also explores promising alternative methodologies that could offer new insights.

Building upon this theoretical groundwork, Chapter 6 presents the data available for this research, Chapter 7 then gives the current research gap, indicating what has been done, and where this research will focus on.

Chapter 2

Problem Statement

Ensuring the fluid and secure movement of air traffic relies heavily on maintaining sufficient distances between aircraft to prevent potential collisions. This responsibility lies primarily with the Air Traffic Service Provider (ANSP), charged with guiding aircraft and ensuring adequate separation. To preserve safe distances and optimal airspace usage, ANSPs establish operational thresholds that ensure safety along with optimal traffic flow. These thresholds, identified as airspace capacity, form the cornerstone of air traffic management.

Upon defining the parameters of airspace capacity, scheduled air traffic gains passage through designated airspace blocks, forming demand. Too much traffic within these blocks can lead to congestion, or traffic overload, thereby escalating the potential for conflicts and disruptions in air traffic flow. In response, ANSPs often employ measures such as implementing holding patterns to accommodate surplus traffic or keeping departing flights on the ground, enabling Air Traffic Controllers (ATCos) to navigate aircraft flow safely. However, holding patterns inevitably cause delays and heightened fuel consumption, endangering safety, escalating emissions, and impacting operational efficiency and economic viability.

In navigating traffic overload, ATCos encounter a spectrum of options, including activating additional runways, redistributing airspace sectors, or regulating traffic demand. However, while effective in select scenarios, regulatory measures may precipitate Air Traffic Flow Management (ATFM) delays, particularly at major hubs like Schiphol. At Schiphol, where timely connections are paramount owing to its hub status, regulatory measures may not fully resolve underlying issues, potentially causing traffic overload despite regulatory interventions.

While traditional demand forecasting methodologies rely on trajectory-based approaches or larger demand prediction systems facilitated by network management initiatives, recent academic research explores alternative methods such as sector flow models. These models give a broad view of the air traffic system, representing the collective interaction of flights across sectors to anticipate demand fluctuations and tackle congestion more effectively.

2.1 Decision support

To the challenges in ATFM, ANSPs adopt a process known as Demand Capacity Balancing (DCB), aimed at optimizing the allocation of airspace capacity to meet demand effectively. Traditionally the tool that gives insight in the traffic load is the Network Manager (NM) Ciflo tool. Recently, the Dutch ANSP, Luchtverkeersleiding Nederland (LVNL) launched a new Decision Support System (DST) intended to improve the DCB process by furnishing precise traffic demand forecasts and insights into available capacity resources. With this tool, ATCos gain decision-making support, allowing them to manage potential traffic overload scenarios proactively.

The DST, as outlined in the specifications provided by the LVNL [9], is designed to enhance the management and operation of air traffic, specifically at Schiphol Airport. The DST has the following requirements, aiming to streamline and improve various aspects of Air Traffic Control (ATC):

- **Enhanced Traffic Load Insight:**
  - The tool is required to provide a more precise and complete understanding of the air traffic load approaching Schiphol. This involves advanced analytics and real-time data processing to monitor and predict traffic flows effectively, ensuring optimal airspace utilization and safety.

- **Increased Runway and Airspace Capacity Insight:**
  - It must also offer detailed insights into the capacity and workload of runways and the surrounding airspace. This includes tracking current usage rates, potential bottlenecks, and performance metrics to manage the airspace more efficiently.
- **Capacity and Demand Management Tools:**
  - **ATFM Regulation Support:** The tool should support the implementation and management of ATFM regulations, helping to ensure that traffic flows remain within the capacity limits of the airspace and airport infrastructure.
  - **Balancing ATFM Regulations and In-FIR Delays:** It should manage the balance between adhering to ATFM regulations and minimizing delays within the Flight Information Region (FIR), optimizing both efficiency and compliance.
  - **Scenario Comparison:** The DST should allow for the composition and comparison of various operational scenarios, such as changes in weather conditions and runway configurations. This feature is crucial for strategic planning and decision-making, allowing operators to assess the impacts of potential changes and choose optimal strategies.
- **Overload Situation Alerts:**
  - Another essential requirement is the ability to signal potential overload situations in advance. This predictive feature enables proactive measures to mitigate risks associated with high traffic volumes, thereby enhancing safety and operational efficiency.

By meeting these specifications, the DST aims to not only predict incoming air traffic loads but also to provide a dynamic, responsive toolset for the Flow Management Position Controler (FMPC). This approach in managing both the current and forecasted situations ensures that Schiphol can maintain efficient, safe operations even under varying and challenging conditions.

The reliability of any tool is intrinsically linked to the accuracy and quality of its input data. For this reason, the DST is heavily reliant on the quality of the forecasting done on various parameters. Among these parameters, traffic load, or demand, stands out as a fundamental metric that needs accurate prediction. Demand, in the context of an airport like Schiphol, refers to the number of aircraft intending to arrive or depart within a specific timeframe. An airport, or airspace sector is constrained by capacity, the maximum number of aircraft that can safely occupy the airspace and landing strips.

2.2 Impact of accurate forecasting

Capacity itself is not a static figure; it can react in response to a variety of factors, including changing weather conditions, runway configuration and the availability of ATCos. Each of these factors can cause the demand to increase or decrease. When demand exceeds this flexible capacity threshold, FMPCs are faced with critical decisions: delay departures, hold arrivals at the gate, or implement stacking procedures for incoming flights. Each option carries its own set of implications. Delaying departures, for example, can lead to aircraft being held at the gate, whereas opting to delay arrivals by stacking them can result in prolonged flight times, and higher fuel consumption, having economic and environmental impacts.

Moreover, these operational decisions often have a cascading effect on the entire aviation ecosystem. impacting airlines economically through increased operational costs and negatively affecting passengers' schedules, including missed connections and extended layovers.

There are a great number of methods to calculate or estimate demand, one of which is Trajectory Based Operations (TBO), further explained in Chapter 3. Currently, the DST uses the arrival times as provided by the NM to make demand estimations, however, these have proven to be inaccurate. If the demand is overestimated, ATCos are likely to regulate traffic, causing flights to be unnecessarily regulated, and held at the gate or holding stack. An underestimation of the demand can result in unsafe situations, or last-minute regulations that lead to sub-optimal airspace usage and aircraft held at holding stacks.

The critical role of ATFM and the DST in managing air traffic flow underscores the urgent need to refine demand forecasting accuracy. Modern airspace management's complexity demands accurate, anticipatory insights into traffic and capacity fluctuations to ensure safety, efficiency, and minimal environmental impact.

The challenges in current demand estimation highlight the necessity for advancements in forecasting methods. Reliance on outdated data, such as NM-provided arrival times, can lead to inefficiencies and safety risks. The adoption of more sophisticated forecasting techniques is essential for improving demand prediction accuracy. Moreover, according to SESAR Joint Undertaking [44], the TBO approach represents the overarching direction for the future development of Air Traffic Management (ATM) systems. This advancement requires a collaborative effort from all ATM stakeholders, including ANSPs, airports, and airlines.

2.3 Role of A-CDM

Vos [50] stated that a large portion of the uncertainty in demand forecasting lies in the pre-departure phase, and therefore it was decided that this research will focus on determining the departure time, or Actual Take Off Time (ATOT). Moreover, the typical FMPC decision horizon currently is three to four hours. This horizon is the sweet spot between relatively reliable information and a sufficient portion of aircraft still on the ground. This enables FMPC to adequately issue restrictive measures whilst still having a manageable effect on the operations. Given the complexity of this problem, the following sections will form a solid base knowledge of the intricates that determine the ATOT.
The overseeing process of which this is a part is Airport Collaborative Decision Making (A-CDM), defined by EUROCONTROL [10] as: 'A-CDM is the concept which aims at improving ATFM at airports by reducing delays, improving the predictability of events and optimising the utilisation of resources'.
A-CDM is an innovative approach in ATM that focuses on improving the efficiency and capacity of airport operations. This methodology is grounded in the principles of collaboration and information sharing among the various stakeholders in the air traffic management ecosystem, which includes airlines, airport authorities, ground handling services, and ANSPs.
The primary objective of A-CDM is to enhance the operational efficiency of airports by fostering better decision-making processes. This is achieved through the real-time exchange of information and collaborative planning among all involved parties. By doing so, A-CDM aims to optimize resource utilization, reduce delays, and improve the predictability and reliability of airport operations.
One of the critical elements of A-CDM is its emphasis on the entire aircraft turnaround process and pre-departure sequencing. It includes various operational phases, from the moment an aircraft lands to its subsequent departure. The integration of these different processes under the A-CDM framework enables more accurate planning and efficient use of airport resources, including gates, baggage systems, and runways.
The implementation of A-CDM involves leveraging advanced technology for data sharing and communication. This includes the establishment of platforms for the real-time exchange of operational data, allowing stakeholders to access and update information regarding flight schedules, aircraft positioning, and airport conditions. Access to this shared data facilitates more informed and coordinated decision-making, leading to enhanced operational effectiveness.
Furthermore, A-CDM plays a significant role in promoting environmental sustainability. By enabling more efficient aircraft taxiing, reducing holding times, and optimizing flight paths, A-CDM contributes to a reduction in fuel consumption and greenhouse gas emissions.
In conclusion, A-CDM represents a significant advancement in ATM, uniting diverse stakeholders through a collaborative framework. Using the power of real-time data and collective efforts, A-CDM aims to streamline airport operations, minimize delays, and improve the overall efficiency and sustainability of air traffic management. Understanding this process is extremely important in predicting the ATOT, and therefore, the following sections will describe the A-CDM process(Eurocontrol [11]).

2.4 Core Times in A-CDM

Key times in A-CDM, as given in Figure 2.1 are crucial for managing the sequence of aircraft departures and enhancing overall airport efficiency. These times are interrelated and collectively contribute to the determination of the ATOT, which is the exact moment an aircraft leaves the ground. The primary times discussed here are the Calculated Take Off Time (CTOT), Estimated Take Off Time (ETOT), Target Take Off Time (TTOT), Estimated Off-Block Time (EOBT), Target Off-Block Time (TOBT), Target Start-Up Approval Time (TSAT), and Actual Off-Block Time (AOBT).

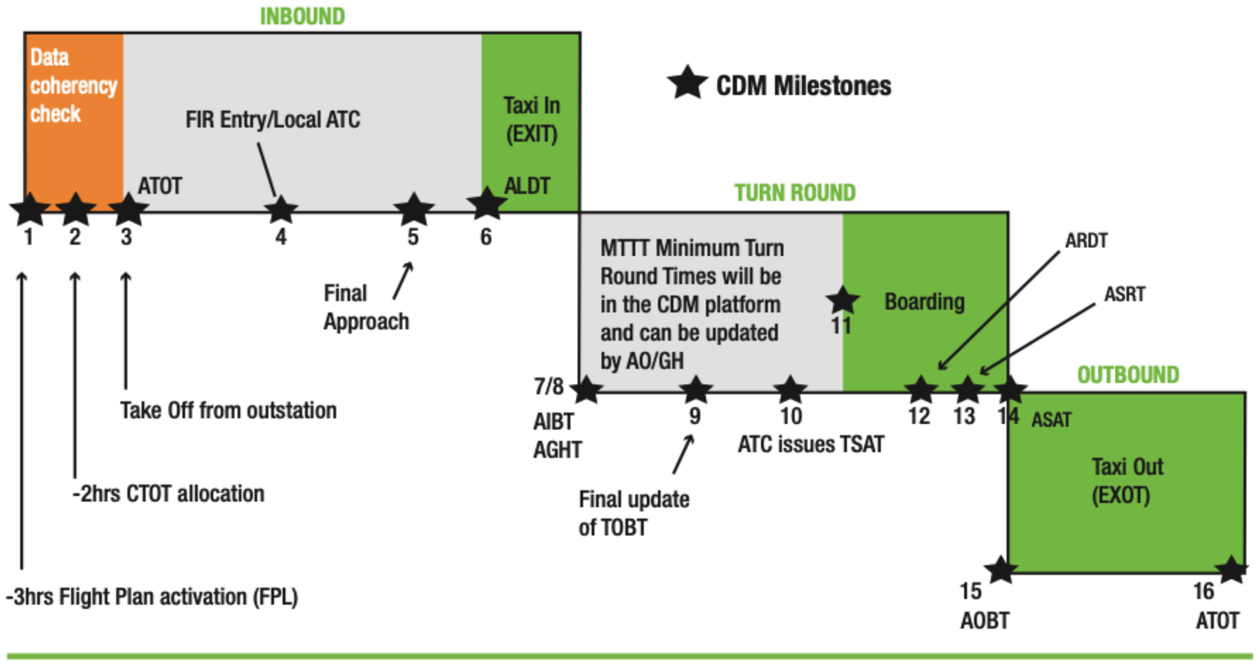


Figure 2.1: A-CDM milestones (Eurocontrol [11])

- **CTOT (Calculated Take-Off Time)**: CTOT is a time assigned to a flight specifying when it can expect to be cleared for takeoff. This time is calculated by ATC to manage the flow of aircraft both departing from airports and in the airspace. CTOT is crucial in minimizing delays and optimizing air traffic sequencing to enhance efficiency in airspace usage.
- **ETOT (Estimated Take-Off Time)**: ETOT refers to the estimated time at which an aircraft will take off. This estimate considers various factors, including CTOT adjustments, airport operations, and the aircraft's readiness. ETOT serves as a dynamic value that adjusts based on real-time data, providing a more accurate forecast of aircraft movement.
- **TTOT (Target Take-Off Time)**: TTOT is the target time established for a flight's takeoff, used by air traffic controllers and airport operators to plan and schedule the sequence of departures. TTOT is influenced by a range of operational factors and aims to synchronize with CTOT to ensure smooth flow of departures.
- **EOBT (Estimated Off-Block Time)**: EOBT is the time at which the aircraft is estimated to commence its movement from the parking stand for the purpose of takeoff. EOBT is a fundamental component in the pre-departure sequence, affecting the calculation of CTOT and TTOT, and is used to plan airport ground handling and air traffic control operations.
- **TOBT (Target Off-Block Time)**: TOBT is set by the aircraft operator or ground handler and indicates the target time at which the aircraft is expected to begin its movement from the parking spot towards the runway. It is a key reference time used by ATC to sequence departures efficiently.
- **TSAT (Target Start-Up Approval Time)**: TSAT is the time at which the aircraft is expected to start its engines and begin taxiing towards the runway. This time is closely coordinated with TOBT to ensure that the aircraft departs within its assigned CTOT window.
- **AOBT (Actual Off-Block Time)**: AOBT is the actual time when the aircraft moves off its parking stand. This time marks the beginning of the aircraft's journey from the gate to the runway and is critical for ATC to track and manage the precise flow of departures.

- **ATOT (Actual Take-Off Time)**: ATOT is the precise moment when the aircraft leaves the ground. Accurate predictions and adjustments of CTOT, ETOT, TTOT, EOBT, TOBT, TSAT, and AOBT are vital for effective A-CDM, ensuring that the sequence of departures is managed efficiently to minimize delays and optimize airport and airspace capacity.

These core times are integral to A-CDM, ensuring efficient airport operations, minimizing delays, and enhancing the predictability of events. By adjusting and predicting these times accurately, A-CDM helps optimize the utilization of resources at the airport and in the airspace.

2.5 ATOT in (A-CDM) airports

Having described the times that determine the ATOT, this section will give more insight into the A-CDM process. Oosterhof [32] conducted a study comparing the communication patterns between A-CDM, Advanced ATC TWR, and Standard Airports. The findings suggest that the type of airport does not significantly affect the quantity of updates communicated; rather, it influences the types and accuracy of those updates. This section will describe the A-CDM process, however, non-A-CDM airports will have similar behaviour.

Starting approximately three hours before the EOBT, when the flight plan is activated, various data points are checked for coherence, ensuring that the data from the airlines, airports, and ATC align. As the aircraft approaches its destination (inbound), key milestones such as Flight Information Region (FIR) Entry or Local ATC and Actual Landing Time (ALDT) are logged. Once on the ground, the aircraft's transition through taxi-in, turn-round, and taxi-out phases are carefully tracked.
The TOBT is updated, taking into account any Minimum Turn Round Times (MTTT) required for that particular flight. This is subject to updates by the Aircraft Operator (AO) or Ground Handler (GH) based on operational requirements. ATC then issues a TSAT, indicating when the aircraft can start its engines and begin taxiing for departure. Boarding times are monitored, and any discrepancies can lead to adjustments in the pre-departure sequence. Once all checks are complete, the ATOT is determined when the aircraft leaves the tarmac. This process ensures that the airport's runway capacity is used optimally, reducing delays and improving efficiency in the overall air traffic management system. EUROCONTROL [10]

Koolen and Suciu [26] give the following updates given as part of the A-CDM process):

1. **Predicted Departure Planning Information (P-DPI)**: This message is intended to provide initial airport data before the A-CDM process begins. It shares relevant information as soon as it becomes available, denoted as 'DPIEXPECTED' in the A-CDM field.

2. **Early Departure Planning Information (E-DPI)**: This message confirms that an airport slot and flight plan for a flight have been correlated at the airport. Its purpose is to prevent duplicate flight plans and eliminate ghost flights. In the A-CDM field, this state is marked as 'ESTIMATED'.

3. **Target Departure Planning Information - Target (T-DPI-t)**: The aim of this message is to provide a Take-Off Time (TOT) based on the Estimated Landing Time (ELDT) and the EXIT of the inbound flight. It also includes estimations for the turnaround, off-block, and taxi times of the outbound flight. This state is indicated as 'TARGETED' in the A-CDM field.

4. **Target Departure Planning Information - Sequenced (T-DPI-s)**: This message contains Take-Off Time information based on calculations from the Pre-Departure Sequencer, using the TOBT and the taxi time. It is shown in the A-CDM field as 'PRESEQUENCED'.

5. **Air Traffic Control Departure Planning Information (ATC-DPI)**: This message informs that the flight has 'off-blocked,' indicating that it is under ATC control and taxiing to take-off. It provides an estimate of the Take-Off Time with higher accuracy than earlier messages. This state is denoted as 'ACTUALOFFBLOCK' in the A-CDM field.

2.6 Partners in A-CDM and their responsibilities

Within the A-CDM framework, the following roles and responsibilities apply as given by IATA [23]:

- Aircraft Operator: The AO is responsible for providing the Flight Plan, and any subsequent updates, such as DLA/CHG messages, providing an accurate TOBT and ensuring that the flight crew is aware of the need to call for start-up at TSAT and TOBT.
- Ground Handling Agent: Providing an accurate TOBT with the Operations Control Centre (OCC), ensuring the flight crew is aware and ready for departure at the TOBT
- Airport Operator: Providing the airport schedule information and gate planning, as well as overall coordination and monitoring of the A-CDM process.
- Air Traffic Service Provider (Tower): Establishing, maintaining and executing the pre-departure sequence, providing TSAT and TTOT, ensuring start-up is issued according to TSAT, ensuring flights depart within their CTOT. Moreover, the Air Traffic Service Provider (ATS) should define and verify Variable Taxi-Times (VTT)s
- Air Traffic Flow Management Unit: ATFM should coordinate the network it is responsible for, disseminate flight plan data, coordinate DCB through the provision of CTOTs, providing the updated arrival information, and share relevant A-CDM data with Network Shareholders.
- De-Icing Operator: provide the De-Icing status of the aircraft, and predict the estimated de-icing times.

The integration and coordination of all partners involved in the A-CDM process are essential for achieving optimized airport operations and ensuring timely departures and arrivals. Each partner's role and responsibilities, as outlined, contribute uniquely to the A-CDM framework, highlighting the importance of collaboration and shared information in the complex ecosystem of airport operations.

The AO plays a pivotal role by providing the initial Flight Plan and any necessary updates, ensuring that the entire operational planning process is based on accurate and current data. The AO's responsibility to provide an accurate TOBT is critical, as it serves as the foundation for the pre-departure sequencing process. By ensuring that the flight crew is promptly informed about the need to call for start-up at the TSAT and TOBT, the AO facilitates a smooth integration into the departure sequence, optimizing the use of available resources and minimizing delays.

GH Agents are key players in preparing the aircraft for departure, directly impacting the accuracy of the TOBT communicated with the OCC. Their role ensures that all ground services, from baggage handling to refuelling, are completed in time for the aircraft to depart as scheduled. The coordination between GH Agents and the flight crew ensures that there are no delays in aircraft readiness, contributing to the overall efficiency of airport operations.

The Airport Operator, through its provision of airport schedule information and gate planning, ensures that the necessary infrastructure is available and allocated efficiently to accommodate the planned operations. Their role in coordinating and monitoring the A-CDM process is vital for managing the complex interactions between multiple stakeholders and ensuring that the airport operates smoothly and efficiently.

ATS Providers, including the TWR, are responsible for the real-time execution of the departure sequence, issuing TSAT and TTOT based on the current operational picture. Their ability to adjust the pre-departure sequence in response to operational changes ensures that flights can depart within their allocated CTOT, maintaining the overall flow of traffic and minimizing disruptions to the network.

The ATFM's role extends beyond the airport, coordinating the broader network to ensure that flights are integrated seamlessly into the overall air traffic system. By sharing flight plan data and coordinating DCB through the provision of CTOTs, the ATFM plays a critical role in managing air traffic flows and preventing congestion both on the ground and in the air.

Lastly, the De-Icing Operator's role is particularly crucial in adverse weather conditions, ensuring the aircraft's safety through de-icing operations. Their ability to provide accurate de-icing status and predict estimated de-icing times is essential for integrating these operations into the pre-departure sequence without causing unnecessary delays.

In summary, the A-CDM framework relies on effective collaboration and information sharing between all partners involved in the airport's operations. By clearly defining the roles and responsibilities of each partner, A-CDM ensures that all aspects of the departure and arrival processes are optimized, leading to improved operational efficiency and reduced delays.

2.7 Horizon

In any research that aims to make a prediction, the horizon—or the timeframe over which predictions are made—is a crucial element that significantly impacts the study's outcomes. The prediction horizon defines not only the scope and scale of the investigation but also influences the complexity of the models used, the reliability of the predictions, and the strategic value of the insights gained.

The length of the prediction horizon directly influences the quantity and type of data required. Longer horizons often necessitate historical data spanning several years to capture long-term trends and cyclic patterns, while shorter horizons may rely more on recent data that reflect current conditions. The quality of data, including its resolution, completeness, and relevance, also plays a critical role in the accuracy of predictions. Complications arising from longer horizons include the increased potential for disruptive events and changes in underlying patterns that historical data may not account for. For instance, technological innovations, policy changes, and unexpected economic shifts can all render long-term forecasts less accurate.

The problem described in this chapter asks for a horizon of 4 hours. This is sufficient for ATFM regulations in the European Region and is the horizon on which decisions are currently made by FMPCs. Before delving into the various reasons that can delay a flight, it is of great importance to list all locations a flight can be within the horizon of this research.

2.7.1 Location of the aircraft 4 hours before departure at the out-station

Understanding the location of an aircraft four hours before departure at the out-station- the station from which the aircraft will depart to Schiphol from- is crucial, despite seeming redundant. This knowledge significantly reduces operational uncertainties and determines the data available for managing departures effectively.

At the Outstation

If the aircraft is already at the out-station four hours prior to departure, it typically undergoes a series of critical pre-flight procedures- the turnaround process. These include maintenance checks, refuelling, cargo loading, and passenger boarding. Each activity must be precisely coordinated to ensure timely departure. Challenges such as adverse weather, technical issues, or airport congestion can complicate these operations, necessitating robust contingency strategies to maintain the schedule. Efficient management during this period is vital for completing the turn-around process promptly.

Airborne

Adding one complicating factor is when an aircraft en is en route to the out-station. This needs careful monitoring to ensure it lands in time for necessary turn-around activities. Delays due to air traffic, weather conditions, or operational constraints can significantly shorten the available time for ground operations, which are essential before the next departure. Collaboration between ATC and the airline's operational team is essential to optimize the route and manage time efficiently while upholding safety standards.

At Schiphol

Another complicating factor arises if the aircraft is still located at Schiphol. In this case, the turnaround process at Schiphol is also uncertain. Schiphol serves as a major hub in the hub and spoke system utilized by many airlines. This system routes passengers through a central airport hub, from which they are transported to their final destinations via the spokes—routes leading to other airports. Schiphol's role as a hub introduces complexities such as higher traffic volumes and potential logistical bottlenecks. Effective management is required to prevent these factors from causing delays in the aircraft's subsequent departure to the outstation. The hub and spoke model can create additional pressure on the turnaround process due to the synchronized arrival and departure times designed to maximize passenger connectivity.

Other Airport

When an aircraft is positioned at a third airport, it encounters challenges similar to those at Schiphol, including managing logistics and ensuring timely departure readiness. The degree of information transparency and sharing at different airports might vary, introducing further uncertainties in departure planning. Ensuring consistent communication and data-sharing practices is critical for minimizing these uncertainties and optimizing departure readiness.

These scenarios underscore the importance of precise location tracking and operational planning for aircraft scheduled for departure. By managing these elements effectively, airlines can enhance operational efficiency, reduce delays, and improve overall satisfaction.

Chapter 3

Demand

Understanding and predicting demand is a fundamental aspect of efficient Air Traffic Management (ATM). This chapter explores various methods used to predict traffic load or demand in air transportation, which is crucial for planning, safety, and optimizing airspace and airport operations. As air travel continues to evolve with increasing traffic volumes and technological advancements, the ability to accurately forecast demand becomes more critical in mitigating congestion and enhancing the operational efficiency of air traffic systems.
Demand prediction in air transportation is a complex field that combines data analysis, modelling, and forecasting techniques to estimate the number and flow of aircraft in different sectors of airspace at any given time. These predictions help air traffic controllers and airport authorities make informed decisions about resource allocation, flight scheduling, and airspace management, ultimately ensuring safety and minimizing delays.
This chapter will delve into the primary methods for predicting air traffic demand, from traditional trajectory-based models to more contemporary approaches like network flow and complexity analysis. Each method offers unique insights and comes with its own set of advantages and limitations, which will be discussed in detail. By examining these methods, this chapter aims to highlight their roles in supporting the dynamic and increasingly complex environment of global air traffic management.
The following sections will break down these methodologies, assess their impact on the management of air traffic, and explore how they are implemented in real-world scenarios to meet the challenges posed by the ever-growing demand for air transportation.

3.1 Demand Prediction Methods

Demand or traffic load predictions are crucial for efficient airspace management. Traditionally, these predictions are derived from aircraft trajectories, which consist of Four Dimensional (4D) positions—latitude, longitude, altitude, and time. By analyzing these positions, air traffic controllers can determine whether an aircraft will occupy a specific sector at a given time, and thus, calculate the total number of aircraft expected in that sector (de Leege et al. [7]). While this trajectory-based method is widely used, including by Air Traffic Controllers (ATCos) at Luchtverkeersleiding Nederland (LVNL) who utilize Network Manager (NM) data from flight plans, alternative approaches like network flow analysis and complexity analysis offer additional insights.

3.1.1 Trajectory-based Demand Prediction

The conventional method for predicting demand involves calculating the 4D trajectories of multiple aircraft, as detailed by Wu and Pan [53]. This approach consists of two main stages:

1. **Pre-Departure Prediction**: Using historical radar data, the model estimates total flying time and predicts future positions and altitudes at various intervals throughout the flight. This stage utilizes statistical regression models to account for variables like traffic flow and weather conditions, which may affect flight duration.

2. **Post-Departure Updates**: After departure, the model continuously updates its predictions based on real-time radar data, ensuring close alignment with the actual flight path. This enhances the reliability and utility of the forecast, adapting to any changes in flight conditions.

These predicted trajectories allow for estimating when an aircraft will enter a specific airspace sector, thereby facilitating demand forecasting. However, this method has its limitations. As Pérez Moreno et al. [36] points out, breaking the trajectory into small segments for precise predictions can be computationally expensive and data-intensive. Despite these challenges, data provided by the Knowledge & Development Centre (KDC) and the current operation run by LVNL supports the feasibility of this approach, building on research by Vos [50]. Additionally, the Single European Sky Aviation Research (SESAR) initiative underscores the importance and belief in the possibility of Trajectory Based Operations (TBO) in future ATM (SESAR Joint Undertaking [44]).

Limitations of Trajectory-Based Demand Prediction

There is a large variety of factors for uncertainty in measuring demand using TBO. Könnemann [25] analysed NM data to quantify the underlying uncertainty in demand predictions for the MUAC upper area control sectors in the Benelux and Germany. The study is based on the MUAC area but uses factors that will apply globally. Following from the study, departure time is the factor that has the greatest influence on sector demand prediction. This follows from the high correlation between departure and arrival time. A departure time is a good indicator of arrival time, although Performance Review Commission [33] shows that aircraft can speed up to compensate for departure time delays as visualised in Figure 3.1. Because of the high impact on demand prediction, this study will focus on departure time prediction, and thus the other factors will remain largely out of scope.

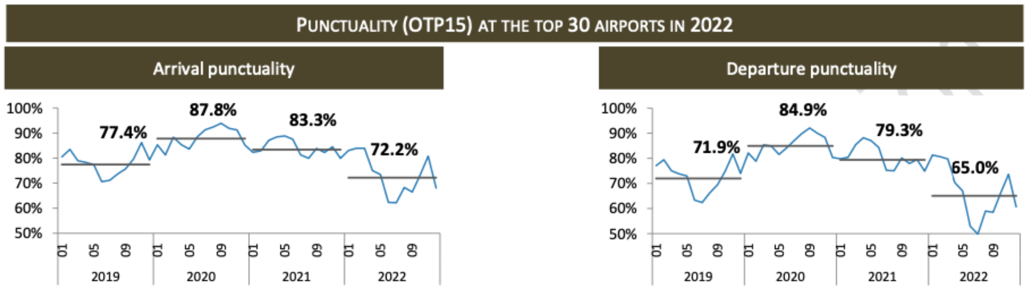


Figure 3.1: Difference between arrival and departure punctuality. (Performance Review Commission [33])

Figure 3.2 shows that the standard deviation in the occupancy count is still around 30%, implying there is quite some room for improvement. Having such high variance in the prediction on demand makes the decisions made by Flow Management Position Controler (FMPC)s harder, which can be the cause for unnecessary restrictions. Könnemann [25] argues that the large prediction error is due to the outflow of traffic, traffic that was supposed to enter the section, but had to re-route. Re-routing can be due to various reasons and is therefore not straightforwardly implemented in an aircraft performance model. Trajectory-based operations prediction should be able to capture outflow.

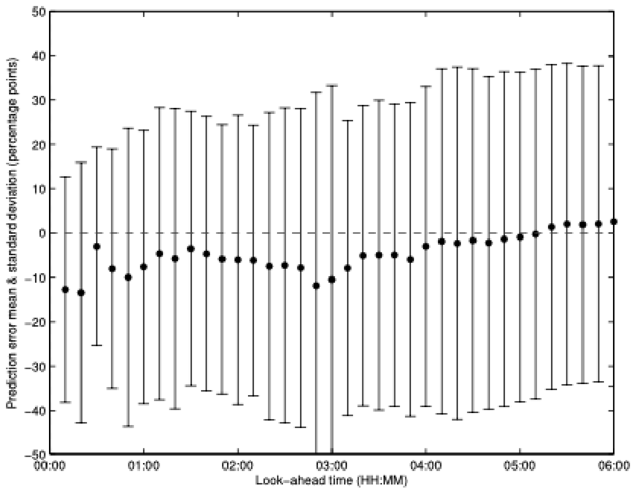


Figure 3.2: Percentage prediction error (Könnemann [25])

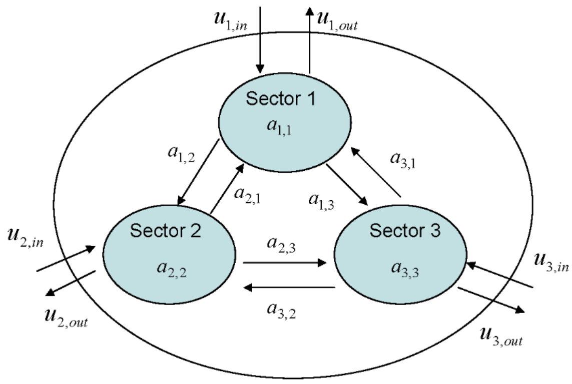


Figure 3.3: Visualisation of the sector flow model (Sridhar et al. [46])

3.1.2 Aggregate model

Traditional demand forecasting in ATM has heavily relied on the analysis of trajectories derived from flight plans. While short-term trajectory predictions for airborne flights exhibit high accuracy, the predictive reliability fades significantly over longer forecast horizons. This is due to a large variety of factors, including Air Traffic Control (ATC) interactions, weather conditions, and ground processes affecting aircraft before they become airborne. Although Trajectory Prediction (TP) and Demand Capacity Balancing (DCB) have seen improvements, a substantial body of research has pivoted towards a new method: aggregate demand forecasting.
One of the frontrunners of this new method was the study by Sridhar et al. [46], who conceptualized the U.S. National Airspace System through a model comprising 22 airspace blocks, alongside an international block. This model, disregarding individual flight trajectories, instead focuses on the flow of traffic through these blocks, predicting aircraft occupation counts as a measure of demand. Utilizing a linear dynamic system, the model uses a state space representation, where each block's state indicates its current aircraft count. A transition matrix, derived from observed traffic flow probabilities within the training dataset facilitates the prediction of traffic flow between sectors at every time step.
The effectiveness of this model was tested by training it with traffic data from two consecutive days, followed by demand prediction for a third day using only the flight schedule as input. Despite slight delays in demand curves and underprediction of peak traffic volumes, Sridhar et al. [46] argue that the model's predictive capacity remains robust enough to give a warning of potential demand overloads.
Building upon this foundation, Sridhar et al. [46] in a subsequent study sought to minimize prediction errors by employing a diverse set of models, each trained on different datasets to account for daily, weekly, and seasonal variations. The integration of a probabilistic hypothesis testing block further refined the model selection process, although the autonomous predictive capability of this system remains a future objective.

Limitations of aggregate model

One of the primary concerns with the aggregate model is its generalization over specificity. While the simplification facilitates easier demand prediction on a macro level, it risks overlooking the behaviours and characteristics of individual flights. This oversight can lead to potential inaccuracies in forecasting, particularly noticeable during irregular operations or within sectors that experience highly variable traffic patterns.
The model's predictive accuracy is also heavily influenced by the quality and completeness of the input data it relies on, such as flight schedules and historical traffic data. Any inaccuracies or gaps in this data can lead to significant deviations between predicted and actual demand, undermining the model's reliability.
Another limitation is the model's reduced effectiveness over extended forecasting horizons. While it shows promise for short- to medium-term predictions, its accuracy tends to wane for long-term forecasts. This reduction is largely due to the model's challenges in accommodating future alterations in flight schedules, air traffic control policies, and unexpected events that could significantly impact traffic flow.
The static nature of the aggregate model further constrains its capacity to adapt to the dynamic and often unpredictable changes that regularly occur in airspace utilization, weather conditions, and ATC interventions. These limitations hinder the model's ability to make real-time adjustments to its predictions.
Addressing these challenges necessitates ongoing research and development efforts aimed at refining the aggregate model's capabilities. The integration of real-time data feeds, advanced analytics, and machine learning

algorithms presents viable pathways to enhance the model's accuracy and responsiveness to the dynamic conditions of ATM. Furthermore, the development of hybrid models, which combine the strengths of both aggregate and trajectory-based approaches, could offer a more nuanced and comprehensive solution to demand prediction in ATM.

3.2 Factors Influencing Demand in Air Transportation

The demand for air transportation is influenced by a complex interplay of factors, which include but are not limited to the capacity constraints at airports, the scheduling strategies of airlines, and external conditions like weather. As highlighted by Barnhart et al. [3], the fundamental challenge facing air transportation is the balance between capacity and demand, particularly at congested airports. This balance is critical not only for maintaining operational efficiency but also for ensuring a high level of service reliability.
Several key factors influence demand in the air transportation sector:

- **Airport Capacity Constraints:** The capacity of an airport, especially during peak hours, significantly impacts demand. Airports with higher capacities can accommodate more flights, thereby potentially increasing demand.
- **Weather Conditions:** Weather is a significant factor affecting both demand and capacity. Poor weather conditions can lead to delays, cancellations, and rerouting of flights, directly impacting demand.
- **Economic Conditions:** The general economic environment influences the demand for air travel. Economic downturns typically lead to a decrease in demand, while economic upswings have the opposite effect.
- **Airline Scheduling Strategies:** Airlines' strategies regarding flight schedules, frequencies, and the choice of aircraft size also influence demand. Efficient scheduling can optimize capacity utilization and meet passenger demand effectively.

Standfuß and Whittome [47] studied the causes of demand fluctuations following the COVID-19 pandemic. Using a regression analysis, the effects of volatility on resilience are quantified, and the aspects that influence the demand are determined.

- **External Shocks**: Highlighting the unprecedented COVID-19 pandemic, financial crises, geopolitical strife, and natural calamities as key disruptors, the study illustrates how such events precipitate abrupt shifts in traffic volume, underscoring the fragility of air traffic demand to sudden global upheavals.
- **Capacity Disturbances, Climate Change, and Safety Issues**: These factors are highlighted as critical challenges to the resilience of ATM, demonstrating how they collectively contribute to the complexity of maintaining a stable air traffic system in the face of ever-evolving external pressures.
- **Inefficient Resource Planning due to Forecast Limitations**: The research highlights the challenge of accurate demand forecasting, a task made even more daunting by air traffic's dynamic nature, severely hindering Air Traffic Service Provider (ANSP)s ability to plan resources effectively.
- **Volatility in Traffic Movements**: The study highlights the growing unpredictability and volatility in traffic movements, caused by both local and global events, which disrupt regular traffic flows and make a recalibration of traffic management strategies necessary.

Standfuß and Whittome [47] analysis is pivotal for exploring the direct and indirect effects of the COVID-19 pandemic on air traffic and its broader implications on the resilience of the aviation industry. By quantitatively assessing the impact of volatility on resilience, the study provides a crucial metric for ANSPs to measure their preparedness and adaptability to future shocks.
Furthermore, the research explores the granularity of operational levels and their influence on demand, offering nuanced insights into volatility's varying impact across different ATM divisions. This granularity reveals the differential challenges faced by various components of the air traffic system, underscoring the need for tailored strategies to enhance resilience at all levels.

Chapter 4

Predictablility of the Actual Take-Off Time

The Actual Take Off Time (ATOT) is a critical factor in the efficiency and reliability of air traffic management systems worldwide. Accurately forecasting ATOT is essential for minimizing delays, optimizing airspace and airport capacity utilization, and improving overall passenger satisfaction. This chapter delves into the complex dynamics that influence ATOT, including the operational, environmental, and systemic factors that can affect the timeliness of flight departures.
Understanding and mitigating the causes of delays are central to enhancing ATOT predictability. This exploration begins with a detailed analysis of the various elements that can precipitate delays—from operational inefficiencies to adverse weather conditions and air traffic congestion. By examining how these factors correlate and impact departure times, stakeholders in the aviation industry can develop more robust strategies to manage and predict delays effectively.
Further, this chapter will explore the sophisticated methodologies and technologies employed to forecast and improve the predictability of ATOT. These include advanced modelling techniques, real-time data analytics, and collaborative efforts between airlines, airports, and air navigation service providers. Through comprehensive case studies and statistical analysis, the effectiveness of these methodologies in various operational contexts will be assessed.
As the global air traffic landscape becomes increasingly complex and crowded, the variety of factors that cause delays improves. The insights provided in this chapter aim to contribute to the ongoing efforts to enhance air traffic efficiency and ensure that the aviation industry can meet the challenges of an ever-evolving operational environment.

First the causes and factors for delays are listed, setting the stage for the rest of the Chapter. This is followed by a more thorough description of each factor. Weather impacts are described in Section 4.2, airport capacity in Section 4.3 and Knock-on delays in Section 4.4. This is followed by the result of all of these delays, and often seen as the cause: Air Traffic Flow Management (ATFM) delays. Finally, the last factor considered in this research is the turn around, in Section 4.6.

4.1 Causes and Factors Affecting Delays

Delays in air traffic management can be attributed to a variety of factors, often influenced by both temporal and spatial correlations. Figures 4.1 and 4.2 depict these correlations, highlighting how delays accumulate over time and are influenced by geographical proximity to major airports.
Figure 4.1 illustrates the daily distribution of delays, showing a strong temporal correlation with delays tending to accumulate throughout the day and decrease overnight. This pattern suggests a reset effect where delays are managed overnight when air traffic is generally lower.
Spatial dependencies are clear as airports near major hubs experience significant delays due to increased air traffic, as shown in Figure 4.2.
A comprehensive analysis by Wang et al. [51] at Beijing Capital International Airport, but widely applicable categorizes several key factors influencing delays:

- **Flight Attributes:** The differentiation in priority and resource allocation based on airline properties, such as the airline's base status at the airport. Additionally, the aircraft's capacity is underscored as a

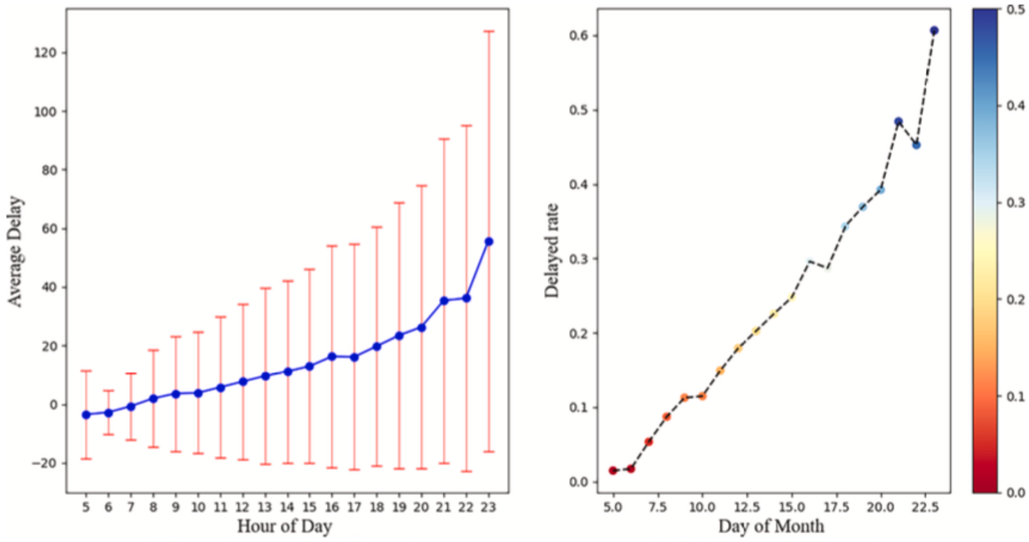
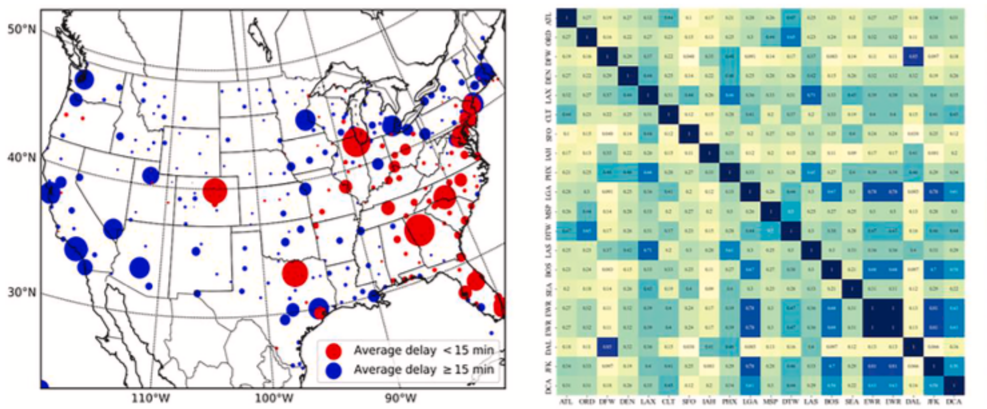


Figure 4.1: Accumulation of delays (Li et al. [29])



(a) Spatial distribution of average delays

(b) Pearson correlations between the top 20 airport delays

Figure 4.2: Spatial distribution of delays in the US (Li et al. [29])

pivotal factor, where larger aircraft are prioritized due to their higher passenger capacity, thus impacting delay patterns. Notably, the cascading effect of delays from previous flights also emerges as a crucial element, indicating a direct linkage between successive flights' punctuality.

- **Weather:** Weather conditions stand as a predominant factor affecting flight delays, highlighting the significant influence of atmospheric conditions at the departure airport on flight timeliness.
- **Periodic Data:** Time-related factors, including the hour of the day, day of the week, and seasonal variations, alongside holidays, are generally identified as influencing flight delay patterns, necessitating sophisticated predictive models to account for these temporal impacts.
- **Arrival/Departure Pressure:** The study delves into the concept of airport congestion, measured through arrival/departure pressure, as a determinant of flight delays, reflecting the operational load and its effects on flight schedules.
- **Cruise Pressure:** Additionally, the conditions during the cruise phase, such as air route congestion and adverse weather, are explored, highlighting the broader operational environment's impact on flight punctuality.

These factors highlight the complex nature of predicting and managing flight delays. The approach suggested by Wang et al. [51] uses ensemble methods to enhance the accuracy of delay predictions, ultimately aiming to improve operational efficiency and passenger satisfaction by reducing delays.

In an effort to quantify the levels of delays, Hanley [21] introduces a metric called NAS, distinguishing between "Low NAS" and "High NAS" states. "Low NAS" refers to periods of minimal delays commonly occurring during low air traffic periods like late at night. Conversely, "High NAS" indicates times of high congestion and is often short-lived but intense, such as the "High ATL" state affecting the Atlanta area. This factor shows that delays are related spatially, but also temporary.

4.2 Weather Impact at Airports

Dalmau et al. [6] discuss the impact of various adverse weather conditions on airport capacity, emphasizing how these conditions necessitate the implementation of ATFM regulations. The key weather phenomena affecting airport operational capacity include:

- **Low Visibility:** Conditions such as fog and heavy rain necessitate increased spacing between aircraft due to reduced visibility, thus decreasing runway throughput.
- **Wind:** High winds, particularly crosswinds, restrict the operational use of runways and necessitate traffic management adjustments. When crosswinds exceed certain thresholds, it becomes unsafe for aircraft to land and take off, effectively reducing the available runway capacity.
- **Thunderstorms and Convective Weather:** These events cause rapid changes in visibility and wind patterns and can introduce hazards like hail. Such conditions frequently lead to the temporary closure of airspace or the rerouting of flights, significantly impacting airport capacity.
- **Snow and Ice:** These conditions extend the time required for aircraft deicing and runway clearing, delaying all airport operations. Ice can also render taxiways and runways unusable, further reducing operational capacity.
- **Fog:** Similar to low visibility scenarios, fog substantially decreases the number of flights that can safely operate, significantly reducing airport capacity, particularly affecting landing and takeoff operations.

These insights are crucial for planning and implementing ATFM regulations to mitigate the adverse effects on airport throughput and ensure safety and efficiency in air traffic operations.
Furthermore, European airports operate within a complex network where each airport handles air traffic, demand calculations, weather impacts, and flow management differently. These operational differences often reflect the geographic locations, traffic volumes, and the sophistication of their Air Traffic Control (ATC) systems. Considering the substantial impact of weather on delays, this section explores how different airports manage the variability in weather impacts.
Rodríguez-Sanz et al. [39] categorize airports using five indicators: operational thresholds, the impact of weather uncertainties, the synthetic index for weather conditions, reaction to delay, and operational strategies. Operational thresholds define the critical points at which weather impacts severely affect airport performance, indicating resilience and vulnerability to various weather events. The impact of weather uncertainties measures an airport's ability to predict and react to uncertainties such as visibility and wind speed, which are critical for operational efficiency. The synthetic index for weather conditions integrates various meteorological parameters to evaluate comprehensively the weather impacts on airport performance.
Both Rodríguez-Sanz et al. [39] and Dalmau et al. [6] note that airports operating close to their maximum capacity are more severely affected by adverse weather, leading to significant delays and reduced throughput. The final indicator is the effectiveness of an airport's operational strategies in response to adverse weather, enabling better-performing airports to maintain high capacity under challenging conditions.

4.3 Capacity at airports

Determining airport capacity is a complex process that involves understanding the theoretical limits of an airport's infrastructure and its practical operational performance. According to He and Pan [22], the process for estimating airport capacity involves combining empirical methods with analytical approaches to capture the airport's actual operational limits as closely as possible. A capacity envelope was constructed to approximate the airport's maximum operational capabilities, reflecting both arrival and departure traffic demands.
This methodology involves observing peak arrival and departure counts over a given time period, under the assumption that these peaks represent the airport's maximum operating capacity. These peaks are used to construct a capacity envelope, which serves as a graphical representation of the airport's practical capacity

across various operational conditions. The study also introduces a collaborative optimization model that reflects the relationship between airport capacity and traffic demand, using an optimization algorithm to allocate flight schedules.
A key component of this method is the introduction of a priority ratio to adjust management preferences dynamically for arrival and departure traffic, revealing the synergy between traffic flow demand and airport capacity. The outcome of this study showed an improvement in flight on-time performance rates by 6% in their case study, demonstrating the efficacy of their proposed method in maximizing airport capacity and traffic flow demand without the need for physical expansion of airport facilities. This approach not only aids in efficiently utilizing existing airport capacities but also provides a model for balancing supply and demand in air traffic management, thereby reducing congestion and delays, enhancing customer satisfaction, and promoting sustainable growth in the aviation industry (He and Pan [22]).

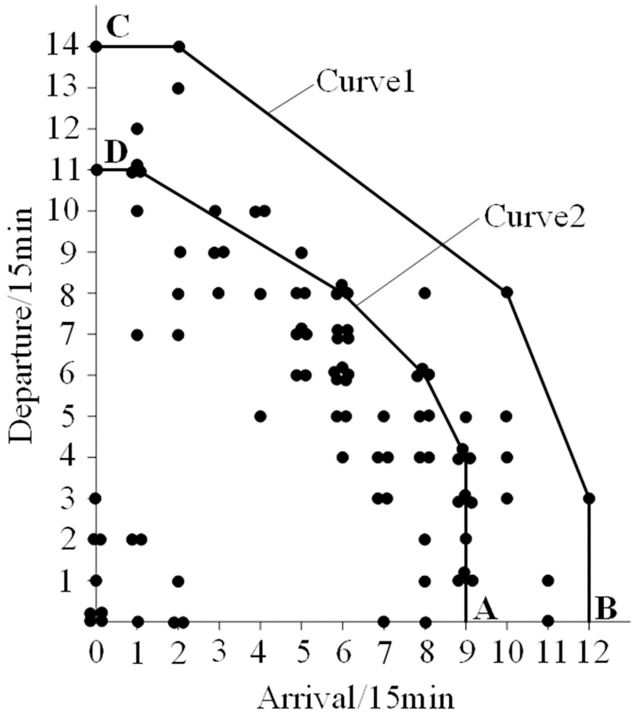


Figure 4.3: Schematic of the determination of airport capacity according to (He and Pan [22])

Rodríguez-Sanz and Rubio Andrada [40] define airport capacity as the ability of the airport infrastructure and operational procedures to process entities such as aircraft, passengers, luggage, vehicles, etc., within a specific interval of time. This capacity is inherently multi-faceted and dynamic, influenced by both available infrastructure and various operational procedures. It's typically measured in both annual and hourly metrics, depending on the planning needs. The article highlights the difficulties in accurately describing airport capacity due to its complexity and the influence of external factors and operational procedures. It points out the challenge of aligning long-term planning, which uses annual capacity figures, with the need for more immediate, hourly throughput measures due to daily and seasonal traffic variations. By examining hourly and annual air traffic volumes at 50 European airports from 2004-2021, the study provides insights into defining a suitable peak hour for capacity evaluation. It emphasizes the need for a detailed understanding of daily traffic distribution patterns and how they vary by the hour of the day. This analysis helps in planning new airport infrastructures by identifying when and where congestion is likely to occur.
Limited airport capacity significantly contributes to delays in air travel. Operating near or at capacity limits, airports face substantial delays from minor disruptions or increased traffic:

- **Bottlenecks:** Physical constraints such as limited runways, gates, or baggage handling facilities can create bottlenecks, slowing aircraft movements and causing delays.
- **Operational Inefficiencies:** High traffic volumes can overburden airport resources and personnel, leading to operational inefficiencies that prolong aircraft processing times.
- **Increased Turnaround Times:** Limited gate availability or delayed services can extend the time aircraft

spend on the ground, affecting their scheduled departure times and causing a ripple effect of delays across the network.

- **Traffic Management Challenges:** Air Traffic Control (ATC) may need to hold aircraft at their origin airport or in the air near the destination to manage traffic flows, directly impacting departure and arrival schedules. This is often referred to as Air Traffic Flow Management (ATFM) delays, which are further discussed in the subsequent section.

Effectively managing and optimizing airport capacity is crucial for minimizing these delays and improving the overall efficiency and reliability of air travel.

4.4 Knock-on Delay

According to EUROCONTROL [14] around 48% of all delays are 'Knock-on' delays, which means that the delay is caused by another aircraft delay. This indicates that flight delays do not often occur in isolation but can propagate through the network, affecting subsequent flights and even other airports. This phenomenon, known as knock-on delays, significantly complicates traffic management and scheduling within airports. Zhang et al. [54] explored the propagation effect of flight delays among airports and introduced a new measure called the propagation index to analyze the interrelationships among airports concerning flight delays. Guo et al. [18] further underscore the importance of knock-on delays and predict departure delays largely influenced by the knock-on effect. This indicates the importance of incorporating the network effects.
Delay propagation occurs when the arrival delay of a flight affects the departure time of the same aircraft for its next segment, potentially causing further delays down the line. While scheduled turnaround times may sometimes absorb these delays, the ripple effect can lead to missed connections for passengers and cascading delays throughout the day. Moreover, because flights of the same airline share ground crew teams and facilities such as runways and gates, the delay of one flight can impact several others, both within the same airline and for other carriers using the same facilities.
The propagation index quantifies the effect of delay propagation by measuring the causality among delay time series, which helps in understanding how delays in one airport can influence operations in another. This metric is crucial for airports to predict and manage delays more effectively, aiming to minimize the broader impact on the air traffic network.
According to Zhang et al. [54], this index has shown a high correlation with observed airport delays, underscoring its potential utility in enhancing the operational strategies of airports to mitigate the effects of delay propagation. The research highlights the need for comprehensive strategies that account for the interconnected nature of airport operations, facilitating better planning and coordination among various stakeholders involved in air traffic management.

4.5 ATFM Delay

Air Traffic Flow Management (ATFM) is an essential component of the global aviation infrastructure, tasked with ensuring the efficient, safe, and orderly movement of aircraft throughout the airspace. Due to increasing congestion and the need for more coordinated air traffic management, the International Civil Aviation Organization (ICAO) proposed a centralized ATFM system to Eurocontrol in 1988. This proposal led to the development of the Central Flow Management Unit (CFMU), designed to harmonize and optimize air traffic flows across Europe's complex airspace structures [35].
The CFMU was officially established in divisions of Western and Eastern Central Executive Units in 1989 and reached full operational capability by 1996. This centralization of pre-tactical and tactical functions within the CFMU significantly advanced European air traffic management, enhancing coordination among air navigation service providers (Air Traffic Service Provider (ANSP)s) and other stakeholders.
The CFMU incorporates several critical systems to effectively perform its functions:

- **Initial Flight Plan Processing System:** This system processes all flight plans to operate within controlled airspace, ensuring compliance with air traffic regulations and feasibility within current traffic and airspace constraints.

- **Tactical Capacity Planning Tool (Tactical Load Factor Calculation and Distribution System (TACT)):** The TACT is instrumental in compiling data for pre-tactical (two days prior to operation) and tactical (day of operation) planning. It considers demand, capacity, and ATFM regulations to generate optimized departure slots, minimizing delays and maximizing airspace utilization.

- **Environmental Database:** This database stores permanent data, including routes, geographical data, airports, and ATC centers. It supports the CFMU's decision-making processes by providing accurate and up-to-date information.

- **Archives System:** This system retains all operational data, enabling continuous improvement of ATFM operations through historical analysis and strategic planning.

These integrated systems ensure that ATFM effectively manages the complex dynamics of air traffic, playing a pivotal role in maintaining efficient and safe airspace operations throughout Europe [35, 27].
EUROCONTROL [13] lists the most common reasons for ATFM delay as seen in Figure 4.4. From this graph, it can be seen that capacity issues are the major factor leading to regulations.

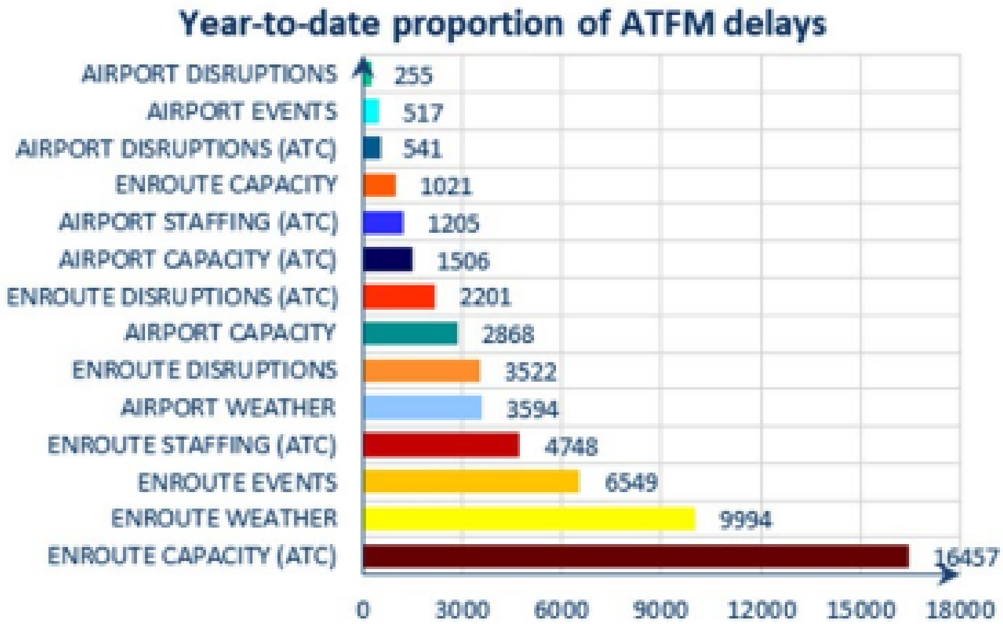


Figure 4.4: Reasons for ATFM delays within Europe in 2022. (EUROCONTROL [13])

An illustrative example of ATFM delay dynamics is depicted in Figure 4.5. The diagram highlights the temporal distribution of airport capacity versus demand based on flight plan data. A notable peak occurs around 5:40, when the demand surpasses the available capacity, primarily due to airlines implementing schedule buffers to accommodate potential delays or operational inefficiencies. This proactive scheduling often leads to a temporary overload of the system.
Conversely, beginning at 6:40, a noticeable decline in scheduled demand falls below the established capacity thresholds. This reduction in demand presents an opportunity for ATFM regulations to redistribute or reschedule the earlier congested flights, thus smoothing out the spikes in demand and optimizing the use of airport and airspace resources throughout the operational day.
Such ATFM interventions, while seemingly causing delays, are strategically implemented to enhance overall traffic flow and resource utilization. Although these adjustments result in apparent delays, they are crucial for maintaining system-wide efficiency and safety. Interestingly, EUROCONTROL [12] reported that in 2018, approximately 40% of ATFM delays ranging from 5 to 15 minutes were effectively absorbed by the airlines themselves. This absorption often involves slight modifications to flight operations that do not significantly impact the overall travel schedule, thus mitigating the perceived inconvenience to passengers.
These findings underscore the complex interplay between scheduled flight plans, actual operational demand, and ATFM regulations. By analyzing such data, stakeholders can better understand how strategic delay management can prevent systemic overloads and improve the punctuality and reliability of air travel [12].
Although ATFM regulations seem like a definite cause for delay, Delgado and Prats [8] focusses on managing ATFM delays through cruise speed reduction, attempting to enhance the efficiency of delay recovery without increasing fuel consumption. This is possible because ATFM delays often result in unnecessary ground holding if the weather conditions or other circumstances change over time. By reducing the cruise speed the delay can be split between ground and airborne delay. A simulation has shown a linear relationship between the cancellation time and the delay recovery, indicating a positive result.

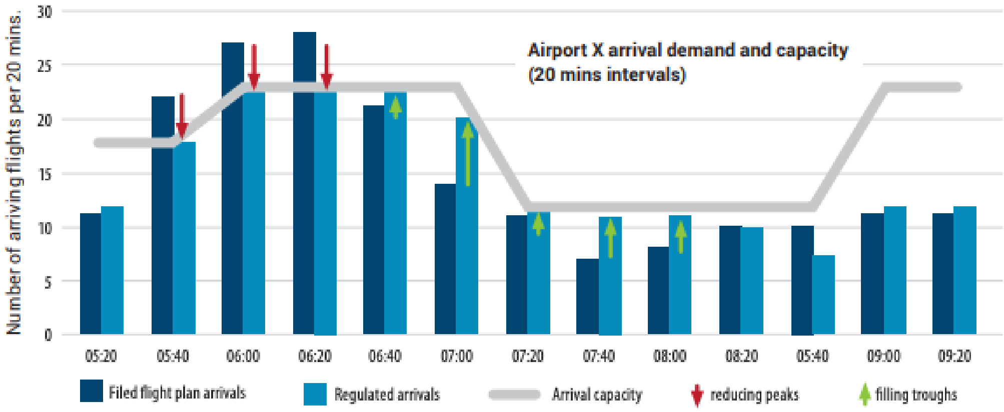


Figure 4.5: Illustration of Demand-Capacity balancing (EUROCONTROL [10])

Departure airport

According to EUROCONTROL [12], the Network Manager Operations Centre (NMOC) acts whenever too many aircraft in the air at the same time at the same place can lead to an unsafe situation. To prevent this from happening, a flight can be 'regulated', by issuing a Calculated Take Off Time (CTOT), or 'slot'. All aircraft try to depart as soon as the turnaround process is finished. However, in the case of a CTOT, the aircraft can only depart between 5 minutes before and 10 minutes after the CTOT. The aircraft is required to be ready for departure, located at the runway during this window. The reason for an ATFM slot can vary between ATFM regulation at the destination airport, in an en-route sector, or at the departure airport, but the result will always be the same.

Airborne

ATFM regulations rarely occur airborne, however, an increasing amount of research is being done on reducing the impact of ATFM regulation while airborne. Rosenow et al. [41] states that airlines increase the cruising speed to compensate for ground delay. However, increasing the cruising speed will increase fuel consumption and thus cost and climate impact. The research finds that increasing the cruising speed can be an effective strategy for reducing reactionary delay costs under certain conditions. The benefit of using this strategy varies with the operational cost scenario and the amount of delay.

Arrival airport

The purpose of ATFM regulations is to prevent regulations from happening during the final stages of the flight. So generally no ATFM regulations happen close to the arrival airport. However, the arrival airport is often the cause of the regulation. According to EUROCONTROL [12] the number of delayed departures exceeds the number of delayed arrivals, and this gap is increasing every year. In 2018, 48.4% of flights had a delay of ≥ 5 minutes, however only 42.8% had such a delay during arrival.

4.6 Turn around times at airports

In the event that there are no restricting measures from outside the aircraft, such as the before-mentioned causes for delay, the aircraft itself has to be ready for departure. Asadi et al. [2] state that the target time of an aircraft turnaround is of major importance for the tactical control of the Air Traffic Management (ATM) process. Moreover, the turnaround time is subject to many random factors, such as passenger behaviour, resource availability and short-notice maintenance. The paper proposes a mathematical optimization model for the turnaround time, taking into account the many uncertainties. The process, visualized in Figure 4.6, shows the 12 standard processes which are split up into up to 150 sub-tasks that involve up to 30 different actors, depending on the characteristics of the airport and airline.
The study by Fricke and Schultz [16] provides an in-depth analysis of the turnaround process, focusing on the identification of technical and procedural processes that hinder operational efficiency. These processes include

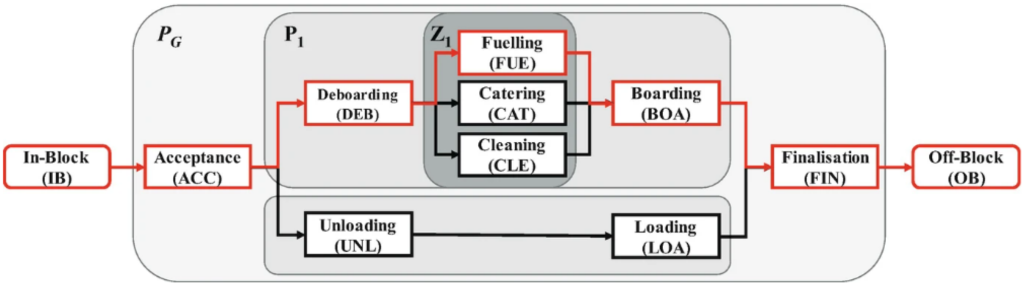


Figure 4.6: Schematic of a standard turnaround process sequence. (Asadi et al. [2]
)

passenger boarding and deboarding, fueling, cargo handling, cleaning, and servicing. One of the study's key findings is that the critical path, which includes boarding, deboarding, fueling, and cargo handling, often suffers from variability due to several technical deficiencies. These deficiencies are mainly associated with the aircraft's design and the interfaces used for ground handling operations.
Fricke and Schultz [16] employed a comprehensive field study across different airport types and aircraft models to collect data on turnaround operations. This data served as the basis for statistical modeling and Monte Carlo simulations, aimed at quantifying the variability in process times and identifying potential improvements. The analysis revealed that specific design features of the aircraft body and ground handling interfaces significantly contribute to the unpredictability and inefficiency of turnaround operations. By addressing these technical deficiencies, the study suggests that it is possible to achieve a more reliable and reduced turnaround time, aligning with Single European Sky Aviation Research (SESAR)'s performance targets for the 2020 Single European Sky initiative SESAR Joint Undertaking [43].
A novel solution to these challenges is the Deep Turnaround system, a state-of-the-art application of computer vision and AI technology. Deep Turnaround enables swift and low-maintenance optimization of turnaround processes through real-time data generation and analysis. This system captures continuous imagery at each stand, processed by an AI model to accurately track and predict turnaround events. (Amsterdam Airport Schiphol [1]) This innovative approach ensures:

1. **Rapid Implementation:** Utilizing a singular Artificial Intelligence (AI) model across airports allows for immediate performance at new stands and drastically shortens the adaptation period, ensuring quick value delivery.

2. **Dynamic Optimization:** Continuous performance assessment and data-centric AI adjustments lead to swift optimization, enhancing data accuracy and operational efficiency.

This technological advancement offers numerous benefits, such as improved on-time performance, optimized stand utilization, accurate Target Off-Block Time (TOBT) setting, and enhanced collaboration between airport and sector partners. Moreover, it contributes to sustainability by streamlining processes to reduce unnecessary engine running times, thereby lowering carbon emissions.

Chapter 5

Forecasting Methods

Accurate forecasting in Air Traffic Management (ATM) is crucial for enhancing operational efficiency and reducing delays, particularly as air travel continues to grow in volume and complexity. This chapter explores a variety of forecasting methods that address the challenges of predicting aircraft departure times from outstations, integrating mathematical models and Machine Learning (ML) techniques to predict and manage the dynamic nature of air traffic systems effectively.
Forecasting the departure times of aircraft involves large amounts of data and considering a variety of variables that can affect timing, from weather conditions and air traffic congestion to mechanical issues and operational inefficiencies. Effective forecasting methods can significantly mitigate delays, optimize the utilization of airspace and airport resources, and improve overall travel efficiency and passenger satisfaction.
The methodologies discussed not only aim to forecast delays more accurately but also seek to provide strategic insights that enable air traffic controllers and airport operators to make more informed decisions, potentially transforming the landscape of ATM through technology and innovation.
The chapter starts by working out the methods that used to be the cornerstones of demand forecasting, stochastic methods in Section 5.1, Poisson distributions in Section 5.2 and Time Series in Section 5.3. Subsequent sections focus on the state of the art methods that are machine learning in Section 5.4. A more detailed description of two techniques is given, first Random Forest in Section 5.5, followed by an extensive exploration of Machine learning models in Section 5.6.

5.1 Stochastics

Stochastic modelling is one of the methods widely applied to prediction in air traffic operations. The application of stochastic methods is particularly suitable for addressing the variability and unpredictability associated with factors such as weather conditions, airspace congestion, and operational disruptions, which can significantly impact departure times.
In the extensive literature review of Shone et al. [45], stochastic models, are described as very well-suited for predicting departure times. These models can incorporate a wide range of probabilistic inputs, including historical data on flight delays, patterns of demand fluctuation, and unpredictable events like adverse weather or technical issues. By analyzing these factors by probability and uncertainty, stochastic models can provide more accurate and reliable departure time forecasts compared to deterministic models, which might overlook or inadequately represent the variability in air traffic systems.
Currently, stochastic modelling is used in departure time forecasting, with varying degrees of implementation across different airports and air traffic control systems. Advanced stochastic models, including queueing theory models and stochastic optimization algorithms, are employed to optimize the scheduling of flight departures, aiming to minimize delays and improve the efficiency of airport operations. These models help in developing dynamic scheduling systems that can adapt to real-time information and changes in the operational environment, thereby enhancing the predictability and reliability.
Furthermore, the integration of stochastic modelling with Decision Support System (DST)s for Air Traffic Controllers (ATCos) and Flow Management Position Controler (FMPC)s is increasingly becoming a focus of research and development efforts in the aviation industry. These tools aim to provide actionable insights based on probabilistic forecasts, allowing for more informed decision-making under uncertainty.
Stochastic models are not only suitable for departure time forecasting but are also actively used in current ATM practices. Their ability to effectively handle the complexity and uncertainty of aviation operations makes them invaluable for improving the accuracy of departure forecasts, optimizing flight schedules, and ultimately

enhancing the overall efficiency and reliability of air traffic systems.

However, in recent years, stochastic modelling has shifted away. This evolution in ATM has been driven by the need to address increasingly complex and dynamic systems that characterize modern air traffic environments. Advances in computational technologies and analytical methods have introduced more scalable, accurate, and real-time capable solutions, essential for today's complex air traffic operations. Moreover, the integration with newer automation and decision support technologies requires a shift towards more adaptable and robust forecasting methods that can interact with these systems. While stochastic models have provided a foundational approach, the state-of-the-art in ATM forecasting has advanced to meet the growing demands for precision, efficiency, and real-time responsiveness, thereby enhancing overall system reliability and safety.

5.2 Poisson distributions

The Poisson distribution is another method frequently used due to its simplicity. Brooker [4] states that arrivals are slightly less random than Poisson predictions. Therefore the predictions can be used as a reasonable first approximation.
The Poisson distribution estimates the probability of a certain number of delays occurring within a specific timeframe. However, the dynamic and interconnected environment of airport operations often violates the Poisson assumptions. Delays are not always independent; they can be highly dependent on preceding events and vary with daily traffic patterns. The model also overlooks important variables such as weather conditions, air traffic control decisions, and the cascading effects of previous delays, simplifying the complex reality of airport operations.
In contrast, ML models offer a more robust and adaptive approach, as found by Rebollo and Balakrishnan [38]. They excel in handling the complex, non-linear relationships typical in airport operations by integrating diverse datasets, including operational constraints, historical delay data, and real-time conditions.

5.3 Time series

Where traditional delay forecasting methods often focus on isolated predictors without fully accounting for the systemic impacts within the airport network, time series methods offer a framework for incorporating such complexities to enhance prediction accuracy. The study by Güvercin et al. [20] introduces an approach termed Clustered Airport Modeling (CAM), which integrates network-based information of airports into time series models. This method aims to leverage the structural features of airport networks to improve the accuracy of delay predictions.
The CAM approach constructs a representative time series model for clusters of airports that exhibit similar characteristics in terms of delay patterns and network centrality metrics, Betweenness Centrality. By clustering airports, the model effectively handles outliers and reduces noise, pooling data from airports with similar operational dynamics.
This methodology highlights the role of the network structure, where each airport is considered a node, and flights between them are the connecting edges. The significance of an airport within this network is quantified using graph-based metrics like hub scores and Betweenness Centrality. Experiments have shown that incorporating network topology and clustering airports based on their delay patterns and connectivity substantially enhances prediction accuracy compared to models that treat each airport independently (Wei et al. [52]).
Future directions could involve exploring the dynamic aspects of airport networks by incorporating real-time data and evolving network structures. Moreover, integrating more detailed data, such as specific flight paths and airline operations, could further enhance the granularity and accuracy of delay forecasts. Despite the advancements in stochastic and Poisson distribution methods, machine learning models have consistently demonstrated superior performance in predicting complex, nonlinear patterns typical in modern air traffic systems (Wei et al. [52]).

5.4 Machine learning techniques

ML algorithms can discover patterns and dependencies in the data that traditional models may miss, and they continuously improve as they learn from new data, adapting to changes in airport operations.
Moreover, ML models can process and utilize large volumes of data from various sources, enhancing the accuracy of delay predictions. This comprehensive data utilization allows these models to dynamically adjust to the ever-

changing environment of airports, providing predictions that are more accurate and reflective of real-world conditions.
As air traffic systems become increasingly complex, the need for more advanced forecasting models becomes imperative. ML techniques have emerged as powerful tools in predicting and managing the dynamic and multifaceted nature of air traffic, particularly in forecasting aircraft departure times. This section explores various machine learning methods, each offering unique approaches and advantages in handling the diverse and often non-linear patterns observed in air traffic data. The most promising techniques are given, but first, a more shallow overview of machine learning techniques is given in Figure 5.1, followed by a brief description.

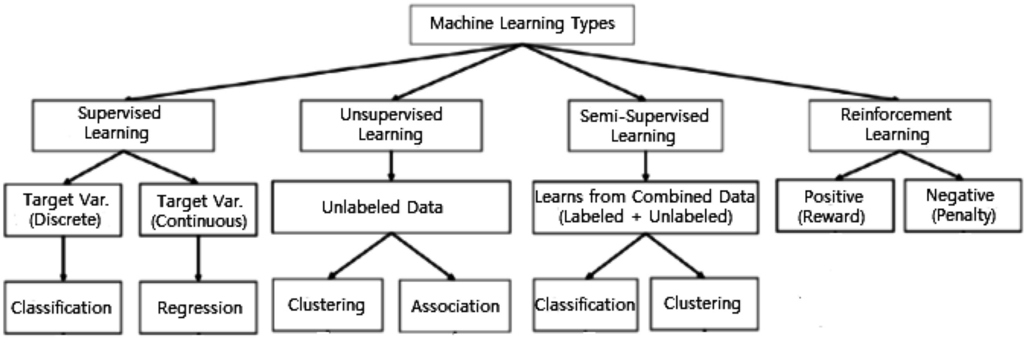


Figure 5.1: Overview of various machine learning techniques (Swana and Doorsamy [49])

Supervised learning is the task within machine learning where the objective is to learn a function that maps input data to corresponding outputs using sample input-output pairs. It relies on labelled training data and a set of training examples to deduce this function. Supervised learning is used when specific objectives need to be achieved from a given set of inputs, thus adapting a task-oriented approach. The primary supervised tasks include classification, which involves categorizing data and regression, which entails fitting the data. However, deep learning and neural networks also fit under supervised learning.

Semi-supervised learning merges supervised and unsupervised methods by utilizing both labeled and unlabeled data, making it ideal for scenarios where labeled data is limited but unlabeled data is plentiful. It aims to surpass the predictive performance achievable with solely labeled data, and is applicable in areas like machine translation, fraud detection, and text classification. Unsupervised learning, on the other hand, focuses on analyzing unlabeled datasets to discover trends and structures, useful for tasks like clustering, density estimation, and dimensionality reduction. Reinforcement learning involves training software agents to make decisions that maximize rewards or minimize risks within a given environment, useful in robotics, autonomous driving, and supply chain logistics, though less suited for simpler tasks like departure time predictions. Together, these learning paradigms address a wide range of complex and varied data-driven challenges.
These methods are widely known and applicable. However, the forecasting of aircraft departure times is a complex problem influenced by a variety of factors, most of which are described in Chapter 4. Traditional forecasting methods have relied heavily on statistical analysis and historical data patterns. However, with the rise of ML, there has been a significant shift towards more dynamic and predictive models capable of accommodating the nonlinearities and variabilitiesr in air traffic systems.
ML offers a transformative approach to predicting aircraft departure times by learning from historical data and identifying patterns that traditional methods might overlook. The application of ML models not only enhances the accuracy of predictions but also contributes to more efficient ATM and reduced delays, thereby improving overall airport operational efficiency. Table 5.1 gives an overview of related literature, found on the usage of ML methods to predict a variety of outcomes, such as individual and aggregate arrival and departure delay and airspace complexity. The papers were selected to give a broad view of the problem, and each paper highlights different aspects of the intricacies of departure delay prediction for a network of airports.
The factors considered in these studies are categorized into three main types:

- **Extrinsic factors (E)**: These include external influences such as weather conditions, air traffic congestion, and regulatory constraints that significantly impact flight operations.
- **Spatial factors (S)**: These factors involve geographical and infrastructural elements such as airport layout, location, distance between gates, and the routes connecting different airports, which are crucial in determining delay propagation.

Table 5.1: Comparison of techniques used in arrival delay, departure delay and airspace complexity prediction studies and their results.

Authors	Factors				Model	Details		
	E	S	T	S		Description	Prediction Horizon	Accuracy
Gopalakrishnan and Balakrishnan [17], 2017	✓	✓		✓	RNN	Aggregate arrival delay	2-hour	Mean error of 4.7 min
Guo et al. [18], 2020	✓	✓	✓	✓	GNN	Departure delays	> 120 min	90.4%
Qu et al. [37], 2021	✓				CNN	Departure delay	None given	93.19% classification
Guo et al. [19], 2022	✓	✓	✓		RF	Departure delay	3-hour	8.73 min MAE, 17.08 min RMSE
Cai et al. [5], 2022	✓			✓	GNN/CNN	Arrival delay	3-hour	±6 min error
Zhou et al. [55], 2022		✓			LSTM	Arrival delay	No horizon	0.42 MAE, 6.07 MAPE
Li et al. [28], 2024		✓	✓		GNN	Airspace complexity	2-hour	83.10%
Li et al. [29], 2024			✓	✓	CNN-LSTM-RF	Classified delay	None given	92.39%

- **Temporal factors (T)**: This category encompasses time-related variables such as time of day, seasonal variations, and day of the week, which influence flight schedules and congestion patterns.
- **Systemic Factors (S)**: This category includes interactions and dependencies within the air traffic network that lead to knock-on delays. These are cascading effects that stem from disruptions in the schedule, impacting operations and are influenced by the connectivity of flight operations, where delays propagate through the network.

The following methods are considered:

- **RNN**: Recurrent Neural Network
- **GNN**: Graph Neural Network
- **CNN**: Convolutional Neural Network
- **LSTM**: Long Short-Term Memory
- **RF**: Random Forest

Table 5.1 illustrates the varied approaches and results in recent research, highlighting the effectiveness of integrating advanced neural network models for predicting flight delays across different time horizons and conditions. All papers will be described below.

5.5 Random Forest

Random Forest (RF) is an ensemble machine learning technique highly applicable for forecasting flight departure times or delays. It constructs numerous decision trees during training and outputs the mode (for classification) or the mean prediction (for regression) across all trees. This methodology is suitable for handling the complex, non-linear interdependencies characteristic of factors influencing flight schedules, such as weather conditions and air traffic congestion. Currently, the DST employs a RF as basis machine learning model.
The strength of Random Forest in flight departure forecasting lies in its ability to mitigate overfitting through its ensemble approach, making it exceptionally robust for variable patterns seen in flight schedules. It ranks input features based on their importance, providing valuable insights into the primary factors affecting departure times or delays. Additionally, its capability to process both categorical and numerical data types ensures comprehensive analysis utilizing diverse datasets related to flight operations.
Utilizing Random Forest involves training on bootstrapped samples of historical flight data, where each tree in the forest predicts the departure time or delay. The aggregate of these predictions forms the final forecast, offering a balance between accuracy and computational efficiency. Despite being computationally demanding

and less interpretable than some models, Random Forest's reliability and predictive power render it a preferred choice for optimizing airline operations and enhancing passenger experience.
In addressing the challenges associated with predicting flight departure delays, Guo et al. [19] propose a hybrid method leveraging Random Forest Regression and Maximal Information Coefficient (RF-MIC). This approach was selected for its superior ability to handle the complexity and non-linearity inherent in flight delay data. The Maximal Information Coefficient (MIC), on the other hand, is employed to quantify the strength of linear or non-linear associations between two variables. This metric is crucial for identifying and including only the most relevant features in the model, thereby enhancing the model's predictive accuracy. MIC excels in discovering intricate patterns within the data that might not be apparent through traditional correlation coefficients, making it particularly suited for the complex nature of flight delay factors which can range from weather conditions to air traffic control constraints. The implementation of this method involved an innovative feature selection process, utilizing a roulette method and prohibitive list based on information consistency, which ensured that the model was both accurate and computationally efficient. Guo et al. [19] demonstrate the effectiveness of the RF-MIC model through a numerical study using flight data from Beijing Capital International Airport. The model showcased superior performance in predicting flight departure delays when compared to conventional models such as linear regression, k-nearest neighbours, and standard Random Forest Regression. This enhanced performance is attributed to the model's ability to precisely identify and incorporate the most predictive features for flight delays, facilitated by the integration of RF and MIC. The application of the RF-MIC method represents a significant advancement in the predictive modelling of flight departure delays, offering a robust and efficient tool for managing the complexities of aviation operations and enhancing passenger experience through more accurate and reliable delay predictions.

5.6 Artificial neural networks

The Artificial Neural Network (ANN) forms the backbone of modern machine learning and artificial intelligence applications. Inspired by the biological neural networks in the human brain, ANNs are composed of interconnected units or nodes, mimicking neurons, that work together to process and interpret complex data inputs. Each neural network consists of three critical components, node character, network topology and learning rules. The character of a node defines the manner in which it processes signals, encompassing aspects like the number of inputs and outputs linked to the node, the weights assigned to each input and output, and the activation function it employs. The organization and connections of nodes are dictated by the network topology. The methods by which weights are initially set and subsequently modified are established by the learning rules. Zou et al. [56]

Node Character

At the heart of an ANN lies the node, or artificial neuron, a fundamental unit designed to mimic the function of biological neurons in the human brain. Each node in an ANN processes signals by receiving inputs through connections that carry specific weights, analogous to the synaptic strength in biological neurons. When the cumulative weighted input surpasses a certain threshold, the node activates, processing the signal through a transfer function and then transmitting it to subsequent nodes. This mechanism is encapsulated in a mathematical model, incorporating elements such as input weights, transfer functions, and threshold values. Transfer functions, especially non-linear ones like the sigmoid function, play a crucial role in enabling the network to tackle complex, non-linearly separable problems.

Network topology

The structure of an ANN is defined by its network topology, which outlines how nodes are arranged and interconnected. Nodes are typically organized into layers, including input, hidden, and output layers, with the possibility of multiple hidden layers. The topology encompasses the number of nodes in each layer, the layering of the network, and the pathways for signal transmission between nodes. This architecture can take various forms, such as feedforward networks, where signals move in one direction, and feedback networks, characterized by dynamic connections that allow for a series of outputs from a single input.

Learning rules

The adaptability of an ANN comes from its ability to learn from data. This learning process involves adjusting the weights of connections between nodes to minimize the difference between the actual output and the desired

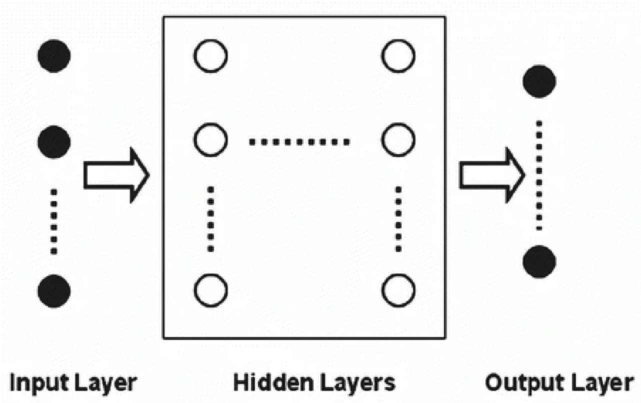


Figure 5.2: General topology of an ANN (Zou et al. [56])

output. Learning in ANNs can be categorized into supervised and unsupervised methods. Supervised learning relies on a predefined set of input-output pairs, adjusting weights to reduce output error. The effectiveness of learning is contingent upon the choice of learning rules, such as error correction methods for supervised learning and nearest neighbour approaches for classification tasks in unsupervised learning. These learning mechanisms enable ANNs to refine their performance over time, enhancing their ability to generalize from training data to new, unseen inputs.

As Artificial Intelligence (AI) is constantly evolving, various types of ANNs have been developed, each tailored to handle different data types and learning tasks more efficiently. Below, we introduce some of the primary types of ANNs, the Convolutional Neural Network (CNN), Graph Neural Network (GNN) and Recurrent Neural Network (RNN).

5.6.1 Graph neural network

GNNs extend the capabilities of ANNs to data that is represented as graphs. This is particularly useful in domains where the data is naturally graph-structured, such as a network consisting of multiple airports or by displaying a set of flights as a graph. GNNs can capture the dependencies not only in the feature space but also in the structure of the data itself. Ma et al. [30] states that GNNs enable better modelling of spatial-temporal features in varying horizons. The core concept of a GNN is visualized in Figure 5.3

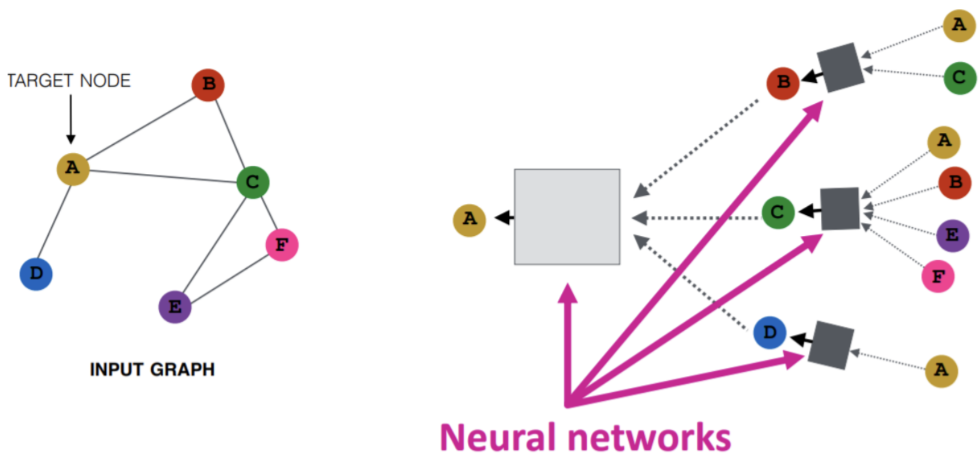


Figure 5.3: Illustration of constructing the multiscale historical delay sequences on graph snapshots (Mehta [31])

GNNs can leverage the graph structure of air traffic networks, enabling the model to understand and utilize the spatial relationships and dependencies between airports. For example, delays at a major hub could propagate to other airports connected to it. By modelling these connections, GNNs can capture the cascading effects of delays more accurately than traditional models that treat each flight or airport independently.

Air traffic is influenced by dynamic factors such as weather conditions, congestion, and operational constraints. GNNs can integrate temporal and dynamic information into the graph model, allowing for the prediction of delays based on current and historical data. This dynamic modelling capability enables GNNs to provide more accurate and timely predictions.
The performance of an airport and its susceptibility to delays can be influenced by various factors, including its connectivity within the network, the volume of traffic it handles, and external conditions affecting it. GNNs can use node-level, edge-level, and global graph attributes to encode such contextual information, enriching the model's understanding and predictive capability.
By using GNNs, it's possible to make predictions not just for individual flights or airports but across the entire network. Moreover, also a set of flights can be displayed as a graph. This view allows for a better assessment of delay probabilities and their potential impact on the network, facilitating better decision-making and management by airlines and airport authorities [42].
Cai et al. [5] investigate flight delay prediction by adopting a deep learning approach that leverages time-evolving graphs. Recognizing the limitations of previous studies that mainly focus on single-airport scenarios and overlook the dynamic spatial interactions within airport networks, their study proposes a novel methodology to model these intricate dynamics. Using a Multiscale Spatial-Temporal Adaptive Graph Convolutional Neural Network (MSTAGCN), the research addresses the challenge of predicting flight delays from a network perspective, encapsulating multiple airports within a unified framework. The MSTAGCN is distinctive for its ability to process both the time-series data of flight delays and the evolving structure of the airport network through graph snapshots, enabling the model to capture the temporal and spatial dependencies inherent in flight operations. The architecture of MSTAGCN consists of two multiscale spatial-temporal adaptive graph convolutional layers followed by an output layer, which collectively aim to forecast flight delays with heightened accuracy. Especially for a long horizon of up to three hours, the model shows great promise. However, it has to be noted that this includes aircraft that are already airborne, which significantly reduces the uncertainty.

In an approach with similar complexity, but different goal, Li et al. [28] propose a Multimodal Adaptive Spatio-Temporal Graph Neural Network for Airspace Complexity Prediction (MAST-GNN) that predicts airspace complexity up to a 120-minute horizon. The MAST-GNN framework operates by first creating a network of airspace sectors represented as nodes in a graph. It processes spatial and temporal data through its two main components. Using Multimodal Adaptive Graph Convolution (MAGCN), spatial data is processed by adapting to different spatial contexts dynamically, using traffic flow and geographic characteristics. Temporal Convolution Network with Attention (TCN-Att) addresses the temporal aspect by employing a self-attention mechanism that adjusts to different temporal scales, crucial for predicting changes in airspace complexity over time. MAST-GNN significantly outperforms traditional methods and contemporary graph-based models across various prediction horizons. Specifically, it shows substantial improvements particularly in longer-term predictions (60 to 120 minutes). The approach of MAST-GNN to integrate multimodal data and its capability to adapt to spatial and temporal variations makes it very suitable for predicting airspace complexity.

When focussing on departure delays, Guo et al. [18] give an interesting take at utilizing the strength of a GNN to predict departure delay at long (>120 minutes) horizon. More specifically, using a Spatio-temporal Graph Dual-Attention Neural Network (SGDAN). The model is designed by first modelling the complex air traffic as graph sequences. Each graph represents flights and their relationships at different times, relationships such as the same departure or arrival airport, or different flights using the same aircraft. Every node is a flight, and the edges represent the relationship, as can be seen in Figure 5.4. Using a dual-attention mechanism, both heterogeneous graph-level attention and sequence level-attention are integrated. The dual-attention mechanism focuses on the spatial aspects by learning the impact of neighbouring flights (those sharing the same departure or arrival airport) on the target flight's delay. It uses multi-head attention mechanisms to weigh the influence of these neighbouring flights differently. Sequence-level attention focuses on the temporal aspects by evaluating how previous flights affect the current flight in the sequence, particularly considering the shared aircraft. This helps to capture the propagation of delays through consecutive flights.

5.6.2 Convolutional neural network

CNN is a type of ANN that is pivotal in the field of deep learning, particularly for tasks involving image and video processing, as well as for some areas of natural language processing and time-series analysis. The architecture of CNNs is inspired by the organization of the animal visual cortex and is particularly adept at recognizing patterns leading to the identification of objects within images. These networks utilize layers of convolutional filters to process data in a hierarchical manner; lower layers might identify edges and textures,

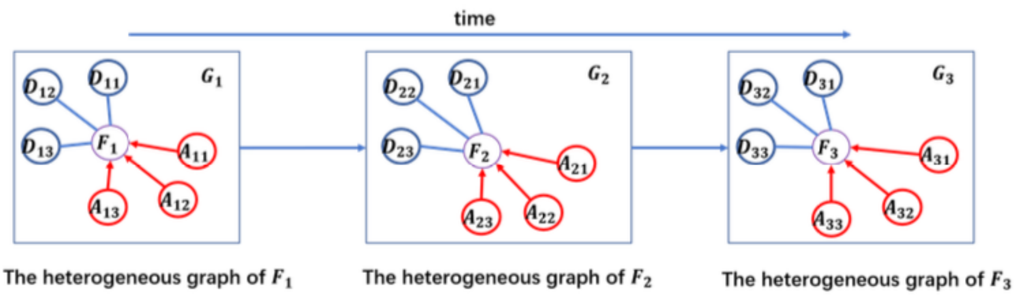


Figure 5.4: Illustration of constructing the multiscale historical delay sequences on graph snapshots (Guo et al. [18])

while deeper layers can recognize more complex structures like faces or objects. This hierarchy allows CNNs to abstract and understand the spatial hierarchies in data effectively.
CNNs uses pooling and activation functions like ReLU to introduce non-linearity and invariance to minor changes in the input data, improving their predictive capabilities. The architecture often concludes with fully connected layers that compile the features extracted by convolutional layers to make final predictions or classifications.
One of the key advantages of CNNs is their efficiency in processing data due to weight sharing and sparsity of connections, reducing the number of parameters compared to fully connected networks. This efficiency not only speeds up the training process but also reduces the computational cost, making it easier to train complex models. Furthermore, CNNs have been at the forefront of deep learning advancements, pushing the boundaries in fields like autonomous vehicles, medical image analysis, and face recognition, showcasing their adaptability and effectiveness in extracting meaningful information from complex datasets. Typically, a CNN has these layers IBM [24] :

Convolutional layer

The convolutional layer, fundamental to CNNs, processes input images (represented in a 3D matrix for RGB images) using filters (or kernels) to produce feature maps. This involves convolving the filter across the image's width, height, and depth, computing dot products to detect specific features. Key parameters affecting the output include the number of filters (determining output depth), stride (the filter's movement step size), and padding (managing border effects). Training adjusts filter weights, while hyperparameters like filter number, stride, and padding type are preset. Post-convolution, a non-linearity (ReLU) is applied to the feature map.

Pooling layer

Pooling layers, also known as downsampling, simplify the input's dimensionality without retaining every detail, primarily through max pooling, which selects the highest pixel value, and average pooling, which calculates the mean value within a specific area. This process, distinct from convolutional layers due to the absence of weighted filters, effectively minimizes computational complexity and the model's susceptibility to overfitting by summarizing the most significant features.

Fully-connected layer

The fully-connected layer links every node from the previous layer to each of its own nodes, enabling the network to classify images based on extracted features and their significance. Unlike convolutional and pooling layers that commonly use the ReLU function, the fully-connected layer typically employs a softmax activation function. This function assists in outputting probabilities for each class, facilitating the precise classification of inputs.
A notable example of a study utilizing a CNN is Qu et al. [37] who present a method to predict flight delays by using a combination of flight data and meteorological information. Two CNN models are proposed: Dual-channel Convolutional Neural Network (DCNN) and Squeeze-and-Excitation Densely Connected Convolutional Network (SE-DenseNet). The method involves preprocessing data to merge relevant flight and weather data, then using the DCNN and SE-DenseNet models to automatically extract features from this fused dataset. The DCNN model integrates features using dual channels to enhance feature matrix transmission and network depth patency, while the SE-DenseNet model incorporates the Squeeze-and-Excitation (SE) module to recalibrate feature importance dynamically. Comparing the CNN, DCNN and SE-DenseNet Experimental results show that incorporating meteorological information improves prediction accuracy by about 1% compared to using

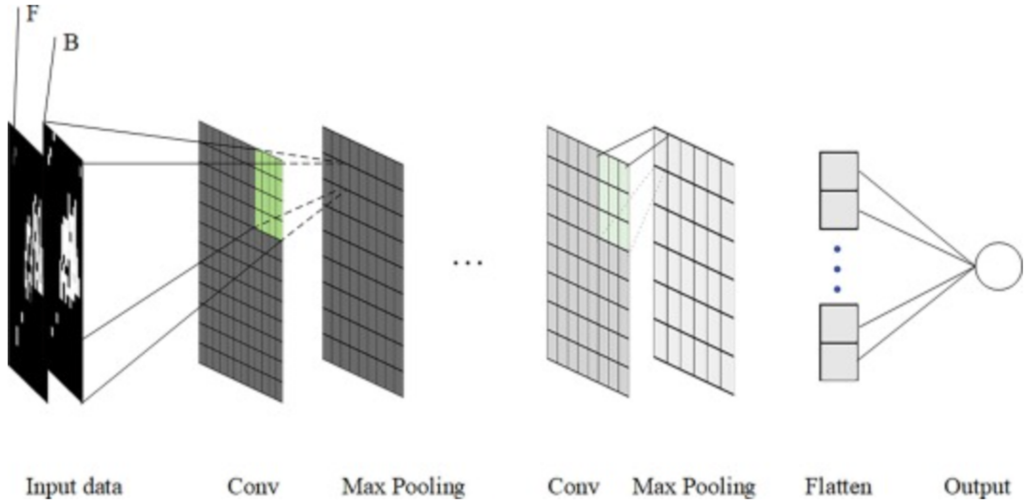


Figure 5.5: Structure of the CNN model (Li et al. [29])

flight data alone, with SE-DenseNet achieving the highest accuracy of 93.19%. The study demonstrates that deep learning methods, particularly when using large datasets that include weather information, can significantly outperform traditional prediction models in accuracy and reliability for flight delay predictions.

5.6.3 Recurrent neural networks

RNNs are designed to recognize patterns in sequences of data, such as text or stock market fluctuations. Unlike feedforward networks, RNNs have "memory" about previous inputs in their internal state, which influences the network's output. This makes them ideal for tasks where context or the sequential order of data points is important. However, RNNs can struggle with long dependency sequences, which led to the development of more advanced RNNs like the Long-Short Term Memory Cell (LSTM) and the Gated Recurrent Unit (GRU).

Long-Short Term Memory networks

A promising technique is the LSTM, which helps modelling sequential data. Pisa et al. [34] describes LSTM cells as particularly useful for handling time-series data due to their unique architecture that allows them to remember and utilize historical data points, making them a suitable choice for modelling the dynamic processes around an airport where current output depends on both the present and past inputs. The LSTM cell accomplishes this through a sophisticated system of gates which are also visualized in Figure 5.6:

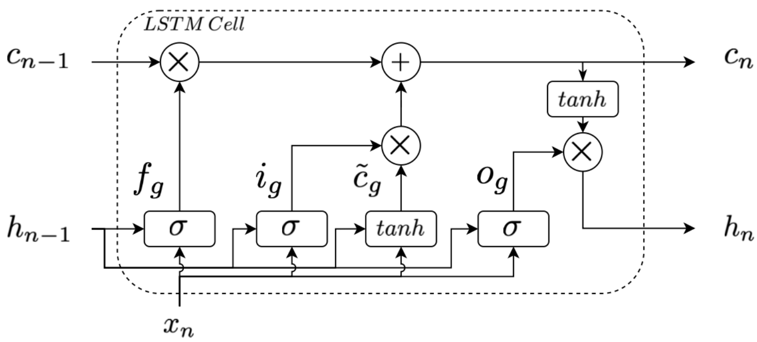


Figure 5.6: Schematic of an LSTM cell. (Pisa et al. [34])

- Forget Gate (fg): This gate decides what information from the cell state should be discarded or retained. It uses the current input and the output from the previous step to generate values between 0 and 1 through a sigmoid layer, with 0 indicating that the information should be forgotten, and 1 indicating that it should be retained and used to modify the hidden state.

- Input Gate (ig): Similar to the forget gate, the input gate decides which of the current input information should be used to update the cell state. It determines the relevance of the input information in the context of the new cell state.
- Cell-State Gate (cg): This gate calculates the candidate values for the new cell state by considering the previous output and the current input. It uses a hyperbolic tangent activation function, which outputs values in the range of -1 to 1, to generate these candidate values. The new cell state is then determined by combining the outputs of the forget gate, input gate, and the cell-state gate.
- Output Gate (og): The output gate decides what the next output should be based on the cell state and the current input. This gate effectively determines the next cell output, which is modified according to the cell state to produce the final output.

The most promising research predicting delays is the LSTM study by Li et al. [29]. This study introduces a two-stage CNN-LSTM-RF model that integrates spatial and temporal data alongside extrinsic features for enhancing flight delay predictions. This hybrid model utilizes the strengths of both CNN and LSTM architectures to capture the complex spatial-temporal correlations that influence flight schedules, further complemented by Random Forest for final delay predictions. The LSTM component plays a pivotal role in capturing temporal dependencies inherent in aviation data, particularly from flight schedules and meteorological conditions: The LSTM networks engage in weather pattern analysis by examining time-sequenced meteorological data leading up to scheduled departures. This analysis helps the networks to comprehend historical weather impacts and effectively predict potential delays caused by similar future conditions. Additionally, the LSTM utilizes its capability to retain relevant historical data over extended periods, thereby extracting essential temporal patterns that significantly influence delay predictions. In coordination with the LSTM, the CNN layers process spatial information about airport congestion and delays. The CNN processes spatial features by mapping airports based on geographical locations and analyzing congestion patterns, which serve as critical predictors of potential delays. The integration of CNN and LSTM outputs enables a comprehensive analysis of how spatial and temporal factors collectively impact flight delays. This spatial-temporal modelling approach combines current and past airport conditions across the network, providing a holistic view of the factors influencing flight times. Furthermore, the features from both CNN and LSTM frameworks are employed as inputs for a RF classifier. Known for its high accuracy, the RF algorithm utilizes an ensemble of decision trees to predict flight delays, capitalizing on the majority voting system to enhance predictive accuracy. The model has achieved an accuracy rate of 92.39%, underscoring its effectiveness in synthesizing various contributing factors to flight delays.

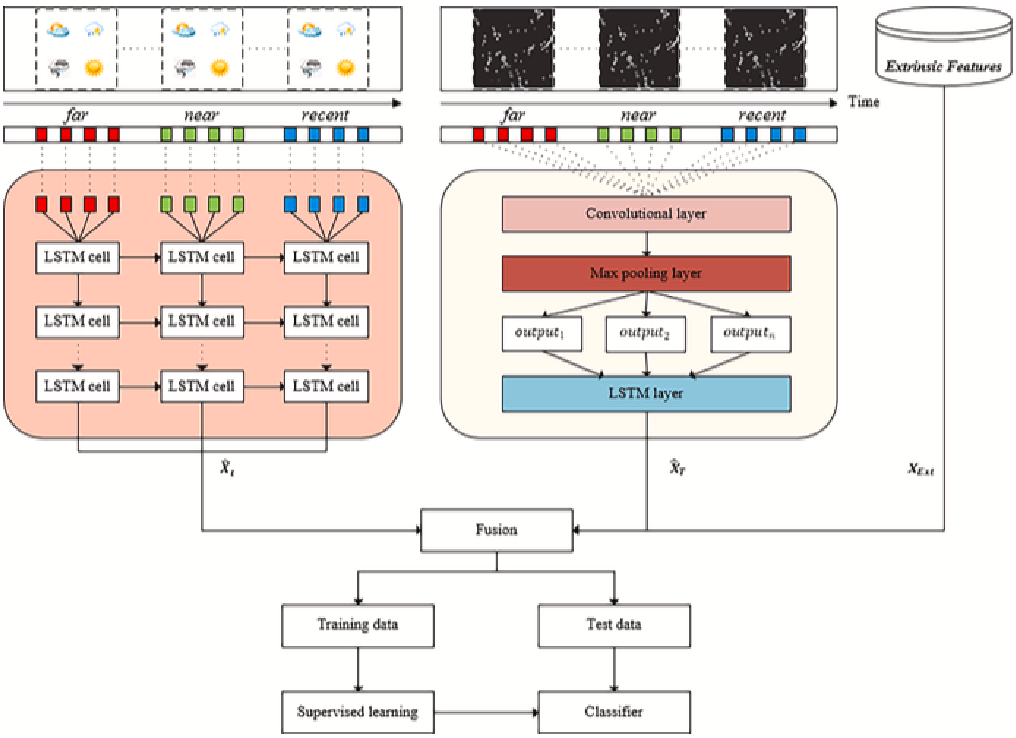


Figure 5.7: The architecture of the CNN-LSTM-Random Forest. (Li et al. [29])

While [29] provide the most promising LSTM research, more results are gained. Sun et al. [48] propose a LSTM network for airport delays and a dynamic spatial-temporal graph attention network (DST-GAT) for network-level predictions. The DST-GAT model, a key innovation, accurately models and predicts delays across airport networks by considering the dynamic relationships between airports. The paper employs two LSTM layers as part of their approach to predict delays at three different scopes: individual flights, airports, and the network of airports. LSTM layers are particularly suitable for this task due to their ability to model temporal dependencies and capture dynamics in time series data, which is inherent in flight delays.

GRU

A notable alternative to the complex LSTM, is a simplified version, the GRU. Zhou et al. [55] proposes a GRU neural network to make departure time estimations. This is done using the flight information, airport, weather and airline data. The model used, a GRU is a model based on the GRU. Unlike the LSTM, the GRU combines the forget gate and input gate into a single update gate and mixes neurons and hidden states to reduce network parameters and improve the training speed.
This LSTM-based structure allows for the detailed and nuanced modelling of time-dependent processes, such as those found in Air Traffic Flow Management (ATFM), by capturing the temporal correlations inherent in the measurement data. The LSTM's ability to maintain and manipulate a memory of past observations over long sequences makes it particularly well-suited for predicting the dynamics within complex systems where historical data significantly influence current conditions. This shows good promise for prediction departure times, however, no forecast horizon is given in the paper, so uncertainty about this method remains.

Chapter 6

Data

The methods described in the previous chapter can only be as good as the provided data. All methods described require extensive data on weather circumstances, Air Traffic Flow Management (ATFM) regulations, Airport Collaborative Decision Making (A-CDM) data among others. As at this given instance, it is still largely uncertain what data is available, this chapter will describe the data sources that are desired.
A key consideration is the desired horizon, as the reliability of data sources is heavily reliant on the amount of time between the prediction and the actual event. For example, flight plans are updated starting from 3 hours before departure. The desired horizon for this research is 4 hours and therefore other data sources have to be found. However, flight plan data of other flights can indicate knock-on delays or other effects, so will be incorporated.

The first type of data that is described is B2B data in Section 6.1, followed by Weather data in Section 6.2. The Chapter is concluded by the description of Automatic Dependent Surveillance-Broadcast (ADS-B) data in section 6.3.

6.1 B2B Data

EUROCONTROL Network Manager (NM), the organisation that monitors the traffic flow for the entire ECAC area, collects and distributes information between all Air Traffic Service Provider (ANSP)s, airlines, airports and other stakeholders in the area. The B2B service is at the core of EUROCONTROL's interoperability strategy and is seen as instrumental in the Single European Sky Aviation Research (SESAR) development. EUROCONTROL [15] gives the following categories of services:

Flight Services

Focused on the facilitation and management of flights within and heading towards the ECAC area, Flight Services offer comprehensive data sharing capabilities that enhance interactions between airspace users and ANSPs. This service enables users to initiate, file, and manage flight plans efficiently. Each flight plan undergoes a thorough validation process and is continuously monitored in coordination with the relevant air traffic control sectors. Additionally, the service incorporates tools for planning departures and arrivals, which are crucial for airports engaged in A-CDM. Air Traffic Flow Capacity Management (ATFCM) slot information forms a critical component of this service, allowing the network manager to effectively balance demand and capacity. Moreover, Flight Services provides updates on the progression of airborne flights, including position reports, status updates, and messages related to system activations. These updates are given in the form of Departure Planning Information (DPI) and Flight Update Messages (FUM). Although these do not include a specific Actual Take Off Time (ATOT) field, a combination of sources can be used to determine the ATOT.

Airspace Services

Airspace Services are dedicated to providing a structured and up-to-date view of operational airspace data, crucial for the smooth functioning of air traffic management. This service includes access to Aeronautical Information Publication (AIP) sourced data such as points, routes, aerodromes, and airspaces, tailored to meet the operational requirements of the NM's flight and flow systems. Not officially published as AIP data, this NM-adapted information helps in strategic planning and execution. The service also covers ATFCM-related airspace data like restrictions, which encompass route availability documentation and profile tuning restrictions. Additionally, electronic Airspace Management Information (e-AMI) is available, facilitating access to the European Airspace Use Plan/Updated Use Plan in compliant formats, further supporting airspace availability and management.

Flow Services

Flow Services are integral to managing and optimizing the flow of air traffic within the network. This section provides access to all relevant regulation information utilized within NM's flow management systems. It encompasses services that allow for a detailed view of the Network Situation, including traffic, delays, causes of delays, and current regulations. Traffic counts by various parameters such as aerodrome, aircraft operator, and airspace are readily accessible. Additionally, this service supports the management of ATFCM's daily operational plans, which include capacity plans and sector configuration plans among others. Flow Services also facilitate scenario management and offer simulation capabilities, allowing stakeholders to assess the impact of different ATFCM measures effectively.

6.1.1 General Information Services

General Information Services serve as a gateway to a wide array of network operations information. These services are vital for maintaining transparency and accessibility in air traffic management operations. Users can access ATFM Information Messages (AIM), which publish general network operations information, and NM B2B Info. B2B data ensures that stakeholders are well-informed and can retrieve necessary operational data and documentation to support their activities within air traffic management.

6.2 Weather Data

The weather has an enormous effect on the capacities of airports. Although forecasts can be unreliable, ATFM regulations can be made according to projections. Therefore, the 4 hour before departure forecast can effect the departure time. Meteorological Aerodrome Reports (METAR) data provides routine observations of weather conditions at airports, offering crucial insights that can significantly impact aircraft departure times. This data is instrumental for machine learning models aimed at enhancing the predictability and efficiency of airport operations. The reports are generated by airport weather stations and include detailed, timely observations on a variety of weather conditions such as wind, visibility, precipitation, cloud cover, temperature, and barometric pressure. These reports are typically issued once every hour, providing a snapshot of the weather conditions at the time of the observation.
Key components of METAR data relevant to machine learning models include:

- **Wind:** Speed, direction, and gusts, which affect decisions on runway usage and can influence takeoff and landing operations.
- **Visibility:** Measured in meters or miles, low visibility conditions can trigger air traffic flow restrictions.
- **Weather Phenomena:** Descriptions of current weather conditions such as rain, snow, fog, or thunderstorms that might delay operations or require additional safety measures.
- **Sky Condition:** Information on cloud cover, including the type and altitude of clouds, which is crucial for flight operations.
- **Temperature and Dew Point:** These factors influence aircraft performance and may impact de-icing operations during cold weather.
- **Pressure:** Barometric pressure readings are essential for aircraft altimeter settings.

For all these components, predictions are made, up to a horizon till 18 hours in advance which makes it useful for this thesis. In conclusion, METAR data is a vital component for machine learning models attempting to predict aircraft departure times. Its detailed and timely observations allow for a nuanced understanding of how weather affects airport operations, enhancing the predictive capabilities of these models and supporting more effective decision-making processes.

6.3 ADS-B data

ADS-B is a sophisticated surveillance technology that enables aircraft to broadcast their location and other state information. This system, which is set to replace traditional secondary surveillance radar systems, is increasingly used in aviation due to its significant benefits in terms of coverage and accuracy. ADS-B signals are straightforward to intercept, which permits not only aircraft but also ground stations and private receivers to process the data easily. As a result, ADS-B has become a cornerstone technology in modern Air Traffic Control (ATC) systems. Its widespread adoption helps improve the security, reliability, and efficiency of airspace usage. Several organizations leverage ADS-B data to enhance aviation services and research. One notable example is OpenSky, which provides open access to real-world air traffic control data. This initiative supports numerous applications in improving airspace security and operational efficiency by making detailed flight data available to the public.
The data transmitted by ADS-B includes essential information such as:

- Aircraft identification code
- Surface position
- Airborne position
- Airborne velocities
- Operational status

While ADS-B data is not the primary data source in this research, it can play a role during the pre-departure phase of flights. Accurate knowledge of an aircraft's position while airborne within the horizon is necessary to predict whether it will arrive at the departure airport on time.

Chapter 7

Research gap

As identified in the previous chapter on forecasting methods in Air Traffic Management (ATM), effective prediction of departure times and management of knock-on delays are critical for enhancing operational efficiency in the aviation industry. However, despite significant advancements in forecasting methods, there are still notable gaps that could be addressed to improve the precision and utility of these predictions, particularly regarding the impact of weather conditions and the extension of the forecast horizon to four hours.

7.1 Extending Forecast Horizons

One of the primary challenges in current forecasting methodologies is the limitation of prediction horizons. Most existing models, including those based on Machine Learning (ML) techniques, typically focus on short-term forecasts—usually up to two hours. Extending the forecast horizon to four hours could significantly enhance the ability of airport operators and airlines to make proactive adjustments, reducing the cascading effects of delays and optimizing overall network efficiency. Longer forecast horizons would allow for better strategic decisions, such as adjusting flight schedules in advance and optimizing Air Traffic Flow Management (ATFM) to anticipate and mitigate potential disruptions.

7.2 Knock-On Delays

Another significant gap is the management and prediction of knock-on delays, which are secondary delays caused by an earlier disruption in the network. Current models often treat flights and airports as independent entities, without fully accounting for the interconnected nature of the aviation network. This oversight can lead to underestimations of delay durations and impacts, particularly in hub airports where delays can ripple across many subsequent flights.

An enhanced focus on knock-on delays would involve developing models that can dynamically interpret and predict the cascading effects of a delay throughout the network. This approach would require better understanding and incorporation of network topologies, interdependencies between flights, and the specific characteristics of hub versus spoke airports.

7.3 Weather Impacts at Different Airports

Weather conditions play a crucial role in the predictability and management of flight schedules. However, the impact of weather can vary significantly between different airports due to their geographic and operational characteristics. Research into the differential impacts of weather on various airports could lead to more customized forecasting models. These models would take into account the unique weather patterns and their typical impacts on specific airports, thereby improving the accuracy of delay predictions. Moreover, integrating real-time weather data into forecasting models can dynamically adjust predictions as weather conditions change.

Chapter 8

Research Objective

This research is driven by the central objective to enhance the accuracy and reliability of demand forecasting used by the Decision Support System (DST) of Luchtverkeersleiding Nederland (LVNL). More specifically by reducing uncertainty during the pre-departure phase (-3 to -4 hours) of flights inbound to Schiphol. The investigation focuses on machine learning techniques to achieve this goal. As a result, the objective of this research is:

To build a model that reduces departure time uncertainty for flights inbound Schiphol at a 4-hour horizon.

8.1 Research question

The primary question guiding this research examines the impact and potential of machine learning to reduce uncertainties in air traffic management at out-stations, specifically within a four-hour flight radius of Schiphol. Therefore the research question is as follows:

To what extent can machine learning be applied to the pre-departure phase at a horizon of four hours?

This can be segmented into three elements, demand forecasting, departure time forecasting and machine learning methods. To further segment this, the following sub-questions are identified.

1. What data features contribute most to the departure time prediction accuracy of flights inbound Schiphol?
2. What are the uncertainties in air traffic at out-stations, located within a 4-hour flight radius within LVNL's operational framework, and how do these uncertainties impact the demand prediction?
3. Which input parameters are available pre-departure and various horizons up to 4 hours, and which features are expected to contribute most to departure time trajectory prediction?
4. How can the results of the model be validated?
5. How does the developed model perform when employed by airports other than Schiphol?

8.2 Scope

The initial scope of this project is focused on predicting the departure times for all flights currently on the ground and destined to depart from Schiphol within four hours. If achieving this objective becomes infeasible, alternative scopes will be considered. These adjustments will be determined based on the availability of data and the progress of the project. The potential alternative focuses include:

1. **Concentrating on Specific Airports:** Narrowing the prediction scope to specific airports may provide more targeted insights and efficient resource utilization.
2. **Altering the Prediction Horizon:** Modifying the forecast horizon from the current four hours to a different duration, which could be shorter or longer depending on data accuracy and completeness.

3. **Focusing on Specific Flights:** Restricting the model to flights using Schiphol as a hub could enhance predictability due to the higher consistency and availability of data, especially for knock-on effects.

These scope adjustments will allow for a more tailored and effective predictive model, capable of adapting to the dynamics of available data and ongoing project insights.

Chapter 9

Conclusion

In Air Traffic Management (ATM), the challenge of balancing demand and capacity is pivotal in ensuring the safe and efficient transit of flights through controlled airspaces. Notably, Luchtverkeersleiding Nederland (LVNL), the Dutch air navigation service provider, has employed a sophisticated decision support tool designed to predict traffic load and facilitate the management of traffic flow by air traffic controller supervisors. While this tool provides a relatively accurate forecast of traffic loads, its effectiveness is occasionally undermined by significant forecast errors that can result in inefficient or even unnecessary traffic management decisions, such as imposing delays on flights preemptively. This research seeks to advance the sector demand forecasting in the tactical domain by leveraging machine learning technologies to enhance trajectory predictions. The core research objective is: 'To build a model, that reduces departure time uncertainty for flights inbound Schiphol at a 4-hour horizon'.

The exploration begins in Chapter 2 with an in-depth analysis of the existing demand forecasting methodologies detailed in earlier chapters of this research. It is noted that these methodologies heavily depend on centralized flight plan data, which inherently carries uncertainties particularly in regards to flight departure times. The thesis critiques these traditional forecasting methods and argues for the adoption of more robust models that can reduce reliance on uncertain and often inaccurate flight plan data.

Subsequently, in Chapter 3 both model-based and data-driven trajectory prediction methods are examined. Through a comparative analysis, it is established that while model-based approaches are grounded in established aircraft performance parameters, they often fall short in dynamically changing conditions where comprehensive historical data and intent information are lacking. In contrast, data-driven approaches, particularly those employing advanced machine learning algorithms, demonstrate a higher capability in recognizing complex patterns and adapting to new data, making them more suited for long-term predictive accuracy.

Moreover, Chapter 4 on the predictability of the Actual Take-Off Time (ATOT) delves deeply into the complexities that influence ATOT and the broader implications for air traffic management systems globally. Key factors such as operational inefficiencies, adverse weather conditions, and air traffic congestion are examined for their impact on ATOT. The Chapter emphasizes the importance of understanding and mitigating the causes of delays to improve ATOT predictability. Methodologies and technologies that are employed to enhance the accuracy of forecasting ATOT are discussed, most importantly Airport Collaborative Decision Making.

Further, various machine forecasting methods that hold relevance to demand forecasting and trajectory prediction within ATM systems. It delves into the functionalities and potential applications of convolutional and graph neural networks, alongside recurrent neural network architectures. These models are evaluated for their effectiveness in generating accurate trajectories and facilitating robust demand forecasts, essential for the proactive management of air traffic.

Finally, a detailed examination of the data utilized in Chapter 6, comprising flight information messages from Eurocontrol's Network Manager and weather data from METAR, optionally extended by actual ADS-B trajectories collected via OpenSky. This dataset supports the development and refinement of the proposed machine learning models. This information is followed by the Research Gap in Chapter 7, and Research Objective in Chapter 8 where the stage of the research is set.

Bibliography

[1] Amsterdam Airport Schiphol (2024). Deep turnaround. `https://www.schiphol.nl/en/aviation-solutions/page/deep-turnaround/`. Accessed: 2024-04-02.

[2] Asadi, E., Evler, J., Preis, H., and Fricke, H. (2020). Coping with uncertainties in predicting the aircraft turnaround time at airports. In Neufeld, J. S., Buscher, U., Lasch, R., Möst, D., and Schönberger, J., editors, *Operations Research Proceedings 2019*, pages 773–780, Cham. Springer International Publishing.

[3] Barnhart, C., Fearing, D., Odoni, A., and Vaze, V. (2012). Demand and capacity management in air transportation. *EURO Journal on Transportation and Logistics*, 1(1-2):135–155.

[4] Brooker, P. (2009). Simple models for airport delays during transition to a trajectory-based air traffic system. *The Journal of Navigation*, 62(4):555–570.

[5] Cai, K., Li, Y., Fang, Y.-P., and Zhu, Y. (2022). A deep learning approach for flight delay prediction through time-evolving graphs. *IEEE Transactions on Intelligent Transportation Systems*, 23(8):11397–11407.

[6] Dalmau, R., Attia, J., and Gawinowski, G. (2023). Modelling the likelihood of air traffic management regulations due to weather at airports. *EUROCONTROL Innovation Hub.* Conference Paper presented at EUROCONTROL.

[7] de Leege, A., Van Paassen, M. M., and Mulder, M. (2013). A machine learning approach to trajectory prediction. *AIAA Guidance, Navigation, and Control (GNC) Conference.*

[8] Delgado, L. and Prats, X. (2011). Simulation of airborne atfm delay and delay recovery by cruise speed reduction. *First SESAR Innovation Days.* 29th November - 1st December 2011.

[9] Dijkstra, F. and Calis, G. (2023). System Subsystem Specification (SSS) Decision Support Tool. Draft, LVNL INTERN. Document number: DST-SSS, Classification: LVNL INTERN.

[10] EUROCONTROL (2017). *Airport CDM Manual.* EUROCONTROL. Accessed: 2024-02-29.

[11] Eurocontrol (2017). *Airport Collaborative Decision-Making (A-CDM) Implementation Manual*, 5.0 edition. PUBLISHED ON 31 March 2017.

[12] EUROCONTROL (2019). Atfm regulation: a power for good. Aviation Intelligence Unit.

[13] EUROCONTROL (2023a). Annual network operations report 2022. EUROCONTROL. The final version of the Annual Network Operations Report 2022 is now available.

[14] EUROCONTROL (2023b). Eurocontrol european aviation overview. `https://www.eurocontrol.int/publication/eurocontrol-european-aviation-overview`. Accessed: 2023-05-07.

[15] EUROCONTROL (2024). Network manager business-to-business web services. https://www.eurocontrol.int/service/network-manager-business-business-b2b-web-services. Accessed: 24-4-23.

[16] Fricke, H. and Schultz, M. (2022). Improving aircraft turn around reliability: Specific aircraft body design parts hamper ground handling and airport performance. *Technische Universität Dresden, Institute of Logistics and Aviation.*

[17] Gopalakrishnan, K. and Balakrishnan, H. (2017). A comparative analysis of models for predicting delays in air traffic networks. In *Twelfth USA/Europe Air Traffic Management Research and Development Seminar (ATM2017).* ATM Seminar. Accessed: 2024-04-13.

[18] Guo, f. Z., Mei, G., Liu, S., Pan, L., Bian, L., Tang, H., and Diansheng, W. (2020). Sgdan—a spatio-temporal graph dual-attention neural network for quantified flight delay prediction. *Sensors*, 20(6433).

[19] Guo, Z., Hao, B. Y. M., Wang, W., Jiang, Y., and Zong, F. (2021). A novel hybrid method for flight departure delay prediction using random forest regression and maximal information coefficient. *Aerospace Science and Technology*, 116:106822.

[20] Güvercin, M., Ferhatosmanoglu, N., and Gedik, B. (2021). Forecasting flight delays using clustered models based on airport networks. *IEEE Transactions on Intelligent Transportation Systems*, 22(5):3179–3189.

[21] Hanley, Z. J. (2015). Delay characterization and prediction in major u.s. airline networks. Master's thesis, Massachusetts Institute of Technology. © Massachusetts Institute of Technology 2015. All rights reserved.

[22] He, P. and Pan, W. (2024). Collaborative optimization model and algorithm for airport capacity and traffic flow allocation. *PLoS ONE*, 19(3):e0298540.

[23] IATA (2018). Airport – collaborative decision making (a-cdm): Iata recommendations. IATA Working Group Document. Recommendations for A-CDM implementation authored by the IATA Airline A-CDM Coordination Group ('AACG'). The AACG is an IATA working group comprising airlines and industry observers with expert knowledge and experience in A-CDM Implementation.

[24] IBM (2023). What are convolutional neural networks? `https://www.ibm.com/topics/convolutional-neural-networks`. Accessed: 2023-10-09.

[25] Könnemann, F. (2014). Performance impact of improved departure time prediction relative to sector demand & arrival time predictability. Master's thesis, Delft University of Technology, Delft, Netherlands.

[26] Koolen, H. and Suciu, I. (2023). *DPI Implementation Guide*. EUROCONTROL, 2.600 edition. APT/USD/DPIImplGuide.

[27] Leal de Matos, P. and Ormerod, R. (2000). The application of operational research to european air traffic flow management – understanding the context. *European Journal of Operational Research*, 123(1):125–144.

[28] Li, B., Li, Z., Chen, J., Yan, Y., Lv, Y., and Du, W. (2024a). Mast-gnn: A multimodal adaptive spatio-temporal graph neural network for airspace complexity prediction. *Transportation Research Part C: Emerging Technologies*, 160:104521.

[29] Li, Q., Guan, X., and Liu, J. (2024b). A cnn-lstm framework for flight delay prediction. *Journal of Intelligent Transportation Systems*, XX(XX):XX–XX. Preprint version. Published online ahead of print.

[30] Ma, C., Alam, S., Cai, Q., and Delahaye, D. (2022). Sector entry flow prediction based on graph convolutional networks. *ICRAT*, pages 1–9. © 2022 ICRAT. All rights reserved. This paper was published in Proceedings of 2022 International Conference on Research in Air Transportation (ICRAT 2022) and is made available with permission of ICRAT. Downloaded on 20 Feb 2024 18:28:58 SGT.

[31] Mehta, P. (2020). Graphs neural networks in nlp. `https://medium.com/neuralspace/graphs-neural-networks-in-nlp-dc475eb089de`. Accessed: 2024-04-11.

[32] Oosterhof, E. (2022). Effect of trajectory prediction uncertainty on a probabilistic debunching concept for inbound air traffic. Master's thesis, Delft University of Technology, Delft, Netherlands. Supervised by Prof. dr. ir. J.M. Hoekstra and F. Dijkstra.

[33] Performance Review Commission (2023). Performance review report. COMMENTS SHOULD BE SENT TO: pru-support@eurocontrol.int before 17h00 on 14 April 2023. An assessment of Air Traffic Management in Europe during the calendar year 2022.

[34] Pisa, I., Morell, A., Lopez Vicario, J., and Vilanova, R. (2020). Denoising autoencoders and lstm-based artificial neural networks data processing for its application to internal model control in industrial environments—the wastewater treatment plant control case. *Sensors*, 20:3743.

[35] Post, P. (2021). Assessment of inbound air traffic flow management delay and total arrival delay for amsterdam airport schiphol. Master's thesis, Delft University of Technology, Delft, The Netherlands. An electronic version of this thesis is available.

[36] Pérez Moreno, F., Gómez Comendador, V. F., Delgado-Aguilera Jurado, R., Zamarreño Suárez, M., Janisch, D., and Arnaldo Valdés, R. M. (2023). Methodology of air traffic flow clustering and 3-d prediction of air traffic density in atc sectors based on machine learning models. *Expert Systems with Applications*, 223:119897.

[37] Qu, J., Zhao, T., Ye, M., Li, J., and Liu, C. (2020). Flight delay prediction using deep convolutional neural network based on fusion of meteorological data. *Neural Processing Letters*, 52:1461–1484.

[38] Rebollo, J. J. and Balakrishnan, H. (2014). Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, 44:231–241.

[39] Rodríguez-Sanz, Á., Cano, J., and Fernandez, B. R. (2022). Impact of weather conditions on airport arrival delay and throughput. *Aircraft Engineering and Aerospace Technology*, 94(1):60–78. Work supported by the program AYUDA PUENTE 2019, URJC, 2020/00006/014. Part of this research was carried out while the last author was visiting the University of Auckland, supported with grants from URJC's mobility programs. Received 30 December 2020; Revised 29 March 2021; 2 May 2021; Accepted 17 May 2021.

[40] Rodríguez-Sanz, Á. and Rubio Andrada, L. (2023). An empirical analysis of airport capacity evaluation: insights regarding air traffic design hours. *Journal of Physics: Conference Series*, 2526:012095.

[41] Rosenow, J., Michling, P., Schultz, M., and Schönberger, J. (2020). Evaluation of strategies to reduce the cost impacts of flight delays on total network costs. *Aerospace*, 7(11):165. This article belongs to the Collection Air Transportation—Operations and Management.

[42] Sanchez-Lengeling, B., Reif, E., Pearce, A., and Wiltschko, A. B. (2021). A gentle introduction to graph neural networks. *Distill*. https://distill.pub/2021/gnn-intro.

[43] SESAR Joint Undertaking (2020a). Blueprint digital european sky. Technical report, Publications Office of the European Union, Luxembourg. Available online at `https://www.schiphol.nl/en/aviation-solutions/page/deep-turnaround/`.

[44] SESAR Joint Undertaking (2020b). *Digitalising Europe's Aviation Infrastructure: MASTER PLAN Executive View 2020 Edition*. Publications Office of the European Union, Luxembourg. Catalogue number: MG-01-20-128-EN-C.

[45] Shone, R., Glazebrook, K., and Zografos, K. G. (2021). Applications of stochastic modeling in air traffic management: Methods, challenges and opportunities for solving air traffic problems under uncertainty. *European Journal of Operational Research*, 292:1–26.

[46] Sridhar, B., Chen, N. Y., and Ng, H. K. (2009). An aggregate sector flow model for air traffic demand forecasting. In *9th AIAA Aviation Technology, Integration, and Operations Conference (ATIO)*, pages AIAA 2009–7129, Hilton Head, South Carolina. American Institute of Aeronautics and Astronautics. NASA Ames Research Center, Moffett Field, CA 94035-1000 and University of California, Santa Cruz, Moffett Field, CA 94035-1000.

[47] Standfuß, T. and Whittome, M. (2020). How to cope with demand fluctuations - resilience as the solution for volatile traffic? In *Proceedings of the InterFAB Research Workshop on Single European Sky and Resilience in ATM*. TU Dresden, Institute of Logistics and Aviation, Chair of Air Transport Technology and Logistics and Functional Airspace Block Europe Central. This is a summarization and the full abstract contains more detailed insights on the resilience in European air transport.

[48] Sun, J., Dijkstra, T., Aristodemou, K., Buzeţelu, V., Falat, T., Hogenelst, T., Prins, N., and Slijper, B. (2022). Designing recurrent and graph neural networks to predict airport and air traffic network delays. In *10th International Conference for Research in Air Transportation*. FAA & Eurocontrol.

[49] Swana, F. and Doorsamy, W. (2021). An unsupervised learning approach to condition assessment on a wound-rotor induction generator. *Energies*, 14:602.

[50] Vos, R. (2023). Machine learning based trajectory prediction to support demand forecasting. Master's thesis.

[51] Wang, X., Wang, Z., Wan, L., and Tian, Y. (2022). Prediction of flight delays at beijing capital international airport based on ensemble methods. *Applied Sciences*, 12(10621). Received: 9 September 2022; Accepted: 17 October 2022; Published: 20 October 2022.

[52] Wei, X., Li, Y., Shang, R., Ruan, C., and Xing, J. (2023). Airport cluster delay prediction based on ts-bilstm-attention. *Aerospace*, 10:580.

[53] Wu, K. and Pan, W. (2008). A 4-d trajectory prediction model based on radar data. In *Proceedings of the 27th Chinese Control Conference*, Kunming, Yunnan, China. IEEE.

[54] Zhang, M., Zhou, X., et al. (2019). Propagation index on airport delays. *Transportation Research Record: Journal of the Transportation Research Board*, 2673(8).

[55] Zhou, H., Li, W., Jiang, Z., Cai, F., and Xue, Y. (2022). Flight departure time prediction based on deep learning. *Aerospace*, 9(394). Academic Editor: Alexei Sharpanskykh.

[56] Zou, J., Han, Y., and So, S.-S. (2009). *Overview of Artificial Neural Networks*, pages 14–22. Humana Press, Totowa, NJ.