

Phase-Decomposed Modelling of Knock-On Delay for Schiphol Demand Prediction

A Case Study at Schiphol Airport

A. Nari

Phase-Decomposed Modelling of Knock-On Delay for Schiphol Demand Prediction

A Case Study at Schiphol Airport

Thesis report

by

A. Nari

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on August 26, 2026 at 13:00

Thesis committee:

Chair:	Ir. O. Stroosma
Supervisors:	Dr.ir. J. Ellerbroek F. Dijkstra
External examiner:	Dr. J. Sun
Place:	Faculty of Aerospace Engineering, Delft
Project Duration:	September, 2025 - August, 2026
Student number:	5882788

An electronic version of this thesis is available at <https://repository.tudelft.nl>.

Cover image by Rob Vesseur on Unsplash.

Preface

With this thesis I close off my time as an Aerospace Engineering student at Delft University of Technology. The research was carried out at Luchtverkeersleiding Nederland, within the Knowledge and Development Centre. It was a demanding year and honestly a bit of a rollercoaster, but I learned a lot from it, both about the work and about myself.

I would like to thank my supervisors for their guidance throughout this project. Ferdinand Dijkstra, thank you for your support and guidance from start to finish. You were always available for me, whether I had a question or just needed to think out loud, and your involvement kept me going throughout the project. What stood out most to me was your passion for Air Traffic Management. Seeing how enthusiastic you are about the field was inspiring, and it motivated me along the way. Joost Ellerbroek, thank you as well for your guidance and supervision. Your critical and structured feedback consistently helped me raise the quality of my work, and your input was valuable in sharpening the direction of the research and keeping it on track.

I would also like to thank my fellow students and colleagues at iLabs. The support within the group meant a great deal to me throughout this period. Whether it was discussing a problem, sharing feedback, or simply talking things through, there was always someone willing to help. The open and collaborative atmosphere made the work more enjoyable, and I am grateful to have spent this time among such a motivated and welcoming group of people.

Finally, I want to thank my family and friends for their support and patience during the busy periods. Above all, I want to thank my parents, who have supported me at every step of my studies, often without me even having to ask. I am truly grateful for everything you have done for me.

This marks the end of my years at Delft University of Technology. Looking back, it has been a time in which I grew a lot, crossed paths with many great people, and found a real interest in aviation along the way. I am thankful for everything these years have brought me, and I am looking forward to my next challenge.

Amine Nari
Delft, August 2026

Contents

List of Figures	v
List of Tables	vi
Nomenclature	vii
1 General Introduction	1
 A Scientific Article	 2
I Introduction	3
II Background	4
II-A Operational Context	4
II-B Airport Collaborative Decision Making	5
II-C Delay Propagation Modelling	5
II-D Phase Level Prediction	6
III Methodology	7
III-A Data Preparation	8
III-B Phase-Decomposed Framework	8
III-C Phase-Level Prediction Models	9
III-D Sequential Prediction Procedure	11
III-E Demand Aggregation	12
IV Experimental Setup	12
IV-A Hypotheses	12
IV-B Assumptions	12
IV-C Dataset	13
IV-D LightGBM Models	13
IV-E Phase-Decomposed Propagation Model Evaluation	14
V Results	14
V-A LightGBM Training Results	15
V-B Phase-Decomposed Propagation Model	16
VI Discussion	19
VI-A Turnaround Delay Propagation	19
VI-B Airborne Delay Propagation	19
VI-C Horizon-Dependent Performance	19
VI-D Aggregated Demand Prediction	20
VI-E Future Work	20
VII Conclusion	20
 B Literature Review & Research Definition	 22
2 Air Traffic Flow and Capacity Management	23
2.1 ATFCM Phases	23
2.2 Airport Collaborative Decision Making	24
2.3 Decision Support Tool	26
3 Problem Statement	28

4	Operational and Data Concepts	29
4.1	Delay Causes	29
4.2	Flight Phases	30
4.3	Weather Data	31
4.4	ADS-B Data	31
5	State of the Art	32
5.1	Economic Models	32
5.2	Statistical Models	33
5.3	Machine Learning	36
5.4	Phase-Level Delay Prediction	40
5.5	Feature Analysis	43
6	Research Proposal	46
6.1	Research Gap	46
6.2	Research Objective and Questions	46
7	Project Plan	48
7.1	Project Overview	48
7.2	Methodology	48
7.3	Uncertainty	49
7.4	Horizon	49
7.5	Plan	49
	References	54
A	Overview of Model Input Features	55
B	SHAP Feature Importance Analysis	56
B.1	SHAP Feature Importance at Schiphol (EHAM)	56
B.2	SHAP Feature Importance at Outstations	58

List of Figures

1	Illustration of the rolling 20-minute demand aggregation.	4
2	A-CDM Milestones.	5
3	Effect of schedule reference choice on available turnaround time and departure delay. . . .	6
4	Overview of the five methodological steps of the proposed phase-decomposed propagation framework.	8
5	Data extraction flow.	8
6	Phase-decomposed framework for milestone prediction through an aircraft rotation.	10
7	SHAP feature importance for the Schiphol OBT model.	15
8	SHAP feature importance for the outstation OBT model.	15
9	SHAP feature importance for the LDT model.	16
10	Single rotation example of the phase-decomposed propagation procedure.	17
11	Prediction error distributions for framework TOT predictions across horizon intervals.	17
12	Prediction error distributions for framework LDT predictions across horizon intervals.	17
13	Aggregated rolling 20-minute Schiphol demand error distributions across horizon intervals. . . .	18
2.1	ATFM Phases [2, 3]	24
2.2	A-CDM Airports [4]	25
2.3	A-CDM Milestones	25
2.4	DST example	27
4.1	Distribution of the average delay per flight [5]	29
5.1	Bayesian Network structure [21]	35
5.2	Decision Tree example [27]	37
5.3	Artificial Neural Network [36]	39
5.4	Correlation heatmap of features for EGLL [40]	44
7.1	Overview of the project plan.	50

List of Tables

I	Hyperparameter search space used for LightGBM optimisation.	14
II	Selected LightGBM hyperparameters for the Schiphol and outstation OBT models.	15
III	Selected LightGBM hyperparameters for the LDT models.	15
IV	TOT error metrics for EHAM departures using the proposed framework and DST.	17
V	LDT error metrics for EHAM arrivals using the proposed framework and DST.	18
VI	Aggregated rolling 20-minute Schiphol demand error metrics for the proposed framework and DST.	18
5.1	Economic models: authors, objective, method, and application level.	33
5.2	Summary of classical statistical models for flight delay propagation.	34
5.3	Summary of BN models for flight delay propagation.	36
5.4	Performance of RF models for delay propagation prediction (Rebollo & Balakrishnan, 2014) [30].	37
5.5	Performance comparison of supervised learning models (Ye et al., 2020).	38
5.6	Feature overview	44

Nomenclature

A-CDM	Airport Collaborative Decision Making	DGLSTM	Directed Graph Long Short-Term Memory
AAT	Available Airborne Time	DMe	Expected Delay Multiplier
ACC-SUP	Area Control Centre Supervisors	DMx	Expected Propagated Delay Multiplier
ADP	ATFCM Daily Plan	DM	Delay Multiplier
ADS-B	Automatic Dependent Surveillance-Broadcast	DPT	Delay Propagation Tree
AIBT	Actual In-Block Time	DST-GAT	Dynamic spatial-temporal Graph Attention Network
ALDT	Actual Landing Time	DST	Decision Support Tool
ANN	Artificial Neural Network	DT	Decision Tree
ANSPs	Air Navigation Service Providers	ECAC	European Civil Aviation Conference
AOBT	Actual Off-Block Time	EGLL	London Heathrow Airport
APR-LSTM	Attention PageRank LSTM-Seq2Seq	EIBT	Estimated In-Block Time
ASAT	Actual Start-Up Approval Time	ELDT	Estimated Landing Time
ASPM	Aviation System Performance Metrics	EM	Expectation Maximization
ASRT	Actual Start-up Request Time	EOBT	Estimated Off-Block Time
ATC	Air Traffic Control	ETFMS	Enhanced Tactical Flow Management System
ATFCM	Air Traffic Flow and Capacity Management	ETOT	Estimated Take-Off Time
ATFM	Air Traffic Flow Management	EXIT	Estimated Taxi-In Time
ATL	Atlanta International Airport	EXOT	Estimated Taxi-Out Time
ATM	Air Traffic Management	FAA	Federal Aviation Administration
ATOT	Actual Take-Off Time	FIML	Full Information Maximum Likelihood
BN	Bayesian Network	FIR	Flight Information Region
CFD	Capacity Forecast DST	FMPCs	Flow Management Position Controllers
CFS	Capacity Forecast Schiphol	GBDT	Gradient Boosting Decision Trees
CODA	Central Office for Delay Analysis	GNNs	Graph Neural Networks
CTOT	Calculated Take-Off Time	IAF	Initial Approach Fix
DCB	Demand Capacity Balancing	IBT	In-Block Time
		JFK	John F. Kennedy International Airport
		LDT	Landing Time

LGA	LaGuardia Airport	PIBT	Predicted In-Block Time
LightGBM	Light Gradient Boosting Machine	PLDT	Predicted Landing Time
LSTM	Long Short-Term Memory	POBT	Predicted Off-Block Time
MAE	Mean Absolute Error	PTOT	Predicted Take-Off Time
MLIL-ANN	Multi-Level Input Layer Artificial Neural Network	RF	Random Forest
MLR	Multiple Linear Regression	RMSE	Root Mean Squared Error
ML	Machine Learning	RNN	Recurrent Neural Network
MSE	Mean Squared Error	SHAP	SHapley Additive exPlanations
MSTAGCN	Multi-Scale Temporal Adaptive Graph Convolution Network	SIBT	Scheduled In-Block Time
MTT	Minimum Turnaround Time	SLDT	Scheduled Landing Time
NAS	National Airspace System	SOBT	Scheduled Off-Block Time
NMOC	Network Manager Operations Centre	STOT	Scheduled Take-Off Time
NM	Network Manager	STPN	Spatio-Temporal Propagation Network
NOAA	National Oceanic and Atmospheric Administration	STT	Scheduled Turnaround Time
NOP	Network Operations Plan	SVM	Support Vector Machine
OBT	Off-Block Time	TMA	Terminal Manoeuvring Area
OLS	Ordinary Least Squares	TOBT	Target Off-Block Time
ORD	O'Hare International Airport	TOT	Take-Off Time
		TSAT	Target Start-Up Approval Time

General Introduction

Air traffic in Europe continues to grow while airport and airspace capacity remain constrained, placing increasing pressure on the efficiency and predictability of air traffic operations. At major hub airports such as Schiphol, a central task is to keep the expected traffic demand compatible with the available capacity, a process known as Demand Capacity Balancing (DCB). This balance is strongly determined by runway capacity, since both inbound and outbound flows depend on the take-off and landing capacity of the active runway configuration. Luchtverkeersleiding Nederland (LVNL) supports this process through a Decision Support Tool (DST) that provides supervisors and flow controllers with a view of the expected future demand and the available capacity. The value of this tool depends directly on how accurately future traffic demand is predicted, particularly three to four hours ahead of operation, when there is still time to prepare regulations and plan runway use for upcoming inbound and outbound peaks.

A major source of uncertainty in this demand prediction is knock-on delay, also referred to as reactionary delay, which occurs when a delay on one flight propagates to later flights. Because an aircraft must first complete one flight leg before it can operate the next, disruptions accumulate over the course of the day, and delayed flights can arrive in concentrated waves that were not visible in the original schedule. These waves create demand spikes that are difficult to anticipate, yet the current DST does not explicitly model how delay propagates through aircraft rotations, which limits its ability to foresee them.

Delay propagation has been studied extensively, but existing approaches generally model delay at the level of complete flights, aircraft rotations, or airport networks, treating each flight as a single unit that departs late and arrives late. This conceals where delay is actually generated, absorbed, or preserved within a rotation, because the operational phases that connect two flight legs, namely the turnaround, taxi-out, airborne, and taxi-in phases, behave very differently from one another. Modelling these phases as a single aggregated quantity therefore provides limited insight into how knock-on delay evolves, and how it ultimately affects the traffic demand at Schiphol.

This study develops a phase-decomposed model of knock-on delay that represents how delay evolves through the successive operational phases and flight legs of an aircraft rotation, and evaluates its contribution to Schiphol demand prediction. Each rotation is represented as a sequence of turnaround, taxi-out, airborne, and taxi-in phases, and the predicted aircraft state is propagated from one phase to the next so that rotational delay can be traced as it is transferred, preserved, or partly absorbed before affecting subsequent flights. The framework is applied to operations at Schiphol, and its milestone predictions are aggregated to assess whether modelling knock-on delay leads to more accurate rolling Schiphol demand predictions.

The thesis is structured in two parts. Part A contains the scientific article, which forms the final research contribution and presents the methodology, results, and conclusions of the study. Part B contains the literature review and research definition from the initial phase of the thesis. It provides the operational, technical, and methodological background on which the article is based, and documents the original research questions and project plan.

Part A

Scientific Article

Phase-Decomposed Modelling of Knock-On Delay for Schiphol Demand Prediction

A. Nari

Supervised by: Dr.ir. J. Ellerbroek, Delft University of Technology,
Dhr. F. Dijkstra, Knowledge & Development Centre Mainport Schiphol

Control & Operations, Faculty of Aerospace Engineering
Delft University of Technology, Delft, the Netherlands

Abstract—Knock-on delay is an important source of uncertainty in air traffic demand prediction. A delay on one flight can propagate to subsequent flights operated by the same aircraft, shifting the timing of later departures and arrivals. This paper develops a phase-decomposed propagation model to predict how knock-on delay evolves through aircraft rotations and how this affects the demand profile for Schiphol traffic. The model represents each aircraft rotation as a sequence of turnaround, taxi-out, airborne, and taxi-in phases. Off-block time and landing time are predicted using LightGBM models, while taxi times are estimated using lookup tables. Predicted milestones are passed from one phase to the next, so that the predicted state of the aircraft remains consistent throughout the remainder of the rotation. The model is trained on July 2025 data and evaluated on August 2025 for KLM, KLM Cityhopper, and Transavia flights. The results show that the proposed framework improves Schiphol take-off time prediction across all evaluated horizons. At the 180–240 minute horizon, the take-off time Mean Absolute Error (MAE) decreases from 16.16 to 13.62 minutes and the Root Mean Squared Error (RMSE) from 23.95 to 21.22 minutes, corresponding to relative reductions of approximately 16% and 11%, respectively. Schiphol landing time prediction also improves, with a lower MAE across all evaluated horizons and improvements in all error metrics from the 15–30 minute horizon onwards. At the 180–240 minute horizon, the landing time MAE decreases from 13.92 to 11.33 minutes and the RMSE from 21.69 to 18.60 minutes, corresponding to relative reductions of approximately 19% and 14%, respectively. These milestone improvements translate into more accurate Schiphol demand prediction across all evaluated horizons. At the 180–240 minute horizon, the rolling demand MAE decreases from 2.97 to 2.53 aircraft and the RMSE from 3.93 to 3.34 aircraft, corresponding to relative reductions of approximately 15% in both metrics. Overall, the results show that modelling rotational knock-on delay improves the prediction of future take-off and landing events and leads to more accurate rolling Schiphol demand prediction.

Index Terms—Knock-on delay, Reactionary delay, Demand prediction, Schiphol Airport

I. INTRODUCTION

Air traffic operations in Europe continue to face increasing pressure as traffic demand grows while both airport and airspace capacity remain constrained [1]. Flight delays reduce the predictability of traffic demand, making it more difficult to anticipate capacity needs and thereby increasing the likelihood of demand–capacity imbalances [1]. Although primary delays originate from factors such as adverse weather, Air Traffic

Flow and Capacity Management (ATFCM) regulations, or airport capacity constraints, the dominant source of disruption in the European network is knock-on delay, also referred to as reactionary delay [2]. Knock-on delay occurs when a delay on one flight propagates to subsequent flights through operational dependencies, such as aircraft rotations, crew connections, or passenger transfers. This paper focuses on rotational knock-on delay, where the same aircraft must complete one flight leg before it can operate the next. According to the EUROCONTROL Central Office for Delay Analysis (CODA) [2], reactionary delay accounted for 46% of all delay minutes in Europe in 2024, translating to 8.0 minutes per flight, making it the largest delay source.

Luchtverkeersleiding Nederland (LVNL), responsible for Air Traffic Control in the Dutch Flight Information Region (FIR), operates the Decision Support Tool (DST) that provides Area Control Center supervisors and Flow Management Position Controllers (FMPCs) with insight into expected traffic demand and available capacity in Dutch airspace. The DST uses simulation and a machine learning method to balance capacity and demand. However, the system does not model knock-on delay. This limits the ability of controllers to anticipate demand spikes that emerge when clusters of delayed flights enter the FIR simultaneously.

A substantial body of research has examined how delay propagates through aircraft rotations using analytical models that identify the mechanisms behind delay propagation [3], [4]. While these models provide valuable insight into the underlying propagation mechanisms, they are not designed to predict future delay evolution. Machine learning models have therefore been developed to fill this gap [5], [6]. However, existing work does not analyse how delay evolves through each individual phase of a rotation, namely the turnaround, taxi-out, airborne, and taxi-in. As a result, it provides limited insight into where delay is generated, absorbed, or preserved within an aircraft rotation, making it more difficult to understand and predict how knock-on delay propagates over time.

The main contribution of this paper is a phase-decomposed knock-on delay model for aircraft rotations at Schiphol Airport, applied to operations of KLM, KLM Cityhopper, and Transavia. The initial filed flight plan serves as the schedule reference for all delay calculations. The model combines

trained prediction models with lookup-table estimates to represent the operational phases of an aircraft rotation. The off-block and landing time models use feature sets that reflect the dynamics of the turnaround and airborne phases, while the taxi-out and taxi-in phases are estimated using lookup tables. The feature design also accounts for the different availability of Airport Collaborative Decision Making (A-CDM) milestone data at Schiphol and at outstations. For each observation time, the predicted take-off and landing times across all active rotations are aggregated to estimate Schiphol arrival and departure demand for LVNL's demand-capacity balancing process.

This paper is structured as follows. Section II provides the background on delay propagation and phase-level modelling. The phase-decomposed methodology is then introduced in Section III, after which Section IV describes the experimental setup. Section V presents the results, which are discussed further in Section VI. Finally, Section VII concludes the paper.

II. BACKGROUND

This section provides the technical and operational background required to understand the phase-decomposed propagation model proposed in this paper. It first introduces the operational context of demand prediction in the Dutch Flight Information Region (FIR). It then describes Airport Collaborative Decision Making and the milestone data used to define the operational phases. Finally, it reviews the literature on delay propagation through aircraft rotations and the state of the art in phase-level delay prediction.

A. Operational Context

Luchtverkeersleiding Nederland (LVNL) is the Air Navigation Service Provider (ANSP) responsible for air traffic control in the Dutch FIR. A central task within this operational environment is Demand Capacity Balancing (DCB), which aims to ensure that expected traffic demand remains compatible with the available airspace and airport capacity. In the Schiphol context, this balance is strongly determined by runway capacity, since both inbound and outbound flows depend on the landing and take-off capacity available for the active runway configuration [7].

Within the European Air Traffic Flow and Capacity Management (ATFCM) network, Air Navigation Service Provider (ANSPs) monitor expected traffic demand and available capacity within their areas of responsibility. When a demand-capacity imbalance is expected in the Dutch FIR, LVNL determines whether an Air Traffic Flow Management (ATFM) regulation is required and coordinates the measure with the Network Manager Operations Centre (NMOC). The NMOC assesses the network-wide impact and implements the regulation in the Network Manager system. Based on the regulation, the Network Manager assigns individual Calculated Take-Off Times (CTOTs) to the affected flights. These flights are required to depart within a tolerance window of five minutes before to ten minutes after the assigned CTOT.

LVNL's Decision Support Tool (DST) supports DCB by providing Flow Management Position Controllers (FMPCs) with insight into future traffic demand and available capacity. In this operational context, traffic demand is evaluated in five-minute time bins using a rolling 20-minute window. For each time bin, the demand represents the number of flights expected within the 20-minute window starting at the beginning of that bin, as illustrated in Figure 1. This gives FMPCs a time-dependent demand profile rather than a set of isolated flight times, making it easier to assess whether the expected traffic load remains compatible with the available capacity. The quality of DCB therefore depends directly on the accuracy of the predicted take-off and landing times from which these demand profiles are constructed. This is especially important three to four hours ahead of operation, because this planning window still leaves time to prepare regulations and plan runway use for upcoming inbound and outbound peaks.

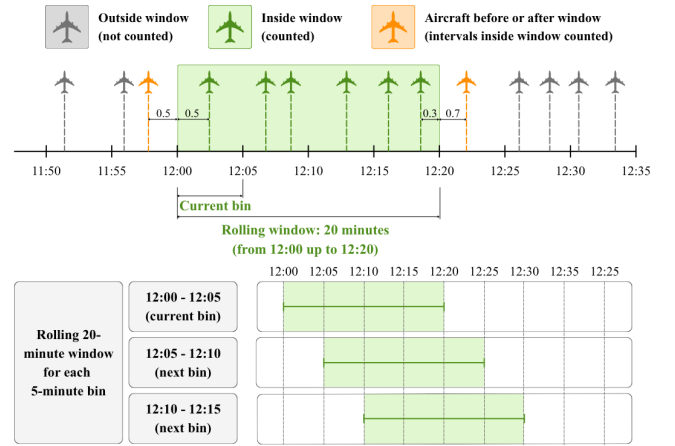


Fig. 1. Illustration of the rolling 20-minute demand aggregation. For the current bin from 12:00 to 12:05, demand is calculated over the 20-minute window from 12:00 to 12:20. Flights inside the window are counted fully, while flights crossing the window boundaries are counted fractionally. In this example, the resulting demand for the current bin is 6.8 aircraft.

If predicted take-off or landing times are shifted by operational delays, the expected demand profile changes. This is particularly relevant for knock-on delay, also referred to as reactionary delay, where a delay on one flight propagates to later flights operated by the same aircraft [2]. When multiple aircraft rotations are affected by earlier disruptions, flights may arrive in the Amsterdam FIR in concentrated waves that were not visible in the original schedule. Although the current DST supports DCB using predicted traffic demand, it does not explicitly model this delay propagation mechanism. This motivates the phase-decomposed propagation model developed in this paper. By representing the operational phases separately, the model can explicitly propagate the predicted aircraft state from one phase to the next. This makes it possible to identify where knock-on delay is absorbed or preserved before it affects subsequent flights and the resulting traffic demand.

B. Airport Collaborative Decision Making

Airport Collaborative Decision Making (A-CDM) has been developed to address challenges that arise from ineffective data and information exchange between different stakeholders at the airport. The aim is to increase the efficiency of airport operations and improve the performance of the overall Air Traffic Management (ATM) network by reducing delays, improving predictability, and optimising the use of resources. A-CDM addresses this by establishing a common milestone process that tracks every flight through its turnaround using a shared set of timestamps, as shown in Figure 2 [8], [9].

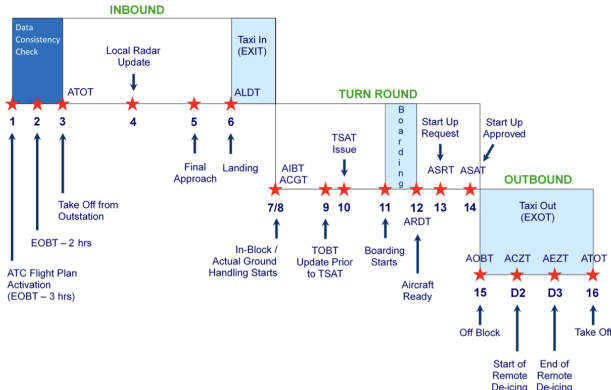


Fig. 2. A-CDM Milestones [10]

It defines sixteen standardised timestamps that describe the progress of an aircraft from scheduled arrival to actual departure. These milestones define the boundaries of the four operational phases considered in this study. Taxi-in is defined from Actual Landing Time (ALDT) to Actual In-Block Time (AIBT), turnaround from AIBT to Actual Off-Block Time (AOBT), taxi-out from AOBT to Actual Take-Off Time (ATOT), and the airborne phase from ATOT to the subsequent ALDT.

The availability of these detailed milestone timestamps provides data for modelling delay propagation at the phase level. Okwir et al. [11] analysed the operational A-CDM data from Madrid Airport and found that A-CDM milestone variables substantially improve the prediction accuracy of the turnaround delay models compared to the models trained without A-CDM data. The SHapley Additive exPlanations (SHAP) analysis, which explains machine learning predictions by measuring the contribution of each input feature to the final prediction, identified the difference between the Actual Start-Up Readiness Time (ASRT) and the assigned Target Start-Up Approval Time (TSAT) slot as the most important predictor in nearly all tested models. This mainly concerns short horizon prediction, because TSAT is a tactical A-CDM variable that is typically calculated shortly before off-block time. For longer horizons, such as three to four hours before departure, TSAT is often not yet available. Nevertheless, this finding is relevant for this study, because the phase-decomposed propagation

model uses a sequential prediction procedure. Short horizon predictions can become inputs for later phases and flight legs, so accurate milestone prediction close to departure can still improve the aircraft state that is propagated through the rest of the rotation. This sequential logic is described in more detail in Section III-D. More generally, the result shows that structured A-CDM information contains useful operational data that are not available in the same standardised form at non A-CDM airports [8], [9].

C. Delay Propagation Modelling

The literature on flight delay propagation modelling can be organised into two perspectives. The first is the flight chain-based perspective, which models how delay propagates sequentially through the legs of a single aircraft rotation, treating each flight and its connecting turnaround as a link in a chain. The second is the airport network-based perspective, which models how delay spreads across airports through the connectivity of routes, treating airports as nodes and delays as signals that propagate through the network [12]. These two perspectives address different levels of the problem: flight chain models capture the operator mechanisms driven by aircraft rotation, crew scheduling, and schedule buffers, while airport network models capture the inter-airport dynamics driven by traffic flow and capacity constraints [6]. This paper adopts the flight chain perspective, as the goal is to trace how an initial delay on a specific aircraft evolves through each successive phase and flight leg of its rotation, ultimately affecting the time at which that aircraft takes off from or lands at Schiphol.

Within the flight chain perspective, one of the earliest approaches was proposed by Beatty et al. [3], who introduced the delay multiplier concept to quantify how an initial disruption propagates through a schedule. AhmadBeygi et al. [13] extended this work using propagation trees on two US carriers, showing that schedules optimised for high resource utilisation leave little slack to absorb disruptions. Kafle and Zou [4] further advanced this line of research by analytically separating propagated delay from newly formed delay within a daily rotation using a Heckman two-step econometric model. Brueckner et al. [14] demonstrated, through a two-flight model, that only the ground buffer of the outbound flight determines whether delay propagates, while the flight buffer of the preceding flight is inconsequential.

These analytical approaches establish the mechanisms through which delay propagates through a flight chain, but they do not provide predictive models of delay propagation. In recent years, machine learning methods have increasingly been applied to fill this gap, demonstrating strong predictive performance because of their ability to capture non-linear relationships and process large operational datasets [6], [12]. Within the flight chain perspective, Dalmau et al. [15] trained a gradient boosted decision tree on EUROCONTROL Enhanced tactical flow management system (ETFMS) data to predict individual take-off time deviations, finding that the Available Turnaround Time (ATT) is the dominant factor.

Building directly on this finding, Dalmau [5] subsequently extended the approach into a probabilistic CatBoost model for rotational reactionary delay, treating the entire rotation as a single prediction unit rather than predicting individual flights in isolation.

Beyond the flight chain perspective, network-level models have been developed to capture how delay spreads across airports. Long Short-Term Memory (LSTM) models exploit temporal dependencies in airport delay sequences, while spatiotemporal graph neural networks represent airports as nodes and routes as edges to model network-wide propagation [12]. Although architectures such as Directed Graph Long Short-Term Memory (DGLSTM) and AG2S-Net have demonstrated strong performance at this scale, they are not designed to trace the mechanisms within an aircraft rotation that drive reactionary delay at the level of an individual flight. This paper therefore focuses on the flight chain perspective.

Across both perspectives, Li et al. [12] identify airport characteristics, temporal features, and weather conditions as the three most frequently used feature categories. Temporal features include the month, day of the week, and time of day. Together, these categories appear in nearly 70% of the reviewed papers and directly inform the feature design adopted in this study. A further methodological consideration specific to flight chain modelling concerns the choice of schedule reference. When a flight is significantly delayed, operators routinely update the filed departure time in the flight plan. If this updated plan is used as the reference, the measured departure delay is reduced and the available turnaround time appears artificially long, because both quantities are computed relative to a reference that has already absorbed the disruption [5]. As illustrated in Figure 3, if an inbound flight arrives 45 minutes late and the operator revises the outbound departure from 10:00 to 10:45, computing against the updated plan yields an available turnaround time of 60 minutes and a departure delay of zero, whereas computing against the initial plan correctly yields 15 minutes and 45 minutes respectively [5]. This research therefore adopts the initial flight plan as the schedule reference.

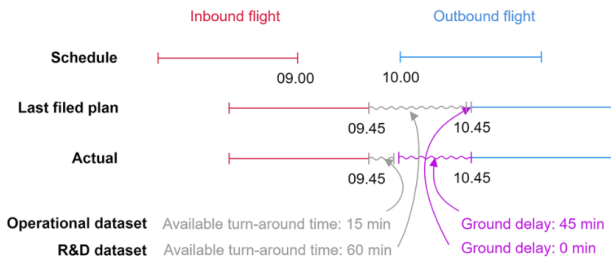


Fig. 3. Effect of schedule reference choice on available turnaround time and departure delay [5].

D. Phase Level Prediction

The following subsections review the state of the art in predicting the duration of each of the four operational phases

that form an aircraft rotation.

1) *Turnaround*: The turnaround is the central mechanism through which delay propagates within an aircraft rotation [4], [14]. It covers all ground handling operations between an aircraft arriving at the gate and pushing back for the next departure, including passenger de-boarding and boarding, baggage handling, cabin cleaning, catering, fuelling, and technical checks [5], [16]. Early work using stochastic modelling of the turnaround process showed that appropriate ground buffer selection can absorb arrival delays while maintaining efficient fleet utilisation [16]. Analysis using field data from German airlines further confirmed that the turnaround follows a critical path structure in which a delay to any single sub-process can extend the total ground time [17].

A key property of the turnaround phase is that its dynamics change depending on whether the inbound flight arrives on time or is delayed. When a delay is present, the Available Turnaround Time (ATT), defined as the difference between the Scheduled Off-Block Time (SOBT) and the Actual In-Block Time (AIBT), shrinks or may even turn negative, reducing the window for ground handling. Knock-on delay arises when this available turnaround time falls below the Minimum Turnaround Time (MTT), that is the shortest feasible ground time for a given aircraft type, airline, and airport [5]. This behaviour motivates treating inbound delay as an explicit input to the turnaround model.

Machine learning models have proven well suited to capturing non-linear relationships without requiring data on individual sub-processes such as catering or boarding. Probabilistic XGBoost models trained to predict turnaround time and Target Off-Block Time (TOBT) at four European Airport Collaborative Decision Making (A-CDM) airports achieved mean absolute errors of four to six minutes in the tactical phase and seven to nine minutes in the pre-tactical phase, representing improvements of 44% to 56% over a scheduled turnaround baseline [18]. The SHapley Additive exPlanations (SHAP) analysis identified available turnaround time as the dominant predictor, followed by Scheduled Turnaround Time (STT), origin distance, aircraft weight, and airline operator, while weather and passenger count contributed negligibly. Notably, these models were trained on nominal operations and explicitly excluded flights with large inbound delays, leaving the behaviour of the turnaround phase under disrupted conditions unaddressed [18]. A subsequent network-wide analysis confirmed that ATT is equally the dominant predictor of departure delay across all operational conditions, and showed that gradient boosted tree models learn the non-linear threshold behaviour around the MTT without being explicitly trained on it [5]. These studies show that machine learning models, in particular gradient boosted tree models, are well suited for modelling turnaround dynamics. Within this context, the turnaround can be interpreted as a mapping from inbound delay to outbound delay, capturing the mechanism through which knock-on delay is generated. This mapping forms a key component of the phase-decomposed model proposed in this paper.

2) *Taxi-Out and Taxi-In*: The taxi-out phase begins at pushback and ends at takeoff, while the taxi-in phase begins at touchdown and ends at gate arrival. Both phases consist of two structural components, an unimpeded component reflecting the minimum time to traverse the route between stand and runway, and a variable component reflecting surface congestion at the time of operation [19]. An empirical analysis at two Korean airports confirmed that both taxi phases are sensitive to weather conditions, with wind speed and direction, take-off and landing direction, and terminal location identified as significant predictors alongside traffic volume [20]. The active runway configuration in particular has a large effect on taxi time. At Schiphol, for example, a flight assigned to runway 18R will have a considerably longer taxi time than a flight assigned to a runway closer to the terminal [21]. This means that even a delayed flight may see part of its delay absorbed if it is assigned to a nearby runway, or accumulate additional time if routed to a distant one. Across both phases, the stand, active runway configuration, time of day, and a measure of surface congestion emerge as the dominant features [19], [20], [22].

Unlike the turnaround phase, there is no evidence in the literature that airlines schedule explicit buffers in either taxi phase to absorb rotational delay. Taxi duration is primarily governed by airport surface conditions at the time of operation. At Schiphol specifically, replacing the current static look-up table with a data-driven predictor increased the fraction of flights predicted within a two-minute error margin from 64% to 68%, a small improvement that suggests the look-up table already captures most of the variation in taxi times [21]. This study therefore adopts a look-up table approach for both taxi phases.

3) *Airborne*: The airborne phase spans from takeoff to touchdown. Operators incorporate a flight buffer in the scheduled block time, defined as the amount by which scheduled flight time exceeds the minimum feasible flight duration, to provide capacity for delay recovery [14]. Flight buffers are smaller and less variable than ground buffers and increase with route distance [4].

The extent to which departure delay is recovered during flight is not uniform. Applied to a dataset of 5.7 million flights crossing European airspace, a trajectory-based analysis found the system to be globally resilient in its en-route phase, able to recover as much delay as it generates on average [23]. However, this resilience is strongly heterogeneous across time, space, and flight phase, with recovery highest during the arrival phase and lowest near departure [23]. Delay recovery is driven by opportunity rather than systematic control, triggered by favourable conditions such as the availability of alternative routes or low traffic density, and is not the result of a deliberate strategy to reduce flight length [23]. Empirical regressions show that the relationship between departure delay and arrival delay is close to one-for-one, and that flying faster provides very little scope to overcome a departure delay, with an extra minute of departure delay reducing airborne time by only between 0.049 and 0.070 minutes [14].

Machine learning models can predict individual flight arrival times with a Mean Absolute Error (MAE) of around five minutes. Using real-time trajectory data at multiple prediction points, machine learning models achieve a MAE of less than 6 minutes after departure and less than 3 minutes after Terminal Manoeuvring Area (TMA) entrance [24]. Even without trajectory data, a random forest trained on schedule-level features including departure delay, departure airport, filed off-block time, filed arrival time, aircraft operator, and airport capacity achieves a MAE of 4.5 minutes and a Root Mean Squared Error (RMSE) of 6.3 minutes across the top European airports [25]. This indicates that departure delay and schedule characteristics form a sufficient feature set to predict arrival time before the flight departs. The airborne model in this study operates under the same constraint, as it must produce a predicted arrival time prior to departure to serve as input to the turnaround model of the next rotation leg. The next section builds on this background by introducing how these operational phases are represented and linked within the proposed phase-decomposed propagation model.

III. METHODOLOGY

This section presents the methodology used to develop the proposed phase-decomposed propagation framework. The framework models knock-on delay by predicting each operational phase of an aircraft rotation sequentially, allowing the predicted aircraft state to be propagated consistently from one flight leg to the next before translating the resulting milestones into Schiphol demand predictions. Figure 4 provides an overview of the five methodological steps, which are described in detail in the following subsections.

The methodology consists of five main steps. First, the operational data are prepared by reconstructing operational milestones, enriching the flight records with weather information, and linking consecutive flights into aircraft rotations. Second, the aircraft rotation is represented as a sequence of operational milestones, allowing modelled delay to be propagated from one phase to the next. Third, phase-level prediction models are defined for the turnaround, taxi-out, airborne, and taxi-in phases. The off-block and landing time milestones are predicted using LightGBM models, while taxi-out and taxi-in durations are estimated using lookup tables. Fourth, these models are applied sequentially from a given observation time, using realised milestones where available and predicted milestones for future events. Finally, the predicted take-off and landing times are aggregated into demand profiles.

The purpose of this methodology is not only to predict individual milestone times, but also to maintain consistency along the aircraft rotation. This is important because a delay on one flight can influence the departure and arrival times of later flights operated by the same aircraft. By explicitly passing predicted milestone times from one phase to the next, the framework captures the mechanism through which knock-on delay changes the temporal distribution of traffic demand.

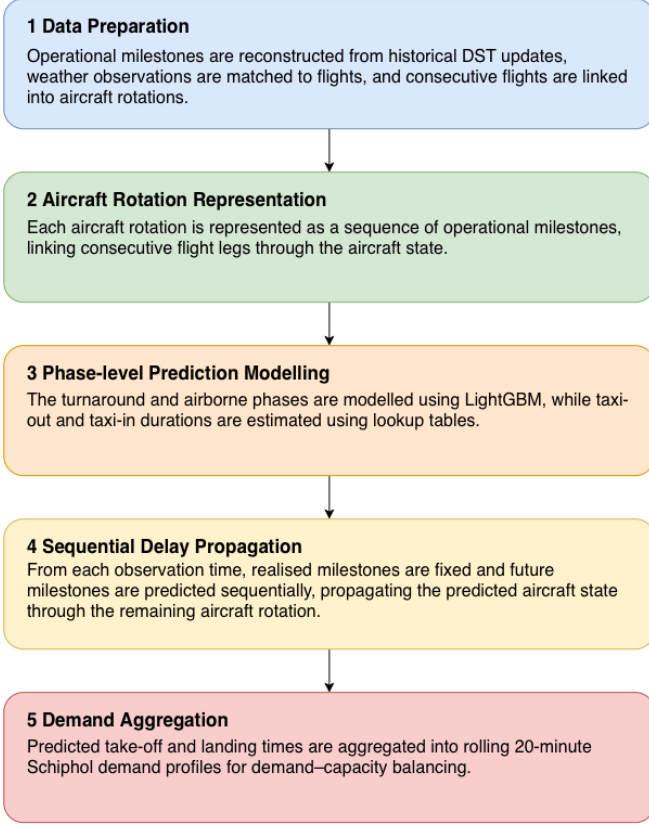


Fig. 4. Overview of the five methodological steps of the proposed phase-decomposed propagation framework.

A. Data Preparation

This study combines multiple operational data sources to construct a flight dataset suitable for modelling knock-on delay propagation within aircraft rotations. The data extraction process is illustrated in Figure 5. The primary data source is the Decision Support Tool (DST) database provided by LVNL. This dataset contains flight plan information, operational milestones, aircraft characteristics, and Air Traffic Flow Management (ATFM)/Airport Collaborative Decision Making (A-CDM)-related information.

A key challenge is that the DST dataset does not provide realised take-off and landing times. Therefore, the operational milestones Actual Take-Off Time (ATOT) and Actual Landing Time (ALDT) are reconstructed from the historical sequence of DST updates. Throughout a flight, the DST system continuously predicts the expected departure and arrival times. Once a predicted event time is earlier than the corresponding message timestamp, the event is assumed to have already occurred. Validation against independent radar data demonstrated that the reconstructed ATOT and ALDT values closely matched the observed operational milestones, indicating that the reconstruction method provides an accurate representation of the realised flight trajectory.

For outstation flights, taxi-in times are estimated using an airport level lookup table because in-block times are not

available in the DST dataset. The lookup table is constructed from observed taxi-in times obtained from Flightradar24. As only a limited amount of Flightradar24 data was available for this study, these observations are used solely to derive representative taxi-in times per airport.

Meteorological information is obtained from historical METAR reports retrieved via Ogimet’s METAR/SPECI/TAF query service [26]. METAR data are collected for Schiphol Airport and the 50 airports with the highest number of flights to and from Schiphol in the dataset. These airports cover approximately 70% of the flights. The reports are decoded into structured weather observations containing variables such as wind direction, wind speed, visibility, temperature, cloud coverage, and wind gusts. For each flight, the most recent weather observation available prior to the corresponding DST message timestamp is assigned to the flight record. This ensures that each prediction is based on information that would have been available at the time of prediction.

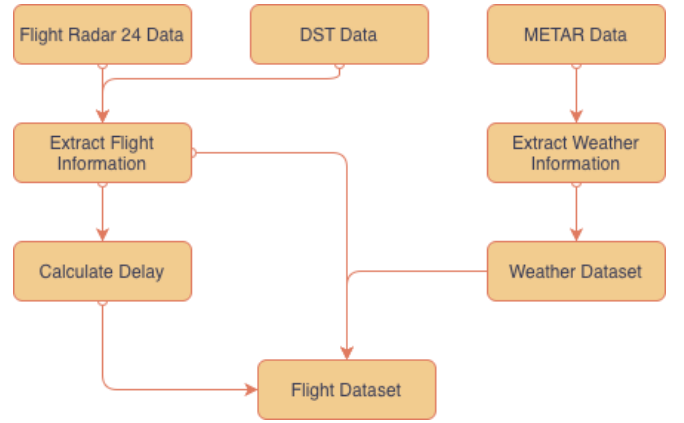


Fig. 5. Data extraction flow

B. Phase-Decomposed Framework

The framework follows each aircraft through the sequence of flight legs it operates within the considered prediction day. This sequence is referred to as an aircraft rotation. For each aircraft registration, flights are ordered chronologically and indexed by $i = 1, 2, \dots, N$, where N denotes the number of flight legs in the reconstructed rotation.

Each flight leg is described by four operational milestones: Off-Block Time (OBT), Take-Off Time (TOT), Landing Time (LDT), and the In-Block Time (IBT). Throughout this paper, the prefixes S , A , and P are used to distinguish between scheduled, actual, and predicted milestone times, respectively. For example, the operational order of the realised milestones for flight i is written as:

$$AOBT_i \rightarrow ATOT_i \rightarrow ALDT_i \rightarrow AIBT_i. \quad (1)$$

The connection between two consecutive flight legs is established through the turnaround. Once flight i has arrived and taxied in, the aircraft becomes available for flight $i + 1$.

Therefore, the IBT of flight i influences the OBT of the next flight:

$$\text{IBT}_i \rightarrow \text{OBT}_{i+1}. \quad (2)$$

This connection is the main mechanism through which knock-on delay is propagated in the framework. Delay is defined relative to the scheduled milestone time. For a given flight i and milestone m , the realised delay is defined as:

$$D_i^m = A_i^m - S_i^m, \quad (3)$$

where A_i^m is the realised milestone time and S_i^m is the corresponding scheduled milestone time. Similarly, the predicted delay is defined as:

$$\hat{D}_i^m(t_o) = P_i^m(t_o) - S_i^m, \quad (4)$$

where $P_i^m(t_o)$ denotes the predicted milestone time at observation time t_o . A positive delay indicates that the milestone occurs later than scheduled, while a negative delay indicates that the milestone occurs earlier than scheduled.

A key aspect of the framework is that predictions are made from a specific observation time t_o . At each observation time, only the information available up to that moment is used. This includes the latest available DST message, the most recent weather information, and the predicted or realised state of earlier milestones in the same aircraft rotation. As a result, the feature vector for a flight depends not only on the flight itself, but also on the observation time and on the upstream state of the aircraft rotation.

The complete framework is illustrated in Figure 6. The figure shows how the predicted state is updated step by step and then passed to the next flight leg until the final flight in the aircraft rotation is reached.

C. Phase-Level Prediction Models

The phase-decomposed propagation model consists of four phase-level components, corresponding to the milestone transitions introduced in the previous subsection. The Off-Block Time (OBT) and Landing Time (LDT) models are implemented as supervised machine learning models, while the taxi-out and taxi-in times are estimated using lookup tables.

LightGBM is selected for the OBT and LDT models because it is a gradient boosting tree method that can model non-linear relationships between operational features and delay. This is relevant for delay propagation, where the effect of variables such as available turnaround time, schedule buffer, and departure delay is not expected to be linear. Previous delay prediction studies have also shown that tree-based gradient boosting models are suitable for turnaround and flight delay prediction tasks [5], [15], [18]. Each predicted milestone is passed to the next component as input, allowing the full operational sequence of a flight leg to be predicted from OBT to In-Block Time (IBT).

The actual milestone times used to train the models, and those which the models are evaluated against, do not correspond to the final realised milestone times. A realised

milestone may incorporate an Air Traffic Flow Management (ATFM) delay that is imposed by Luchtverkeersleiding Nederland (LVNL) through the Demand Capacity Balancing (DCB) process, for instance by holding a flight on the ground under an assigned Calculated Take-Off Time (CTOT). Training and evaluating the models on these realised times would therefore include the effect of the regulatory intervention into the learned demand, rather than representing the underlying demand to which the intervention responds. This would yield a distorted representation of the demand that LVNL seeks to anticipate.

To avoid this, the models are trained on the realised milestone time with the assigned ATFM delay subtracted, and evaluated on this corrected milestone time with the ATFM delay known at the observation time added back. Let ATFM_i denote the ATFM delay assigned to flight i , and $\text{ATFM}_i(t_o)$ the portion of this delay known at observation time t_o .

1) *Off-Block Time Prediction:* The OBT model represents the turnaround between two consecutive flight legs. For each turnaround, the model combines information from the inbound flight $i - 1$ and the outbound flight i to describe the aircraft state before departure. Separate models are used for Schiphol and outstation turnarounds because more detailed operational data are available for Schiphol. The learning target is the off-block delay, defined as the difference between the Actual Off-Block Time (AOBT) and the Scheduled Off-Block Time (SOBT):

$$D_i^{\text{OBT}} = (\text{AOBT}_i - \text{ATFM}_i) - \text{SOBT}_i. \quad (5)$$

During prediction, the model estimates the off-block delay using the feature vector available at observation time t_o :

$$\hat{D}_i^{\text{OBT}}(t_o) = f_{\text{OBT}}(X_i^{\text{OBT}}(t_o)), \quad (6)$$

where $\hat{D}_i^{\text{OBT}}(t_o)$ is the predicted off-block delay for outbound flight i , f_{OBT} denotes the trained OBT model, and $X_i^{\text{OBT}}(t_o)$ is the feature vector constructed for the turnaround leading to flight i .

The turnaround state is constructed from the predicted or realised milestones available at observation time t_o . To obtain the in-block time, the estimated taxi-in duration is added to the landing time of the inbound flight.

This in-block time is then used to calculate the Available Turnaround Time (ATT), defined as the time between the inbound in-block time and the SOBT of the outbound flight. Scheduled Turnaround Time (STT) is calculated by taking the difference from the scheduled in-block time of the inbound flight and the SOBT of the outbound flight. For both Schiphol and outstation turnarounds, the feature set also includes the ratio between Available Turnaround Time (ATT) and Scheduled Turnaround Time (STT), together with the available buffer, defined as the difference between STT and ATT.

The representation of the minimum turnaround requirement differs between Schiphol and the outstations. At Schiphol, realised in-block times were available. This made it possible to derive an airport-specific Minimum Turnaround Time (MTT)

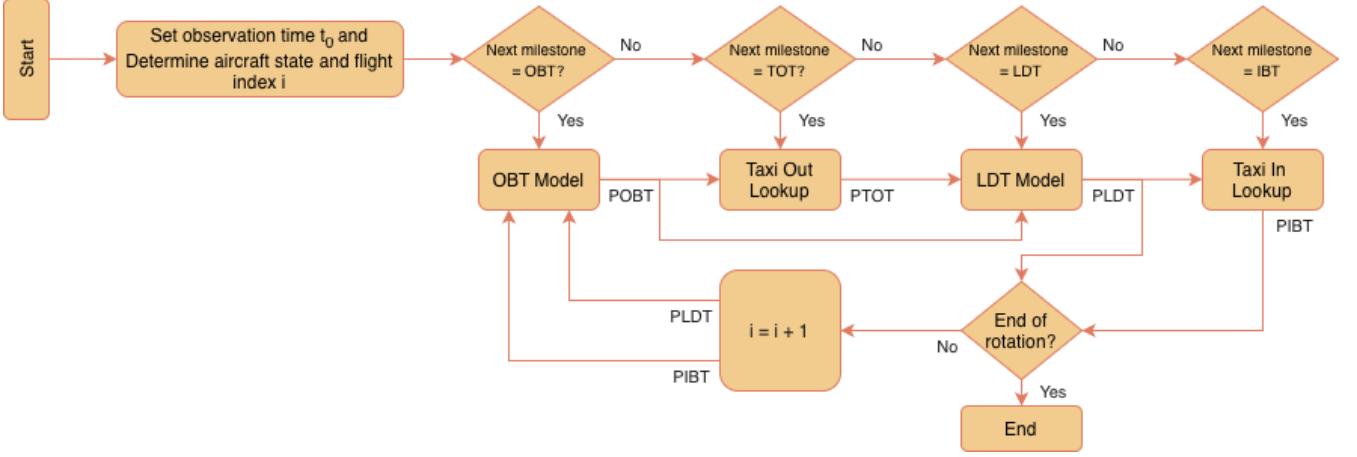


Fig. 6. Phase-decomposed framework for milestone prediction through an aircraft rotation.

for each aircraft type from observed turnaround times, defined as the time between Actual In-Block Time (AIBT) and AOBT. Since the MTT is intended to represent a minimum turnaround rather than a typical turnaround, the lower whisker of the observed turnaround-time distribution was used. The Schiphol feature set therefore includes the MTT and the Scheduled Buffer, defined as the difference between STT and MTT.

At the outstations, realised in-block times were not available. The minimum required turnaround interval could therefore not be determined from the time between in-block and off-block. Instead, the outstation feature set uses the minimum observed time between the inbound LDT and the outbound OBT, determined by airport and aircraft group. The corresponding buffer represents the time between the earliest feasible off-block time based on this minimum interval and the outbound SOBT.

The remaining features describe the flight schedule, aircraft and airport characteristics, operational state, and weather conditions. Airport congestion is included only for Schiphol because the available Decision Support Tool (DST) data provide a complete view of local departures and arrivals at this airport. The same feature could not be constructed consistently for outstations, where the dataset only contains flights connected to Schiphol. Airport Collaborative Decision Making (A-CDM) features are used only at A-CDM airports, while weather variables are taken from the most recent METAR observation available before t_o .

The complete feature sets for the Schiphol and outstation models are shown in Appendix A. The predicted delay is converted into a predicted milestone time by adding it to the SOBT. The Predicted Off-Block Time (POBT) is therefore obtained as:

$$\text{POBT}_i(t_o) = \text{SOBT}_i + \hat{D}_i^{\text{OBT}}(t_o) + \text{ATFM}_i(t_o). \quad (7)$$

2) *Landing Time Prediction:* The LDT model represents the airborne phase of a flight leg. For each flight i , the model uses information about the departure state, the planned and

predicted arrival times, route characteristics, aircraft characteristics, weather conditions, and the operational flight state. The learning target is the landing delay, defined as the difference between the Actual Landing Time (ALDT) and the Scheduled Landing Time (SLDT):

$$D_i^{\text{LDT}} = (\text{ALDT}_i - \text{ATFM}_i) - \text{SLDT}_i. \quad (8)$$

During prediction, the model estimates this landing delay using the feature vector available at observation time t_o . This is formulated as:

$$\hat{D}_i^{\text{LDT}}(t_o) = f_{\text{LDT}}(X_i^{\text{LDT}}(t_o)), \quad (9)$$

where $\hat{D}_i^{\text{LDT}}(t_o)$ is the predicted landing delay for flight i , f_{LDT} denotes the trained LDT model, and $X_i^{\text{LDT}}(t_o)$ is the feature vector constructed for flight i at observation time t_o .

The departure state is constructed from the predicted or realised milestones available at observation time t_o . It includes the departure delay, taxi-out time, and take-off milestone. When a required milestone has not yet occurred, the corresponding prediction from the preceding phase-level model is used.

The Scheduled Airborne Time (SAT) is defined as the time between Scheduled Take-Off Time (STOT) and SLDT. The Available Airborne Time (AAT) represents the time remaining between the take-off time and the SLDT. The feature set also includes the ratio between AAT and SAT, together with the available buffer, defined as the difference between SAT and AAT.

The model also uses the latest flight-plan arrival estimate and the latest DST arrival prediction, both expressed relative to the SLDT. Route-related features include the origin and destination airports and the distance between them. Weather variables are taken from the most recent METAR observations available before t_o at both airports. The remaining features describe the aircraft operator, aircraft type, flight state, and prediction horizon.

The complete feature set is shown in Appendix A. The predicted delay is converted into a predicted milestone time by adding it to the SLDT. The Predicted Landing Time (PLDT) is therefore obtained as:

$$\text{PLDT}_i(t_o) = \text{SLDT}_i + \hat{D}_i^{LDT}(t_o) + \text{ATFM}_i(t_o). \quad (10)$$

3) *Taxi-Time Formulation*: In contrast to the OBT and LDT models, the taxi models do not predict delay relative to a scheduled milestone. Instead, they estimate the duration between two consecutive operational milestones. The taxi-out model estimates the time between OBT and Take-Off Time (TOT), while the taxi-in model estimates the time between LDT and IBT. These estimated durations are then used to convert the predicted off-block and landing times into the next predicted milestones in the flight sequence. For taxi-out, the estimated taxi-out duration $T_i^{\text{out}}(t_o)$ is added to the POBT to obtain the Predicted Take-Off Time (PTOT):

$$\text{PTOT}_i(t_o) = \text{POBT}_i(t_o) + T_i^{\text{out}}(t_o). \quad (11)$$

For taxi-in, the estimated taxi-in duration $T_i^{\text{in}}(t_o)$ is added to the PLDT. The Predicted In-Block Time (PIBT) is therefore obtained as:

$$\text{PIBT}_i(t_o) = \text{PLDT}_i(t_o) + T_i^{\text{in}}(t_o). \quad (12)$$

4) *Taxi Time Estimation at Schiphol Airport*: For Schiphol Airport, taxi times are estimated using lookup tables that account for the active runway configuration. This is important because the assigned runway determines the length of the surface route and therefore has a strong influence on the duration of the taxi. Runway 18R/36L can require substantially longer taxi routes than runways located closer to the terminal area.

For taxi-out at Schiphol, the lookup table uses the take-off runway, destination region, and aircraft operator. The destination region separates flights within and outside the European Civil Aviation Conference (ECAC) area. This is included because flights to different regions may use different gate areas at Schiphol. When more than one take-off runway is active, the most likely runway for an individual departure is determined from the planned departure direction. This runway is then used to select the corresponding taxi-out duration.

For taxi-in at Schiphol, the lookup table uses the landing runway, origin region, and aircraft operator. When two landing runways are active, the assigned runway of an individual aircraft is not directly available. Therefore, the stack is used to determine the most likely landing runway. In this study, the stack refers to one of the three Initial Approach Fix (IAF)s through which flights approach Schiphol: ARTIP, RIVER, or SUGOL. Historical distributions between these IAFs and the active landing runways are then used to select the landing runway.

A specific adjustment is included for taxi routes to and from runway 18R/36L. These routes depend on whether runway 18C/36C is active for landing or departure, because runway-crossing and routing restrictions can prevent aircraft from using the shortest taxi route. Therefore, runway variants

such as 18R_W, 18R_Y, and 18R_Z are included in the lookup table. The suffixes W, Y, and Z represent different taxi-routing cases. A similar distinction is applied for departures from runway 36L. These variants allow the lookup table to account for changes in taxi route caused by the active runway configuration.

5) *Taxi Time Estimation at Outstations*: For outstations, the available runway-configuration information is more limited than for Schiphol because of the lack of data. Therefore, a different lookup strategy is used. For both taxi-in and taxi-out, the assigned runway is not directly available, so wind direction is used as a proxy for the likely runway configuration.

The wind direction is divided into fixed directional bins, and for each airport and wind-direction bin the most likely landing or take-off runway is determined from historical data. The corresponding taxi-in or taxi-out duration is then obtained from the historical taxi time associated with that airport-runway combination. If no wind-based runway estimate is available, the model falls back to a median taxi-in time per arrival airport or a median taxi-out time per departure airport. This fallback does not account for runway configuration, but provides a robust estimate when the more detailed wind-based lookup cannot be applied.

Together, these phase-level components make it possible to predict each future milestone in the aircraft rotation. The next subsection explains how these milestone predictions are linked sequentially, so that the output of one phase can be used as input for the next phase and flight leg.

D. Sequential Prediction Procedure

The phase-level models are applied in a sequential prediction procedure at each observation time t_o . First, the latest available Decision Support Tool (DST) message before t_o is selected for each flight. This creates a snapshot of the information available to the model at the time of prediction. Flights are then grouped by aircraft registration and ordered by their Scheduled Off-Block Time (SOBT) to reconstruct the aircraft rotation within the prediction day. The resulting sequence is checked using the departure and arrival airports, so that consecutive flights are only linked when the arrival airport of one flight matches the departure airport of the next flight. For each aircraft, the current position in the rotation is determined by comparing the realised milestone times with the observation time. Milestones that have already occurred before t_o are fixed to their realised values and are not predicted again. This prevents the model from overwriting known operational outcomes. Only milestones that have not yet occurred are predicted. As a result, the first prediction step depends on the state of the aircraft at t_o . Once this state has been identified, the next upcoming milestone in the rotation is predicted. The resulting predicted milestone time is then used as input for the subsequent prediction step. This recursive procedure is repeated until all remaining flights in the rotation have been assigned predicted milestone times. The result is a complete predicted timeline for each aircraft from the observation time until the end of the day. Because realised milestones are

used where available and predicted milestones are propagated only for future events, the procedure reflects the information that would have been available in an operational setting. Figure 6 illustrates how realised milestones are fixed before t_o , while future milestones are predicted sequentially and passed forward through the aircraft rotation.

E. Demand Aggregation

After the sequential predictions, the predicted milestones are aggregated into Schiphol demand profiles. Only EHAM departures and arrivals are included, because these are the milestones relevant for the Luchtverkeersleiding Nederland (LVNL) demand prediction. For each observation time t_o , demand is evaluated on a sequence of five-minute time bins from t_o until the end of the operational day. For each five-minute time bin with start time t_b , demand is calculated over the 20-minute interval from t_b to $t_b + 20$ minutes. This produces a rolling demand profile rather than a count of flights within isolated five-minute intervals. The rolling-window formulation is used because it better represents the traffic load relevant for flow management than individual point counts. The rolling demand is calculated consistently for the predicted, realised, and DST demand profiles. A flight contributes to the demand profile according to the position of its take-off or landing time relative to the rolling window. Events that fall fully inside the window contribute to the demand count, while the intervals crossing the start and end of the window are treated fractionally. This interval-based aggregation, illustrated in Figure 1, reduces sensitivity to bin boundaries and gives a smoother representation of the demand profile.

This aggregation step links the phase-level delay propagation model to the operational use case of demand prediction. Delays that propagate through aircraft rotations shift the predicted Take-Off Time (TOT) and Landing Time (LDT) values of subsequent flights. These shifts can create local increases or decreases in the number of flights expected within a rolling demand window. The resulting demand profile therefore captures not only individual milestone errors, but also the effect of knock-on delay on the temporal distribution of Schiphol traffic demand.

IV. EXPERIMENTAL SETUP

This section describes how the proposed model is evaluated. It first presents the hypotheses and modelling assumptions. It then describes the dataset, the LightGBM model setup, the model interpretation method, and the evaluation approach for the phase-decomposed propagation model. The numerical performance evaluation focuses on the chained model, because this reflects the operational setting in which future milestones are predicted recursively through the aircraft rotation.

A. Hypotheses

Based on the phase-decomposed framework introduced in Section III and the literature discussed in Section II, this study

formulates four hypotheses regarding delay propagation within aircraft rotations.

- H1: Available Turnaround Time (ATT) is the dominant driver of delay propagation in the turnaround phase. As the ATT decreases with increasing inbound delay, the likelihood of off-block delay on the next flight increases.
- H2: The airborne phase can absorb part of the upstream delay, but this recovery is limited. As a result, part of the departure delay is expected to remain visible at landing.
- H3: A phase-decomposed propagation framework provides insight into where delay is generated, absorbed, or preserved within an aircraft rotation.
- H4: Improved prediction of Schiphol take-off and landing milestones leads to improved aggregated demand prediction, since the demand profiles are constructed from these milestones.

B. Assumptions

The proposed phase-decomposed propagation model relies on several assumptions regarding the schedule reference, rotation reconstruction, milestone accuracy, and the propagation mechanism. These assumptions define the scope within which the results should be interpreted.

- 1) **Schedule reference:** The initial filed flight plan is used as the schedule reference for all delay calculations. Flight plans are often updated after disruptions occur, which can reduce or remove measured delay if the latest filed plan is used. Fixing the reference to the initial schedule ensures that the propagated delay is measured consistently across flights.
- 2) **Rotation reconstruction:** Aircraft rotations are reconstructed from the Decision Support Tool (DST) data using aircraft registration and the chronological ordering of flights. The DST data only contain flights that are inbound to or outbound from Schiphol, and therefore do not provide the complete daily operation of every aircraft outside the Schiphol network. This means that consecutive flights can only be reliably linked when the relevant rotation legs are present in the available dataset. The analysis is therefore restricted to KLM, KLM Cityhopper, and Transavia operations, because these operators have Schiphol as a main hub or operating base and provide sufficiently complete aircraft rotations in the DST data.
- 3) **Milestone accuracy:** Operational milestone timestamps are assumed to represent the phase boundaries of the aircraft rotation. The timestamps Actual Off-Block Time (AOBT), Actual Take-Off Time (ATOT), Actual Landing Time (ALDT), and Actual In-Block Time (AIBT) are therefore treated as the transitions between the turnaround, taxi-out, airborne, and taxi-in phases.
- 4) **Propagation mechanism:** Delay propagation is assumed to occur primarily through aircraft rotations. Other propagation mechanisms, such as crew connections or passenger transfers, are not explicitly modelled. This allows the analysis to focus on rotational knock-on delay, but excludes other sources.

C. Dataset

The Off-Block Time (OBT) and Landing Time (LDT) models are developed using data from July 2025 and evaluated within the chained framework on August 2025. The July data are split chronologically into a training and validation set:

- Training set: 1–25 July 2025;
- Validation set: 26–31 July 2025;
- Evaluation period: 1–31 August 2025, excluding 23 and 24 August because of known DST data gaps.

The training set is used to fit the LightGBM models, while the validation set is used for hyperparameter selection. The August data are kept separate from model training and are used to evaluate the complete phase-decomposed propagation model framework.

The taxi-time lookup tables are constructed separately from the LightGBM models. For Schiphol, the tables are based on historical taxi-time data from 2025 and use the lookup variables described in Section III-C. For outstations, taxi-time estimates are derived from Flightradar24 data from May 2026. During the August evaluation, the trained OBT and LDT models and the fixed taxi-time tables are applied sequentially to the reconstructed aircraft rotations.

D. LightGBM Models

As defined in Section III-C, the OBT and LDT models are formulated as delay prediction tasks. The OBT model predicts off-block delay relative to Scheduled Off-Block Time (SOBT), while the LDT model predicts landing delay relative to Scheduled Landing Time (SLDT). The predicted delays are then converted into predicted milestone times using Equations 7 and 10.

The OBT and LDT models are not evaluated as standalone operational predictors. Some input features depend on upstream milestones, such as the landing time of the inbound flight for the OBT model and the departure state of the same flight for the LDT model. In a standalone evaluation, the OBT and LDT models could use realised upstream milestones from the historical data. For example, the LDT model could use the ATOT of the same flight, even if that take-off time would not yet be known at the observation time. In the phase-decomposed propagation model, this is not allowed. If an upstream milestone has not yet occurred, the model must use the predicted milestone from the previous step. Therefore, a standalone evaluation would not reflect the operational setting of the model. For this reason, the numerical performance evaluation focuses on the full phase-decomposed propagation model.

The evaluation data are processed as a sequence of operational snapshots. For each observation time, the latest available DST message is used to represent the information that would have been available in real time. The proposed model and the DST baseline are therefore evaluated from the same observation times and against the same reference ATOT and ALDT values, corrected for the assigned Air Traffic Flow Management (ATFM) delay as described in Section III-C.

This ensures that differences in performance are caused by the prediction method itself, rather than by differences in data availability or evaluation timing.

1) *Model Features*: The input features are designed to capture the main operational drivers of the OBT and LDT prediction tasks. They include schedule information, aircraft and airport characteristics, upstream delay information, Airport Collaborative Decision Making (A-CDM) variables, weather observations, and traffic-related variables. The complete feature sets are listed in Appendix A. The airport congestion feature is only used for Schiphol turnarounds. As described in the rotation reconstruction assumption in Section IV-B, the available DST data only provide a complete operational picture for flights inbound to or outbound from Schiphol. Local traffic demand at outstations therefore cannot be reconstructed consistently from the available data.

2) *Model Configuration*: Separate OBT models are trained for Schiphol and outstation turnarounds. This distinction is made because Schiphol has more detailed local data, while the outstation cases are spread across many different airports with less airport-specific information. For both groups, six horizon-specific models are trained using the horizon with respect to SOBT: (0, 15], (15, 30], (30, 60], (60, 120], (120, 180], and (180, 240] minutes. In addition, two boundary models are trained for each group to cover horizons that fall outside these intervals during chained prediction: one for scheduled horizons below 0 minutes and one for scheduled horizons above 240 minutes.

For the LDT prediction task, no distinction is made between Schiphol and outstation cases. Each flight in the dataset either departs from or arrives at Schiphol, and the features used by the LDT model are available for both flight directions. Six horizon-specific LDT models are trained using the same interval boundaries, based on the horizon with respect to SLDT. Just like the OBT models, two additional boundary models are trained: one for horizons below 0 minutes and one for horizons above 240 minutes.

The horizon-specific models are selected using scheduled horizons, because these horizons are known at prediction time. During chained prediction, the scheduled horizon can become negative or exceed 240 minutes because of accumulated delay. In these cases, the below-0 or above-240 boundary model is used, respectively.

3) *Model Optimisation*: The LightGBM hyperparameters are tuned using random search, which provides a simple and effective alternative to grid search for hyperparameter optimisation, especially when only a subset of hyperparameters strongly affects model performance [27]. For each model, 100 hyperparameter combinations are sampled from the search space shown in Table I. Each separate model is trained on the training set and evaluated on the validation set using MAE. The hyperparameter combination with the lowest validation MAE is selected.

The search space includes parameters that control tree complexity, learning rate, and binning. The objective function is set to L_1 regression and the evaluation metric is Mean Absolute Error (MAE), so that the optimisation directly focuses on reducing the average absolute error in minutes.

TABLE I
HYPERPARAMETER SEARCH SPACE USED FOR LIGHTGBM OPTIMISATION

Hyperparameter	Search interval	Step size / values
Number of leaves	[32, 256]	32
Maximum depth	[8, 64]	8
Number of trees	[64, 512]	64
Learning rate	[0.001, 1.000]	0.001 and 0.05 increments
Maximum bin	[64, 256]	64

4) *Model Interpretation*: Although the standalone OBT and LDT models are not evaluated as final operational predictors, their feature importance is analysed to understand which variables drive the phase-level predictions. SHapley Additive exPlanations (SHAP) values are used for this purpose. SHAP is based on Shapley values from cooperative game theory and provides an additive feature attribution method [28].

In the SHAP summary plots, each point represents one prediction. The horizontal position of a point shows the contribution of that feature to the predicted delay. A positive SHAP value indicates that the feature increases the predicted delay, while a negative value indicates that the feature decreases the predicted delay. The colour of each point represents the corresponding feature value relative to the values observed for that feature. Blue points indicate low feature values, while red points indicate high feature values. Features are ordered by their mean absolute SHAP value, so features shown near the top have the largest overall influence on the model predictions.

E. Phase-Decomposed Propagation Model Evaluation

The full performance evaluation is performed on the phase-decomposed propagation model, because this reflects the operational setting in which future milestones are predicted recursively through the aircraft rotation. The model combines the OBT and LDT LightGBM models with the taxi lookup tables, as described in Section III-D.

The phase-decomposed propagation framework is evaluated using three metrics: MAE, Root Mean Squared Error (RMSE), and the standard deviation of the prediction error. The corresponding error distributions are also visualised using boxplots to compare the spread, central tendency, and outliers of the framework and DST errors across horizon intervals. For the milestone-level evaluation, the prediction error for flight i is defined as:

$$e_i = \hat{y}_i - y_i, \quad (13)$$

where $\hat{y}_i$ is the predicted milestone time and y_i is the realised milestone time. The same definition is used for the DST baseline. The MAE is the main metric because it expresses the average absolute prediction error directly in minutes:

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|. \quad (14)$$

The RMSE is included because it gives more weight to larger errors:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}. \quad (15)$$

The standard deviation of the prediction error is used to describe the spread of the errors around their mean:

$$\sigma_e = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (e_i - \bar{e})^2}. \quad (16)$$

For the aggregated demand evaluation, the same metrics are applied to demand errors expressed in aircraft rather than minutes.

The evaluation is performed at three levels. First, one aircraft rotation is examined as an illustrative example. This example is used only to show how the model updates the aircraft state and propagates predicted milestones through successive flight legs. It is not used to assess the overall prediction accuracy of the framework.

Second, the milestone prediction accuracy is evaluated quantitatively across all relevant Schiphol flights in the evaluation dataset. For each observation time, Predicted Take-Off Time (PTOT) is compared with ATOT for EHAM departures, while Predicted Landing Time (PLDT) is compared with ALDT for EHAM arrivals. The corresponding DST take-off and landing predictions are evaluated in the same way and serve as the operational baseline.

Third, the operational impact is evaluated by comparing aggregated demand profiles. The predicted, realised, and DST demand profiles are constructed using the rolling 20-minute demand aggregation described in Section III-E. This evaluation assesses whether the model improves the representation of future Schiphol demand.

V. RESULTS

This section presents the results of the phase-decomposed propagation framework. First, the selected LightGBM hyperparameters and the feature importance results are reported for the separate Off-Block Time (OBT) and Landing Time (LDT) models. These results are used to inspect which variables drive the fitted models. The operational performance is then evaluated using the full phase-decomposed propagation framework, first through a single-rotation example, then at milestone level, and finally at the level of aggregated rolling Schiphol demand.

A. LightGBM Training Results

The OBT and LDT models are the trained components of the phase-decomposed propagation model. Before evaluating the full model, this subsection inspects their selected LightGBM configurations and feature-importance patterns. These

results provide context for interpreting the later milestone and demand evaluation, without analysing the standalone models as separate operational predictors.

1) *Selected Hyperparameters*: Table II presents the selected hyperparameters for the Schiphol and outstation OBT models. Separate horizon-specific models are trained for Schiphol and outstation turnarounds using the scheduled horizon with respect to Scheduled Off-Block Time (SOBT). During chained prediction, boundary models are used when the scheduled horizon falls below 0 or above 240 minutes.

TABLE II
SELECTED LIGHTGBM HYPERPARAMETERS FOR THE SCHIPHOL AND OUTSTATION OBT MODELS.

Model	Num Leaves	Max depth	Num trees	Learning rate	Max Bin
Schiphol $(-\infty, 0)$	32	24	512	0.05	128
Schiphol $[0, 15)$	224	8	320	0.05	128
Schiphol $[15, 30)$	224	40	192	0.05	256
Schiphol $[30, 60)$	32	24	512	0.05	128
Schiphol $[60, 120)$	32	24	512	0.05	128
Schiphol $[120, 180)$	224	8	320	0.05	128
Schiphol $[180, 240)$	192	8	256	0.10	256
Schiphol $[240, \infty)$	32	24	512	0.05	128
Outstations $(-\infty, 0)$	32	24	512	0.05	128
Outstations $[0, 15)$	160	64	320	0.05	192
Outstations $[15, 30)$	160	64	320	0.05	192
Outstations $[30, 60)$	32	24	512	0.05	128
Outstations $[60, 120)$	192	8	256	0.10	256
Outstations $[120, 180)$	224	8	320	0.05	128
Outstations $[180, 240)$	32	24	512	0.05	128
Outstations $[240, \infty)$	32	24	512	0.05	128

The selected hyperparameters for the LDT models are shown in Table III. In contrast to the OBT prediction task, the LDT model is not separated into Schiphol and outstation cases. The same horizon intervals are used, but the horizon is defined with respect to Scheduled Landing Time (SLDT), as explained in Section IV-D2.

TABLE III
SELECTED LIGHTGBM HYPERPARAMETERS FOR THE LDT MODELS.

Model	Num Leaves	Max depth	Num trees	Learning rate	Max Bin
$(-\infty, 0)$	160	48	512	0.05	256
$[0, 15)$	160	48	512	0.05	256
$[15, 30)$	160	64	320	0.05	192
$[30, 60)$	160	48	512	0.05	256
$[60, 120)$	160	48	512	0.05	256
$[120, 180)$	160	48	512	0.05	256
$[180, 240)$	160	48	512	0.05	256
$[240, \infty)$	160	48	512	0.05	256

2) *Feature Importance*: The phase-level models are interpreted using SHapley Additive exPlanations (SHAP) feature importance, as described in Section IV-D4. The analysis focuses on the $[180, 240)$ minute horizon because it corresponds to the operational planning window for flow management. Shorter horizons can be dominated by operational target variables, such as Target Off-Block Time (TOBT) and Target Start-Up Approval Time (TSAT). The corresponding SHAP summary plots for the remaining horizons are provided in Appendix B. Figure 7 shows the SHAP feature importance for the Schiphol OBT model. Available Turnaround Time (ATT)

has the largest mean absolute SHAP value and represents the time between the inbound arrival and the scheduled outbound departure. Other important variables include the destination airport, origin airport, the horizon with respect to SOBT, and the ratio between ATT and Scheduled Turnaround Time (STT). The results suggest that the model uses both the available turnaround capacity and the planned turnaround structure when predicting off-block delay.

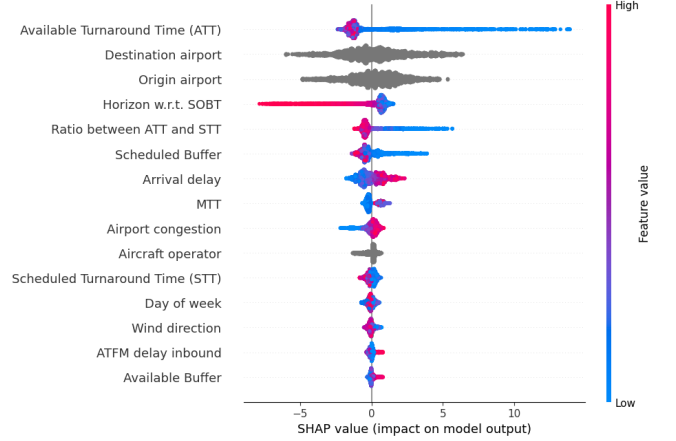


Fig. 7. SHAP feature importance for the Schiphol OBT model in the $[180, 240)$ minute horizon with respect to SOBT.

Figure 8 shows the feature importance for the outstation OBT model. Similar to the Schiphol model, the outstation model gives high importance to variables that describe how much time remains for the turnaround, especially ATT and the ratio between ATT and STT. The airport, operator, hour of the day, and scheduled prediction horizon are also important. This suggests that the model accounts not only for the available turnaround time, but also for differences between outstations, such as airport-specific operating patterns and airline schedules.

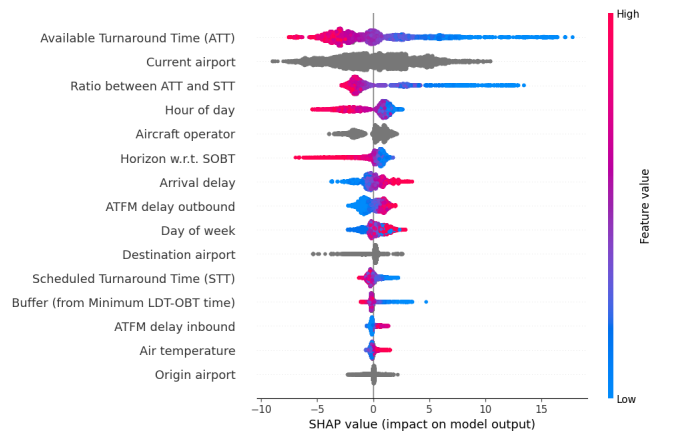


Fig. 8. SHAP feature importance for the outstation OBT model in the $[180, 240)$ minute horizon with respect to SOBT.

Figure 9 shows the SHAP feature importance for the

LDT model. The most important feature is departure delay, followed by variables that describe available airborne time, the ratio between available and scheduled airborne time, and available buffer. Route related variables, including departure airport, destination airport, and distance, also contribute to the prediction. These features show that the LDT model combines the upstream departure state when predicting landing delay.

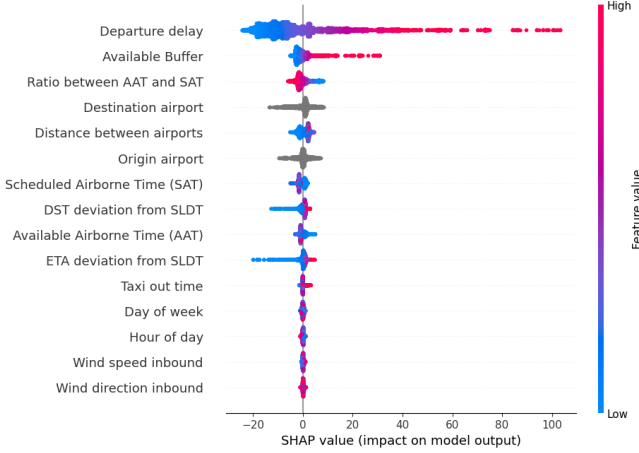


Fig. 9. SHAP feature importance for the LDT model in the [180, 240) minute horizon with respect to SLDT.

B. Phase-Decomposed Propagation Model

The phase-decomposed propagation model is evaluated using the sequential procedure introduced in Section III-D. At each observation time, realised milestones are used where available, while future milestones are predicted and passed forward through the remaining aircraft rotation. This reflects the operational setting in which the predicted state of one flight leg becomes input to the following phases and flight legs.

The evaluation first uses a single aircraft rotation to illustrate how milestones are propagated through successive flights. It then assesses the accuracy of Schiphol take-off and landing predictions across the full evaluation dataset, while outstation milestones serve as intermediate states in the propagation process. Finally, it evaluates how these milestone predictions translate into rolling Schiphol demand profiles.

1) *Single Rotation Example:* Before evaluating the framework across all flights, a single aircraft rotation is analysed to illustrate how predicted milestones are propagated through consecutive flight legs. This example is not intended as a statistical performance measure, but as a qualitative demonstration of how the framework updates the aircraft state and carries delay forward through the remaining rotation.

Figure 10 shows the scheduled, actual, Decision Support Tool (DST), and the framework timelines for one aircraft rotation. The vertical black line indicates the observation time, which lies between 09:00 and 10:00 UTC. At this moment, the first two flight legs have already taken place, while the remaining part of the rotation still has to be predicted.

The second flight leg, from EHAM to LSZH, arrives later than scheduled. This reduces the available turnaround time

before the next outbound flight. Both the DST and the phase-decomposed propagation framework adjust the predicted take-off time of the following flight, indicating that the immediate effect of the delay is captured. The main difference appears in the final flight leg. In the DST prediction, the propagated effect of the earlier delay is no longer visible, and the final departure and arrival are predicted close to their scheduled times. In contrast, the framework still predicts a later departure and arrival for the final flight, because the predicted milestones of the previous leg are used as input for the next prediction step.

The example also shows why the sequential structure matters beyond the immediately affected flight. A change in one predicted milestone changes the starting condition for the following phase. A later landing time reduces the available turnaround time, which can shift the next off-block and take-off times. The resulting departure state then influences the predicted landing time of the subsequent flight. The framework therefore represents the cumulative effect of the rotation rather than treating the milestones of each flight leg as independent predictions.

This example illustrates the operational behaviour of the framework. Rather than predicting each flight independently, it maintains a consistent aircraft state along the rotation. In this case, the framework preserves the knock-on effect of the earlier delay on the final flight, while this effect is not retained in the original DST prediction.

2) *Milestone Results:* The milestone results evaluate how accurately the framework predicts take-off and landing times for Schiphol. As described in Section IV-E, these milestones are evaluated because they directly determine the Schiphol demand profile. Milestones at outstations are still predicted as part of the sequential propagation procedure, but they serve only as intermediate states through which delay is carried towards subsequent flights. They are therefore not included in the milestone error analysis. The corresponding DST predictions are used as the operational baseline.

Table IV shows that the framework improves Take-Off Time (TOT) prediction across all evaluated horizon intervals. In the shortest [0,15) minute horizon, the Mean Absolute Error (MAE) decreases from 4.62 to 3.78 minutes, the Root Mean Squared Error (RMSE) from 10.43 to 7.66 minutes, and the standard deviation from 10.33 to 7.64 minutes. The framework continues to outperform the DST baseline across the remaining horizons. In the [180,240) minute horizon, the MAE decreases from 16.16 to 13.62 minutes, the RMSE from 23.95 to 21.22 minutes, and the standard deviation from 21.52 to 21.04 minutes. These changes correspond to relative reductions of approximately 16%, 11%, and 2%, respectively.

Figure 11 supports these results. Across all evaluated horizons, the framework generally produces a more compact prediction error distribution than the DST baseline, with the largest visual differences occurring at the shorter and medium prediction horizons.

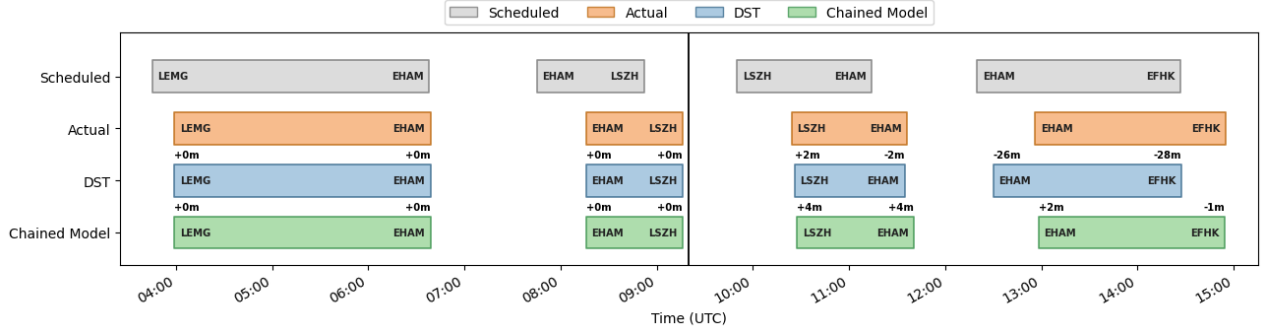


Fig. 10. Single rotation example of the phase-decomposed propagation procedure. The observation time is between 09:00 and 10:00 UTC, after which the remaining milestones are predicted sequentially through the aircraft rotation.

TABLE IV
TOT ERROR METRICS FOR EHAM DEPARTURES USING THE PROPOSED FRAMEWORK AND DST.

Horizon	n	Propagation Model			DST		
		MAE	RMSE	σ	MAE	RMSE	σ
[0, 15)	9835	3.78	7.66	7.64	4.62	10.43	10.33
[15, 30)	11916	5.99	9.17	9.05	6.53	10.71	10.71
[30, 60)	23476	7.64	11.23	11.22	8.43	12.55	12.41
[60, 120)	46534	9.81	14.51	14.49	11.79	16.45	14.99
[120, 180)	42394	11.88	18.36	18.28	14.45	20.99	19.20
[180, 240)	39223	13.62	21.22	21.04	16.16	23.95	21.52

the RMSE and standard deviation remain slightly higher than those of the DST in this interval, the framework improves all three error metrics from the [15,30) minute horizon onwards. In the [180,240) minute horizon, the MAE decreases from 13.92 to 11.33 minutes, the RMSE from 21.69 to 18.60 minutes, and the standard deviation from 21.40 to 18.60 minutes. These changes correspond to relative reductions of approximately 19%, 14%, and 13%, respectively. Figure 12 supports these results, showing that the framework generally produces more compact arrival-time error distributions than the DST baseline, particularly from the [15,30) minute horizon onwards.

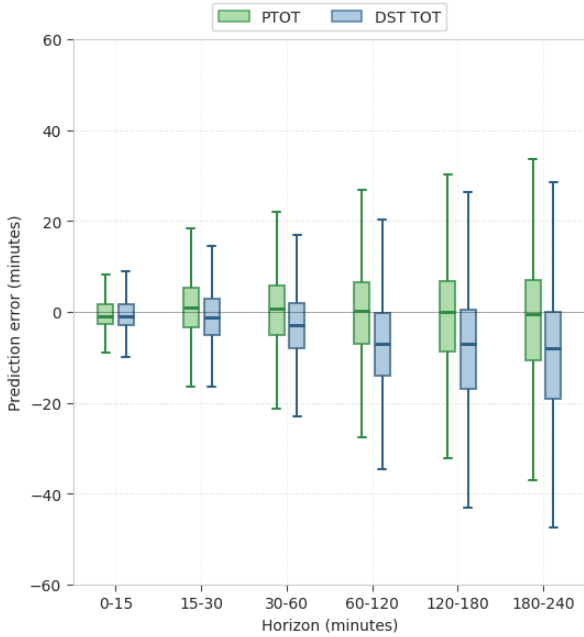


Fig. 11. Prediction error distributions for framework TOT predictions and the DST baseline for EHAM departures across horizon intervals.

For EHAM arrivals, Table V shows a similar pattern. The framework achieves a lower MAE than the DST across all evaluated horizon intervals. In the shortest [0,15) minute horizon, the MAE decreases from 2.64 to 2.27 minutes. Although

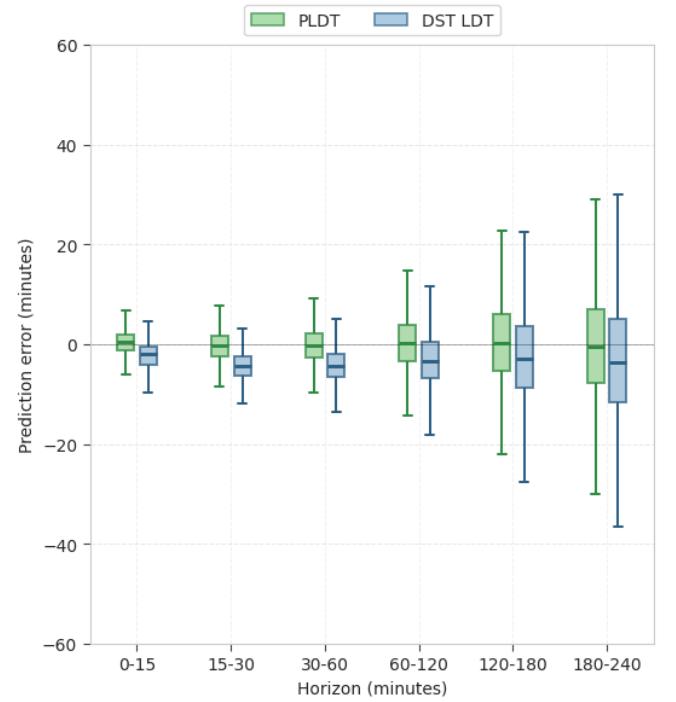


Fig. 12. Prediction error distributions for framework LDT predictions and the DST baseline for EHAM arrivals across horizon intervals.

TABLE V
LDT ERROR METRICS FOR EHAM ARRIVALS USING THE PROPOSED
FRAMEWORK AND DST.

Horizon	n	Propagation Model			DST		
		MAE	RMSE	σ	MAE	RMSE	σ
[0, 15)	13739	2.27	4.24	4.22	2.64	4.10	3.26
[15, 30)	13778	2.71	4.71	4.71	4.55	5.69	3.89
[30, 60)	27282	3.35	6.75	6.75	5.10	7.89	6.84
[60, 120)	50875	5.81	10.74	10.70	7.03	11.83	11.55
[120, 180)	46196	8.89	14.66	14.64	10.77	17.00	16.89
[180, 240)	45009	11.33	18.60	18.60	13.92	21.69	21.40

3) *Aggregated Demand Results*: The aggregated demand results evaluate how the milestone predictions translate into Schiphol traffic demand. For each observation time, EHAM departures and arrivals are aggregated into rolling 20-minute demand values. The demand error is expressed in aircraft, where a positive value indicates an overprediction of demand and a negative value indicates an underprediction.

In contrast to the milestone results, the horizon in this subsection refers to the time between the observation time and the start of the evaluated demand bin. The demand value for that bin is calculated over the following 20-minute window.

Table VI shows the aggregated Schiphol demand error metrics for the framework and the DST baseline. The framework achieves lower MAE, RMSE, and standard deviation values across all evaluated horizon intervals. In the shortest [0, 15) minute horizon, the MAE decreases from 2.05 to 1.66 aircraft, the RMSE from 2.67 to 2.18 aircraft, and the standard deviation from 2.45 to 2.18 aircraft. In the longest [180, 240) minute horizon, the MAE decreases from 2.97 to 2.53 aircraft, the RMSE from 3.93 to 3.34 aircraft, and the standard deviation from 3.92 to 3.34 aircraft. These changes correspond to relative reductions of approximately 19%, 18%, and 11% in the shortest horizon, and 15%, 15%, and 15% in the longest horizon, respectively.

TABLE VI
AGGREGATED ROLLING 20-MINUTE SCHIPHOL DEMAND ERROR METRICS
FOR THE PROPOSED FRAMEWORK AND DST.

Horizon	n	Propagation Model			DST		
		MAE	RMSE	σ	MAE	RMSE	σ
[0, 15)	8358	1.66	2.18	2.18	2.05	2.67	2.45
[15, 30)	8358	1.92	2.52	2.52	2.13	2.75	2.72
[30, 60)	16716	2.13	2.81	2.79	2.55	3.36	3.24
[60, 120)	33432	2.34	3.08	3.08	2.68	3.52	3.52
[120, 180)	33432	2.47	3.27	3.27	2.76	3.65	3.65
[180, 240)	34797	2.53	3.34	3.34	2.97	3.93	3.92

The framework also achieves a lower standard deviation than the DST baseline across all evaluated horizons. This indicates that the framework consistently produces a more compact demand error distribution, suggesting that larger demand prediction errors occur less frequently.

The error distributions in Figure 13 support the numerical results. Across all evaluated horizons, the framework errors are generally centred closer to zero and exhibit a smaller spread than the DST baseline. The improvement is visible throughout

the evaluation, although the visual difference remains more modest than for the milestone predictions.

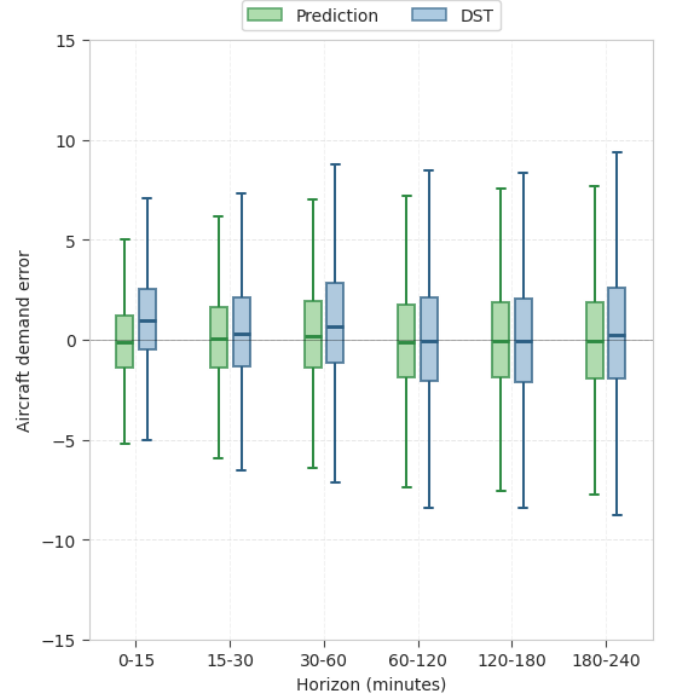


Fig. 13. Aggregated rolling 20-minute Schiphol demand error distributions for the proposed framework and the DST baseline across horizon intervals.

Overall, the demand results show that the phase-decomposed propagation framework improves the rolling 20-minute Schiphol demand prediction across all evaluated horizons. This improvement is observed alongside consistent improvements in take-off time prediction and arrival-time results at medium and long horizons.

VI. DISCUSSION

The results show that the phase-decomposed propagation model consistently improves milestone and demand prediction across the evaluated prediction horizons. For Schiphol departures, the framework improves all Take-Off Time (TOT) error metrics across every evaluated horizon. For Schiphol arrivals, the framework achieves a lower Mean Absolute Error (MAE) across all horizons and improves all three error metrics from the [15, 30) minute horizon onwards. Only in the shortest arrival horizon does the Decision Support Tool (DST) retain a slightly lower Root Mean Squared Error (RMSE) and standard deviation. At the demand level, the framework improves all evaluated error metrics across every horizon, which directly supports the operational objective of improving rolling Schiphol demand prediction.

A. Turnaround Delay Propagation

The results support Hypothesis H1, which states that ATT is the dominant driver of delay propagation during the turnaround phase. For both the Schiphol and outstation Off-Block Time

(OBT) models, Available Turnaround Time (ATT) is one of the most important features. This is consistent with earlier studies that identify available turnaround time as a dominant predictor of turnaround and departure delay [5], [18]. It also matches the expected turnaround propagation mechanism: when an inbound aircraft arrives late, the time available before the next scheduled departure decreases, and the probability of outbound delay increases.

The importance of ATT is relevant because it shows that the model does not only use the presence of inbound delay, but also the remaining ground time available to absorb it. A delayed inbound flight does not necessarily lead to a delayed outbound flight when sufficient turnaround time remains. Conversely, when the available turnaround time becomes limited, delay is more likely to propagate to the next flight. This pattern is visible for both Schiphol and outstation turnarounds. Schiphol provides more detailed local information, such as Airport Collaborative Decision Making (A-CDM) related variables and airport specific features, while the outstation model has to learn from flights distributed across many different airports. The fact that variables describing the available turnaround time remain important in both cases suggests that the turnaround phase is a general mechanism for rotational knock-on delay.

B. Airborne Delay Propagation

The Landing Time (LDT) feature importance results support H2, which proposed that the airborne phase can absorb part of the upstream delay, but that this recovery is limited. The results show that departure delay is the most important feature in the LDT model. This means that delay at take-off often remains visible at landing. The airborne phase can therefore absorb part of the upstream delay, but it does not fully remove the effect of a delayed departure. This is also supported by the importance of airborne time variables, such as Available Airborne Time (AAT), available buffer, and the ratio between available and scheduled airborne time. These variables describe how much time is available between departure and the scheduled landing time.

The milestone results show that this airborne modelling step improves the timing of Schiphol arrivals. The model achieves a lower LDT MAE than the DST baseline across all horizon intervals. This means that, on average, the predicted landing times are closer to the realised landing times. From the [15, 30) minute horizon onwards, the framework also improves the RMSE and standard deviation. Beyond the shortest prediction horizon, the improvement is therefore visible not only in the average error, but also in the spread of the errors.

At short arrival horizons, the framework improves the average absolute landing-time error, but some larger individual errors remain. This is reflected in the lower MAE, but higher RMSE and standard deviation. These larger errors are likely related to tactical arrival effects that are difficult to capture from the aircraft rotation alone, such as sequencing, runway changes, or local traffic conditions near Schiphol. Such effects can change close to arrival and may only become visible in the

operational data shortly before landing. This explains why the framework still performs slightly worse in terms of RMSE and standard deviation in the shortest horizon, while the remaining horizons benefit from modelling the propagated aircraft state.

C. Horizon-Dependent Performance

The milestone results show that the added value of the phase-decomposed propagation framework varies with the prediction horizon. Although the framework improves the take-off prediction across all evaluated horizons and the landing prediction across nearly all evaluated metrics, the relative improvements are generally smaller at shorter horizons. Close to take-off or landing, recent operational updates reduce the remaining uncertainty, leaving less room for improvement from a propagation model than at longer prediction horizons.

This short-horizon comparison should also be interpreted in light of the milestone reconstruction method used in this study. Since Actual Take-Off Time (ATOT) and Actual Landing Time (ALDT) are reconstructed from historical DST updates, the realised milestone times are not fully independent of the DST update process. This may favour the DST baseline at short horizons, where the latest operational estimate is already close to the reconstructed realised event time. The effect is less relevant at longer horizons, where the comparison mainly concerns whether upstream delay is propagated before it becomes visible in later operational updates.

At longer horizons, the phase-decomposed propagation model provides more value. This is because later flights in an aircraft rotation may still appear close to schedule in the DST baseline, even when the same aircraft has already accumulated delay earlier in the day. The proposed model uses the predicted state of the aircraft rotation as input for the next flight leg, allowing this upstream delay to be propagated before it becomes visible in later operational updates. This explains why the relative improvements generally become more apparent at medium and long horizons, even though the framework already improves most evaluation metrics at shorter horizons.

The results also support H3, which proposed that a phase-decomposed propagation framework provides insight into where delay is generated, absorbed, or preserved within an aircraft rotation. By modelling the turnaround, taxi-out, airborne, and taxi-in phases separately, the framework makes it possible to identify how delay evolves throughout the rotation. The turnaround phase mainly determines whether inbound delay is transferred to the next outbound flight, while the airborne phase can absorb part of the upstream delay but often preserves a remaining delay until landing. The taxi phases translate OBT and LDT predictions into TOT and In-Block Time (IBT) using surface movement estimates. Together, these phase-level components provide insight into where delay is transferred, preserved, or partly absorbed within the aircraft rotation.

D. Aggregated Demand Prediction

The final evaluation step considers whether the improved milestone predictions also lead to better rolling demand

profiles. Across all horizon intervals, the phase-decomposed propagation model achieves lower demand errors than the DST baseline. These results support H4, which proposed that improved prediction of Schiphol take-off and landing milestones leads to improved aggregated demand prediction. The results show that more accurate milestone predictions indeed translate into more accurate rolling demand profiles.

However, the improvement at demand level is less pronounced than the milestone improvement. In the milestone results, especially for TOT, the boxplots show a clear reduction in both the interquartile range and the whiskers at medium and long horizons. In the aggregated demand results, the model errors are generally centred closer to zero than the DST errors, but the difference in spread is smaller. The boxplots still show overlap between the model and DST error distributions, indicating that the improvement in individual milestone predictions is only partly reflected in the aggregated demand profile.

This can be explained by the aggregation step. Demand is evaluated in rolling 20-minute windows, so not every improvement in an individual flight time changes the demand value. If a flight is predicted more accurately but remains within the same rolling window, the aggregated demand value may stay the same. At the same time, small timing shifts of several flights can affect the demand profile when they move flights into or out of a window.

The reduction in the spread of the demand errors is also operationally relevant. Figure 13 shows that the framework generally has shorter whiskers than the DST baseline, especially at medium and long horizons. This indicates that large demand errors occur less often. Large positive errors may lead to an overestimation of the future traffic load, while large negative errors may hide an emerging demand peak. Although regulation decisions are not simulated in this study, reducing these larger demand errors can therefore support more reliable flow management decision making.

The main operational benefit is therefore that the model provides a better temporal placement of Schiphol traffic demand across the evaluated horizons, but with a smaller visible improvement than at milestone level. By keeping aircraft rotations consistent, the model captures part of the effect of rotational knock-on delay on future departures and arrivals. The remaining overlap in the demand-error distributions shows that the translation from individual milestone predictions to aggregated demand can still be improved.

E. Future Work

Future work could further analyse how milestone prediction errors translate into aggregated demand errors. The results show that improved TOT and LDT predictions can improve demand prediction, but the improvement at demand level is smaller than the milestone-level improvement. A more detailed analysis of the translation from milestone predictions to demand predictions could show why some improvements in individual TOT and LDT predictions have only a limited effect on the rolling demand profile. This would make it

clearer when flight-level prediction improvements lead to operationally relevant demand improvements.

The model could also be extended by using information from other flights when predicting a specific flight. In the current setup, each flight is mainly predicted from the state of its own aircraft rotation. This captures rotational knock-on delay, but it does not use information from other flights that may be affected by the same local conditions. For example, if earlier flights departing from the same airport show increasing delay, this may indicate a local constraint such as runway capacity, weather impact, or surface congestion. Including such information from recent or similar flights could help the model capture airport-level effects that are not visible from the aircraft rotation alone.

The modelling of outstation operations could be improved if more complete flight data become available. The current DST data only contain flights that are inbound to or outbound from Schiphol. As a result, the model cannot observe all local traffic at outstations and cannot always reconstruct complete aircraft rotations for operators that do not have Schiphol as a main hub. If data for all flights at the relevant outstations were available, the model could include local traffic conditions, runway use, and airport-specific delay patterns more directly. It would also make it possible to reconstruct more complete rotations for other operators, reducing the need to restrict the analysis to KLM, KLM Cityhopper, and Transavia. This could improve the prediction of delays that are generated or preserved at outstations and later propagate back to Schiphol.

The training period could also be extended. The current OBT and LDT models are trained on July 2025 and evaluated on August 2025. Training the models on a longer period would expose them to a wider range of traffic levels, weather conditions, runway configurations, and disruption patterns. This could improve robustness and make it possible to evaluate whether the results remain stable across seasons.

VII. CONCLUSION

This study shows that the phase-decomposed propagation model improves the prediction of Schiphol Take-Off Time (TOT) and Landing Time (LDT). The framework consistently improves milestone prediction across the evaluated horizons, with the relative improvements generally becoming larger at medium and long prediction horizons, where delay accumulated earlier in an aircraft rotation is not yet fully reflected in the Decision Support Tool (DST) baseline. These improved milestone predictions also lead to more accurate rolling Schiphol demand profiles.

The results show that the amount of remaining turnaround time is a key factor in determining whether inbound delay is carried into the next departure. In both the Schiphol and outstation Off-Block Time (OBT) models, Available Turnaround Time (ATT) is one of the most important features, indicating that the available ground time strongly influences the extent to which delay propagates. Departure delay is the dominant feature in the LDT model, showing that upstream delay often remains visible at landing. This suggests that the airborne

phase can absorb part of the delay, but generally does not remove it completely. Together, these findings provide insight into how delay is propagated and partly recovered within an aircraft rotation.

Schiphol departure predictions improve across all evaluated horizons. At the [180,240) minute horizon, the Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and standard deviation are reduced by approximately 16%, 11%, and 2%, respectively. Schiphol arrival predictions also improve, with a lower MAE across all evaluated horizons and improvements in all three error metrics from the [15,30) minute horizon onwards. At the longest horizon, the LDT MAE, RMSE, and standard deviation are reduced by approximately 19%, 14%, and 13%, respectively. These results confirm that the phase-decomposed structure consistently improves milestone prediction while providing the greatest relative benefit before propagated delay becomes visible in later operational updates.

The improved milestone predictions also lead to lower rolling Schiphol demand errors across all evaluated horizons. Although the relative improvement at demand level is smaller than at milestone level, because more accurate individual flight times do not always change the number of flights within a rolling 20-minute window, the framework consistently improves the MAE, RMSE, and standard deviation of the rolling demand prediction. This results in a more accurate temporal placement of future demand and fewer large demand prediction errors.

Overall, the results support all four hypotheses. The phase-decomposed propagation model captures the main rotational delay mechanisms, improves take-off and landing predictions, and translates these improvements into more accurate Schiphol demand profiles. Its main operational value lies in identifying future demand changes several hours before operation, when delay propagation is still insufficiently represented in the operational baseline.

REFERENCES

- [1] Single European Sky ATM Research 3 Joint Undertaking, "European atm master plan – making europe the most efficient and environmentally friendly sky to fly in the world – 2025 edition," 2024.
- [2] C. Walker, "Central Office for Delay Analysis - CODA," 2024.
- [3] R. Beatty, R. Hsu, L. Berry, and J. Rome, "Preliminary Evaluation of Flight Delay Propagation through an Airline Schedule," *Air Traffic Control Quarterly*, vol. 7, no. 4, pp. 259–270, Oct. 1999.
- [4] N. Kaffle and B. Zou, "Modeling flight delay propagation: A new analytical-econometric approach," *Transportation Research Part B: Methodological*, vol. 93, pp. 520–542, Nov. 2016.
- [5] R. Dalmau, "Probabilistic and explainable tree-based models for rotational reactionary flight delay prediction," *CEAS Aeronautical Journal*, vol. 15, no. 4, pp. 1157–1173, Oct. 2024.
- [6] N. Li and H. G. Yao, "A Review of Research on Flight Delay Propagation: Current Situation and Prospect," *Journal of Advanced Transportation*, vol. 2025, no. 1, p. 4851103, Jan. 2025.
- [7] J. Zuidberg and K. Vinkx, "Capacity demand at schiphol airport in 2030: And the possibilities to free up capacity in the inbound and outbound peaks," SEO Amsterdam Economics and To70, Amsterdam, The Netherlands, Tech. Rep. 2018-35, Apr. 2018, commissioned by the Ministry of Infrastructure and Water Management. ISBN: 978-90-6733-919-3.
- [8] EUROCONTROL, "Airport collaborative decision making (a-cdm)," <https://www.eurocontrol.int/concept/airport-collaborative-decision-making>, 2025, accessed: 2025-01-27.
- [9] Y. Alexopoulos, "A-CDM Manual Schiphol Airport," Feb. 2024.
- [10] ASSIRE. Guillaume, "EUROCONTROL Specification for Airport Collaborative Decision Making (A-CDM)."
- [11] S. Okwir, K. Amouzgar, and A. H. Ng, "Exploring prediction accuracy for optimal taxi times in airport operations using various machine learning models," *Journal of Air Transport Management*, vol. 122, p. 102684, Jan. 2025.
- [12] C. Li, J. Mao, L. Li, J. Wu, L. Zhang, J. Zhu, and Z. Pan, "Flight delay propagation modeling: Data, Methods, and Future opportunities," *Transportation Research Part E: Logistics and Transportation Review*, vol. 185, p. 103525, May 2024.
- [13] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," *Journal of Air Transport Management*, vol. 14, no. 5, pp. 221–236, Sep. 2008.
- [14] J. K. Brueckner, A. I. Czerny, and A. A. Gaggero, "Airline delay propagation: A simple method for measuring its extent and determinants," *Transportation Research Part B: Methodological*, vol. 162, pp. 55–71, Aug. 2022.
- [15] R. Dalmau, F. Ballerini, H. Naessens, S. Belkoura, and S. Wangnick, "An explainable machine learning approach to improve take-off time predictions," *Journal of Air Transport Management*, vol. 95, p. 102090, Aug. 2021.
- [16] C.-L. Wu and R. E. Caves, "Modelling and simulation of aircraft turnaround operations at airports," *Transportation Planning and Technology*, vol. 27, no. 1, pp. 25–46, Feb. 2004.
- [17] H. Fricke and M. Schultz, "Delay impacts onto turnaround performance - optimal time buffering for minimizing delay propagation," 06 2009.
- [18] P. De Falco, J. Kubat, V. Kuran, J. Rodriguez Varela, S. Plutino, and A. Leonardi, "Probabilistic prediction of aircraft turnaround time and target off-block time," in *Proceedings of the 13th SESAR Innovation Days (SIDs 2023)*, Seville, Spain, 2023.
- [19] I. Simaiakis and H. Balakrishnan, "A Queuing Model of the Airport Departure Process," *Transportation Science*, vol. 50, no. 1, pp. 94–109, Feb. 2016.
- [20] D. K. Park and J. K. Kim, "Influential factors to aircraft taxi time in airport," *Journal of Air Transport Management*, vol. 106, p. 102321, Jan. 2023.
- [21] C. Vakaet, "Taxi time prediction with classical and auto machine learning at schiphol airport," Master's thesis, Delft University of Technology, Delft, The Netherlands, 2020, accessed: 2026-04-03. [Online]. Available: <https://resolver.tudelft.nl/uuid:efba9763-cb49-4f89-b19b-80ad57c65803>
- [22] T. Diana, "Can machines learn how to forecast taxi-out time? A comparison of predictive models applied to the case of Seattle/Tacoma International Airport," *Transportation Research Part E: Logistics and Transportation Review*, vol. 119, pp. 149–164, Nov. 2018.
- [23] S. Belkoura, J. M. Peña, and M. Zanin, "Beyond Linear Delay Multipliers in Air Transport," *Journal of Advanced Transportation*, vol. 2017, pp. 1–11, 2017.
- [24] O. Basturk and C. Cetek, "Prediction of aircraft estimated time of arrival using machine learning methods," *The Aeronautical Journal*, vol. 125, no. 1289, pp. 1245–1259, Jul. 2021.
- [25] J. Sun, T. Dijkstra, K. Aristodemou, V. Buzetelu, T. Falat, T. Hogenelst, N. Prins, and B. Slijper, "Designing recurrent and graph neural networks to predict airport and air traffic network delays," 2022, 10th International Conference for Research in Air Transportation, ICRAT 2022 ; Conference date: 19-06-2022 Through 23-07-2022. [Online]. Available: <https://www.icrat.org/previous-conferences/10th-international-conference/>
- [26] OgiMET, "METAR/SPECI/TAF Reports Selection Query," <https://www.ogimet.com/metars.phtml.en>, accessed: 6 July 2026.
- [27] J. Bergstra and Y. Bengio, "Random Search for Hyper-Parameter Optimization," *Journal of Machine Learning Research* 13, 2012.
- [28] S. Lundberg and S.-I. Lee, "A Unified Approach to Interpreting Model Predictions," Nov. 2017.

Part B

Literature Review & Research Definition

Air Traffic Flow and Capacity Management

Air traffic demand continues to grow [1], while the capacity of the airspace remains fixed because it is a limited and fixed volume resource. Airports face similar physical and operational constraints, as they have limited capacities and resources and expanding airports is not always possible. As a result, both airports and airspace have become increasingly congested, resulting in an imbalance between demand and available capacity. To regulate the flow of air traffic as efficiently as possible in order to avoid congestion, Air Traffic Flow and Capacity Management (ATFCM) is introduced. It is one of the three main services of Air Traffic Management, the other two being Airspace Management (ASM) and Air Traffic Control (ATC). Within Europe, ATFCM services are provided by EUROCONTROL's Network Manager Operations Centre (NMOC), which coordinates demand and capacity across all member states of the European Civil Aviation Conference (ECAC). When traffic demand approaches or exceeds the available capacity, measures are required to prevent overload and maintain safety and efficiency. ATFCM plays a central role in balancing demand and capacity by combining strategic and tactical planning with collaborative processes that align stakeholders across the network [2].

This chapter is structured into three main parts. First, Section 2.1 introduces the four ATFCM phases, outlining how demand and capacity are managed from the strategic point. Next, Section 2.2 discusses Airport Collaborative Decision Making (A-CDM), with particular attention to the milestone approach that defines the operational timeline relevant for analyzing delay propagation. Finally, Section 2.3 presents LVNL's Decision Support Tool (DST), describing its role in predicting traffic demand, identifying capacity imbalances, and supporting operational decision making.

2.1. ATFCM Phases

ATFCM can be divided into four phases, the first is the Strategic flow management phase, which takes place seven days or more before the day of operations. This is followed by the Pre-tactical flow management phase, applied during the six days leading up to the day of operations. The Tactical flow management phase occurs on the day of operations itself, after which the process concludes with the Post operational analysis phase [2, 3]. An overview of these phases is shown in Figure 2.1.

Phases	Strategic	Pre-tactical	Tactical	Post operational
Objective	Early identification of major demand / capacity imbalances	Optimising efficiency by refining plans and balancing predicted demand and capacity	Resolving real-time demand / capacity issues	Evaluating performance, identifying causes of imbalance
Actions	- Capacity / demand unbalancing detection - Prepare ATFCM measures - Creation of NOP	- Update plan if there is new data such as weather data - Publication of ADP	- Real-time Capacity / demand monitoring - Adjusting the ADP when required	- Analysis of delays, regulations and capacity use. - Feedback for improving future strategic and pre-tactical planning
Time	Up to 7 days	6 to 1 day	Day of operation	After Operation

Figure 2.1: ATFM Phases [2, 3]

Strategic flow management

The strategic flow management phase takes place six months to seven days before the day of operations. It includes research, planning and coordination activities that support the early identification of demand–capacity imbalances. The main output of this phase is the weekly and daily Network Operations Plan (NOP), which identifies measures aimed at optimizing the utilization of available capacity [2].

Pre-tactical flow management

The Pre-tactical management phase takes place six days before the day of operation. It consists of planning and coordination activities aiming to fine tune the original plan made in the strategic flow management phase. The main objective is to optimize efficiency and balance demand and capacity through effective organization such as sector configuration. The output of this phase is the ATFCM Daily Plan (ADP), which is a set of tactical ATFCM measures prepared and agreed between all partners concerned to optimize the European Network [2].

Tactical flow management

The tactical flow management phase takes place on the day of operations and focuses on monitoring real-time events that affect the ADP and adjusting it when necessary. The aim is to ensure that disturbances such as weather, staffing issues or other unexpected constraints result in only minimal changes, keeping the required measures as close as possible to those defined in the preceding phases [2].

Post operational analysis

The post-operational analysis phase takes place after the day of operations and focuses on reviewing how the network performed compared to the plans and measures established in the earlier phases. In this stage, stakeholders analyse traffic flows, applied regulations, delay patterns and capacity usage to determine whether the strategic, pre-tactical and tactical measures were effective. The objective is to identify lessons learned and understand the causes of any remaining imbalances or inefficiencies [2].

2.2. Airport Collaborative Decision Making

Airport Collaborative Decision Making (A-CDM) has been developed to address challenges that arise from ineffective data and information exchange between different stakeholders. The aim is to increase the efficiency of airport operations and improve the performance of the overall ATM network by reducing delays, improving predictability and optimising the use of resources. A-CDM is currently adopted at 34 airports across Europe shown in Figure 2.2, and it consists of six main elements. For this thesis, the milestone approach is the primary element of interest, as it defines the operational timeline on which delay

propagation is analysed.

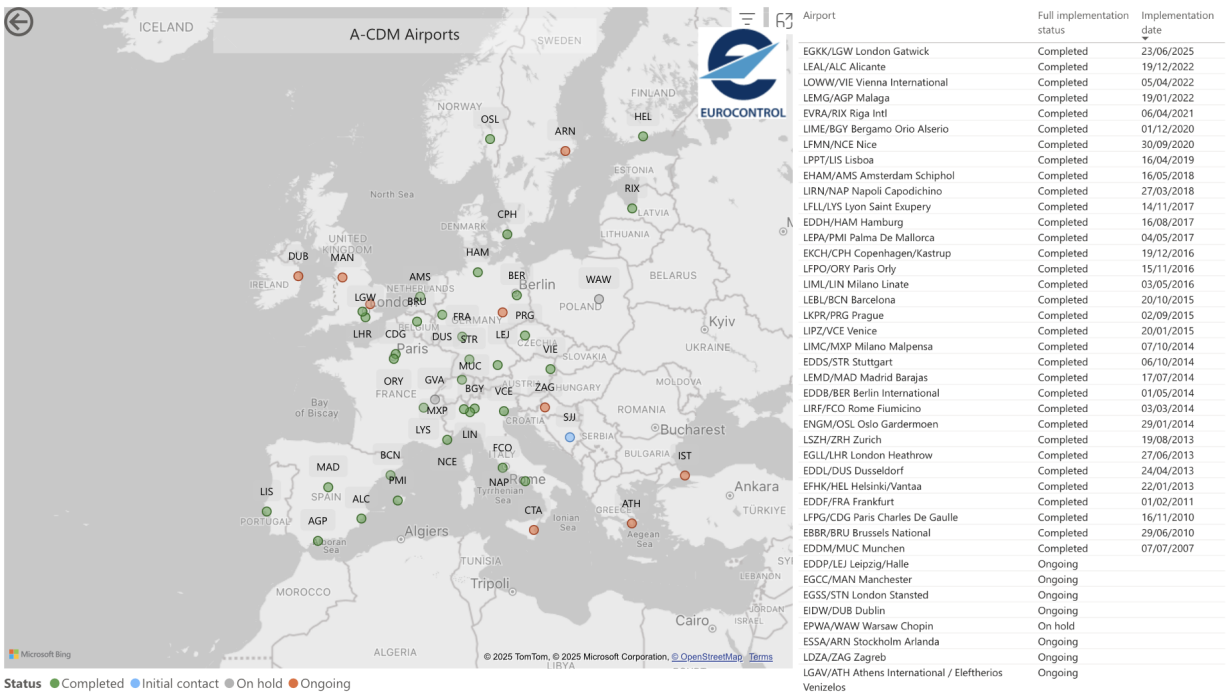


Figure 2.2: A-CDM Airports [4]

Milestone approach

To ensure consistency across the network, a common milestone process has been established that aligns with the key events of the flight process. Each A-CDM partner is responsible for specific milestones, with the goal of combining them into one seamless process. There are a total of 16 milestones, these milestones are a core part of A-CDM. Figure 2.3 illustrates these milestones in different flight phases. The goal of the milestone approach is to improve the situational awareness across the complete flight process.

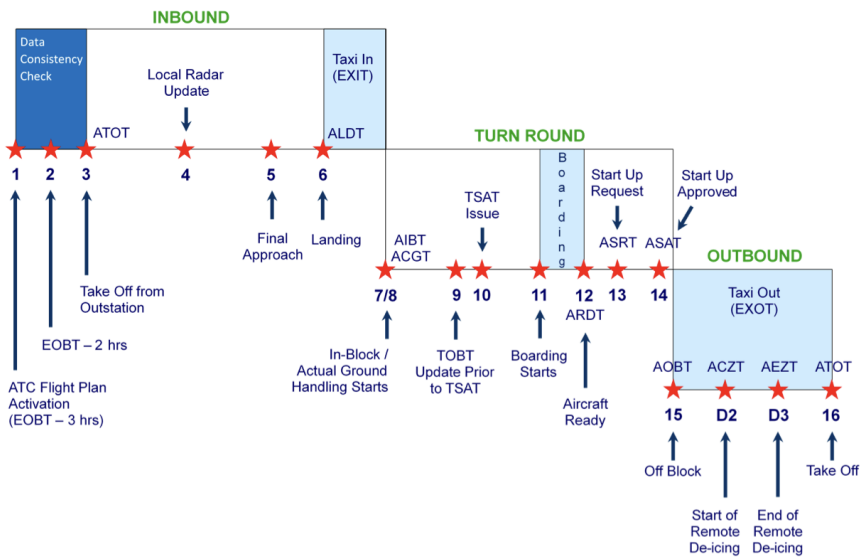


Figure 2.3: A-CDM Milestones

The A-CDM process begins at Milestone 1, three hours before the Estimated Off-Block Time (EOBT), where the flight plan is filed and the flight becomes visible in the A-CDM system. One hour later, at Milestone 2, full flight data becomes available to all A-CDM partners and, if demand exceeds capacity in a certain region, the Network Manager assigns a Calculated Take-Off Time (CTOT). The process continues with Milestone 3, which is triggered when the aircraft takes off from the origin airport. Milestone 4 occurs when the aircraft enters the relevant Flight Information Region (FIR) of the destination airport and becomes visible to ATC surveillance systems, for Schiphol Airport this corresponds to entry into the Dutch FIR. Milestone 5 marks the moment the aircraft enters final approach based on ATC instructions, followed by Milestone 6 which is the Actual Landing Time (ALDT), when the aircraft touches down. Milestone 7 is the moment the aircraft arrives at the gate, representing the Actual In-Block Time (AIBT). This also marks the beginning of the turnaround phase. Milestone 8 denotes at the same moment the start of ground-handling activities such as catering, cleaning and baggage services, this milestone is not implemented at Schiphol Airport. During the turnaround phase, the aircraft operator or ground handler updates the Target Off-Block Time (TOBT), this moment is defined as Milestone 9. Milestone 10 follows when ATC issues the Target Start-Up Approval Time (TSAT). Milestone 11 marks the beginning of boarding as passengers, and Milestone 12 indicates that the aircraft is technically ready for departure. The process continues with Milestone 13, when the pilot requests start-up clearance (ASRT), and Milestone 14, when ATC approves this request (ASAT), which is not used at Schiphol. Milestone 15 records the Actual Off-Block Time (AOBT) when the aircraft pushes back from the stand, and finally Milestone 16 captures the Actual Take-Off Time (ATOT), marking the aircraft's departure and the completion of the A-CDM sequence.

2.3. Decision Support Tool

The Area Control Center Supervisors (ACC-SUP) and the Flow Management Position Controllers (FMPC) at LVNL have gained an increasingly important role in day to day operational decision making due to the growth in air traffic over recent years. They are required to make frequent decisions to align the expected traffic demand with the available capacity. Several systems provide insight into predicted traffic demand and sector workload, such as the Network Manager (NM) CIFLO tool. However, the use of these systems comes with several limitations: supervisors cannot steer on measures other than ATFM regulations, they lack the ability to estimate airspace or stack delay under changing conditions, scenario testing options are extremely limited, and there is no clear insight into how simulated measures would affect controller workload. These constraints make it difficult to evaluate decisions in advance and to assess whether chosen measures were effective afterwards. With these shortcomings in mind, LVNL has introduced a new tool called the Decision Support Tool (DST), intended to provide ACC-SUP and FMPC with better decision-support information and help them make more effective use of the available capacity.

To illustrate how DST provides supervisors with insight into both traffic demand and available capacity, an example of the monitoring view is shown below. Figure 2.4 shows the DST monitoring view for all sectors in the Dutch airspace. The display consists of two graphs that together indicate whether the expected traffic demand fits within the available capacity. The upper graph shows the expected traffic load, where each bar represents the number of flights that are expected to be in the Dutch airspace during that specific time interval. The bars are divided into airborne, planned and regulated traffic, giving supervisors a clear view of the total load. The capacity is shown using two forecast lines: the CFS (Capacity Forecast Schiphol), representing the planned capacity based on declared runway and sector rates, and the CFD (Capacity Forecast DST), which is dynamically calculated by DST. When the total load exceeds one of these capacity lines, it indicates that traffic demand is higher than the available runway or sector capacity. The lower graph displays In-FIR delay, representing the number of aircraft that are expected to experience airborne delay within the Dutch airspace. These delays typically increase when the load in the upper graph exceeds the available capacity.

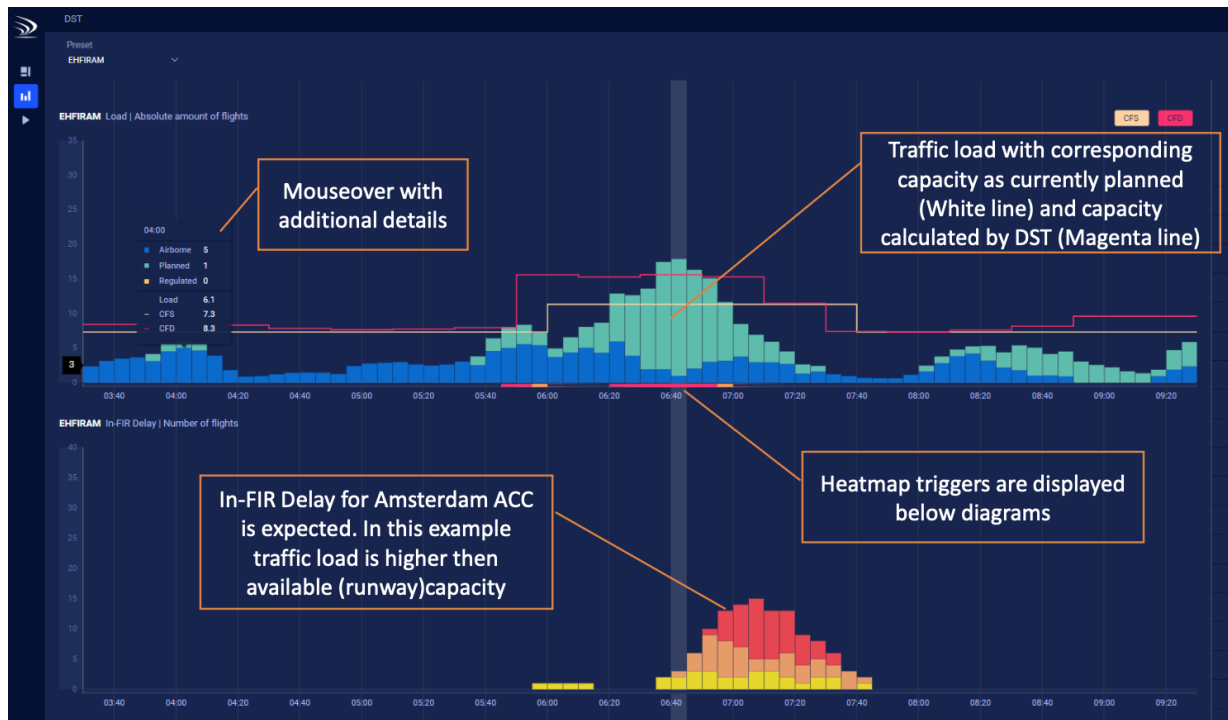


Figure 2.4: DST example

DST uses rolling time windows to represent traffic demand. In the example shown in Figure 2.4, a 20 rolling 5 setting is used. This means that demand is evaluated in five-minute time bins, where the demand of each bin is calculated over the 20-minute window that starts at the beginning of that bin. For example, the demand of the bin from 04:00 to 04:05 is calculated over the window from 04:00 to 04:20, while the demand of the next bin, from 04:05 to 04:10, is calculated over the window from 04:05 to 04:25. Flights that fall entirely within a window are counted fully, while flights that cross the start or end of the window are counted fractionally. Because consecutive windows overlap, a single aircraft can contribute to multiple consecutive bins. This rolling window approach reduces sensitivity to bin boundaries and results in a smoother and more realistic representation of the traffic density.

Although the DST provides supervisors with a forward-looking view of traffic demand and available capacity, its demand predictions are constructed from the latest available flight-plan estimates for each flight individually. The tool does not explicitly model how delay propagates through aircraft rotations, so delays accumulated earlier in a rotation are not necessarily reflected in the predicted take-off and landing times of later flight legs. This limitation forms the starting point of the problem statement in Chapter 3.

Problem Statement

According to EUROCONTROL [5], the average delay per flight in 2024 was 17.5 minutes, while arrival punctuality reached 72.4 percent, meaning that nearly three quarters of all flights arrived within 15 minutes of their scheduled arrival time. Although delay may originate from different sources such as weather conditions, ATC regulations or airport capacity constraints, the dominant contributor to total delay is reactionary delay, which accounts for 46 percent of all delay minutes and corresponds to approximately 8 minutes per flight. Even though reactionary delay represents the majority of all delays in Europe, LVNL currently has no model that predicts how delay propagates over time. The propagation of delays is a complicated process that depends on a number of variables. From an airline perspective, delay propagation is driven by the structure and interconnectedness of flight chains. Aircraft rotations, crew shifts and tightly scheduled connections create dependencies where delay on one leg can carry over to the next flights. From an airport perspective, several operational factors influence how delays propagate. Airport capacity, weather conditions, route congestion and runway configuration are often identified as key drivers of delay propagation between airports, as they determine how much delay can be absorbed before operations become constrained. Airports with high connectivity or tightly coordinated arrival and departure waves are particularly sensitive to disturbances because concentrated traffic and limited capacity leave little room to absorb delay times.

A further complication is that delay does not change uniformly across the different phases of a flight. Each phase has its own scheduling buffer, and the size of these buffers can vary considerably. This variation is influenced by factors such as aircraft type, airline schedules, airport procedures and operational conditions [6]. Because each phase can absorb delay differently, predicting how delay will propagate becomes more complex.

Schiphol is one of Europe's major hub airports, and hub airports experience higher levels of delay than non-hub airports [7]. This makes delay propagation particularly relevant at Schiphol. As one of the 34 A-CDM airports in Europe, EUROCONTROL provides Schiphol and LVNL with A-CDM data for these airports. However, there are more than 300 European airports, which means that there is no A-CDM data available for a part of the traffic. This requires a different approach for those flights and makes the modelling task more complex. These challenges lead to the following problem statement:

There is currently no model at LVNL that predicts how delay propagates over time within the European network, with specific focus on Schiphol Airport, to support area control center supervisors and the flow management position controllers in making better operational decisions.

Operational and Data Concepts

This chapter introduces the operational and data concepts used in this thesis. Section 4.1 explains how delays are classified using the standardized CODA categories. Section 4.2 describes the three main flight phases namely: turnaround, taxi, and airborne which are essential for understanding where delay can be absorbed or amplified. Section 4.3 discusses weather data, and section 4.4 presents ADS-B data as a source for analyzing airborne behaviour.

4.1. Delay Causes

EUROCONTROL and IATA use a standardized delay coding framework which is applied by the Central Office for Delay Analysis (CODA) to classify all causes of delay across the European network. The framework divides delay into seven categories namely: Airline, Airport, ATFM En-route, Governmental, Weather, Miscellaneous, and Reactionary. Figure 4.1 illustrates the distribution of average delay per flight for the year 2023 and 2024 [5].

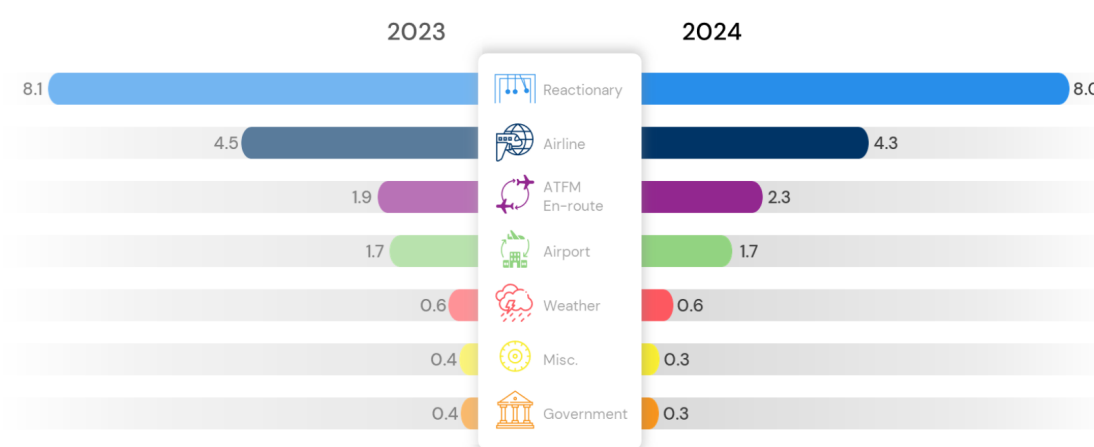


Figure 4.1: Distribution of the average delay per flight [5]

Each primary delay category reflects a different source of operational disruption within the European network. Airline delay arises from factors under the airline responsibility, such as ground handling efficiency, aircraft turnaround performance, technical issues, or crew related problems. Airport delay originates from local aerodrome conditions, including capacity limitations, gate or stand shortages, runway availability, and other infrastructure restrictions. ATFM en-route delay occurs when airspace sectors cannot meet planned demand due to limited capacity, staffing issues, military activity, or system transitions, often creating wide area effects because multiple traffic flows are affected at the same time. Weather delay refers to meteorological conditions such as strong winds, low visibility, thunderstorms or snow that reduce the performance of aircraft and limit the capacity of both airports and airspace. Governmental and miscellaneous delays represent a smaller part of the total delay and include regulatory checks, security

processes or other exceptional circumstances. Together, these primary delay types form the basis for the reactionary delays that develop in the network.

4.2. Flight Phases

A flight consists of several operational phases, all with their own characteristics and sources of delay. These phases can absorb or amplify delays in different ways. In this section, the flight phases are discussed separately.

4.2.1. Turnaround Phase

During the turnaround phase, a lot of ground handling activities take place. Once the aircraft arrives at the gate, passengers and baggage are unloaded, the cabin is cleaned, catering is restocked, technical checks are performed and the aircraft is prepared and boarded for the next flight. Turnaround time can be described in different ways depending on the time reference used. The first one is the scheduled turnaround time, which refers to the interval between the Scheduled In-Block Time (SIBT) and the Scheduled Off-Block Time (SOBT). This duration is determined in the airline schedule planning and is the turnaround time that the airline intends to achieve under normal operational conditions. The second definition is the expected turnaround time, which is based on the Estimated In-Block Time (EIBT) and Estimated Off-Block Time (EOBT), the expected values change dynamically depending on the operations. Lastly, the actual turnaround time reflects the true operational duration, it starts from the Actual In-Block Time (AIBT) and ends at the Actual Off-Block Time (AOBT). This value captures how long the aircraft actually remained at the stand. The three turnaround time definitions are formulated below.

$$\text{Scheduled Turnaround Time} = \text{SOBT} - \text{SIBT} \quad (4.1)$$

$$\text{Expected Turnaround Time} = \text{EOBT} - \text{EIBT} \quad (4.2)$$

$$\text{Actual Turnaround Time} = \text{AOBT} - \text{AIBT} \quad (4.3)$$

Airlines also include buffer time in their turnaround planning to account for delay. This buffer is the margin between the scheduled turnaround time and the minimum turnaround time required. In addition, the delay in the turnaround phase can be determined by comparing the actual turnaround duration with the scheduled one. Both buffer time and delay are described in the functions below.

$$\text{Turnaround Buffer} = \text{Scheduled Turnaround Time} - \text{Minimum Turnaround Time} \quad (4.4)$$

$$\text{Turnaround Delay} = \text{Actual Turnaround Time} - \text{Scheduled Turnaround Time} \quad (4.5)$$

4.2.2. Taxi Phase

Taxi operations differ between airports. At smaller airports, aircraft often follow the shortest path to the runway with minimal need for strategic planning. This is not the case at larger and more complex airports such as Schiphol, where high traffic density, intersecting taxiways and multiple runway configurations require coordinated sequencing and routing. These operational characteristics result in a wide variation of taxi times across flights.

Within the A-CDM framework, taxi times are defined using two indicators. The arrival taxi time is represented by the Estimated Taxi-In Time (EXIT), and the departure taxi time is represented by the Estimated Taxi-Out Time (EXOT). Both can be formulated as follows:

$$\text{EXIT} = \text{EIBT} - \text{ELDT/ALDT} \quad (4.6)$$

$$\text{EXOT} = \text{ETOT/TTOT/CTOT} - \text{EOBT/TOBT} \quad (4.7)$$

The assignment of gates has a direct impact on taxi times because it determines the destination of the aircraft within the airport layout [8]. Airlines also have preferred gates for operational or commercial reasons, creating patterns in taxi routes for specific airline–airport combinations [9].

4.2.3. Airborne Phase

The airborne phase represents the time the aircraft spends in the air between take off and landing. This phase is less constrained by airport layout compared to the turnaround and taxi phases, however airborne time can still vary significantly. Factors such as wind conditions, weather deviations, airspace structure, ATC instructions and traffic density all influence the duration of the flight. As a result, the airborne phase can either absorb delay or generate additional delay. Airborne time can be expressed through scheduled, expected and actual values. The actual airborne time is the interval between the ATOT and the ALDT. The scheduled airborne time is defined in the airline schedule using the Scheduled Take-Off Time (STOT) and Scheduled Landing Time (SLDT). The expected airborne time is based on the estimated flight plan values, such as ETOT and ELDT, which reflect the predicted time.

These three different expressions are summarized below:

$$\text{Actual Airborne Time} = \text{ALDT} - \text{ATOT} \quad (4.8)$$

$$\text{Scheduled Airborne Time} = \text{SLDT} - \text{STOT} \quad (4.9)$$

$$\text{Expected Airborne Time} = \text{ELDT} - \text{ETOT} \quad (4.10)$$

Airborne delay is determined by comparing the actual airborne time with the scheduled airborne time. And is formulated as follows:

$$\text{Airborne Delay} = \text{Actual Airborne Time} - \text{Scheduled Airborne Time} \quad (4.11)$$

4.3. Weather Data

Weather is one of the primary delay categories defined by CODA [5]. Numerous studies identify weather as one of the dominant drivers of both primary and reactionary delay across the air transport system, influencing taxi operations, airborne trajectories, runway capacity, and ATFM regulations [10, 11, 12]. Meteorological Aerodrome Reports (METAR) give hourly observations of the local weather conditions at airports. The reports are generated by weather stations and contain detailed information on atmospheric conditions that can directly affect airport and airspace operations. For this thesis, the following METAR variables are of particular interest:

- **Wind:** Speed, direction and gusts, which affect runway configuration and can constrain take-off and landing operations.
- **Visibility:** Reduced visibility often leads to operational restrictions.
- **Weather phenomena:** Current conditions such as rain, snow, fog or thunderstorms, each of which can introduce delays or require additional procedures.
- **Sky condition:** Cloud coverage and cloud base altitude, both relevant for approach and departure performance.
- **Temperature and dew point:** Important for aircraft performance and for determining when de-icing procedures are required.

4.4. ADS-B Data

Automatic Dependent Surveillance–Broadcast (ADS-B) is a surveillance technology in which aircraft transmit their position and movement information. Because these messages are broadcast directly from the aircraft and can be received by a wide network of ground stations, ADS-B provides high resolution and high frequency observations of airborne behaviour [13]. In the context of delay research, ADS-B is widely used to analyze airborne trajectories and approach or departure behaviour. This makes it a useful data source for estimating how long an aircraft will remain in the air, whether it is likely to arrive on time, and how airborne conditions influence overall delay evolution. Recent studies in trajectory prediction and delay forecasting often rely on ADS-B because its continuous positional updates allow accurate tracking of airborne performance throughout the en-route and arrival phases.

State of the Art

In this chapter, the current state of the art in flight delay propagation modelling is reviewed. The focus is on how existing methods approach the propagation of delay across flights, operational phases, airports, and wider air traffic networks. The literature shows a broad range of modelling strategies ranging from analytical formulations to data driven and network based approaches, each offering different levels of interpretability, scalability, and predictive capability.

Across the reviewed work, several methods emerge. Early analytical and econometric models provide insight into fundamental propagation mechanisms but lack the ability to represent network complexity. Statistical approaches capture temporal patterns and distributions of delay but remain limited in modelling leg-to-leg interactions. Bayesian networks introduce probabilistic causality and can represent operational dependencies, though they rely heavily on predefined structures. More recent machine learning and deep learning methods exploit large historical datasets to model non-linear and spatial-temporal relationships, including airport interactions, network spillovers, and dynamic propagation patterns.

This chapter is structured as follows. First, Section 5.1 presents economic models that characterize delay propagation using simplified mechanisms. Section 5.2 then discusses statistical approaches, including classical regression, survival models, distribution based methods and Bayesian networks. Finally, Section 5.3 reviews machine learning and deep learning methods including supervised learning, recurrent neural networks, and graph-based models followed by a feature overview that summaries the input variables commonly used in the reviewed literature.

5.1. Economic Models

Economic models are used in economics to describe how different factors are related and to analyse how changes in one variable can influence another. They combine economic theory, mathematics, and statistical methods to study and measure relationships within an economic system. The goal of economic models is to make theoretical ideas measurable and to better understand real-world economic behaviour [14].

In the field of flight delay propagation, the term economic model is used in a broader and more applied way. These models do not study markets or prices, but are used to describe and explain how delays develop and spread within the air transport system [15]. Economic models often use real flight data and simple econometric methods, such as regression analysis, to validate relationships between operational factors and the extent of delay propagation. The main goal of these models is not to predict exact delay values but to understand why and how delays are transferred from one flight to another within the network [16].

Beatty et al. (1999) [17] introduced one of the first quantitative models of delay propagation by developing the Delay Multiplier (DM). This indicator measures how an initial delay spreads to later flights in an airline schedule. Using flight data from United Airlines, Beatty et al. showed that early delays in the day cause more reactionary delay, while later delays are often absorbed. The DM provided the first simple way to describe how flight delays propagate through the network. Eighteen years later Belkoura et al. (2017) [18] improved this idea by introducing a nonlinear delay multiplier. Using European flight data, they showed that the relationship between inbound and outbound delays is not always proportional.

Kafle and Zou (2016) [6] advanced the study of delay propagation by developing an econometric model that distinguishes between propagated and newly formed delays within an aircraft's daily flight sequence.

Their approach combines an analytical framework, which defines how delays move through successive flight legs depending on flight and ground buffers, with an econometric model that estimates the influence of operational and external factors on this process. Using flight data from the U.S. Bureau of Transportation Statistics, they identified buffers as the key variable affecting the extent to which delays are absorbed or passed on. To avoid the bias of simple regressions, the authors applied a latent variable framework based on a discrete choice model and estimated the coefficients using both the Heckman two-step and Full Information Maximum Likelihood (FIML) methods. The results show that about two-thirds of departure delays originate from previous flights, while the rest are newly formed due to external factors such as weather or congestion. Propagation effects also differ between airports, and whether it is arriving or departing, suggesting that local network structures and operational policies play a major role in how delays spread.

As summarized in Table 5.1, economic models remain useful because they make the basic mechanisms of delay propagation clear, including delay multipliers and the distinction between propagated and newly formed delay. Their interpretability helps explain how delays spread through the air transport system. However, their simplifying assumptions, limited scalability, and lack of explicit network representation make them unsuitable for operational delay propagation prediction. Overall, these models serve mainly as conceptual insight rather than as a practical framework for detailed or real-time forecasting.

Authors / Year	Objective	Method	Application
Beatty et al. (1999)	Measure how initial delay spreads through airline schedules.	Delay Multiplier (linear econometric indicator).	Single airline (United Airlines); propagation across sequential flights.
Belkoura et al. (2017)	Make delay-propagation measurement more realistic using European data.	Nonlinear Delay Multiplier; regression of inbound-outbound relationships.	Network / multiairport (EUROCONTROL); multi-airline interactions.
Kafle & Zou (2016)	Differentiate propagated vs. newly formed delays within aircraft rotations.	Analytical/econometric model with latent variables (Heckman two-step / FIML).	From aircraft rotations to airport/network propagation; not real-time.

Table 5.1: Economic models: authors, objective, method, and application level.

5.2. Statistical Models

While economic models focus on explaining the underlying mechanisms of delay propagation, statistical models aim to quantify relationships and identify patterns using probabilistic methods. These models are mainly used to measure the impact of different factors on delays and to describe how delay propagation follows certain statistical distributions.

5.2.1. Classical Statistical Models

Classical statistical models describe flight delay propagation in terms of patterns and probabilities rather than causal mechanisms. These models use statistical techniques to identify how delay build up and persist over time, and to estimate the influence of operational and environmental factors.

Tu et al. (2008) [19] developed one of the first statistical approaches to study how departure delays change over time. They separated the total delay into three parts: a long-term seasonal trend, a daily pattern showing how delays increase throughout the day, and a random component. Using data from United Airlines at Denver Airport, they found clear seasonal and daily variations, with higher delays in summer and an accumulation of delays during the afternoon. This study showed how delays tend to build up as the day progresses, although it did not measure how one flight's delay directly affects the next.

Wong and Tsai (2012) [10] introduced a different approach based on survival analysis. Instead of focusing on average delay times, they examined how long a delay tends to last within the daily schedule of a single aircraft. Their model linked each flight's departure delay to the arrival delay of the previous flight and to the available turnaround time. Using one year of operational data from UNI Air in Taiwan, they found that larger turnaround buffers increase the chance of recovering from a delay, while bad weather and technical problems make recovery less likely. Later, Kim and Bae (2021) [20] used a similar survival analysis for South Korean flights. They confirmed that turnaround time plays a crucial role in delay recovery. When the actual turnaround time is short, having a larger buffer helps reduce the persistence of delays. They also observed that international flights tend to have longer delays than domestic ones, showing differences in operations and scheduling flexibility.

Zheng et al. (2020) [11] adopted a quantile regression approach to study how different factors affect not only average delays but also short and extreme delay situations. Ordinary Least Squares (OLS) is a regression method that estimates the average relationship between variables, while quantile regression allows researchers to focus on different points in the delay distribution, such as very long delays. Using data from Shanghai Pudong Airport, they found that ground buffers help reduce short delays but are less effective for longer delays. In those cases, factors like weather and congestion dominate.

Nayak and Zhang (2011) [12] analysed how delays at one airport can influence others in the United States. They used a multivariate statistical model to estimate how delays spread between 34 major airports. Their results showed strong spillover effects, meaning that a delay at one airport does not remain isolated but spreads to other airports through interconnected flight schedules, especially under bad weather or heavy congestion.

Overall, classical statistical models focus on describing how delays behave, how long they last, how they vary across conditions, and how they are distributed, rather than on explaining why they occur. Compared with economic models, they provide a more data driven and probabilistic view of delay propagation, focusing on measurable relationships instead of theoretical ones. The statistical models discussed above are summarized in Table 5.2.

Authors / Year	Objective	Method	Application / Key findings
Tu et al. (2008)	Study how departure delays vary over time.	Nonparametric spline estimation of seasonal and daily trends with random residuals.	United Airlines (Denver). Clear seasonal pattern and daily build-up of delays; not direct leg-to-leg propagation.
Wong & Tsai (2012)	Examine how long delays persist within daily aircraft schedules.	Survival analysis (Cox model) linking previous arrival delay and turnaround time.	UNI Air (Taiwan). Larger turnaround buffers help recovery; weather and technical issues extend delays.
Kim & Bae (2021)	Develop a survival-based model to explain and predict departure delay duration using operational and flight characteristics.	Survival analysis with actual turnaround buffer as main variable.	South Korea. Shorter turnaround increases persistence, international flights show longer delays.
Zheng et al. (2020)	Analyze how factors affect different levels of delay.	Ordinary Least Squares (OLS) and quantile regression.	Shanghai Pudong Airport. Buffers help short delays, severe delays dominated by weather and congestion.
Nayak & Zhang (2011)	Evaluate how airport delays influence each other across the system.	Multivariate regression model (system-wide).	U.S. network. Significant spillover between airports, weather and congestion amplify propagation.

Table 5.2: Summary of classical statistical models for flight delay propagation.

5.2.2. Bayesian Networks

Bayesian networks (BN) are probabilistic graphical models that represent dependencies between different operational factors. Each node in the network corresponds to a variable such as arrival delay, turnaround delay, or weather, while directed links describe how these variables influence one another. This makes BN particularly suitable for modelling flight delay propagation, where delays evolve through multiple interconnected processes [15].

Typically, a BN is developed in three main steps [15]. The first step is the creation of variables, defining which operational components and external factors are included. The second step is structure learning, where the causal links between variables are determined. This can be based on expert judgment or on data driven algorithms such as K2. The third step is parameter learning, which estimates the conditional probabilities between variables. This process is often implemented through the Expectation Maximization (EM) algorithm. Most studies build their models in Netica, a software package commonly used for probabilistic network modelling [21, 22, 23, 24].

Xu et al. (2005) [25] were among the first to apply Bayesian networks to model how flight delays propagate between airports in the U.S. National Aviation System. Their study analysed flight connections between Chicago O'Hare (ORD) and LaGuardia (LGA) as origin airports and Atlanta (ATL) as destination airport. Using delay related variables such as departure delay, arrival delay, weather, time of day, and flight cancellations, three estimation approaches were compared. The first used the Dirichlet multinomial model to estimate the conditional probability distributions of each node in the network directly from data, representing a purely data driven form of Bayesian parameter learning. The second applied regression based estimation to incorporate prior knowledge about relationships between delay variables at connected airports. The third approach combined both methods by using regression results as prior expectations and

refining them through Dirichlet updating based on historical flight data. This combined approach provided the highest predictive accuracy. The model, showed that departure delays at hub airports such as ORD strongly influence downstream arrival delays, while adverse weather and cancellations at the destination airport ATL further amplify delay propagation across the network.

Laskey et al. (2012) [21] refined the Bayesian approach by developing a structured network that captures how delays evolve through different operational stages of a flight, as illustrated in Figure 5.1. The model incorporated influencing factors such as weather and traffic volume, with the structure defined using expert knowledge and the parameters estimated in Netica through Dirichlet multinomial learning.

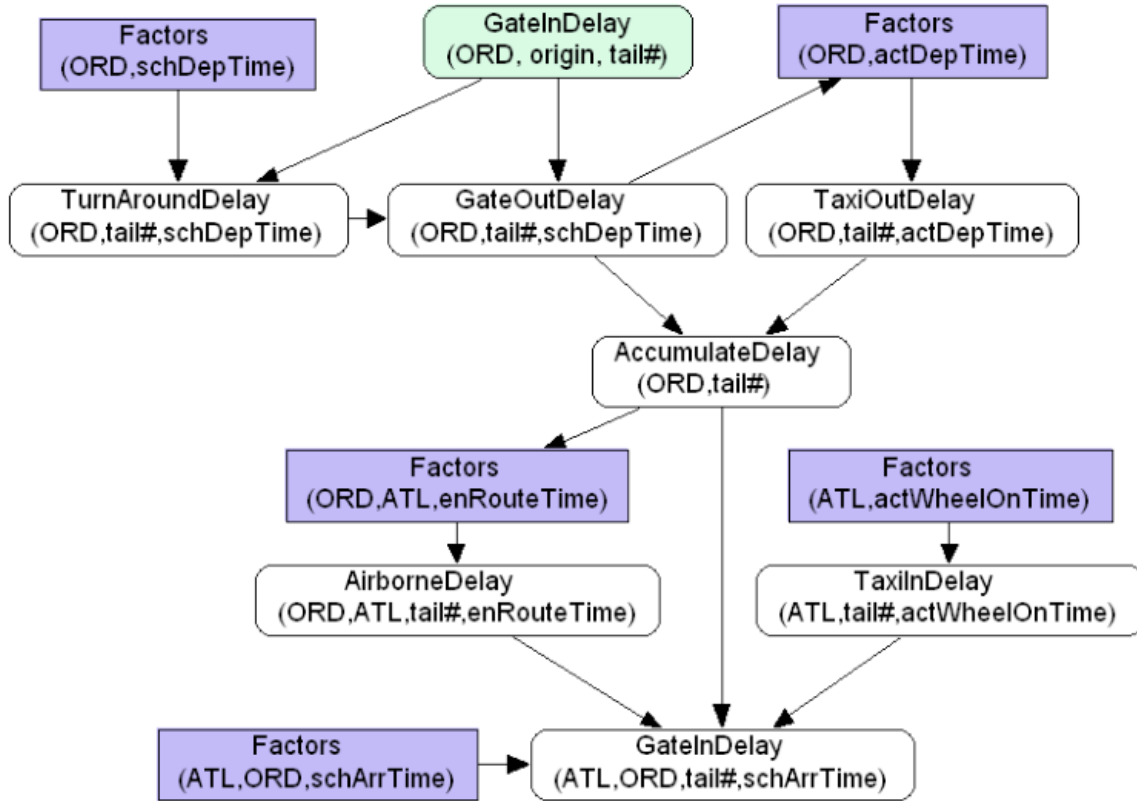


Figure 5.1: Bayesian Network structure [21]

Liu and Ma (2008) [22] studied how delays are transferred between consecutive flight legs operated by the same aircraft. They developed a Bayesian network model for a reduced flight chain and compared three different ways of training it. The first method only estimated the parameters, the second learned the structure of the network using the K2 algorithm, and the third combined both methods. The combined learning approach first determined the structure of the network from the data and then estimated the parameters using the EM algorithm. This approach achieved the highest accuracy. The results showed that Bayesian networks can not only predict delay propagation but also identify key influencing factors, such as turnaround time and airport congestion. This demonstrates how learning based models can uncover relationships between operational variables that are not directly visible in raw data.

More recent studies by Wu and Wu (2018) [23] and Wu and Law (2019) [24] extended Bayesian approaches to include multiple types of resources and probabilistic delay profiles. Wu and Wu integrated Bayesian networks with a Delay Propagation Tree (DPT), modelling delay as a probability distribution rather than a single value. This approach captures how variations in upstream delays influence the probability of downstream disruptions and allows for scenario analysis such as testing different buffer settings. Wu and Law expanded this framework further by incorporating multiple propagation sources like aircraft, crew, cabin crew, and passengers within one unified BN. They introduced two new measures: the Expected

Delay Multiplier (DMe), representing the expected downstream delay given a certain condition, and the Expected Propagated Delay Multiplier (DMx), which quantifies how updated information changes those expectations. These developments demonstrate how Bayesian networks can model complex resource interconnections and reveal weak links where small disruptions can escalate into larger system wide delays.

Bayesian networks are effective at representing dependencies between operational factors that drive delay propagation. Their probabilistic structure makes it possible to reason about uncertainty and to trace how one disruption can influence others. However, the models rely on probability distributions and predefined causal structures, which limits their ability to capture the full complexity and variability of real operations. In practice, they function more as probabilistic explanation tools than predictive models. The main Bayesian models are summarized in Table 5.3.

Authors / Year	Objective	Method	Application / Key findings
Xu et al. (2005)	Model delay propagation between operational phases.	Bayesian network with regression-based priors; EM parameter learning in Netica.	ORD–ATL flights. Early departure and weather effects strongly influence downstream delays.
Laskey et al. (2012)	Extend the phase-based BN to multiple airports within the NAS.	Hybrid structure combining expert knowledge and data learning.	U.S. network. Integration of regression priors improves accuracy and interpretability.
Liu & Ma (2008)	Predict delay transfer within aircraft flight chains using data-driven learning.	K2 structure learning combined with EM parameter estimation.	Chinese hub airport. Mixed learning improves accuracy; turnaround and congestion are key drivers.
Wu & Wu (2018)	Represent probabilistic delay propagation and buffer effects.	Delay Propagation Tree (DPT) integrated with Bayesian inference.	Single airline network. Probabilistic modelling identifies heterogeneous buffer impacts.
Wu & Law (2019)	Analyze multi-resource delay propagation in airline operations.	Extended BN including aircraft, crew, and passenger connections.	Airline network. New DMe and DMx metrics identify weak links and major propagation paths.

Table 5.3: Summary of BN models for flight delay propagation.

5.3. Machine Learning

In recent years, machine learning (ML) has emerged as one of the dominant approaches for modelling and forecasting flight delay propagation. Traditional statistical and simulation based methods often struggle to capture the nonlinear and highly interconnected nature of the air traffic system. As highlighted by Li et al. (2024) [15], “In the selection of modelling methods for flight delay propagation, aligning with available data types is paramount. Machine learning and deep learning excel with large, complex datasets, adept at revealing hidden patterns through their capacity to analyse vast and intricate data structures.”

Machine learning models are well suited for this context because they can learn relationships directly from historical data without requiring physical or probabilistic assumptions about how delay propagates. These models can detect nonlinear dependencies between airports, flight schedules, and environmental factors, which makes them powerful tools for prediction [26]. However, a common drawback of machine learning models is that they often function as a “black box”, which means that their internal decision process is not directly interpretable. Although these models can achieve high predictive accuracy, they provide limited insight into the causal mechanisms that drive delay propagation [15, 16]. This section is divided in two subsections, namely supervised learning and deep learning.

5.3.1. Supervised Learning

A common supervised learning algorithm is the Decision Tree (DT). A DT is a non parametric model that divides data into smaller subgroups based on a series of decision rules derived from input features. The structure of a DT includes a root node, internal nodes, and leaf nodes. The root node represents the first split of all data, internal nodes correspond to subsequent splits, and the leaf nodes contain the final prediction for each subset of data. Each branch connecting the nodes represents a possible outcome of a rule and can be interpreted as an if–then statement. The algorithm continues to split the data until stopping criteria are met. The result is the final tree that represents a set of simple decision rules that map input conditions to a predicted outcome [27]. An example of a DT is shown in Figure 5.2.

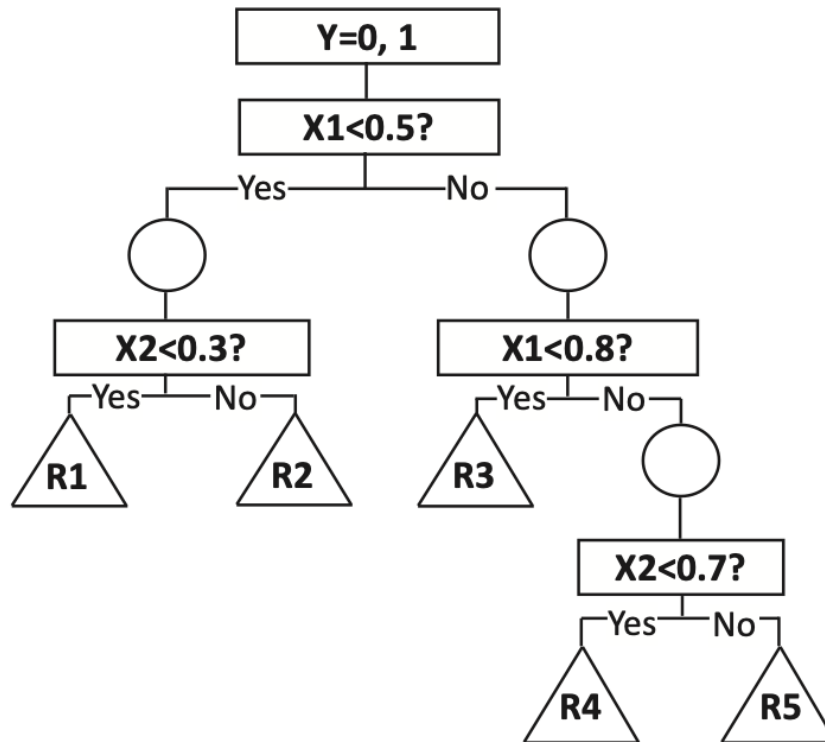


Figure 5.2: Decision Tree example [27]

To improve predictive power and robustness, ensemble methods have been developed that combine multiple trees into a single model, reducing overfitting and increasing accuracy. One such method is the Random Forest (RF), which builds multiple Decision Trees, each trained on different subsets of features and data samples. The predictions from all trees are then combined through a voting process, where the majority vote for the final prediction, resulting in a more stable and accurate model [28]. Another ensemble method is the Gradient Boosted Decision Tree (GBDT), in which multiple Decision Trees are arranged in series rather than built independently. Each tree is trained to correct the errors made by the previous ones, allowing the model to progressively improve its predictions [29].

Rebollo and Balakrishnan (2014) [30] developed one of the first supervised learning models for predicting flight delay propagation across the U.S. National Airspace System (NAS). Their approach was a network based one, predicting the average departure delay for specific origin-destination routes rather than for individual flights or single airports. The input data were derived from the FAA's Aviation System Performance Metrics (ASPM) database and included temporal features (time of day, day of week, and season) as well as network level delay indicators such as the current NAS delay state and delay levels at neighboring airports. The model's output represented the expected average departure delay for a given route, using forecast horizons of 2 and 24 hours. Two model types were tested: a RF regression model predicting the number of delay minutes, and a RF classification model determining whether a delay would exceed 60 minutes. The models were trained using 80% of the data and evaluated on the remaining 20%. Table 5.4 presents the main performance results for both the regression and classification models.

Model Type	Horizon	Median Error (min)	Classification Error (%)
Random Forest (Regression)	2 hours	21.0	–
Random Forest (Regression)	24 hours	27.0	–
Random Forest (Classification)	2 hours	–	19.0

Table 5.4: Performance of RF models for delay propagation prediction (Rebollo & Balakrishnan, 2014) [30].

Chen and Li (2019) [31] developed a mixed approach model that combines a RF algorithm with a chained delay propagation mechanism to predict delays on a flight level. Their model uses the predicted delay of one flight as input for the next flight operated by the same aircraft, allowing it to capture how delays propagate through consecutive operations. The input data consisted of flight, weather, and operational features derived from the Bureau of Transportation Statistics (BTS), NOAA, and ASPM databases, including tail number, flight schedule, weather conditions, and previous delay states. The output was the predicted delay in minutes for each flight, updated iteratively across the daily rotation of the aircraft. The authors compared the chained RF model with a non-chained baseline and found that incorporating the propagation mechanism improved prediction accuracy by approximately 9 to 12%. One of the key findings of the Feature importance analysis revealed that the delay of the previous flight, airport congestion, and weather conditions were the strongest predictors of future delays.

Ye et al. (2020) [32] developed an airport based supervised learning framework to predict the average hourly departure delay at Nanjing Lukou International Airport. Instead of focusing on individual flights, their model aggregates flight data at the airport level, using a one hour forecasting horizon. To evaluate performance, the authors trained and compared four supervised learning algorithms, namely: Multiple Linear Regression (MLR), Support Vector Machine (SVM), Extremely Randomized Trees (ExtraRT), and Light Gradient Boosting Machine (LightGBM). All models were trained using 80% of the data and tested on the remaining 20%. The performance of each model, including mean absolute error (MAE), mean squared error (MSE), and validation accuracy (R^2), is summarized in Table 5.5.

Model	MAE (min)	MSE	Validation Accuracy (R^2)
MLR	6.80	83.47	0.8625
SVM	6.69	82.56	0.8595
ExtraRT	6.82	83.24	0.8623
LightGBM	6.64	81.27	0.8653

Table 5.5: Performance comparison of supervised learning models (Ye et al., 2020).

Dalmau et al. (2021) [33] developed a LightGBM model for EUROCONTROL to predict the difference between ATOT and ETOT within the European ATFM network. The model uses operational data from the ETFMS, including milestones such as EOBT, AOBT, ETOT, ATOT, and various operational, regulatory, and weather related features. The model improved prediction accuracy compared to ETFMS, reducing MAE from 14 to 23 min to about 9 to 15 min depending on the forecast horizon, with the largest improvement of 34% for horizons of three to four hours before departure. The most influential features were the ATFM delay and available turnaround time, suggesting that the model is able to capture reactionary delay effects, as a higher delay on the previous leg led to later predicted take off times relative to ETFMS, demonstrating its capability to model propagation across consecutive flights.

In summary, LightGBM was found to provide the best results among the reviewed supervised learning models. In these studies, LightGBM was chosen for its ability to handle large datasets and its speed in both training and prediction, making it suitable for real-time delay forecasting. LightGBM's high efficiency and accuracy make it particularly effective for predictive modelling of flight delays, where rapid data updates are essential. However, a potential limitation is that tuning the model for optimal performance can be complex and requires significant expertise. Moreover, LightGBM is known to exhibit hyperparameter sensitivity, where small changes in configuration can substantially affect model performance [34].

5.3.2. Deep Learning

Artificial Neural Networks

Artificial Neural Networks (ANN) are computational models inspired by the way biological neurons process information. The model consist of interconnected nodes organized into an input layer, one or more hidden layers, and an output layer, as illustrated in Fig. 5.3. As information passes through these layers, the network adjusts its internal weights during training, allowing it to learn complex and nonlinear relationships between variables. ANNs are widely used in areas such as pattern recognition, classification, regression, and forecasting, largely because they can generalize from examples and perform well even when data relationships are difficult to model with traditional methods [35].

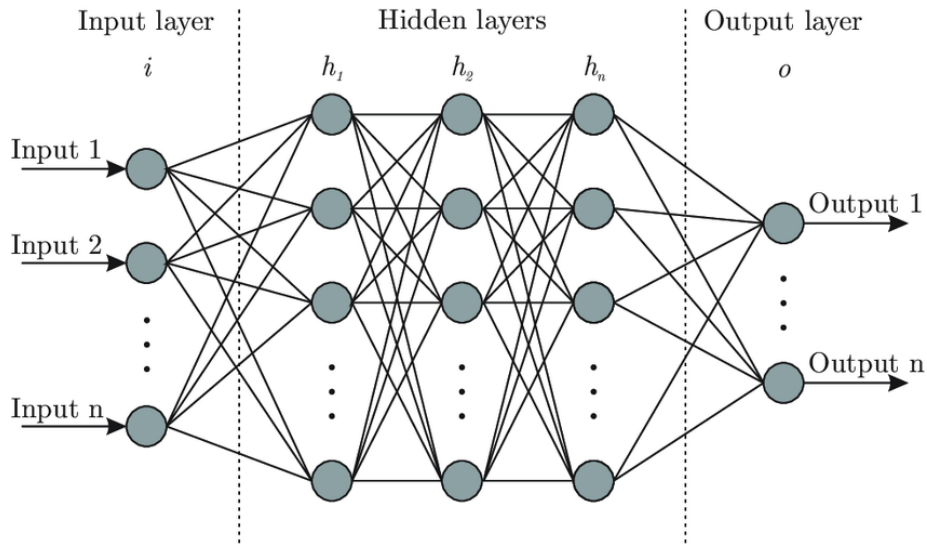


Figure 5.3: Artificial Neural Network [36]

Khanmohammadi et al. (2016) [37] developed a Multilevel Input Layer Artificial Neural Network (MLIL-ANN) model to predict the delay of inbound flights at John F. Kennedy (JFK) Airport. The model introduces a multilayered input structure that groups related features such as day of the week, airport ID, and delay type variables into separate input layers to better capture their relationships. The goal was to predict the arrival delay. However, the input also included both scheduled and actual arrival times, which already define the delay. This leads to data leakage, as the model effectively receives the answer it is meant to predict. The model achieved a Root Mean Squared Error (RMSE) of 0.1366, but it was evaluated on the same dataset used for training, making the result unreliable and influenced by overfitting.

Two years later, Pamplona et al. (2018) [38] further developed the Multilevel Input Layer Artificial Neural Network (MLIL-ANN) introduced by Khanmohammadi et al. (2016). While the original model predicted the arrival delay in minutes, Pamplona reformulated the problem as a binary classification, determining whether a flight would arrive on time or more than 15 minutes late. Similar to the previous study, this model also suffered from data leakage, as it included variables that indirectly contained information about the output. In addition, no weather data were considered, meaning the model captured only operational factors rather than environmental influences.

Recurrent Neural Networks

A Recurrent Neural Network (RNN) is a neural network designed to capture temporal dependencies in sequential data such as time series, text, or speech. Unlike feedforward networks, RNN models use feedback connections so that information from previous time steps influences the current output, giving the model a form of memory [39]. However, when RNNs are trained using backpropagation, the gradients can become extremely small or large as they propagate through many time steps. This leads to vanishing and exploding gradients, which prevents standard RNN models from learning long term patterns effectively. To overcome this limitation, the Long Short Term Memory (LSTM) model was introduced. LSTM models include a dedicated memory mechanism that stabilizes gradient flow, allowing the model to retain information over longer periods and learn long term dependencies much more reliably than traditional RNNs [39].

Sun et al. (2022) [40] applied an LSTM model to predict the average arrival and departure delays for a given airport. The model is designed to capture how delays evolve during daily operations by using multiple past time steps of operational data. Each feature, such as the number of arriving and departing flights, is represented as a time series, allowing the LSTM to learn the temporal patterns that influence delay evolution. The LSTM predicts arrival and departure delays for several future time steps, from 30 to 180 minutes ahead. Their results show that the LSTM performs consistently well: at London Heathrow Airport (EGLL), the MAE is approximately 4.6 to 5.4 minutes for arrival delays and around 2.2 to 2.3 minutes for departure delays across the different horizons. These results demonstrate that LSTM models can capture short-term airport-level delay dynamics.

Zhang et al. (2021) [41] developed an Attention PageRank LSTM-Seq2Seq (APR-LSTM) model to predict multi step airport delays. To model spatial dependencies between airports, they first applied the PageRank algorithm to quantify how delay at one airport influences delay across the network. The resulting spatial weights were then fed into an LSTM network to learn temporal dependencies, enabling the model to capture how delays evolve over time. An attention mechanism further highlighted which past time steps contributed most to current delay predictions. The model used features such as time of day, weather, and holiday indicators, and was compared with several baseline models including ARIMA, XGBoost, ConvLSTM, and ST-ResNet. The APR-LSTM achieved the best performance with a MAE of 3.93 minutes and RMSE of 9.24 minutes.

Graph Neural Networks

Graph Neural Networks (GNNs) are designed to learn from data that are structured as graphs, making them particularly suitable for air traffic networks where airports form nodes and flight connections represent edges. Unlike standard neural networks, which assume that the input samples are independent, GNNs allow the information to propagate through the graph from node to node so that the prediction of each node depends not only on its own features, but also on the state of its neighbors. This ability to model spatial dependencies makes GNNs well suited for delay propagation problems, where disruptions at one airport can influence delays across the network through connecting flights and shared airspace [42].

In the same study discussed in the previous subsection, Sun et al. (2022) [40] also developed a Dynamic spatial-temporal Graph Attention Network (DST-GAT) to predict airport arrival and departure delays across the European network. In this model, airports are represented as nodes, and the edges between the nodes encode spatial proximity and flight connections. A graph attention mechanism assigns adaptive weights to neighbouring airports, enabling the model to learn which nodes are most influential at each moment. Combined with a temporal modelling component, the model forecasts delays 30 to 180 minutes ahead and achieves a MAE of around or below 5 minutes for arrival delays and around 4 minutes for departure delays across the top 50 European airports, while additionally enabling the study of delay propagation at the network level.

Wu et al. (2024) [43] proposed a Spatiotemporal Propagation Network (STPN) to model how delays spread over time across large airport networks. Their approach incorporates multi graph convolution layers to capture different types of spatial relationships such as geographic distance, flight frequency, and operational interactions while temporal convolutions learn short- and long term delay dynamics. A self attention mechanism is added to capture shifting propagation patterns under different traffic conditions. The model predicts network wide delays several hours ahead and shows clear improvements over LSTM, ConvLSTM, and static GCN based approaches.

Cai et al. (2022) [44] introduced a Multi-Scale Temporal Adaptive Graph Convolution Network (MSTAGCN) to forecast delays on airport graphs that change over time. Unlike static graph models, MSTAGCN updates the graph structure dynamically to reflect real time traffic patterns. The model combines multi scale temporal convolution layers with an adaptive graph convolution layer that learns how influence between airports changes throughout the day. This dynamic formulation significantly improves predictive performance, achieving lower MAE and RMSE values than baseline models such as standard GCN, ConvLSTM, and spatiotemporal residual networks.

5.4. Phase-Level Delay Prediction

The models reviewed in Sections 5.1 to 5.3 treat delay propagation at the level of complete flights, aircraft rotations, or airport networks. In most of these approaches, a flight is represented as a single unit that departs late and arrives late, and the operational processes between those two events are not represented explicitly. Although the Bayesian network of Laskey et al. [21] decomposes delay into phase-level components, it is not embedded in a predictive framework that propagates the aircraft state through successive flight legs. This is a limitation for the present study, because the amount of delay that is transferred from one flight leg to the next is determined by processes that behave very differently from one another. Ground handling is bounded by a critical path of sub-processes, taxi duration is governed by airport surface conditions, and airborne time is constrained by aerodynamic and airspace limitations. Treating these processes as one aggregated quantity conceals where delay is generated, where it is absorbed, and where it is simply passed on.

A separate body of literature addresses the individual operational phases directly. These studies are usually not framed as delay propagation research, but as prediction problems for a specific phase duration, such as turnaround time, taxi time or arrival time. They are nevertheless directly relevant, because a phase-decomposed propagation model is constructed from exactly these components. This section therefore reviews the state of the art for each of the four phases that constitute an aircraft rotation, namely the turnaround, taxi-out, airborne and taxi-in phases. Section 5.4.1 discusses the turnaround, Section 5.4.2 the two taxi phases, and Section 5.4.3 the airborne phase.

5.4.1. Turnaround Phase

The turnaround is the mechanism through which delay is transferred between two consecutive flight legs operated by the same aircraft, and it is consequently the phase that determines whether rotational knock-on delay occurs at all. Brueckner et al. (2022) [45] formalised this in an operational definition: a propagated departure delay occurs when the arrival delay of the inbound flight exceeds the ground buffer of the subsequent flight. Applying this definition to more than 4.2 million United States domestic flights, they quantified how large a share of arrival delay is attributable to propagation. The same principle underlies the analytical-econometric formulation of Kafle and Zou (2016) [6], in which the ground buffer determines how much of an incoming delay is absorbed before it is passed on, and the earlier network analysis of AhmadBeygi et al. (2008) [46], which showed that schedules optimised for high resource utilisation retain little slack and therefore propagate root delays further. The quantity that governs the turnaround phase is thus not the turnaround duration in itself, but the time remaining between the actual arrival of the aircraft and the scheduled departure of the next leg.

The process behind this behaviour has been characterised in two complementary ways. Wu and Caves (2004) [47] modelled the turnaround as a stochastic process using a Markovian simulation calibrated on operational data from a European airline, and showed that departure punctuality is a direct function of the scheduled ground time and of the buffer allocated within it. Fricke and Schultz [48] subsequently characterised the internal structure of the turnaround empirically using field data from German airline operations, identifying a critical path in which de-boarding is followed by fuelling, catering and cleaning in parallel and concluded by boarding. Because the sub-processes are arranged along this critical path, a delay in any single sub-process extends the total ground time, while ground handling can absorb part of an inbound delay by accelerating individual processes. This absorption capacity is limited and decreases as the magnitude of the inbound delay grows. The relationship between inbound delay and outbound delay is therefore non-linear, and reproducing this non-linearity is the central requirement for a phase-level turnaround model.

Machine learning approaches have been applied to this problem more recently, and consistently identify the same dominant variable. De Falco et al. (2023) [49] developed a two-stage probabilistic model in which a gradient boosted regression predicts the turnaround duration and a subsequent classifier converts the residual distribution into a full probability distribution for the Target Off-Block Time. Using operational Airport Collaborative Decision Making data from four European airports, they reported mean absolute errors of approximately four to six minutes in the tactical phase, corresponding to an improvement of roughly 44 to 56 percent over a baseline that assumes the scheduled turnaround duration. A SHAP analysis identified the available turnaround time as the dominant predictor, ahead of the scheduled turnaround time, origin distance, aircraft weight and airline operator, while weather and passenger count contributed only marginally. An important limitation for the present study is that the applied data filters removed flights with large inbound delays, so the behaviour of the model in exactly the disrupted regime that generates knock-on delay was not evaluated.

Dalmau (2024) [50] addressed rotational reactionary delay directly, applying categorical and natural gradient boosting to EUROCONTROL data and treating the aircraft rotation rather than the isolated flight as the unit of prediction. This study confirmed the available turnaround time as the dominant predictor across all operational conditions, and additionally showed that gradient boosted tree models learn the non-linear threshold behaviour around the minimum turnaround time without being explicitly instructed to do so. The same study establishes a methodological point that applies to any propagation model, namely the choice of schedule reference. When an operator updates the filed departure time after a disruption, both the measured departure delay and the apparent available turnaround time are computed against a reference that has already absorbed the disruption, which systematically understates propagation. The initial filed flight plan is therefore the appropriate reference against which turnaround delay should be

measured.

The value of milestone information for this phase has also been demonstrated directly. Dalmau et al. (2021) [33], discussed in Section 5.3, found the available turnaround time to be one of the two most influential features when predicting take-off time deviations from Enhanced Tactical Flow Management System data, which indicates that a model trained on flight-level milestones captures reactionary delay effects rather than only local departure variability. Okwir et al. (2025) [51] evaluated several supervised learning methods on operational Airport Collaborative Decision Making data from Madrid Airport and observed that Collaborative Decision Making variables are frequently omitted from delay prediction studies while carrying substantial predictive information for turnaround and taxi operations. Their results also show that linear models remained competitive with ensemble methods on this dataset, which indicates that the benefit derives primarily from the availability of the milestone variables rather than from model complexity. A practical consequence is that the most informative milestones are tactical values that are established only shortly before departure, so their contribution is concentrated at short prediction horizons.

Taken together, the turnaround literature supports three requirements for a phase-level model. The available turnaround time, defined relative to the initial filed schedule, is the variable that determines whether inbound delay is transferred to the next departure. The relationship between the two is non-linear, with limited absorption near the minimum turnaround time and progressively less recovery as the inbound delay grows. And the operational milestones exchanged within Airport Collaborative Decision Making provide the information required to estimate this relationship, which implies that separate treatment is needed for airports where these milestones are not available.

5.4.2. Taxi-Out and Taxi-In Phases

The taxi-out phase begins at pushback and ends at take-off, while the taxi-in phase begins at touchdown and ends at gate arrival. Simaiakis and Balakrishnan (2016) [52] established the decomposition that underlies most subsequent work on these phases. Their model of the airport departure process separates taxi-out time into an unimpeded component, determined by the geometry of the route between stand and runway, and a congestion-dependent component represented as a transient queue at the departure runway. Validated on operational data from Newark, the model predicts runway schedules, take-off times, taxi-out times and the associated queuing delay. The distinction is important for delay propagation, because only the second component varies with the operational situation at the moment of execution, while the first is a structural property of the stand and runway combination.

The variables that govern taxi duration follow directly from this decomposition. Park and Kim (2023) [53] analysed both taxi-in and taxi-out at Gimpo and Jeju using operational data from 2019 and identified wind speed, air pressure and precipitation as significant meteorological factors, alongside arrival and departure flight volume, runway occupancy time, take-off and landing direction, and terminal position. They also found that taxi times differ significantly between the two airports, which indicates that the relationship between these variables and taxi duration is airport-specific rather than transferable. For Schiphol this is directly relevant, because the distance between stands and the active runway varies considerably with the runway configuration in use, so that the same flight can experience a substantially different taxi duration depending on the runway to which it is assigned.

Whether a learned model is required for this phase is less clear than for the turnaround. Diana (2018) [54] compared ensemble machine learning methods against classical regression for taxi-out prediction at Seattle-Tacoma across two periods representing different stages of operational modernisation. Ordinary least squares and ridge regression performed best in the first period and gradient boosting in the second, leading to the explicit conclusion that no single algorithm fits the data best in all cases and that the choice should follow from the bias-variance trade-off rather than from model sophistication. Okwir et al. (2025) [51] reached a comparable result on Airport Collaborative Decision Making data from Madrid, where linear regression and elastic nets achieved the highest accuracy among the evaluated methods, which included regression trees, random forests and a multilayer perceptron.

The most directly comparable evidence comes from Schiphol itself. Vakaet (2020) [55] developed a data-driven taxi time predictor for Schiphol in support of start-up planning, and increased the fraction of flights predicted within a two-minute error margin from approximately 64 to 68 percent relative to the predictor in operational use at LVNL. The improvement is real but modest, which indicates that the operational reference already captures most of the systematic variation in taxi duration, and that the residual error

is dominated by surface conditions at the moment of execution rather than by the functional form of the estimator.

Two conclusions follow for a phase-decomposed propagation model. First, the taxi phases are governed primarily by structural properties of the stand and runway combination together with the prevailing surface conditions, and unlike the turnaround they contain no scheduled buffer that is deliberately available to absorb rotational delay. Second, the marginal accuracy gain from a separately trained taxi model over a well-stratified reference estimate is limited, provided that the estimate is stratified by the variables that the literature identifies as dominant, namely the stand, the active runway configuration and the prevailing traffic conditions. Within the error budget of a complete aircraft rotation, the residual uncertainty of the taxi phases is therefore small in comparison with that of the turnaround.

5.4.3. Airborne Phase

The airborne phase determines the extent to which a departure delay is still present at landing. The econometric evidence indicates that the recovery capacity of this phase is small in relation to the delay that must be absorbed. Brueckner et al. (2022) [45] analysed more than 4.2 million United States domestic flights and found that an additional minute of departure delay produces approximately 0.876 additional minutes of arrival delay, with the propagated component transferring at approximately 0.765 minutes per minute. The reduction in airborne time achieved per additional minute of departure delay is an order of magnitude smaller than the delay itself, which means that flying faster offers very little scope to recover a late departure. Kafle and Zou (2016) [6] treat the flight buffer and the ground buffer as distinct quantities in their analytical formulation, with the flight buffer scaling with route distance. For a rotation model dominated by European short-haul operations the available flight buffer is therefore limited, and the airborne phase should be treated as a largely conservative transfer of upstream delay rather than as a mechanism for recovery.

This transfer is nevertheless not proportional. Belkoura et al. (2017) [18] showed, using a historical dataset of flights crossing European airspace, that the relationship between inbound and outbound delay separates into a linear regime for small delays and a non-linear regime for large ones, and that individual airports respond heterogeneously to extreme delays depending on their structural properties. The implication for a phase-level model is that the amount of delay carried through the airborne phase and into the next turnaround cannot be represented by a single constant recovery factor, and that a model which can capture non-linear responses is appropriate.

Regarding predictive accuracy, a clear distinction exists between models that have access to in-flight trajectory information and models that must produce a prediction before departure. Basturk and Cetek (2021) [56] combined flight information, trajectory data and meteorological data in random forest and deep neural network models for estimated time of arrival prediction, and achieved a mean absolute error below six minutes after departure and below three minutes after entry into the terminal manoeuvring area. Such trajectory information is not available when the landing time must be predicted before departure in order to serve as input to the turnaround of the subsequent flight leg, which is the constraint that applies within a sequential propagation framework.

5.5. Feature Analysis

Table 5.6 provides an overview of the main feature groups found in the reviewed literature. Many studies use similar types of features, such as temporal or weather variables, but the selection of features in each study depends on the research focus, meaning whether the model is built at a flight, airport, or network level, and which phase of the operation it aims to predict.

Temporal Features	Spatial & Network Features	Operational Features	Disruption Features
Hour of day	Airport (origin and destination)	Airline	Weather
Day of week	Route / distance	Aircraft type	ATC regulations
Month	Airport capacity	Scheduled timestamps (A-CDM)	Runway closures
Season	Airport demand	Runway configuration	Special events
—	Network delay state	Previous leg delay	—
—	Connectivity in the network	Tail numbers (aircraft rotations)	—

Table 5.6: Feature overview

Sun et al. (2022) [40] performed a correlation analysis for the features used in their airport and network delay prediction models, based on a subset of data from London Heathrow Airport (EGLL). The correlations are presented in a heatmap, shown in Figure 5.4. One example visible in this figure is that month shows almost no correlation with average arrival or departure delay. This contrasts with other studies, such as Tu et al. (2008) [19], which reported clear seasonal patterns in delay behaviour. These differences show that feature relationships are context dependent, and therefore need to be analysed specifically for this thesis.

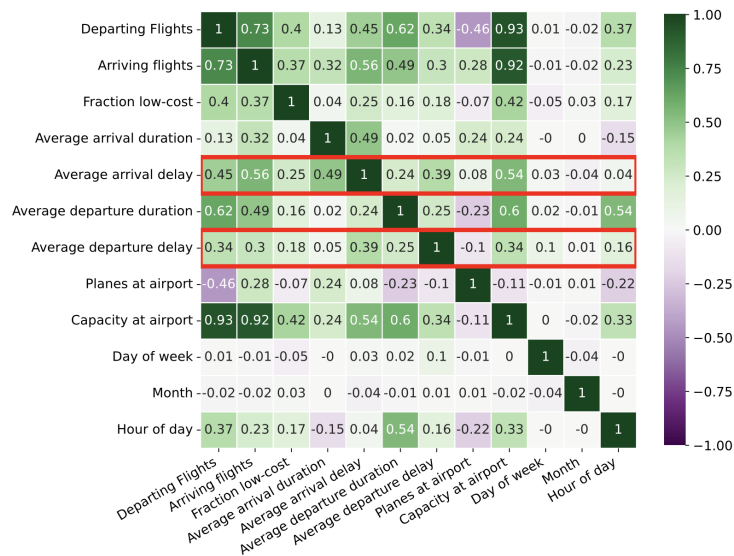


Figure 5.4: Correlation heatmap of features for EGLL [40]

The four feature groups in Table 5.6 play different roles in the reviewed models. Temporal features such as hour of day and day of week capture recurring demand patterns and appear in nearly all reviewed studies, although their predictive contribution varies between airports, as illustrated by the contrast between Tu et al. (2008) and Sun et al. (2022) discussed above. Spatial and network features describe the position of a flight within the wider system and dominate in network-level models, where airport connectivity and the current network delay state drive propagation. Operational features, including the airline, aircraft type, scheduled timestamps, and the delay of the previous flight leg, are the primary inputs of flight-chain models, with the A-CDM milestones providing the most detailed operational timeline where available. Disruption features such as weather and ATC regulations act as external drivers that explain delay which cannot be attributed to the schedule itself.

For a phase-decomposed model, the relevance of these feature groups differs per phase. The turnaround literature identifies the available turnaround time and the inbound delay as the dominant variables, the taxi literature points to the runway configuration, stand location, and surface congestion, and the airborne literature emphasises the departure delay and the scheduled flight buffer. This phase-dependence provides guidance for the phase-specific feature selection in this thesis, and indicates that the difference in milestone availability between Schiphol and the outstations needs to be taken into account when defining the feature sets.

Research Proposal

6.1. Research Gap

In the previous chapters, several approaches were reviewed and compared. Although these studies examine delay propagation at the level of flight chains, airports, or wider networks, limited research has investigated rotational knock-on delay specifically in relation to Schiphol demand prediction. Some studies have developed prediction models for Schiphol-related operations, such as Overtveld (2019) [57], who developed a deep learning model to predict delay estimation errors, and Ruskamp (2023) [58], who predicted take-off times at outstations. However, these studies do not trace predicted delay through the successive operational phases and flight legs of an aircraft rotation, nor do they evaluate how this propagation affects the Schiphol demand.

A second gap concerns the representation of the operational phases within a predictive delay-propagation framework. Existing approaches generally model delay at the level of complete flights, aircraft rotations, or airport networks. The Bayesian network model developed by Laskey et al. (2012) decomposes delay into gate, taxi-out, airborne, and taxi-in components, demonstrating the value of a phase-level representation. However, the reviewed literature does not combine separate representations of the turnaround, taxi-out, airborne, and taxi-in phases within a sequential framework in which the predicted aircraft state is passed from one phase to the next and subsequently to later flight legs. This limits the ability to follow how rotational delay propagates throughout the remainder of an aircraft rotation.

Finally, the DST currently has no explicit mechanism for modelling rotational knock-on delay. Although it provides predictions of future traffic demand and available capacity, it does not propagate the predicted aircraft state consistently through successive future flight legs. As a result, delays occurring earlier in an aircraft rotation may not yet be fully reflected in the predicted take-off and landing times from which future Schiphol demand profiles are constructed.

Based on these observations, the following three knowledge gaps are identified:

- Limited research has investigated rotational knock-on delay specifically in relation to Schiphol demand prediction.
- No reviewed study combines phase-level modelling with sequential propagation of the aircraft state through successive flight legs.
- LVNL's Decision Support Tool does not explicitly account for rotational knock-on delay when predicting future Schiphol demand.

6.2. Research Objective and Questions

The main objective of this research is formulated below, it describes the aim of this study and forms a basis for the research questions that follow.

Research Objective

The objective of this research is to develop a knock-on delay propagation model that represents how delay evolves through aircraft rotations and to evaluate its contribution to Schiphol demand prediction.

To achieve this objective, the following main research question is formulated:

Research Question

How can knock-on delay be modelled through the operational phases of an aircraft rotation to improve Schiphol demand prediction?

The research question is subdivided into four sub questions formulated below.

Sub Questions

Sub question 1: How can the turnaround, taxi-out, airborne, and taxi-in phases be modelled separately to represent knock-on delay propagation within an aircraft rotation?

Sub question 2: How can these phase-level models be combined into a propagation framework that maintains a consistent aircraft state throughout an aircraft rotation?

Sub question 3: To what extent does modelling knock-on delay improve the prediction of operational milestone times compared with the current DST prediction method?

Sub question 4: To what extent do the improved milestone predictions translate into more accurate Schiphol demand prediction?

Project Plan

The research questions, objective, and operational background have been established in the previous chapters. This chapter presents the approach that will guide the remainder of the thesis, describing how a phase-decomposed model of rotational knock-on delay will be developed and evaluated.

7.1. Project Overview

The aim of this thesis is to represent how rotational knock-on delay evolves through aircraft rotations and to evaluate its contribution to Schiphol demand prediction. An aircraft rotation will be represented as a sequence of four operational phases: turnaround, taxi-out, airborne, and taxi-in. Each phase will be modelled separately using an approach suited to its operational characteristics and the available data. The phase-level components will subsequently be connected so that the aircraft state resulting from one phase can be passed to the following phase and to later flight legs within the same rotation.

The project will proceed in several stages. First, the required operational data will be prepared and consecutive flights will be linked to reconstruct aircraft rotations. Second, suitable modelling approaches will be developed for the individual operational phases, taking into account differences in data availability between Schiphol and outstations. Third, the phase-level components will be combined into a sequential propagation framework that predicts the remaining operational timeline of an aircraft from a given observation time. Finally, the framework will be evaluated by comparing its operational milestone predictions with the current DST predictions and by examining whether these predictions lead to more accurate rolling Schiphol demand profiles.

7.2. Methodology

Data acquisition and preparation

The first step consists of collecting and preparing the required operational data. This includes scheduled, estimated, and actual timestamps, together with temporal, spatial, aircraft-related, and weather-related information. The available data will be examined to assess their quality and availability and to determine which variables can be used for each operational phase. Consecutive flights operated by the same aircraft will then be linked to reconstruct aircraft rotations.

Phase-specific modelling

The turnaround, taxi-out, airborne, and taxi-in phases will each be represented separately. For every phase, a suitable modelling approach will be selected based on its operational characteristics and the available data. The phase-level components will represent the transitions between consecutive operational milestones and will account for how the upstream aircraft state influences the timing of later events. Differences in data availability between Schiphol and outstations will be considered when defining the corresponding feature sets and modelling approaches.

Sequential propagation framework

The phase-level components will be combined into a sequential propagation framework for complete aircraft rotations. At a given observation time, operational milestones that have already occurred will be fixed to their realised values, while future milestones will be predicted. Each predicted milestone will then

become input to the following phase and, where applicable, to the next flight leg operated by the same aircraft. This procedure will maintain a consistent predicted aircraft state throughout the remainder of the rotation.

7.3. Uncertainty

Flight delay propagation is influenced by several sources of uncertainty. Studies consistently show that operational variability, including schedule buffers, turnaround performance, airport congestion, and weather-related disruptions, introduces unpredictability into how delay evolves across flights. Econometric and regression-based analyses demonstrate that even after accounting for known predictors, a considerable proportion of delay variance remains unexplained, highlighting the stochastic nature of delay propagation [17, 18, 6]. Weather conditions affect both airborne and ground operations and are repeatedly identified as important contributors to delay variability [10, 11, 12]. Probabilistic models such as Bayesian networks further emphasise that delay propagates through complex operational dependencies involving weather, traffic conditions, turnaround time, and flight connections [25, 22, 21, 23].

Explicit probabilistic modelling of these uncertainties is beyond the scope of this thesis. The proposed framework will therefore produce deterministic predictions. Predictive variability and robustness will instead be evaluated using separate validation and evaluation periods and by examining prediction errors across different horizon intervals. In addition to average prediction accuracy, the spread and occurrence of larger errors will be considered to assess how consistently the framework performs under varying operational conditions.

7.4. Horizon

The reviewed literature uses a wide range of forecasting horizons, from short-term predictions of 30–180 minutes [40, 41] to medium- and long-term horizons extending to a full day [30, 32]. In general, shorter horizons tend to produce more accurate predictions because they rely on more recent operational information, while longer horizons involve greater uncertainty.

For this thesis, the operational focus is on predictions up to four hours ahead. This horizon is particularly relevant to LVNL's demand capacity balancing process, as it provides sufficient time to identify future demand changes and prepare flow-management measures before operation. The framework will therefore be evaluated across multiple prediction-horizon intervals between the observation time and the relevant future milestone or demand period. This makes it possible to assess how the value of modelling knock-on delay changes as the operation approaches.

7.5. Plan

The project is structured into four main parts, each addressing a specific stage of the modelling process. The first part focuses on data acquisition and preprocessing, including aircraft rotation reconstruction, feature engineering, and the construction of phase-specific datasets. The second part concerns the development and validation of modelling approaches for the turnaround, taxi-in, taxi-out, and airborne phases. In the third part, the phase-specific models are integrated into an aircraft-rotation propagation framework, in which sequential propagation is implemented across consecutive flight legs. The final part evaluates the operational milestone predictions against the current DST predictions, aggregates these predictions into Schiphol demand profiles, and evaluates the resulting demand predictions. Figure 7.1 illustrates the project plan.

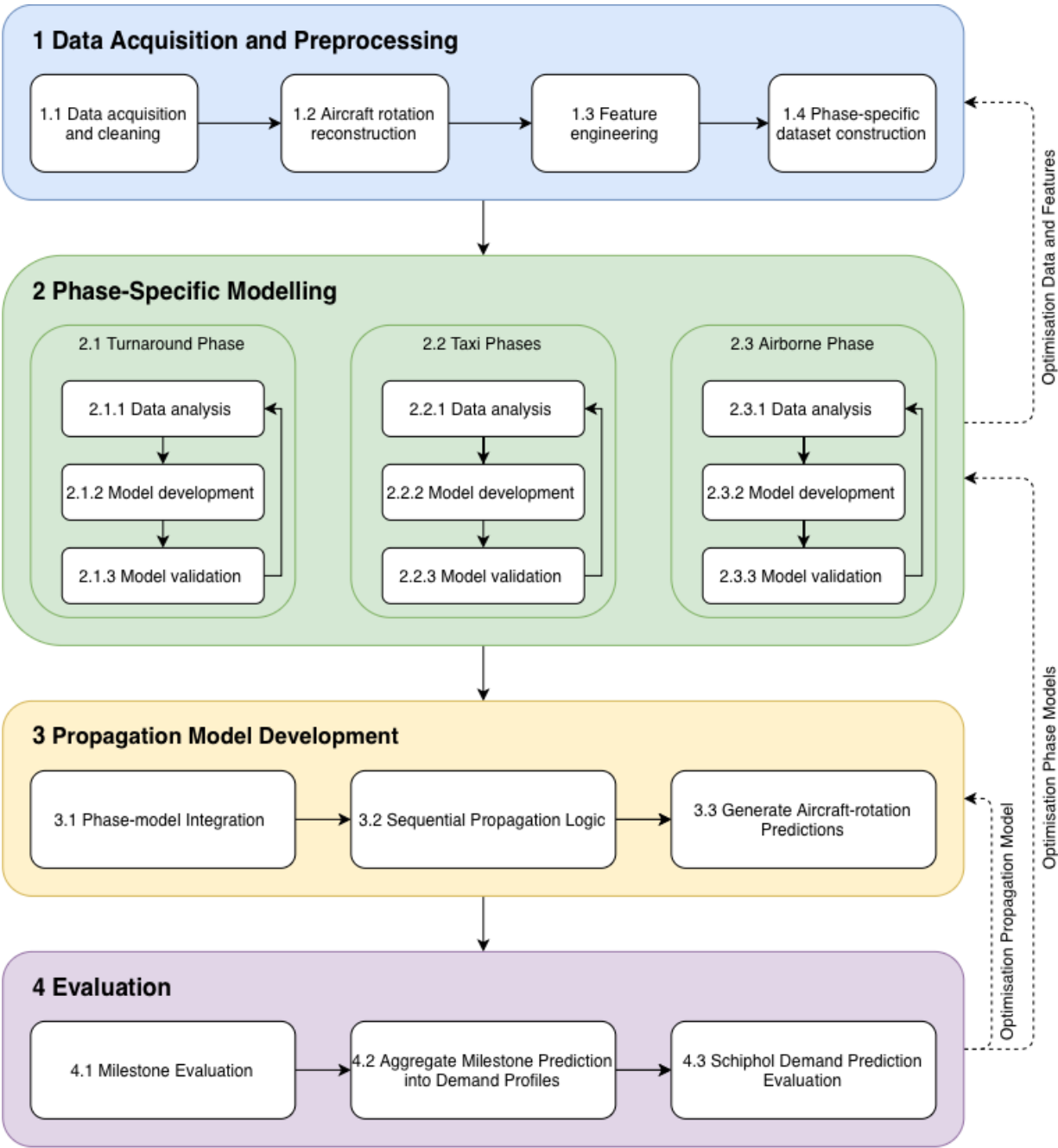


Figure 7.1: Overview of the project plan.

References

- [1] EUROCONTROL. *European Aviation in 2040: Challenges of Growth. Annex 1: Flight Forecast to 2040*. Tech. rep. Prepared by the EUROCONTROL Statistics and Forecast Service (STATFOR). Brussels, Belgium: EUROCONTROL, 2018.
- [2] Stéphane Pierre. *ATFCM Operations Manual. Network Manager. Wave 2.1*. Classification: Public. EUROCONTROL. Nov. 4, 2025. URL: <https://www.nm.eurocontrol.int/>.
- [3] Li Weigang et al. “Air Traffic Flow and Capacity Management”. In: *International Encyclopedia of Transportation*. Elsevier, 2021, pp. 380–389. DOI: 10.1016/B978-0-08-102671-7.10350-1. (Visited on 11/19/2025).
- [4] EUROCONTROL. *Airport Collaborative Decision Making (A-CDM)*. <https://www.eurocontrol.int/concept/airport-collaborative-decision-making>. Accessed: 2025-01-27. 2025.
- [5] Charles Walker. “Central Office for Delay Analysis - CODA”. In: (2024).
- [6] Nabin Kafle et al. “Modeling Flight Delay Propagation: A New Analytical-Econometric Approach”. In: *Transportation Research Part B: Methodological* 93 (Nov. 2016), pp. 520–542. DOI: 10.1016/j.trb.2016.08.012. (Visited on 10/01/2025).
- [7] Georgina Santos et al. “Determinants of Delays at European Airports”. In: *Transportation Research Part B: Methodological* 44.3 (Mar. 2010), pp. 392–403. DOI: 10.1016/j.trb.2009.10.007. (Visited on 10/06/2025).
- [8] John Behrends et al. “An Integrated Solution for the Aircraft Taxi and Gate Assignment Problems”. In: *Cogent Engineering* 4.1 (Jan. 2017). Ed. by Jun Guo, p. 1413722. DOI: 10.1080/23311916.2017.1413722. (Visited on 11/29/2025).
- [9] Aircraft Process Management Royal Schiphol Group Airport Operations & Aviation Partnerships. *RASAS: Regulation Aircraft Stand Allocation Schiphol*. Tech. rep. Version 1.0. Valid from 1 January 2024. Amsterdam: Royal Schiphol Group, 2023.
- [10] Jinn-Tsai Wong et al. “A Survival Model for Flight Delay Propagation”. In: *Journal of Air Transport Management* 23 (Aug. 2012), pp. 5–11. DOI: 10.1016/j.jairtraman.2012.01.016. (Visited on 10/14/2025).
- [11] Zhe Zheng et al. “How Late Does Your Flight Depart? A Quantile Regression Approach for a Chinese Case Study”. In: *Sustainability* 12.24 (Dec. 2020), p. 10553. DOI: 10.3390/su122410553. (Visited on 10/14/2025).
- [12] Nagesh Nayak et al. “Estimation and Comparison of Impact of Single Airport Delay on National Airspace System with Multivariate Simultaneous Models”. In: *Transportation Research Record: Journal of the Transportation Research Board* 2206.1 (Jan. 2011), pp. 52–60. DOI: 10.3141/2206-07. (Visited on 10/15/2025).
- [13] Xuzhou Yang et al. “Aircraft Trajectory Prediction Using ADS-B Data”. In: (2022). Ed. by Jérôme Louveaux et al., pp. 113–122.
- [14] Sam Ouliaris. *Finance and Development*. <https://www.imf.org/external/pubs/ft/fandd/2011/06/basics.htm>. (Visited on 10/11/2025).
- [15] Chi Li et al. “Flight Delay Propagation Modeling: Data, Methods, and Future Opportunities”. In: *Transportation Research Part E: Logistics and Transportation Review* 185 (May 2024), p. 103525. DOI: 10.1016/j.tre.2024.103525. (Visited on 09/30/2025).

- [16] N. Li et al. "A Review of Research on Flight Delay Propagation: Current Situation and Prospect". In: *Journal of Advanced Transportation* 2025.1 (Jan. 2025). Ed. by Jaeyoung Jay Lee, p. 4851103. DOI: 10.1155/atp/4851103. (Visited on 09/30/2025).
- [17] Roger Beatty et al. "Preliminary Evaluation of Flight Delay Propagation through an Airline Schedule". In: *Air Traffic Control Quarterly* 7.4 (Oct. 1999), pp. 259–270. DOI: 10.2514/atcq.7.4.259. (Visited on 10/10/2025).
- [18] Seddik Belkoura et al. "Beyond Linear Delay Multipliers in Air Transport". In: *Journal of Advanced Transportation* 2017 (2017), pp. 1–11. DOI: 10.1155/2017/8139215. (Visited on 10/13/2025).
- [19] Yufeng Tu et al. "Estimating Flight Departure Delay Distributions—A Statistical Approach With Long-Term Trend and Short-Term Pattern". In: *Journal of the American Statistical Association* 103.481 (Mar. 2008), pp. 112–125. DOI: 10.1198/016214507000000257. (Visited on 10/14/2025).
- [20] Myeonghyeon Kim et al. "Modeling the Flight Departure Delay Using Survival Analysis in South Korea". In: *Journal of Air Transport Management* 91 (Mar. 2021), p. 101996. DOI: 10.1016/j.jairtraman.2020.101996. (Visited on 10/15/2025).
- [21] Kathryn Blackmond Laskey et al. *Propagation of Delays in the National Airspace System*. June 2012. DOI: 10.48550/arXiv.1206.6859. arXiv: 1206.6859 [cs]. (Visited on 10/16/2025).
- [22] Yujie Liu et al. "Modeling and Estimating for Flight Delay Propagation in a Reduced Flight Chain Based on a Mixed Learning Method". In: *2008 International Symposium on Knowledge Acquisition and Modeling*. Wuhan, China: IEEE, Dec. 2008, pp. 491–495. DOI: 10.1109/KAM.2008.95. (Visited on 10/17/2025).
- [23] Weiwei Wu et al. "Enhanced Delay Propagation Tree Model with Bayesian Network for Modelling Flight Delay Propagation". In: *Transportation Planning and Technology* 41.3 (Apr. 2018), pp. 319–335. DOI: 10.1080/03081060.2018.1435453. (Visited on 10/01/2025).
- [24] Cheng-Lung Wu et al. "Modelling the Delay Propagation Effects of Multiple Resource Connections in an Airline Network Using a Bayesian Network Model". In: *Transportation Research Part E: Logistics and Transportation Review* 122 (Feb. 2019), pp. 62–77. DOI: 10.1016/j.tre.2018.11.004. (Visited on 10/20/2025).
- [25] Ning Xu et al. "Estimation of delay propagation in the national aviation system using bayesian networks". In: (2005).
- [26] Aravinda Jatavallabha et al. *Deciphering Air Travel Disruptions: A Machine Learning Approach*. Aug. 2024. DOI: 10.48550/arXiv.2408.02802. arXiv: 2408.02802 [cs]. (Visited on 10/24/2025).
- [27] Yan-Yan Song et al. "Decision Tree Methods: Applications for Classification and Prediction". In: *Shanghai archives of psychiatry* 27.2 (Apr. 2015), pp. 130–5. DOI: 10.11919/j.issn.1002-0829.215044.
- [28] Leo Breiman. "Random Forests". In: *Machine Learning* 45.1 (Oct. 2001), pp. 5–32. DOI: 10.1023/A:1010933404324. (Visited on 10/28/2025).
- [29] Alexey Natekin et al. "Gradient Boosting Machines, a Tutorial". In: *Frontiers in Neurorobotics* 7 (2013). DOI: 10.3389/fnbot.2013.00021. (Visited on 10/28/2025).
- [30] Juan Jose Rebollo et al. "Characterization and Prediction of Air Traffic Delays". In: *Transportation Research Part C: Emerging Technologies* 44 (July 2014), pp. 231–241. DOI: 10.1016/j.trc.2014.04.007. (Visited on 10/21/2025).
- [31] Jun Chen et al. "Chained Predictions of Flight Delay Using Machine Learning". In: *AIAA Scitech 2019 Forum*. San Diego, California: American Institute of Aeronautics and Astronautics, Jan. 2019. DOI: 10.2514/6.2019-1661. (Visited on 10/02/2025).
- [32] Bojia Ye et al. "A Methodology for Predicting Aggregate Flight Departure Delays in Airports Based on Supervised Learning". In: *Sustainability* 12.7 (Apr. 2020), p. 2749. DOI: 10.3390/su12072749. (Visited on 10/26/2025).

- [33] Ramon Dalmau et al. "An Explainable Machine Learning Approach to Improve Take-off Time Predictions". In: *Journal of Air Transport Management* 95 (Aug. 2021), p. 102090. DOI: 10.1016/j.jairtraman.2021.102090. (Visited on 10/20/2025).
- [34] *Challenges with LightGBM*. <https://www.byteplus.com/en/topic/399996?title=challenges-with-lightgbm-navigating-performance-issues-and-optimization-strategies>. (Visited on 10/28/2025).
- [35] Roheen Qamar et al. "Artificial Neural Networks: An Overview". In: *Mesopotamian Journal of Computer Science* 2023 (Aug. 2023), pp. 130–139. DOI: 10.58496/MJCSC/2023/015.
- [36] Facundo Bre et al. "Prediction of Wind Pressure Coefficients on Building Surfaces Using Artificial Neural Networks". In: *Energy and Buildings* 158 (Nov. 2017). DOI: 10.1016/j.enbuild.2017.11.045.
- [37] Sina Khanmohammadi et al. "A New Multilevel Input Layer Artificial Neural Network for Predicting Flight Delays at JFK Airport". In: *Procedia Computer Science* 95 (2016), pp. 237–244. DOI: 10.1016/j.procs.2016.09.321. (Visited on 11/03/2025).
- [38] Daniel Alberto Pamplona et al. "Supervised Neural Network with Multilevel Input Layers for Predicting of Air Traffic Delays". In: *2018 International Joint Conference on Neural Networks (IJCNN)*. Rio de Janeiro: IEEE, July 2018, pp. 1–6. DOI: 10.1109/IJCNN.2018.8489511. (Visited on 11/03/2025).
- [39] Robin M. Schmidt. *Recurrent Neural Networks (RNNs): A Gentle Introduction and Overview*. Nov. 2019. DOI: 10.48550/arXiv.1912.05911. arXiv: 1912.05911 [cs]. (Visited on 11/15/2025).
- [40] Junzi Sun et al. "Designing Recurrent and Graph Neural Networks to Predict Airport and Air Traffic Network Delays". English. In: *10th International Conference for Research in Air Transportation*. Ed. by Dave Lovell. 10th International Conference for Research in Air Transportation, ICRAT 2022 ; Conference date: 19-06-2022 Through 23-07-2022. FAA & Eurocontrol, 2022. URL: <https://www.icrat.org/previous-conferences/10th-international-conference/>.
- [41] Hao Zhang et al. "A Multi-step Airport Delay Prediction Model Based on Spatial-temporal Correlation and Auxiliary Features". In: *IET Intelligent Transport Systems* 15.7 (July 2021), pp. 916–928. DOI: 10.1049/itr2.12071. (Visited on 11/09/2025).
- [42] Zonghan Wu et al. "A Comprehensive Survey on Graph Neural Networks". In: *IEEE Transactions on Neural Networks and Learning Systems* 32.1 (Jan. 2021), pp. 4–24. DOI: 10.1109/TNNLS.2020.2978386. arXiv: 1901.00596 [cs]. (Visited on 11/15/2025).
- [43] Yuankai Wu et al. "Spatiotemporal Propagation Learning for Network-Wide Flight Delay Prediction". In: *IEEE Transactions on Knowledge and Data Engineering* 36.1 (Jan. 2024), pp. 386–400. DOI: 10.1109/TKDE.2023.3286690. (Visited on 10/01/2025).
- [44] Kaiquan Cai et al. "A Deep Learning Approach for Flight Delay Prediction Through Time-Evolving Graphs". In: *IEEE Transactions on Intelligent Transportation Systems* 23.8 (Aug. 2022), pp. 11397–11407. DOI: 10.1109/TITS.2021.3103502. (Visited on 11/15/2025).
- [45] Jan K. Brueckner et al. "Airline Delay Propagation: A Simple Method for Measuring Its Extent and Determinants". In: *Transportation Research Part B: Methodological* 162 (Aug. 2022), pp. 55–71. DOI: 10.1016/j.trb.2022.05.003. (Visited on 10/02/2025).
- [46] Shervin AhmadBeygi et al. "Analysis of the Potential for Delay Propagation in Passenger Airline Networks". In: *Journal of Air Transport Management* 14.5 (Sept. 2008), pp. 221–236. DOI: 10.1016/j.jairtraman.2008.04.010. (Visited on 10/01/2025).
- [47] Cheng-Lung Wu et al. "Modelling and Simulation of Aircraft Turnaround Operations at Airports". In: *Transportation Planning and Technology* 27.1 (Feb. 2004), pp. 25–46. DOI: 10.1080/0308106042000184445. (Visited on 04/03/2026).
- [48] Hartmut Fricke et al. "Delay Impacts onto Turnaround Performance - Optimal Time Buffering for Minimizing Delay Propagation". In: June 2009.
- [49] Paolino De Falco et al. "Probabilistic Prediction of Aircraft Turnaround Time and Target Off-Block Time". In: *Proceedings of the 13th SESAR Innovation Days (SIDs 2023)*. Seville, Spain, 2023.

- [50] Ramon Dalmau. "Probabilistic and Explainable Tree-Based Models for Rotational Reactionary Flight Delay Prediction". In: *CEAS Aeronautical Journal* 15.4 (Oct. 2024), pp. 1157–1173. DOI: 10.1007/s13272-024-00750-w. (Visited on 03/28/2026).
- [51] Simon Okwir et al. "Exploring Prediction Accuracy for Optimal Taxi Times in Airport Operations Using Various Machine Learning Models". In: *Journal of Air Transport Management* 122 (Jan. 2025), p. 102684. DOI: 10.1016/j.jairtraman.2024.102684. (Visited on 04/03/2026).
- [52] Ioannis Simaiakis et al. "A Queuing Model of the Airport Departure Process". In: *Transportation Science* 50.1 (Feb. 2016), pp. 94–109. DOI: 10.1287/trsc.2015.0603. (Visited on 04/03/2026).
- [53] Dong Kie Park et al. "Influential Factors to Aircraft Taxi Time in Airport". In: *Journal of Air Transport Management* 106 (Jan. 2023), p. 102321. DOI: 10.1016/j.jairtraman.2022.102321. (Visited on 04/05/2026).
- [54] Tony Diana. "Can Machines Learn How to Forecast Taxi-out Time? A Comparison of Predictive Models Applied to the Case of Seattle/Tacoma International Airport". In: *Transportation Research Part E: Logistics and Transportation Review* 119 (Nov. 2018), pp. 149–164. DOI: 10.1016/j.tre.2018.10.003. (Visited on 04/03/2026).
- [55] Christophe Vakaet. "Taxi Time Prediction with Classical and Auto Machine Learning at Schiphol Airport". Accessed: 2026-04-03. Master's thesis. Delft, The Netherlands: Delft University of Technology, 2020. URL: <https://resolver.tudelft.nl/uuid:efba9763-cb49-4f89-b19b-80ad57c65803>.
- [56] O. Basturk et al. "Prediction of Aircraft Estimated Time of Arrival Using Machine Learning Methods". In: *The Aeronautical Journal* 125.1289 (July 2021), pp. 1245–1259. DOI: 10.1017/aer.2021.13. (Visited on 04/03/2026).
- [57] Christopher Overtveld et al. "Flight Delay Error Analysis for Amsterdam Airport Schiphol Using a Deep Neural Network (DNN)". In: Apr. 2019.
- [58] Tex Ruskamp et al. "Prediction of Traffic Take-Off Times at Out-Stations: A Case Study at Schiphol Airport". In: *IEEE Open Journal of Intelligent Transportation Systems* 6 (2025), pp. 1391–1411. DOI: 10.1109/OJITS.2025.3621578.



Overview of Model Input Features

Table A.1: Features used by the OBT model

Feature	Type
Day of week	Categorical
Hour of day	Numerical
Observation hour	Numerical
Arrival delay	Numerical
ETA arrival delay (Flightplan)	Numerical
Aircraft operator	Categorical
AC Type	Categorical
Widebody aircraft	Categorical
Current airport	Categorical
Origin airport region	Categorical
CDM airport type	Categorical
Flight state inbound	Categorical
Flight state outbound	Categorical
Stack	Categorical
WTC	Categorical
Taxi in time	Numerical
Scheduled Turnaround Time (STT)	Numerical
Available Turnaround Time (ATT)	Numerical
Ratio between ATT and STT	Numerical
Available Buffer	Numerical
Minimum LDT–OBT time	Numerical
Buffer (from Minimum LDT–OBT time)	Numerical
TSAT deviation from SOBT	Numerical
TOBT deviation from SOBT	Numerical
EOBT deviation from SOBT	Numerical
EOBT Schedule Shift	Numerical
CTOT deviation from SOBT	Numerical
Horizon w.r.t. SOBT	Numerical
ATFM delay inbound	Numerical
ATFM delay outbound	Numerical
Wind direction	Numerical
Wind speed	Numerical
Visibility	Numerical
Wind gust	Numerical
Air temperature	Numerical
Cloud coverage	Numerical

Table A.2: Features used by the LDT model

Feature	Type
Day of week	Categorical
Hour of day	Numerical
Departure delay	Numerical
ETA arrival delay (Flightplan)	Numerical
DST arrival delay (DST prediction)	Numerical
Scheduled Airborne Time (SAT)	Numerical
Available Airborne Time (AAT)	Numerical
Available Buffer	Numerical
Ratio between AAT and SAT	Numerical
Taxi out time	Numerical
Horizon w.r.t. SLDT	Numerical
Aircraft operator	Categorical
Origin airport	Categorical
Destination airport	Categorical
Distance between airports	Numerical
ATFM delay	Numerical
Flight state	Categorical
Wind direction outbound	Numerical
Wind speed outbound	Numerical
Visibility outbound	Numerical
Wind gust outbound	Numerical
Air temperature outbound	Numerical
Cloud coverage outbound	Numerical
Wind direction inbound	Numerical
Wind speed inbound	Numerical
Visibility inbound	Numerical
Wind gust inbound	Numerical
Air temperature inbound	Numerical
Cloud coverage inbound	Numerical
Minutes until estimated arrival	Numerical

SHAP Feature Importance Analysis

B.1. SHAP Feature Importance at Schiphol (EHAM)

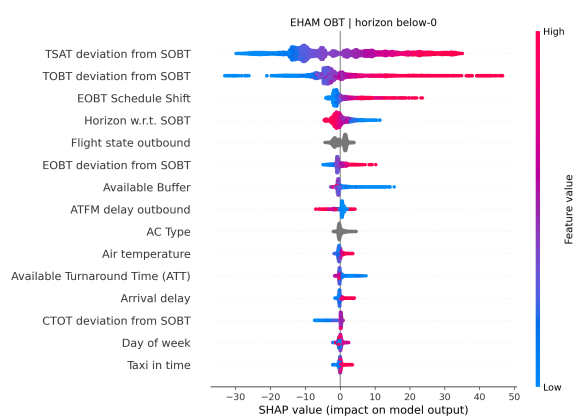


Figure B.1: SHAP beeswarm plot at Schiphol (EHAM), horizon below 0 min.

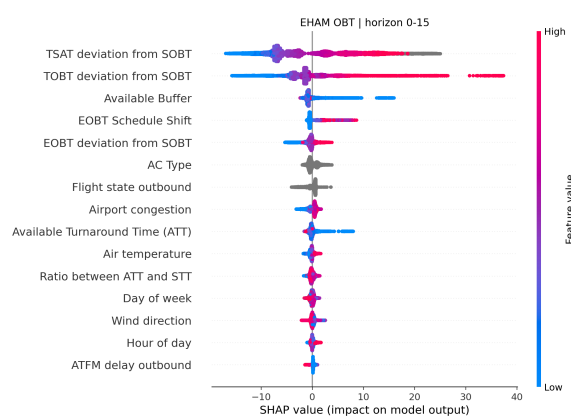


Figure B.2: SHAP beeswarm plot at Schiphol (EHAM), horizon 0–15 min.

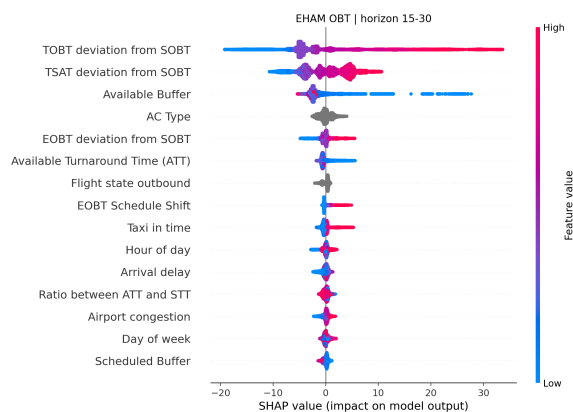


Figure B.3: SHAP beeswarm plot at Schiphol (EHAM), horizon 15–30 min.

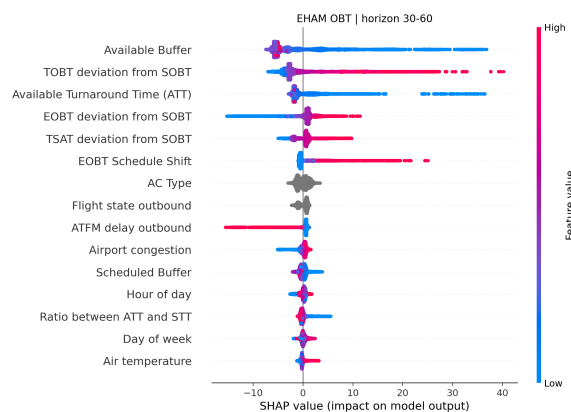


Figure B.4: SHAP beeswarm plot at Schiphol (EHAM), horizon 30–60 min.

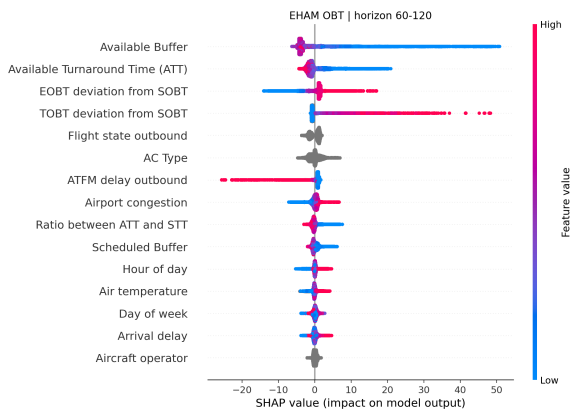


Figure B.5: SHAP beeswarm plot at Schiphol (EHAM), horizon 60–120 min.

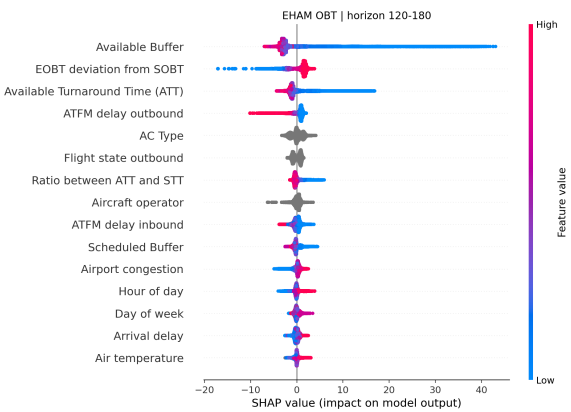


Figure B.6: SHAP beeswarm plot at Schiphol (EHAM), horizon 120–180 min.

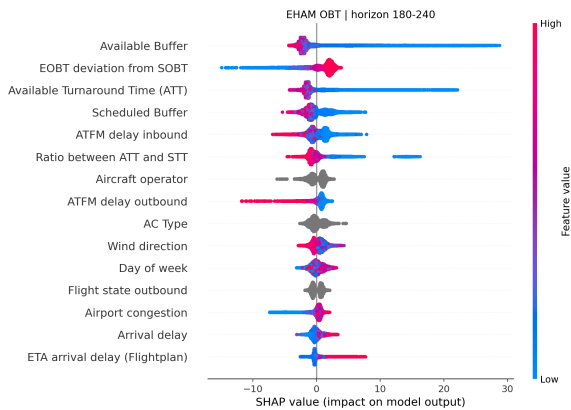


Figure B.7: SHAP beeswarm plot at Schiphol (EHAM), horizon 180–240 min.

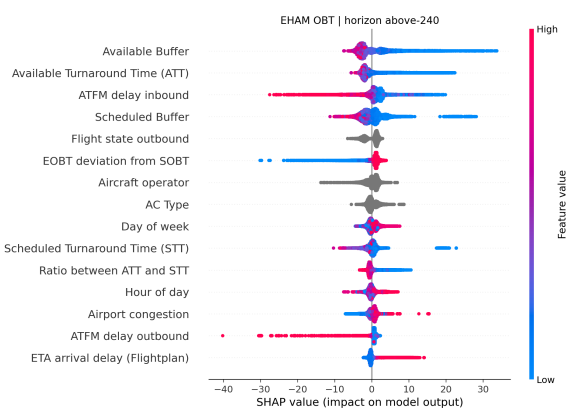


Figure B.8: SHAP beeswarm plot at Schiphol (EHAM), horizon above 240 min.

B.2. SHAP Feature Importance at Outstations

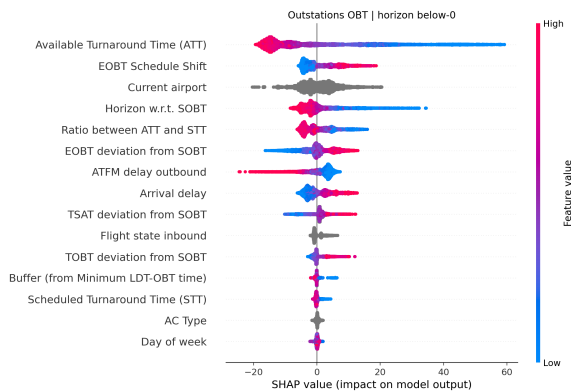


Figure B.9: SHAP beeswarm plot at outstations, horizon below 0 min.

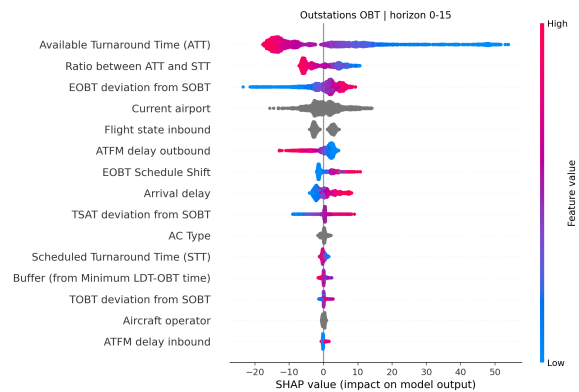


Figure B.10: SHAP beeswarm plot at outstations, horizon 0–15 min.

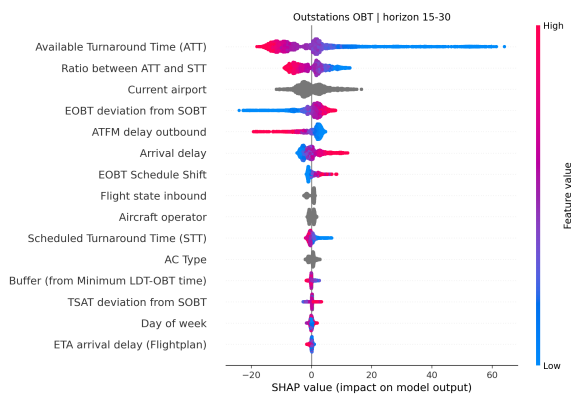


Figure B.11: SHAP beeswarm plot at outstations, horizon 15–30 min.

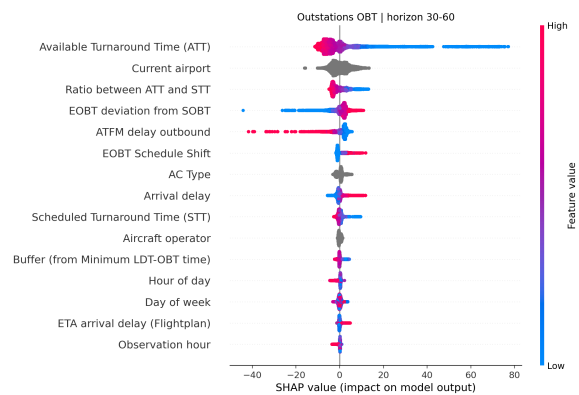


Figure B.12: SHAP beeswarm plot at outstations, horizon 30–60 min.

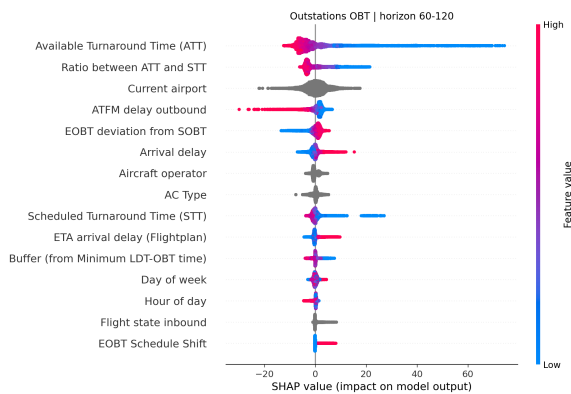


Figure B.13: SHAP beeswarm plot at outstations, horizon 60–120 min.

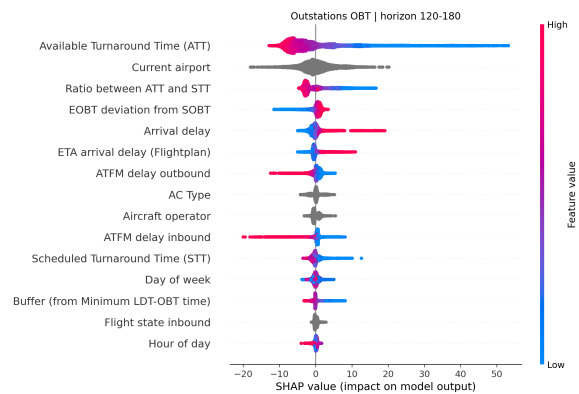


Figure B.14: SHAP beeswarm plot at outstations, horizon 120–180 min.

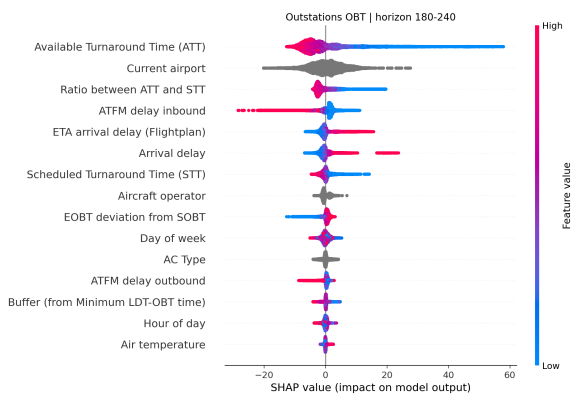


Figure B.15: SHAP beeswarm plot at outstations, horizon 180–240 min.

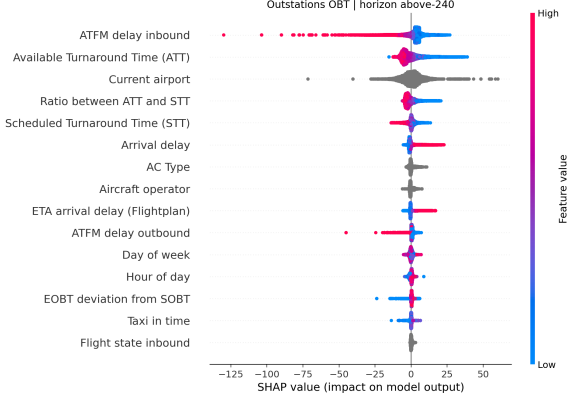


Figure B.16: SHAP beeswarm plot at outstations, horizon above 240 min.