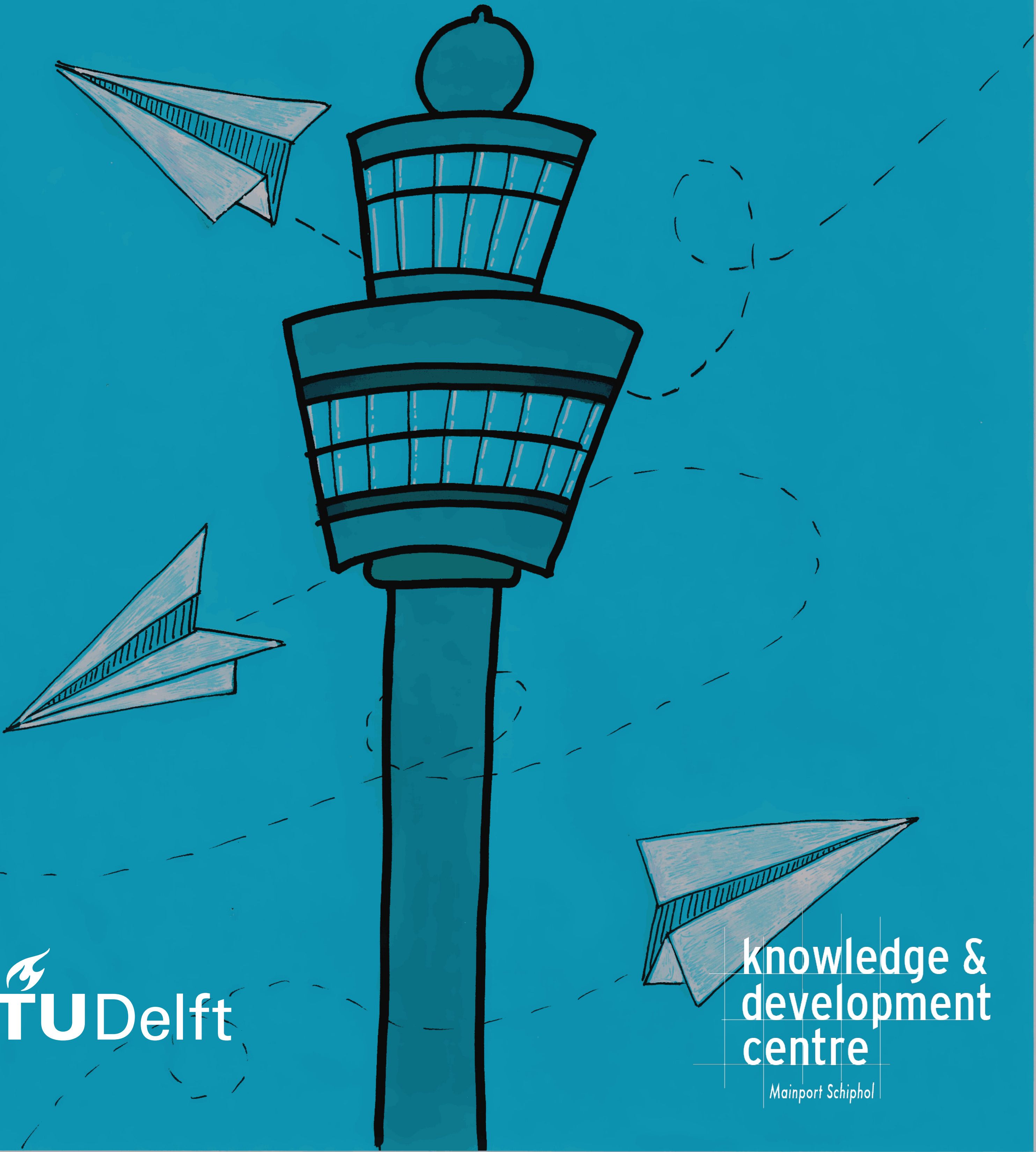
# Departure Manager improvement through Vision-based Predicted End of Ground handling Time

**J.P.H. Bremer** | Master of Science Thesis Aerospace Engineering



knowledge & development centre
Mainport Schiphol

# Departure Manager improvement through Vision-based Predicted End of Ground handling Time

## A case study at Amsterdam Airport Schiphol

Thesis report

by

J.P.H. Bremer

to obtain the degree of Master of Science
at the Delft University of Technology
to be defended publicly on March 17, 2026 at 13:00

| *Thesis committee*: | |
|---|---|
| Chair: | Dr.ir. C. Borst |
| Supervisors: | Dr.ir. J. Ellerbroek |
| | Dhr. F. Dijkstra |
| External examiner: | Dr. X. Wang |
| Place: | Faculty of Aerospace Engineering, Delft |
| Project Duration: | April, 2025 - March, 2026 |
| Student number: | 5019834 |

*Cover image credit: Drawing by Dr. A.P.M. Van den Elzen (my mother <3 )*

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

Faculty of Aerospace Engineering · Delft University of Technology

# Preface

*In theory, there is no difference between theory and practice. In practice, there is.*
**- Jan L. A. van de Snepscheut**

This thesis concludes my Master's degree in Aerospace Engineering at Delft University of Technology. The research was carried out at Luchtverkeersleiding Nederland (LVNL) for the Knowledge & Development Centre (KDC), under the supervision of Ferdinand Dijkstra and Joost Ellerbroek.

Their guidance made this eleven-month project both engaging and insightful. The project allowed me to gain a deep understanding of the operational complexity of a major airport operations environment, particularly at Amsterdam Airport Schiphol.

I am also grateful to my fellow student colleagues at iLabs. The daily discussions about each other's work, and the plane-watching (especially the break-offs and go-arounds) made iLabs a place I genuinely looked forward to. You made it a no-brainer to take the train in every day; thank you for making this such a good time.

A highlight of my time at Schiphol was the visit to both the control tower and the radar room. Seeing airport operations from that perspective was remarkable, and observing the controllers' calm precision, sharp situational awareness, and rapid decision-making only deepened my respect for the profession. It also strengthened my motivation to pursue a career in air traffic control.

This work also marks the closing chapter of an academic journey spanning six and a half years. Over this period, I have grown significantly, both academically and personally. I am grateful for the opportunities and friendships that shaped this time. Ahead lies a new and exciting challenge: pursuing a career as an air traffic controller.

# Contents

# Nomenclature

**A-CDM** airport collaborative decision making
**ANSP** air navigation service provider
**ATC** air traffic control
**ATCo** air traffic controller
**ATFM** air traffic flow management
**BZO** beperkt zicht omstandigheden
**CDM** collaborative decision making
**CPDSP** collaborative pre-departure sequence planner
**DMAN** departure manager
**EASA** European Union Aviation Safety Agency
**ICAO** International Civil Aviation Organization
**IQR** interquartile range
**KPI** key performance indicator
**LVNL** Luchtverkeersleiding Nederland (Air Traffic Control the Netherlands)
**MDI** minimum departure interval
**MTOW** maximum take-off weight
**NM** network manager
**OTP15** on-time performance within ±15 min
**RECAT-EU** European wake turbulence re-categorisation
**RIE** reject if earlier (filter)
**SID** standard instrument departure
**STW** slot tolerance window
**TMA** terminal maneuvering area
**VTT** variable taxi time
**WTC** wake turbulence category

## A-CDM time abbreviations

**AEGT** actual end of ground-handling time
**AIBT** actual in-block time
**AOBT** actual off-block time
**ATOT** actual take-off time
**CTOT** calculated take-off time
**PEGT** predicted end of ground-handling time
**SOBT** scheduled off-block time
**TOBT** target off-block time
**TSAT** target start-up approval time
**TTOT** target take-off time

# 1 General Introduction

This report presents the results of the thesis project conducted to obtain the degree of Master of Science in Aerospace Engineering at the Faculty of Aerospace Engineering, Delft University of Technology. The project was carried out in collaboration with Air Traffic Control the Netherlands.

More passengers are flying each year, increasing the baseline utilisation of Schiphol's infrastructure (with demand expected to continue growing). At the same time, airlines are upgauging across nearly all fleet categories: smaller narrow-body types are increasingly replaced by larger variants (e.g., Boeing 737 replaced by Airbus A321neo) to carry more passengers per flight movement. In some cases, airlines even deploy wide-body aircraft on relatively short intra-European routes.

Under regulatory caps on annual flight movements, this shift is a rational response to passenger demand growth, but it increases operational burden. Together, these effects tighten the constraints on departure runway capacity. Because the airport does not plan to add new runways, any capacity gains must come from more efficient use of existing infrastructure.

A key enabler for departure efficiency is the accurate prediction of when an aircraft will be ready to leave the gate. Currently, the departure manager relies on the Target Off-Block Time (TOBT), a human-reported estimate that is updated throughout the turnaround. Recently, Schiphol Aviation's *Deep Turnaround* project has introduced a camera-based system at each gate that uses machine-learning models to estimate gate readiness in real time, producing a Predicted End of Ground handling Time (PEGT). This sensor-derived signal offers an independent, continuously updated view of turnaround progress.

This study investigates how PEGT can be fused with TOBT to produce a more accurate readiness estimate and, consequently, a more effective departure sequence without destabilizing the outbound sequence. The approach is designed as an addition to the existing system rather than a full redesign, so that it remains operationally implementable. The research encompasses a statistical characterisation of PEGT behaviour, a data-fusion framework that merges PEGT and TOBT, a reconstruction of Schiphol's departure sequencer, and a simulation-based evaluation of the resulting departure sequences.

The report is structured as follows. Part A presents the scientific article, which contains the core methodology, results, and conclusions. Part B collects the supporting appendices: a detailed PEGT data analysis, a meteorological screening for nominal operating conditions, and an analysis of SID-based inter-departure separations. Part C provides the literature review and research definition developed during the initial phase of the project.

# Part A

## Scientific Article

# Departure Manager improvement through Vision-based Predicted End of Ground handling Time

A case study at Amsterdam Airport Schiphol

**J.P.H. Bremer**
Under the supervision of Dr.ir. J. Ellerbroek (TU Delft) and Dhr. F. Dijkstra (LVNL)
*Faculty of Aerospace Engineering, Delft University of Technology*

***Abstract*—Departure management at major airports relies on Target Off-Block Time (TOBT), a human-declared readiness estimate that is prone to inaccuracy. Sensor-derived predictions from computer-vision turnaround monitoring offer a complementary signal, but their higher update frequency and distinct error profile risk destabilising the departure sequence. This study evaluates whether Predicted End of Ground handling Time (PEGT) predictions can improve departure sequencing at Amsterdam Airport Schiphol without sacrificing schedule stability. A characterisation of operational PEGT data shows that PEGT becomes more accurate than TOBT within approximately 27 minutes of departure, but produces nearly twice as many updates and exhibits pessimistic bias in the final minutes before off-block. These properties motivate the design of selective acceptance filters. Using a reconstructed rule-based Departure Manager and counterfactual replay of 21,152 departures across 31 operating days (August 2024), 230 configurations of five conjunctive, interpretable acceptance filters were evaluated via Latin hypercube sampling. Results show that unrestricted PEGT adoption reduces vacated slots by 22.6% but increases late resequencing by 18.6%, confirming that improved accuracy alone does not guarantee operational improvement. However, selective filtering, predominantly through suppression of frequent and late-stage updates, identifies a regime of 55 configurations (24% of those tested) that simultaneously improve all five metrics (resequencing (−0.6%), late resequencing (−6.6%), vacated slots (−13.3%), TSAT delay (−1.6%), and on-time performance (+0.2%)) relative to the TOBT-only baseline. These configurations improve both the TOBT-only and naive unrestricted-PEGT baselines on every tested metric, demonstrating that composite use of TOBT and selectively filtered PEGT can transcend the baseline stability–slot adherence trade-off. The results are based on one month of nominal operations at Amsterdam Airport Schiphol; generalisation to disrupted conditions and other departure management architectures requires further investigation.***

## I. Introduction

Aviation is an important sector in Europe for economic and social connectivity, supporting nearly 5 million jobs and contributing approximately €300 billion (2.1%) to European GDP (European Commission, 2025). With European air traffic projected to grow by 1–2% annually through 2030 (slower than the global average due to market saturation). Airports and airspace will increasingly operate near capacity limits. In this mature, capacity-constrained environment, predictable operations become essential: unlike emerging markets that can expand infrastructure, European aviation must optimise existing resources.

Despite traffic volumes almost back at pre-pandemic levels, punctuality has not recovered. In 2024, European network departure punctuality was 66.2% (–6.5 pp vs. 2019), while average departure delay reached 17.5 min per flight (+35% vs. 2019) (Eurocontrol, 2025; Walker, 2025). At Amsterdam Airport Schiphol, departure punctuality stood at 67%, similarly below 2019 levels (Eurocontrol, 2025). Part of this degradation coincided with a 50% increase in weather-disrupted days compared to 2023.

Importantly, the comparison across major European hubs in Table I shows that high traffic volumes do not inherently imply low punctuality. Airports with comparable daily movements are able to achieve a higher departure punctuality than Schiphol. This persistent performance gap motivates targeted intervention in the departure phase, where surface and runway-side decisions directly shape throughput, delay propagation, and schedule predictability.

TABLE I
MAJOR EUROPEAN AIRPORTS (EXCEEDING 1,000 DAILY MOVEMENTS) AND THEIR DEPARTURE PUNCTUALITY (PERCENTAGE OF FLIGHTS DEPARTING WITHIN 15 MIN OF SCHEDULED TIME) OVER 2024.

| Airport | daily movements [#] | | dep. punctuality [%] | |
|---|---|---|---|---|
| | median | mean | median | mean |
| Istanbul (LTFM) | 1400.5 | 1401.1 | 77.2 | 74.9 |
| Madrid (LEMD) | 1162.0 | 1148.4 | 71.8 | 70.9 |
| London (EGLL) | 1313.5 | 1301.5 | 70.2 | 68.7 |
| Amsterdam (EHAM) | 1350.5 | 1336.3 | 67.4 | 65.7 |
| Frankfurt (EDDF) | 1250.0 | 1204.5 | 61.7 | 60.5 |
| Paris (LFPG) | 1296.5 | 1275.3 | 59.6 | 58.2 |

Source: https://www.eurocontrol.int/Economics/DailyPunctuality-Airports.html.

### A. Departure management and A-CDM

At airports, departures can be constrained by runway capacity, separation requirements (wake vortex and route interactions), and surface limitations (pushback conflicts, taxiway congestion, de-icing queues). To manage these constraints efficiently, modern airports employ a pre-departure sequencing system that implements *virtual queuing*: aircraft are held at their gates until a runway slot becomes available, rather than physically queuing at holding points. This reduces fuel burn and surface congestion while improving predictability for air traffic control (ATC) and airport stakeholders (SESAR-JU, 2011).

Virtual queuing operates within the Airport Collaborative Decision Making (A-CDM) framework, which supports shared planning between airlines, ground handlers, the airport operator, and air traffic control. Runway access is managed via a sequence of target times. Each flight is assigned a Target Take-Off Time (TTOT), the planned time at which it should depart from the runway. Based on the expected taxi-out duration, this is converted into a Target Start-up Approval Time (TSAT), the planned time at which the aircraft should receive start-up approval (push back) so it can taxi and reach the runway in time. The effectiveness of this planning chain depends largely on a single uncertainty: when the aircraft will actually be ready to leave the gate.

### *B. Aircraft readiness estimation*

The Target Off-Block Time (TOBT) is the declared estimate of when an aircraft will be ready for pushback. That is, when the turnaround is complete and wheel chocks can be removed. TOBT is the primary interface between the stochastic, multi-actor turnaround process and deterministic runway-slot planning (Snijders, 2024).

This interface is inherently difficult to manage. The turnaround comprises numerous interdependent tasks (e.g., crew positioning, fuelling, catering, baggage handling, boarding, and technical checks), yet operational readiness is represented by a single scalar time that must be revised as uncertainty resolves. If TOBT is optimistic or updated late, runway capacity is reserved for aircraft that cannot utilise it, causing unused slots and reactionary delay propagation to other flights. Conversely, if TOBT updates are frequent or erratic, TSAT assignments become volatile, increasing the coordination workload and reducing predictability for airlines, ground handlers, and ATC. Over time, such instability can erode stakeholder confidence and reduce adherence to the target times, thereby undermining the coordination benefits of the A-CDM process.

### *C. Schiphol's departure sequencer*

At Schiphol, the current pre-departure planning tool, the Collaborative Pre-Departure Sequence Planner (CPDSP) has been criticised for producing volatile outputs that are difficult to anticipate. In response, LVNL (Air Traffic Control the Netherlands) is implementing a new sequencer designed to improve stability and incentivise timely, reliable readiness information.

The proposed new sequencer at Schiphol is called Departure Manager (DMAN), which organises departures into fixed 10-minute planning bins instead of one continuous planning. It applies a lexicographical priority rule set to resolve demand-capacity imbalances (Frequentis, 2021, 2022). This rule-based and binning approach trades some theoretical optimality for reduced schedule volatility. If flights provide timely, reliable readiness information, the runway utilisation is maximised, and little delay has to be incurred.

### *D. Sensor-derived readiness signal: PEGT*

Recent developments provide an additional readiness estimate alongside the human-declared TOBT. Schiphol has deployed camera-based turnaround monitoring (*Deep Turnaround*), which uses machine vision to produce a Predicted End of Ground handling Time (PEGT). Like TOBT, PEGT estimates when the aircraft will be ready for push-back; however, whereas TOBT is declared by ground-handling staff based on their planning intent, PEGT is continuously updated based on observed turnaround sub-process states at the stand (including fuelling, cargo, catering, cleaning, and boarding) rather than on stakeholder declarations (Aviation Solutions, 2023; Schäfer & Alexopoulos, 2024). The term 'ground handling' in PEGT is used as an umbrella for all stand-side turnaround activities. This new data stream has limited operational evidence; its accuracy, update stability, and failure modes have not yet been characterised in the departure-planning context.

Conceptually, PEGT offers an independent, sensor-derived signal that could become more accurate or respond more consistently to certain delays than TOBT updates. In contrast, TOBT reflects the reported intention of when the aircraft will be ready, declared by experienced ground personnel. However, it may lag behind operational changes or reflect strategic behaviour.

Because the proposed Departure Manager consumes a single readiness input per flight, introducing PEGT as a second, parallel signal would require substantive changes to the planning interface and decision logic. This study therefore treats PEGT integration as a retrofit: TOBT and PEGT are combined into one *fused readiness time* that can be passed to the scheduler without restructuring the DMAN. The fusion is defined by an explicit, interpretable policy that specifies when the human-declared intent signal (TOBT) or the sensor-derived estimate (PEGT) should govern the operational plan.

### *E. Research gap*

Prior work established that the proposed Schiphol DMAN is sensitive to the timing of readiness information: replaying historical operations with TOBT updates shifted 5–10 minutes earlier reduced unused capacity and improved schedule quality (Snijders, 2024). Notably, these gains were achieved without changing the underlying accuracy of TOBT, indicating that earlier availability of readiness information can be operationally valuable.

However, this experiment assumes that readiness updates could have been known and communicated earlier, whereas in practice, part of that information could not have been available at the time. While some improvement may be achievable through better procedures, training, or incentives for the personnel declaring TOBT. The deployment of sensor-derived PEGT therefore introduces a fundamentally different readiness signal that could fill this information gap. The resulting question is therefore how PEGT's distinct dynamics affect a rule-based, fixed-bin planner when used for operational decision making.

*The central knowledge gap is: how does a sensor-derived readiness signal (PEGT), with its own accuracy profile and update dynamics, interact with the proposed DMAN's fixed-bin planning and priority-based sequencing, and under what fusion policies, if any, can it improve departure performance without destabilising the outbound sequence?*

Addressing this gap is challenging for two reasons. First, the impact is systemic: slot assignments depend on the relative ordering of multiple competing flights rather than on a single flight's readiness estimate in isolation. Consequently, a signal that is more accurate on average can still degrade performance if it increases resequencing and reduces schedule stability. Second, PEGT's operational characteristics (e.g., bias, variability, update behaviour, and failure modes) have not yet been characterised in literature, making it unclear if (and under which conditions) PEGT can be preferred over the human-declared TOBT.

### *F. Research objective*

This paper investigates whether, and under what conditions, incorporating PEGT into departure planning at Schiphol improves DMAN performance relative to TOBT-only operation. The study uses a simulation-based replay of historical operations to compare TOBT-only, PEGT-substitution, and Composite TOBT scenarios, evaluating trade-offs between slot efficiency and sequence stability.

This paper makes the following contributions:

1) A quantitative characterisation of PEGT prediction behaviour (accuracy, bias, update frequency, and limitations) relative to TOBT in operational data.
2) A set of interpretable acceptance and filtering policies for integrating PEGT updates into a Composite TOBT-PEGT.
3) A comparative evaluation of DMAN performance under TOBT-only, PEGT-only, and composite TOBT-PEGT scenarios, using metrics for slot utilisation, TSAT delay, sequence stability, and departure punctuality.
4) Operational guidance on when and why PEGT should be accepted or rejected, informing future deployment decisions.

### *G. Approach*

The study performs a case study at Amsterdam Airport Schiphol using operational data from August 2024. A Python-based reconstruction of the proposed DMAN replays historical flight records under counterfactual information scenarios. Performance is assessed using four metric categories: slot utilisation (vacated slots), start-up delay (TSAT delay), schedule stability (resequencing frequency), and departure punctuality (OTP15).

### *H. Paper structure.*

Section II elaborates on: A-CDM, pre-departure sequencing, turnaround prediction, and positioning PEGT within readiness-estimation research. Section III compares PEGT and TOBT across their operational behaviour. Section IV details the data sources, DMAN reconstruction, taxi-time modelling, PEGT characterisation, and Composite TOBT-PEGT policy design. Section V presents experimental results and their operational interpretation. Section VI discusses findings and outlines directions for future research.

## II. Background and Related Work

This section defines the concepts and constraints required to understand the PEGT analysis, fusion methodology, and evaluation framework presented in subsequent sections.

### *A. System Boundary and Operating Context*

At Amsterdam Airport Schiphol, pre-departure sequencing determines when each aircraft may push back and start taxiing to its assigned runway. The quality of this sequencing directly affects runway throughput, network slot compliance, and schedule predictability for airlines, handlers, and air traffic control (ATC). The system boundary of this study spans the A-CDM information chain from declared aircraft readiness (ground handler) to runway-slot assignment (ATC: outbound planner) to start-up approval (ATC: delivery controller) to taxiing (ATC: ground controller) and finally take-off clearance (ATC: runway controller). Upstream processes (arrival management, stand allocation, turnaround execution, runway configuration usage) and downstream operations (en-route flow) are treated as boundary conditions rather than decision variables.

### *B. A-CDM Vocabulary and Update Discipline*

Airport Collaborative Decision Making (A-CDM) coordinates information exchange among airlines, ground handlers, airport operators, ATC, and the network manager through a standardised set of target times (Eurocontrol, 2009). The timestamps relevant to this study are defined in Table II.

TABLE II
A-CDM timestamps used in this study.

| Acronym | Definition |
|---|---|
| TOBT | Target Off-Block Time: declared time at which the aircraft is expected to be ready for pushback upon clearance. |
| TSAT | Target Start-up Approval Time: time at which ATC intends to approve engine start and pushback, derived from TTOT. |
| TTOT | Target Take-Off Time: planned take off time, result of the sequencer & outbound planner. |
| CTOT | Calculated Take-Off Time: externally imposed slot constraint from Eurocontrol when demand exceeds en-route or destination capacity. TTOT must be within CTOT−5 min and CTOT+10 min |
| AOBT, ATOT | Actual Off-Block Time and Actual Take-Off Time: recorded execution times used for performance evaluation. |
| AEGT | Actual End of Ground handling Time: recorded time at which the aircraft completes all ground-handling tasks. |
| SOBT | Scheduled Off-Block Time: published departure time from the airline schedule, used for strategic planning. Initial reference before operational updates. |

TOBT is the primary readiness input to the pre-departure sequencer: it represents the aircraft operator's commitment that the turnaround will be complete and pushback can occur immediately upon clearance. The sequencing procedure operates as follows: expected taxi-out time is added to the declared TOBT to obtain the earliest feasible take-off time, which then enters the runway sequencer. The sequencer assigns a TTOT by resolving conflicts arising from runway capacity, CTOT constraints, and optionally wake-vortex separation and SID divergence requirements. The TSAT is subsequently derived by subtracting the expected taxi-out duration from the assigned TTOT. The relationship between these timestamps is illustrated in Figure 1.

Within the A-CDM framework, stakeholders follow structured update procedures intended to stabilise the pre-departure plan and limit late-stage schedule churn. When TOBT information is timely and credible, the resulting TSAT and sequencing decisions support more efficient runway utilisation and improved predictability. However, because TOBT is a stakeholder-declared intent time, it can be subject to incentive misalignment: actors may report optimistic TOBTs to obtain favourable target times and only revise them once delay becomes unavoidable. Such behaviour reduces information reliability, increases downstream volatility, and can undermine system predictability. This motivates the use of additional, independent readiness signals that are less sensitive to strategic reporting.

### C. DMAN Philosophy: Heuristics over Optimisation

Pre-departure sequencing can be formulated as a constrained optimisation problem minimising delay or maximising throughput. However, optimisation-based approaches tend to produce frequent schedule adjustments as new information arrives, undermining the predictability that airlines, handlers, and controllers require for effective coordination (SESAR-JU, 2012).

The proposed Schiphol DMAN adopts a rule-based, fixed-bin approach that prioritises operational acceptability over mathematical optimality. Departures are grouped into 10-minute planning intervals, and conflicts are resolved through lexicographically ordered priority rules that favour flights with constrained CTOT windows, earlier TOBT declarations, and longer waiting times. This design reflects the principle of *best planned, best served* instead of the traditional *first come, first served* (Frequentis, 2022).

The current pre-departure sequencer at Schiphol has been criticised for producing volatile outputs that are difficult for stakeholders to anticipate. Schedule stability is therefore a primary design objective for the proposed DMAN. Any integration of PEGT must preserve this stability requirement. However, achieving stability is non-trivial: because priority depends on relative ordering across flights, a single readiness update can trigger resequencing of multiple flights.

This cascade sensitivity establishes the operational requirement that any form of PEGT integration must balance prediction accuracy against update frequency: more accurate readiness estimates are valuable only if they do not induce excessive resequencing.

### D. Departure Sequencing Literature

Departure sequencing concerns assigning runway take-off slots and, when integrated with surface management, corresponding pushback/release times for outbound flights. The problem has been studied extensively through mathematical optimisation. Mixed-integer formulations model sequencing as a constrained optimisation problem (e.g., minimising delay) subject to operational constraints such as minimum separation, runway interactions, slot capacity, and Calculated Take-Off Time (CTOT) compliance constraints (Furini et al., 2015; Ma et al., 2024). To achieve real-time applicability when exact optimisation becomes computationally too expensive, heuristic and metaheuristic approaches (e.g. rolling-horizon rules, tabu search, genetic algorithms) are widely used (Bikir et al., 2024; Furini et al., 2015). In parallel, queueing-theoretic models represent the departure runway system as a service facility driven by stochastic pushback demand, enabling analysis and prediction of queue delay, congestion, and effective capacity (Bäuerle et al., 2007; Simaiakis & Balakrishnan, 2016).

The reconstructed DMAN used in this study differs fundamentally from these optimisation-based schedulers. Rather than seeking a mathematically optimal sequence, it allocates departures in fixed 10-minute bins using lexicographic priority rules. Explicitly trading optimality for stability and operational robustness. This reflects the practical requirement that airlines, ground handlers, and ATC benefit more from predictable, stable sequences than globally optimal but frequently changing ones (Alexopoulos, 2024; Frequentis, 2022). Consequently, the optimisation literature provides useful context, but its performance benchmarks are not directly comparable to the stability-oriented constraints and objectives of this work.

### E. Deep Turnaround and PEGT as an information source

Deep Turnaround is a vision-based turnaround monitoring system developed at Schiphol that uses two ultra-wide apron cameras to detect and track ground handling activities (Aviation Solutions, 2023). The system recognises over 70 distinct turnaround events across approximately 30 processes (e.g., bridge connection, fuelling start/end, cargo door status, pushback vehicle arrival) and continuously updates a Predicted End of Ground-handling Time (PEGT) based on observed progress relative to historical patterns.

PEGT differs from TOBT in three key respects:

- **Information source**: PEGT is derived from automated visual observation of physical turnaround state; TOBT is a human declaration reflecting the operator's intent and planning.
- **Update dynamics**: PEGT updates continuously as new visual observations arrive (typically every three minutes once available); TOBT updates occur only at operator-initiated intervals.
- **Availability window**: PEGT becomes available only after sufficient visual features are observable, typically 25–30 minutes before departure; TOBT is available from flight-plan activation (approximately 3 hours prior).

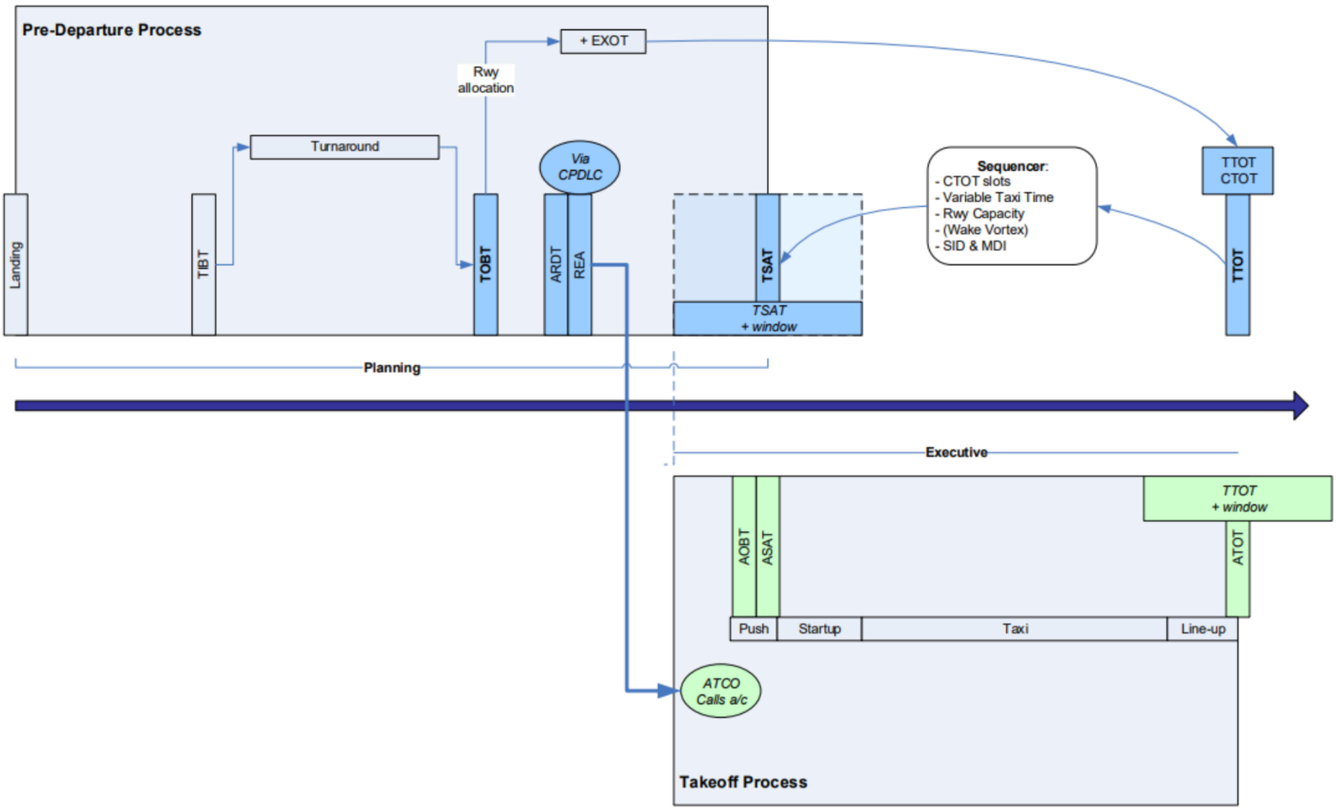



Fig. 1. Schematic diagram of departure manager process overview. The flow contains two distinct phases: the pre-departure phase and the executive phase. (LVNL, 2019)

At Schiphol, PEGT is currently used as an advisory signal for turnaround stakeholders. When PEGT and the declared TOBT diverge, turnaround coordinators receive an alert and are presented with the PEGT estimate together with an operational explanation of the detected cause (e.g., baggage operations started late or take longer). Coordinators are then asked to revise their TOBT. This human-in-the-loop workflow can introduce dependence between PEGT and subsequent TOBT updates of that same flight.

Deep Turnaround is not deployed at all stands, coverage depends on camera availability and line-of-sight constraints. In August 2024, 61 of 142 (43%) stands were fully live, 42 had cameras installed but not yet operational, and 39 had no camera installation. As a result, PEGT was available for only a subset of turnarounds. Also, the PEGT dataset over-represents narrow-body operations due to (i) camera deployment across stand types and (ii) higher stand turnover associated with shorter turnarounds.

This partial coverage imposes an important design constraint: the fusion strategy must accommodate a mix of PEGT-available and TOBT-only flights within the same departure sequence. Because operators cannot control whether their stand has Deep Turnaround available, the system must not systematically advantage or disadvantage either group.

Integrating PEGT is non-trivial. The DMAN's cascade sensitivity means a single readiness update can shift multiple flights if it changes bin or priority rank. More accurate predictions are valuable only if they do not induce excessive resequencing. Previous work has shown that earlier TOBT availability improves schedule quality (Snijders, 2024), but the distinct accuracy profile, update dynamics, and failure modes of PEGT have not been characterised yet in the departure-planning context.

### *F. Fusion Constraints and Design Requirements*

Integrating PEGT into the proposed DMAN is subject to architectural and operational constraints that define the design space for fusion policies:

1) **Single-input architecture**: The DMAN accepts one readiness timestamp per flight. TOBT and PEGT must be fused externally into a single *Composite TOBT-PEGT* before entering the sequencer.
2) **Asynchronous updates**: TOBT and PEGT refresh at different rates and horizons. The fusion logic must produce a consistent output whenever either source updates.
3) **Graceful degradation**: When PEGT is unavailable (due to coverage gaps or system failures), the fusion must default to TOBT without disrupting operations.

4) **Interpretability and auditability**: The fusion logic must remain transparent to human operators and auditable for safety certification. Black-box machine-learning approaches are explicitly excluded to ensure maintainability as Deep Turnaround evolves.
5) **Stability preservation**: The fused readiness signal should not induce excessive TSAT volatility. Frequent or large updates cascade through the priority logic and can destabilise the departure sequence.
6) **Constraint compliance**: Throughput and CTOT adherence cannot be sacrificed. Any fusion policy must preserve the DMAN's ability to meet network slot requirements.

## III. PEGT Characterisation Study

To assess the performance of Deep Turnaround, PEGT is compared against TOBT using operational data from August 2024, for all turnaround where both readiness estimates were available. The absolute prediction error relative to actual end of ground handling time (AEGT) is computed at each lead time. As shown in Figure 2, a crossover in prediction accuracy occurs at approximately 27 minutes before departure: from this point onward, PEGT exhibits lower error than TOBT in terms of mean, median, and interquartile range. Prior to 27 minutes, PEGT accuracy is substantially worse, often unusable for scheduling purposes.

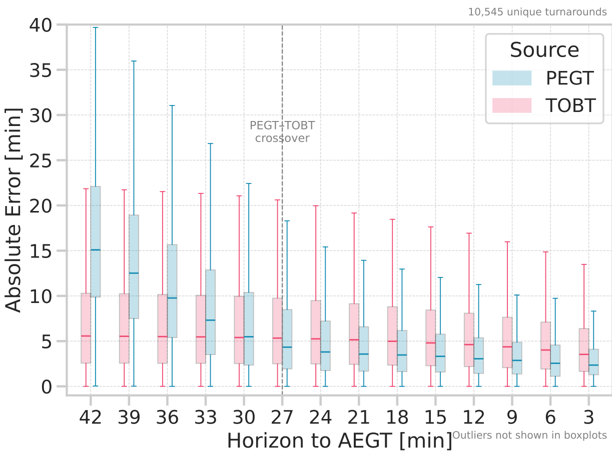



Fig. 2. Absolute prediction error of PEGT and TOBT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

This 27-minute crossover is explained by the first-availability distribution shown in Figure 3. TOBT appears approximately 3 hours before departure, when each flight plan is activated in the system. PEGT, by contrast, becomes available only once sufficient visual features are observable, with a median first availability of 28 minutes before AEGT. By 19 minutes before AEGT, 90% of turnarounds have received a PEGT prediction.

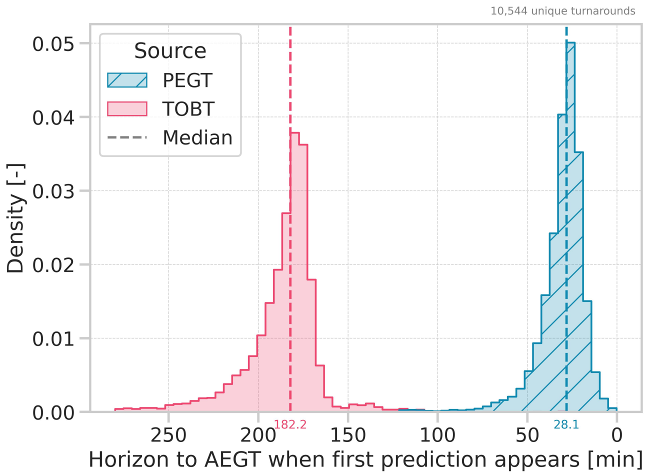



Fig. 3. First availability of TOBT and PEGT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

The late first availability of PEGT has an important operational implication: PEGT cannot replace TOBT for early planning purposes (e.g., determining runway pressure and anticipating the need to open additional runways). However, combined with the accuracy crossover, this finding supports adopting PEGT for readiness estimation once it becomes available, since almost all PEGT predictions at that point will be more accurate than the TOBT.

Comparing the first PEGT prediction against the TOBT confirms this: 49.4 % of first PEGT values have lower absolute error than TOBT at this time of availability, 30.5 % have higher error, and 20.1 % are tied. Thus, in approximately 70 % of cases, adopting the first PEGT does not degrade accuracy and frequently provides an immediate improvement. This fraction increases closer to departure: at 20 minutes before AEGT, PEGT is more accurate than TOBT in 85 % of cases.

The number of updates per source also differs significantly. As shown in Figure 4, PEGT receives on average nearly double the number of updates compared to TOBT. Notably, 48 % of turnarounds receive only a single TOBT prediction. Combined with the observation that TOBT first appears 3 hours prior to departure and exhibits larger error, this implies that nearly half of all turnarounds rely on a TOBT value determined long before departure.

Despite its accuracy advantages, PEGT exhibits a notable bias pattern. As shown in Figure 5, the median bias (median of estimate minus AEGT) shifts from optimistic (predicting earlier readiness than actual) early in the observation window to pessimistic (predicting delay) just before departure.

This shift in bias stems from a limitation of Deep Turnaround's vision system. Individual turnaround analysis revealed that PEGT values can spike upward (i.e. predicting delay) within the final 5 minutes before AEGT. This occurs when the passenger boarding bridge remains connected late into the turnaround. Because Deep Turnaround observes only

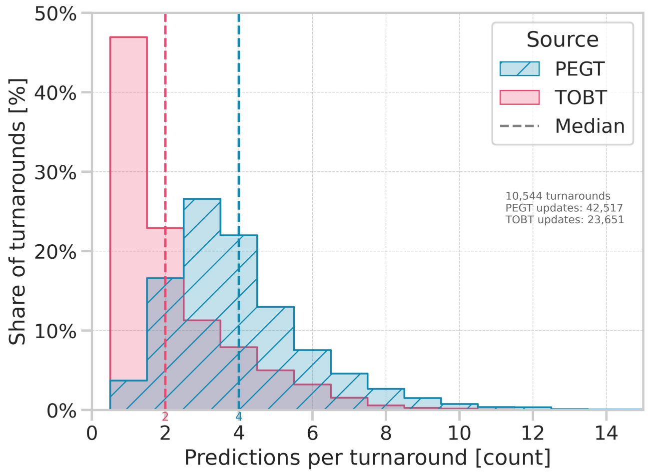



Fig. 4. Histogram of the number of readiness predictions per turnaround (TOBT vs. PEGT), for August 2024 at Amsterdam Airport Schiphol.

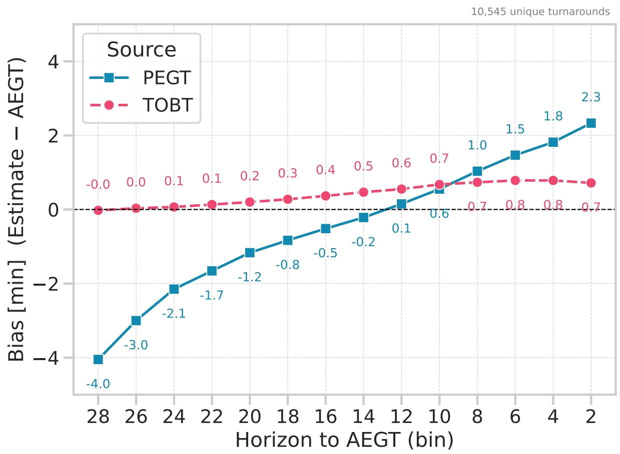



Fig. 5. Bias of TOBT and PEGT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

binary states (e.g., bridge connected vs. disconnected) without information about passenger flow, the system interprets a late-connected bridge as evidence of ongoing boarding and predicts delay accordingly. In reality, boarding may be complete while the bridge remains attached for other reasons. This represents a case where the human-declared TOBT has an advantage: the turnaround coordinator can observe whether passenger boarding has completed and typically knows why the bridge remains connected. The binary nature of the vision system thus has inherent drawbacks, as it classifies observations against a database where late-connected bridges were often associated with delayed boarding, even when this is not the case. Deep Turnaround is addressing this by augmenting the binary bridge-state cue with additional turnaround signals to overcome the inherent limits of a single binary detector, and thus improve operational usability. This phenomenon motivates the 'Reject If Close To Event' filtering policy, which suppresses late-stage PEGT updates that are susceptible to this bias.

### *A. Implications for Fusion Design*

The data analysis yields four key findings that inform the PEGT integration strategy:

1) **PEGT is superior within 27 minutes of departure**: Once available, PEGT should be preferred over TOBT for readiness estimation.
2) **PEGT availability is late**: PEGT cannot support early planning; the fusion design must default to TOBT before PEGT becomes available.
3) **PEGT accuracy improves closer to departure**: The benefit of PEGT over TOBT increases as departure approaches, supporting the adoption of successive PEGT updates.
4) **Late-stage PEGT may be pessimistically biased**: Updates in the final minutes should be treated with caution due to binary information limitation.

## IV. Methodology

This section describes the experimental framework used to evaluate PEGT integration into departure sequencing. The methodology proceeds in five steps, each addressed in a dedicated subsection. First, the data sources and preparation are described (IV-A). Second, the rule-based Departure Manager used to generate departure schedules is reconstructed (IV-B), including its sequencing logic, priority rules, and taxi-time model. Third, the PEGT integration and fusion design is presented (IV-D): two baseline approaches are defined, after which a set of five conjunctive acceptance filters is introduced to selectively admit PEGT updates. Fourth, the five evaluation metrics used to assess schedule quality are defined (IV-E). Finally, the experimental design and the assumptions and limitations of the study are stated (IV-F).

### *A. Data Sources and Preparation*

The primary dataset consists of historical A-CDM records from Amsterdam Airport Schiphol covering 1–31 August 2024. Deep Turnaround records for the same period were provided by Schiphol Aviation Solutions. After preprocessing, the dataset comprises 21,152 departures across 31 operating days, with runway configurations using 18L, 24, and 36L. These runways together account for more than 85 % of departures. To verify that the analysis period reflects nominal operating conditions, the airport meteorological observations were analysed for indicators of capacity-restricting conditions (e.g., severe weather, low-visibility procedures, sustained high crosswind). Using a representative crosswind threshold of $<$25 kt (typical for a Boeing 737) and a visibility threshold of $>$2,000 m, no days exhibited periods outside these limits that would plausibly trigger operational disruptions.

### B. Departure Manager Reconstruction

The experiment emulates the logic of the proposed Departure Manager (DMAN) in a modular Python framework. The proposed DMAN accepts a single scalar readiness estimate per flight and produces a sequence in which aircraft may use the available runway(s), which is then converted to TSATs and TTOTs. The reconstruction separates sequencing from prediction, allowing controlled substitution of different readiness signals while keeping all other system behaviour constant.

*1) Sequencing Logic:* Time is discretised at one-minute resolution. At each minute $t$, the scheduler recomputes the departure plan over the remaining planning horizon, conditional on the information available up to that time. The sequencing procedure consists of the following steps:

1) **Freeze executed flights**: Flights with a planned take-off time earlier than $t$ are considered executed, removed from the active planning set.
2) **Apply information updates**: All readiness (TOBT/PEGT), network (CTOT), cancellation, and runway-related updates time-stamped at $t$ are applied to the remaining flights.
3) **Revive deferred flights**: Flights previously frozen that receive an updated readiness or network constraint placing their earliest feasible departure after $t$ are reintroduced into the active set. The originally assigned slot is recorded as a vacated slot.
4) **Reconstruct the take-off sequence**: For each runway, an initial take-off order is constructed based on earliest feasible readiness times, forming the unconstrained demand profile over future slots.
5) **Resolve capacity conflicts**: If the demand within a 10-minute slot exceeds the declared capacity, flights are iteratively deferred to subsequent slots according to the lexicographic priority rules described in the next section.
6) **Derive operational times**: From the final slot assignment, Target Take-Off Times (TTOT) and corresponding Target Start-Up Approval Times (TSAT) are computed.

To limit unnecessary recomputation, the sequencing logic is executed only at minutes where at least one update is received; otherwise, the previously computed plan is retained.

*2) Lexicographic priority rules:* For each runway and day, the DMAN constructs a departure schedule using fixed 10-minute planning intervals, each with a predefined capacity. When the number of flights assigned to a bin exceeds its capacity, the scheduler resolves the conflict by ranking all candidate flights according to a fixed set of ordered priority rules (PR1-PR8). These rules are applied sequentially: if flights cannot be distinguished at a given priority level, the next rule in the hierarchy is evaluated until a strict ordering is obtained.

The resulting priority ranking determines which flights retain their position in the constrained slot and which flights are deferred to subsequent slots. In this way, the rule set operationalises scheduling trade-offs between network constraints, schedule stability, and declared readiness, while ensuring deterministic and reproducible sequencing behaviour under excess demand.

**PR1:** Flights with more than 10 Calculated Take-Off Time (CTOT) updates. Frequent CTOT changes indicate heavy network regulation and limit flexibility.

**PR2:** Flights with more than two previous shifts in the pre-departure sequence, ensuring that already disrupted flights are stabilised.

**PR3:** Flights whose CTOT would violate the Slot Tolerance Window (STW) if shifted. Shifts that would force wheels-off outside the $[-5, +10]$ minute window are avoided.

**PR4:** Flights with a cancelled CTOT that would be replanned later than the cancelled time if shifted further.

**PR5:** Flights with one or two prior shifts, prioritising moderately disrupted flights.

**PR6:** Flights with an earlier Target Off-Block Time (TOBT), sorted in ascending order.

**PR7:** Flights with a smaller difference between TOBT and Scheduled Off-Block Time (SOBT), also sorted ascending, favouring flights closer to their scheduled time.

**PR8:** If none of the above rules differentiate flights, a deterministic tie-break is applied based on flight identifier.

### C. Taxi-Time Modelling

Pre-departure sequencing requires an estimation of the taxi-out duration from each stand to the assigned runway. At Schiphol, taxi times vary substantially: the remote Polderbaan (18R/36L) has significantly higher taxi time compared to runways adjacent to the terminal complex. Errors in taxi time estimation propagate directly into TTOT accuracy and can cause slot violations even when off-block prediction is correct.

Empirical taxi-time distributions are derived for each stand-runway pair. For each historical departure, the taxi-out time is computed as the interval between actual off-block and arrival at the runway holding point. Table III reports taxi-time medians aggregated at the pier-runway level to provide an intuitive overview of typical taxi distances and runway-dependent variability. In the DMAN sequencing calculations, however, a static lookup table is constructed for each individual stand-runway pair for the highest available resolution. The median was chosen over the mean to minimise the influence of outliers. The taxi time lookup table is static, dynamic congestion effects are thus not modelled. It only distinguishes whether 18C/36C (Zwanenburgbaan) is in use, since typical taxi routes (and thus taxi times) differ substantially between these configurations at Schiphol.

### D. PEGT Integration and Fusion Design

The introduction of an additional readiness estimate, the PEGT, alongside the TOBT, offers improved accuracy in predicting turnaround completion. This improvement comes at the cost of a substantially higher update rate: the median number of readiness updates per turnaround increases from 2 to 4, which can induce volatility in the departure sequence. The objective of the fusion design is therefore to exploit

TABLE III
MEDIAN TAXI TIME (ROUNDED) TO EACH RUNWAY PER PIER AND PERCENTAGE OF DEPARTURES PER RUNWAY AT AMSTERDAM AIRPORT SCHIPHOL

| Runway | Pier | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| *(share of departures)* | A | B | C | D | E | F | G | H |
| 18L *(22.1 %)* | 10 | 10 | 10 | 10 | 12 | 13 | 13 | 11 |
| 36L *(14.7 %)* | 14 | 16 | 17 | 18 | 19 | 18 | 17 | 16 |
| 24 *(48.8 %)* | 7 | 8 | 8 | 9 | 12 | 13 | 13 | 12 |

the improved accuracy of PEGT while limiting unnecessary resequencing and TSAT volatility.

*1) Baseline Integration Approaches:* Two baseline approaches are considered. The first uses TOBT exclusively, reflecting current operational practice. The second switches to PEGT as soon as it becomes available and ignores all subsequent TOBT updates. While this 'all PEGT' approach benefits from improved prediction accuracy, it reacts to every PEGT update and therefore almost doubles the number of readiness revisions for turnarounds with PEGT available. Moreover, PEGT updates are not uniformly well-behaved: the data show occasional late-stage instability, with some trajectories exhibiting abrupt and erroneous jumps in the final minutes before off-block (see section III). This behaviour can amplify short-horizon volatility precisely when the outbound plan is most sensitive to change.

As shown later in Section V, this leads to substantial resequencing and TSAT variability. Importantly, the relationship between update frequency and amount of resequencing is complex: a single readiness update can have a cascading effect. These dynamics motivate the need for a more selective integration strategy.

*2) Design Objective:* The fusion design must balance predictive accuracy against operational stability. Excessive resequencing undermines trust among airlines, ATC, groundhandlers, and network partners, rendering a naive 'all PEGT' approach operationally unattractive. The primary objective is therefore to admit only those PEGT updates that provide material predictive value, while suppressing low-value updates that introduce churn without improving schedule quality.

A second, equally important objective is to preserve full interpretability and operational controllability of the fusion mechanism. PEGT is generated by a system that is still evolving and expanding, and future improvements or recalibrations are expected to change its statistical properties. The fusion mechanism must therefore be robust to such changes and allow rapid adaptation without retraining or revalidation of complex models. Simple, rule-based filters provide this flexibility: thresholds can be tuned, enabled, or disabled at runtime, and their effects remain transparent to human operators. This design choice also facilitates operational validation in a safety-critical environment, where opaque machine-learning models are difficult to certify and explain.

*3) PEGT Update Filtering Policies:* To suppress low-value PEGT updates, a set of operationally motivated filtering policies is introduced. Each policy applies a simple, interpretable condition to determine whether an incoming PEGT update is accepted or rejected. The candidate policies, their parameters, and the tested value ranges are summarised in Table IV.

TABLE IV
PEGT FILTERING POLICIES: ABBREVIATIONS, PARAMETERS, AND TESTED VALUE RANGES.

| Abbr. | Filter Name | Parameter | Range |
|---|---|---|---|
| OW | Outside Window | $\lvert\mathrm{PEGT} - \mathrm{TOBT}\rvert$ threshold [min] | [10–40] |
| SD | Small Delta | Min. change magnitude [min] | [0–15] |
| RIE | Reject if Earlier | $\mathrm{PEGT} < \mathrm{TOBT}$ | 0, 1 |
| TF | Too Frequent | Min. time between updates [min] | [0–10] |
| CTE | Close To Event | Min. time before PEGT [min] | [0–15] |

Each filter addresses a specific source of low-value updates. These filter thresholds form the independent variables of the experiment:

**OW (Outside Window)** rejects updates where PEGT deviates excessively from TOBT. Large deviations (e.g., $\lvert\mathrm{PEGT} - \mathrm{TOBT}\rvert > 40$ min) are likely erroneous or reflect anomalous turnarounds for which the vision system lacks sufficient training data.

**SD (Small Delta)** rejects updates where the magnitude of change from the previous accepted readiness estimate falls below a threshold. Small corrections (e.g., $< 5$ min) may be operationally irrelevant yet still trigger resequencing.

**RIE (Reject if Earlier)** optionally rejects updates that shift readiness earlier than the current TOBT. From an operational perspective, flights departing earlier than planned are generally less disruptive than delays: they do not propagate reactionary delay through the network. However, earlier predictions can still destabilise the local sequence by pulling flights forward and cascading through downstream slot assignments. The RIE filter therefore trades a modest information gain (earlier readiness detection) for reduced resequencing overhead.

**TF (Too Frequent)** enforces a minimum time interval between accepted updates for the same flight. Rapid successive updates (e.g., every 3 minutes) can induce oscillatory behaviour in the schedule, where a flight repeatedly shifts position before stabilising.

**CTE (Close To Event)** rejects updates occurring within a specified time before the predicted off-block. Late-stage changes are particularly disruptive, as crews and handlers have insufficient time to adapt. Moreover, the TSAT compliance window ($\pm 5$ min) provides operational tolerance for minor timing deviations, reducing the need for last-minute corrections. By suppressing updates close to departure, CTE avoids cascading resequences triggered by changes that the system can already absorb.

*a) Filter Combination Logic:* The five filters are applied conjunctively: an incoming PEGT update is accepted if and only if it passes all active filter conditions. Formally, for an incoming PEGT update at time $t$ for flight $f$, let $\mathrm{PEGT}_f(t)$ denote the new prediction, $\mathrm{TOBT}_f$ the current declared readiness, and $\hat{R}_f$ the last accepted readiness estimate. $\theta$ represents the filter parameter setting. The update is accepted if all of the following conditions hold:

1) $\left|\text{PEGT}_f(t) - \text{TOBT}_f\right| \leq \theta_{\text{OW}}$ (Outside Window)
2) $\left|\text{PEGT}_f(t) - \hat{R}_f\right| \geq \theta_{\text{SD}}$ (Small Delta)
3) $\theta_{\text{RIE}} = 0 \lor \text{PEGT}_f(t) \geq \text{TOBT}_f$ (Reject if Earlier)
4) $t - t_{\text{last},f} \geq \theta_{\text{TF}}$ (Too Frequent)
5) $\text{PEGT}_f(t) - t \geq \theta_{\text{CTE}}$ (Close To Event)

The variable $t_{\text{last},f}$ denotes the timestamp of the last accepted PEGT update for flight $f$; for the first PEGT update of a flight, condition 4 is satisfied by default. If any condition fails, the update is rejected and the fused readiness estimate $\hat{R}_f$ remains unchanged.

When a PEGT update is accepted, the fused readiness estimate is set to $\hat{R}_f \leftarrow \text{PEGT}_f(t)$; otherwise, the system retains the previous estimate (either TOBT or the last accepted PEGT). This conjunctive design ensures that only updates satisfying all quality criteria influence the departure sequence, providing conservative protection against noisy or disruptive predictions.

A total of 228 filter configurations were evaluated, sampled via Latin Hypercube design, which ensures uniform coverage of the five-dimensional parameter space at manageable computational cost. Together with the two baselines (NO_PEGT and ALL_PEGT), this yields 230 distinct configurations.

*4) Filter Hypotheses:* Each filter targets a specific mechanism by which PEGT updates may degrade schedule stability. The following hypotheses motivate their inclusion:

- **Outside Window (OW):** When PEGT deviates substantially from TOBT (e.g. >30 min), the update likely reflects a sensor anomaly or outlier rather than genuine readiness information. Filtering such updates is expected to improve robustness, particularly given the system's operational immaturity.
- **Small Delta (SD):** Very small changes (e.g. <3 min) contribute limited new information but can trigger resequencing. Given the TSAT acceptance window of ±5 min, updates below this threshold often have negligible operational impact; suppressing them is expected to reduce volatility without meaningful information loss.
- **Reject If Earlier (RIE):** A prediction of earlier readiness (PEGT < TOBT) typically has less operational impact than a delay, because earlier readiness can be absorbed by existing schedule slack. Rejecting "earlier" updates is hypothesised to reduce volatility with minimal operational penalty.
- **Too Frequent (TF):** Rapid successive updates increase plan volatility disproportionately relative to the new information they provide. Imposing a minimum inter-update interval is expected to improve stability by admitting only updates that reflect meaningful state changes.
- **Close To Event (CTE):** Updates arriving close to the predicted event time can cause large disruptions to an already-committed sequence. The TSAT acceptance window and inherent detection latency limit the actionability of late updates; suppressing them is expected to reduce late resequencing and improve operational acceptability.

### *E. Evaluation Metrics*

The performance of the departure scheduler is evaluated using five dependent measures that capture complementary aspects of schedule quality. Because the DMAN was reconstructed for this study, the full scheduling process is observable and traceable at each decision step, including every readiness update, slot assignment, and resequencing action. This transparency enables consistent computation of both outcome-based and process-level metrics. These five metrics serve as the primary objectives in the multi-objective analysis:

**Resequencing per flight** The mean number of TSAT changes observed per flight throughout its presence in the departure queue. Each time a flight's TSAT is modified after initial assignment, it constitutes a resequencing event. Frequent resequences degrade predictability and trust in the system. *Lower is better.*

**Late resequencing (within 20 minutes)** The total count of TSAT changes occurring within 20 minutes of the actual off-block time (AOBT) across all flights. Late resequences are particularly disruptive as they leave insufficient time for crews to adapt. *Lower is better.*

**Vacated slots** The total count of TSAT violations, where a flight cannot make its assigned TSAT because the underlying readiness estimate shifted after the TSAT window became active. Vacated slots waste runway capacity. *Lower is better.*

**TSAT delay** The sum of differences between the last assigned TSAT and the last known TOBT across all flights, measured in minutes. This captures ground holding time imposed by sequencing constraints. *Lower is better.*

**Departure punctuality (OTP15)** The fraction of flights whose scheduled TSAT falls within ±15 minutes of the Scheduled Off-Block Time (SOBT). This is a standard industry benchmark for on-time performance. *Higher is better*.

Together, these metrics capture complementary aspects of schedule quality: resequencing metrics address operational stability, vacated slots indicate capacity wastage, TSAT delay quantifies ATC inefficiency, and OTP15 provides a high-level punctuality indicator.

### *F. Experimental Design*

The experiment employs counterfactual replay to compare sequencing strategies under controlled conditions. Historical operations are replayed through the reconstructed DMAN using different readiness inputs while holding all other factors constant.

For each configuration, the reconstructed DMAN iterates through August 2024 minute by minute. At each minute $t$, all readiness updates time-stamped at or before $t$ are applied, the sequencer computes the departure plan, and the five metrics are recorded. Identical historical update sequences are used across all configurations to ensure comparability. No future information is used, the system is strictly causal.

### G. Assumptions and Limitations

The study operates under the following explicit assumptions and limitations:

- Time is discretised to a one-minute resolution.
- CTOT windows are strictly enforced. No negotiation beyond the $[-5, +10]$ window is possible; Network Manager not modelled.
- Taxi-time estimates are fixed per flight using a static lookup table; dynamic surface congestion effects are not modelled.
- Tactical interventions by ATC are not modelled (e.g. last-minute manual resequencing, visual assessment, radio conversations, or internal ATC coordination).
- Late TOBT, PEGT, CTOT changes may revive previously frozen flights; the vacated slot is retained as a synthetic lost record to assess slot utilisation.
- Runway allocation is treated as an upstream process and is held constant across simulations; runway reallocation and closures are not considered.
- Airline dispatcher-driven delays are not modelled explicitly; such effects are only reflected insofar as they are embedded in historical TOBT updates.
- All comparisons are made against the TOBT-driven DMAN schedule; direct comparison against realised scheduling decisions is not possible due to the absence of process-level operational data.
- The analysis covers a single representative month of nominal operations with PEGT data available (August 2024).

### H. Validation

In the absence of an operational DMAN in departure scheduling, direct validation against the target system is not possible. Instead, validation is performed by assessing whether the reconstructed scheduler reproduces the observed throughput structure of real-world operations. Validation focuses on runway-level throughput dynamics rather than individual flight-level accuracy, due to the above-mentioned limitations. The reconstructed scheduler is driven solely by historical TOBT inputs, identical to those ingested by the operational CPDSP, but applies a fundamentally different sequencing philosophy. As a result, some discrepancies with realised throughput are expected.

*a) Qualitative Assessment:* For each day in the analysed dataset, scheduled and realised throughput were visually inspected per runway. Across days, consistent qualitative agreement was observed in the temporal structure of operations, including primary runway dominance, secondary runway activation during peak periods, and the timing and magnitude of departure peaks. An example day is shown in Figure 6, where scheduled throughput remains within the same operational envelope as the realised throughput. No day exhibited systematic discrepancy between scheduled and observed patterns.

*b) Quantitative Assessment:* To complement the qualitative inspection, Pearson correlation coefficients were computed between the TOBT-driven DMAN scheduled throughput and the observed (wheels-off) throughput for each runway on each day. Across the 31 days in August 2024, the daily per-runway correlations exhibited a median of 0.885 with an interquartile range of [0.860, 0.905]. These values indicate strong linear agreement between the temporal profiles of scheduled and realised departures, confirming that the reconstructed DMAN captures the dominant throughput dynamics of the operational system.

*c) Sources of Discrepancy:* Observed differences are primarily attributable to structural modelling choices rather than fundamental misrepresentation of departure demand. First, the use of fixed 10-minute scheduling bins, as opposed to CPDSP's continuously replanning horizon, limits fidelity at fine temporal scales and produces sharper transitions in scheduled throughput. Second, tactical interventions by ATC, airline dispatchers, and Network Manager processes are not modelled, reducing the ability to replicate local adaptations and efficiency recoveries.

A notable source of discrepancy arises from runway opening and closure times, which are not modelled. In practice, secondary runways are activated dynamically based on traffic demand and operational constraints. The DMAN reconstruction assumes runways are available throughout the day, which can result in scheduled throughput appearing on runways earlier than operationally observed, particularly during the morning ramp-up when secondary departure runways are opened.

Finally, TOBT records represent declared readiness rather than executed intent. Operational factors such as dispatcher decisions, pilot-controller interaction, and situational awareness are only indirectly reflected, further limiting replication of last-minute operational adaptations.

*d) Validation Conclusion:* The combination of qualitative agreement in peak timing, runway role allocation, and throughput envelope, together with quantitative confirmation via high Pearson correlations (median $r = 0.885$), indicates that the simplified DMAN captures the dominant throughput characteristics of the realised system despite known modelling limitations. This level of agreement is sufficient for the comparative analyses presented in this study, which focus on relative differences between filter configurations rather than absolute scheduling accuracy.

## V. Results

This section presents the experimental evaluation of PEGT filtering strategies for the Departure Manager at Amsterdam Airport Schiphol. The analysis quantifies (i) the baseline trade-off between stability and slot adherence, (ii) the effect of individual filter parameters on each objective, and (iii) the existence of configuration clusters exhibiting distinct performance characteristics. All comparisons are expressed as percentage

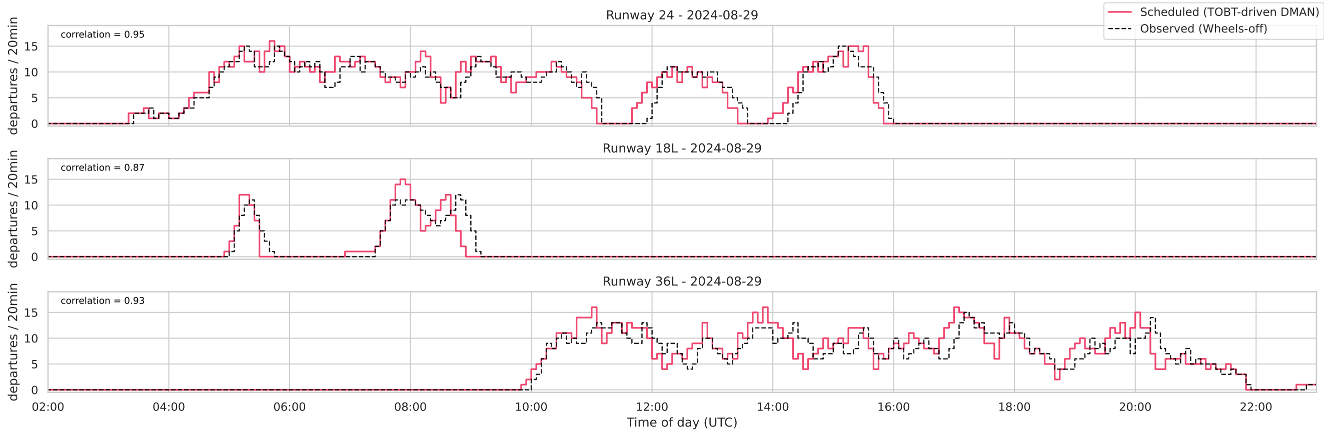



Fig. 6. Comparison of reconstructed DMAN schedule and observed departure throughput per runway on 29 August 2024. Throughput is evaluated using a trailing 20-minute window sampled every 5 minutes. Pearson correlation coefficients per runway are annotated in each subplot.

changes relative to the NO_PEGT baseline unless stated otherwise.

### A. Experimental Setup

A total of 230 configurations were evaluated: two baselines (NO_PEGT and ALL_PEGT) and 228 filtered configurations sampled via Latin hypercube design over the five-dimensional filter space defined in Section IV-D (Table IV). Each configuration was replayed over 31 operating days (August 2024) using the reconstructed DMAN and evaluated on the five metrics defined in Section IV-E.

### B. Baseline Comparison

Table V presents the performance difference between the two baselines. The ALL_PEGT configuration accepts all PEGT updates without filtering, while NO_PEGT (i.e. TOBT-only) ignores PEGT entirely. All percentage changes ($\Delta\%$) throughout this chapter are computed relative to NO_PEGT, enabling direct comparison of how each configuration deviates from the stability-oriented baseline.

TABLE V
BASELINE COMPARISON: ALL_PEGT RELATIVE TO NO_PEGT. FOR ALL METRICS EXCEPT OTP15, LOWER VALUES ARE DESIRABLE.

| Metric | NO_PEGT | ALL_PEGT | $\Delta\%$ |
|---|---|---|---|
| Reseq/flight | 1.043 | 1.110 | +6.5% |
| Late reseq ($\leq$20m) | 19,642 | 23,297 | +18.6% |
| Vacated slots | 2,184 | 1,690 | −22.6% |
| TSAT delay [min] | 69,344 | 63,363 | −8.6% |
| OTP15 | 0.671 | 0.657 | −2.0% |

Compared with NO_PEGT, accepting all PEGT updates reduces vacated slots by 22.6% and TSAT delay by 8.6%, demonstrating that PEGT's improved accuracy enables better slot adherence. However, this comes at the cost of increased resequencing (+6.5%), substantially more late resequencing (+18.6%), and degraded on-time performance (−2.0%). These results establish the fundamental stability–slot adherence trade-off that filtering seeks to navigate.

### C. Individual Filter Effects

To quantify how each filter parameter affects performance, Figure 7 presents Spearman correlations ($\rho$) between filter parameter values and the percentage change in each objective relative to NO_PEGT. For minimisation objectives (resequencing, vacated slots, TSAT delay), negative correlations indicate improvement; for OTP15, positive correlations indicate improvement. Note that correlation magnitude does not directly translate to effect size: a filter with low correlation may still produce substantial effects at specific threshold values.

The correlations provide a first-order ranking of which parameters matter most in isolation. TF and CTE show the strongest associations across stability- and punctuality-related metrics, consistent with their role in suppressing late-stage revisions. SD is primarily associated with slot-adherence penalties (higher TSAT delay and more vacated slots), indicating that aggressive filtering can reduce responsiveness when off-block shifts are real. OW and RIE exhibit weak correlations (both $|\rho| \leq 0.19$) and therefore appear to have limited influence when varied alone. However, the Spearman analysis is not well-suited to RIE since it is a binary function.

However, the operationally relevant behaviour emerges when filters are combined. In practice, filters interact: timing/frequency constraints (TF, CTE) can stabilise the plan but also risk feasibility losses if applied without safeguards, while amplitude-based smoothing (SD) can be beneficial only when paired with mechanisms that still allow genuine shifts to pass. As filtering becomes more restrictive, performance tends to converge toward the NO_PEGT baseline, but the configuration space contains combinations that avoid this collapse by trading small accuracy losses for disproportionate stability gains. The next subsection therefore shifts from marginal, single-parameter effects to multi-parameter patterns by grouping configurations with similar joint performance profiles.

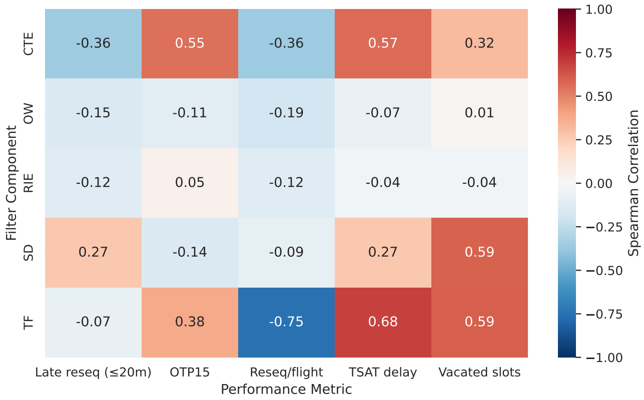



Fig. 7. Spearman correlation between filter parameters and $\Delta\%$ versus NO_PEGT ($n = 228$)

## D. Configuration Clusters

To identify distinct operational strategies within the configuration space, hierarchical clustering with Ward linkage (Wang et al., 2025) was applied to the 228 filtered configurations based on their normalised performance profiles across all five objectives. Analysis of the silhouette score across different cluster counts indicated 6 clusters as an appropriate partition, balancing interpretability with cluster cohesion.

Figure 8 presents the performance profile of each cluster, showing the mean percentage change relative to NO_PEGT for each objective. Table VI summarises the cluster characteristics and typical filter settings.

TABLE VI
TYPICAL FILTER SETTINGS PER CLUSTER (MEDIAN VALUES)

| Cluster | $n$ | OW | SD | RIE | TF | CTE |
|---|---|---|---|---|---|---|
| C0 | 11 | 25 | 0 | 0 | 0 | 0 |
| C1 | 9 | 15 | 3 | 0 | 2 | 3 |
| C2 | 26 | 20 | 5 | 0 | 3 | 2 |
| C3 | 55 | 25 | 5 | 0 | 6 | 10 |
| C4 | 46 | 20 | 12 | 0 | 9 | 8 |
| C5 | 81 | 25 | 8 | 0 | 6 | 6 |

The clustering reveals a clear gradient from stability-oriented to feasibility-oriented configurations:

*Cluster 0: Unfiltered regime ($n = 11$):* With effectively no active filtering (TF = 0, CTE = 0, SD = 0), this cluster replicates ALL_PEGT behaviour within measurement noise (+6.5% resequencing, +18.5% late resequencing, −22.6% vacated slots, −8.5% TSAT delay, −2.0% OTP15). Its inclusion confirms that the Latin hypercube sample spans the full parameter space, but demonstrates that negligible filtering offers no operational advantage over accepting all PEGT updates.

*Cluster 1: Slot-adherence–dominant regime ($n = 9$):* Minimal filtering (TF ≈ 2, CTE ≈ 3) retains the majority of PEGT updates, yielding the largest vacated-slot reduction (−24.3%) and substantial TSAT delay improvement (−7.8%). Resequencing is modestly lower than ALL_PEGT (+4.4% vs. +6.5%), but OTP15 still degrades (−1.6%). This regime is appropriate only when slot compliance is the overriding operational objective.

*Cluster 2: Throughput-oriented regime ($n = 26$):* With low temporal thresholds (TF ≈ 3, CTE ≈ 2), this cluster admits most PEGT updates and achieves vacated-slot reduction (−18.4%) approaching ALL_PEGT levels, at the cost of increased overall resequencing (+1.8%). Late resequencing, however, is already reduced (−3.9%), indicating that even light filtering suppresses the most disruptive schedule changes while preserving the bulk of PEGT's accuracy benefit.

*Cluster 3: Simultaneous improvement ($n = 55$):* This cluster achieves simultaneous improvement on all five metrics relative to NO_PEGT: resequencing −0.6%, late resequencing −6.6%, vacated slots −13.3%, TSAT delay −1.6%, and OTP15 +0.2%. Typical settings (TF ≈ 6, CTE ≈ 10) apply moderate-to-high filtering that admits PEGT updates reflecting genuine readiness changes while rejecting those that introduce volatility. The elevated CTE threshold is operationally significant: it prevents updates from entering the DMAN during the final minutes when the sequence is effectively committed. Because these configurations dominate both baselines on every metric, Cluster 3 represents the recommended operational strategy and is analysed further in Section V-E.

*Cluster 4: Conservative regime ($n = 46$):* Aggressive thresholds (SD ≈ 12 min, TF ≈ 9 min) suppress nearly all PEGT updates, producing only marginal deviations from NO_PEGT (−0.2% resequencing, −1.6% vacated slots). This regime effectively recovers baseline behaviour and is suited to operators who prioritise schedule predictability over optimisation.

*Cluster 5: Moderate-gain regime ($n = 81$):* The largest cluster occupies an intermediate position (TF ≈ 6, CTE ≈ 6), achieving meaningful vacated-slot reduction (−5.7%) and TSAT delay improvement (−1.5%) while remaining neutral on resequencing (−0.1%). For operators seeking incremental improvement with minimal risk, this regime offers a conservative alternative to Cluster 3.

## E. Key Finding: Simultaneous Improvement

The central finding of this study is that PEGT filtering enables simultaneous improvement across all five objectives: a result that defies the conventional Pareto trade-off assumption. In a typical multi-objective optimisation problem, the two baseline configurations (NO_PEGT and ALL_PEGT) would define the Pareto frontier, with all intermediate configurations representing trade-offs between stability and slot adherence. Any gain on one dimension would necessitate a loss on another.

The experimental results demonstrate that this assumption does not hold. Cluster C3 achieves: improved stability, reduced

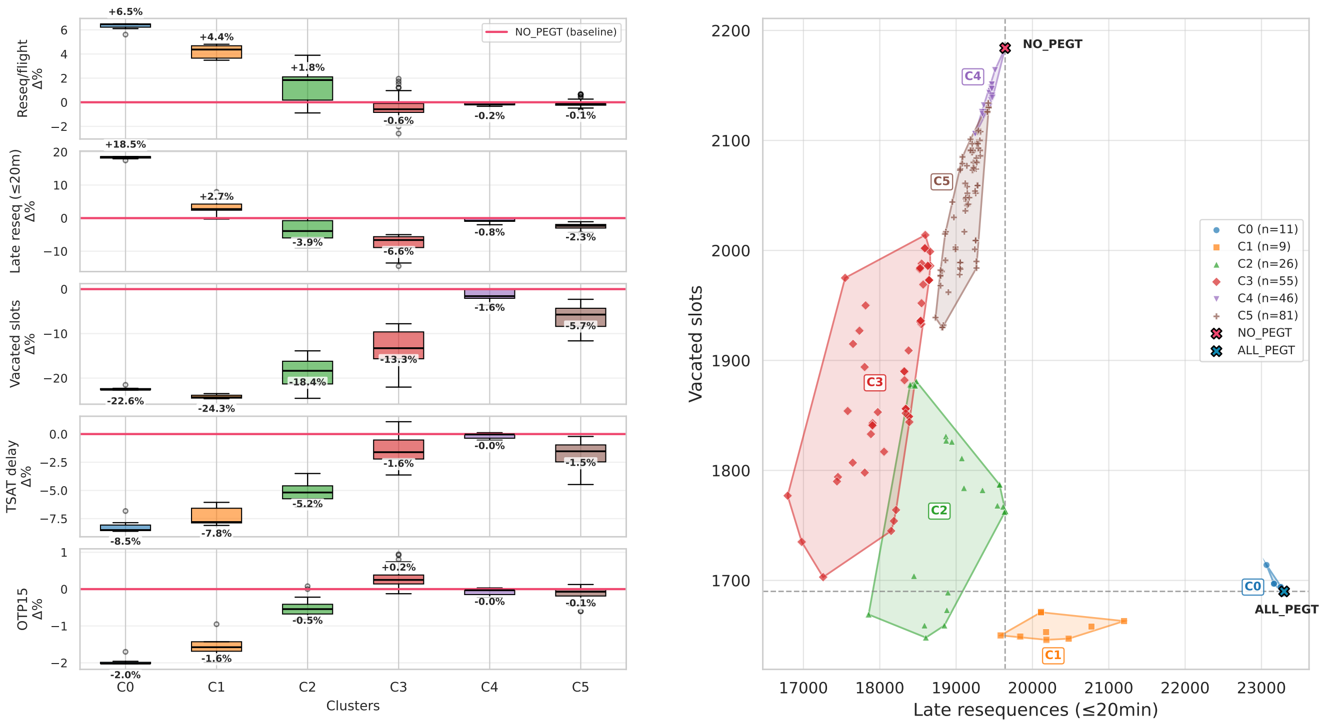



Fig. 8. Cluster performance profiles. Left: bar charts showing median Δ% versus NO_PEGT for each metric across the six clusters and baselines. Right: scatter plot positioning clusters in the resequencing–vacated slots trade-off space, with colour/marker indicating cluster membership.

disruption, improved slot adherence, reduced ground holding, improved punctuality

This outcome is not a marginal boundary case: 55 of 228 filtered configurations (24%) achieve simultaneous improvement on all metrics. The mechanism is intuitive, selective filtering rejects PEGT updates that introduce volatility (frequent small corrections, updates close to departure) while admitting updates that reflect genuine changes in aircraft readiness. The rejected updates represent noise that degrades schedule stability without improving accuracy; the admitted updates represent signal that enhances slot utilisation.

From an operational perspective, this finding has two implications. First, it demonstrates that PEGT integration does not need to be an all-or-nothing decision: carefully tuned filtering can extract the accuracy benefits of PEGT while preserving the predictability benefits of the original TOBT-based planning. Second, it suggests that the static filter approach evaluated here represents a lower bound on achievable performance. Adaptive or context-aware filtering, adjusting thresholds based on traffic density, time-to-departure, or historical update patterns, could potentially expand the ‘dominated’ region further, yielding additional gains beyond what static parameters achieve.

### *F. Robustness*

A complementary finding concerns the robustness of PEGT integration: no configuration among the 228 tested degrades performance on all five metrics simultaneously. All filtered configurations improve vacated slots relative to NO_PEGT, with reductions ranging from $-1.2\%$ to $-24.2\%$. Even the most conservative filtering strategy still provides a meaningful feasibility improvement. This indicates that any reasonable PEGT filtering parameterisation delivers some benefit to slot adherence, reducing the risk of misconfiguration.

### *G. Summary of Quantitative Findings*

The experimental evaluation of 228 configurations yields the following quantitative findings:

1) **Baseline trade-off**: ALL_PEGT reduces vacated slots by 22.6% versus NO_PEGT, but increases resequencing by 6.5% and late resequencing by 18.6%.
2) **Primary filter effects**: TF and CTE are the dominant stability levers; SD primarily affects slot adherence; OW and RIE have negligible impact.
3) **Simultaneous improvement is achievable**: Cluster C3 (55 configurations, 24% of total) achieves improvement on all five metrics simultaneously, demonstrating that filtering can capture the ‘better than both worlds’ rather than forcing a trade-off.
4) **Six distinct clusters**: Hierarchical clustering with $k = 6$ identifies different operational strategies, with Cluster C3 showing the most potential. With settings TF≈6, CTE≈10 achieving $-0.6\%$ resequencing, $-13.3\%$ vacated slots, $-6.6\%$ late resequencing, and +0.2% OTP15 versus NO_PEGT.

5) **Robustness**: All 228 filtered configurations improve vacated slots ($-1.2\%$ to $-24.2\%$), and no configuration degrades all five metrics. PEGT integration provides consistent slot adherence benefits regardless of filter configuration.
6) **Implications for DMAN development**: The existence of a cluster improving on all tested metrics suggests that further research into PEGT integration could yield additional performance improvements beyond the static filter approach evaluated here.

## VI. Discussion & Conclusion

### A. Discussion

Integrating a more accurate readiness signal into a departure sequencer is not straightforward: because slot assignments depend on the relative ordering across flights, a locally accurate update can degrade system-level performance if it increases resequencing. The results of this study confirm this tension and show how selective filtering resolves it.

The central finding of this study is that 55 filter configurations simultaneously improve all five schedule-quality metrics. This has direct implications for the operational integration of sensor-derived readiness predictions into departure management. The reconstructed DMAN prioritises schedule stability and stakeholder predictability over mathematical optimality, reflecting real-world operational requirements. Within this stability-oriented framework, the results demonstrate that PEGT can be integrated without the destabilising effects observed under unrestricted adoption.

The dominant role of timing-based filters (TF, CTE) over amplitude-based filters (SD, OW) suggests that *when* an update enters the DMAN matters more than *how large* it is. This aligns with operational practice: once a sequence is committed and communicated to stakeholders, even an accurate correction can be disruptive. The CTE threshold of approximately 10 minutes in the best-performing cluster (C3) effectively prevents updates from entering the planning window after the sequence has stabilised, while the TF threshold of approximately 6 minutes dampens high-frequency oscillation that would otherwise propagate through the cascade. In contrast, clusters C0–C2 show that admitting frequent or late updates (even when individually accurate) degrades stability, confirming the hypothesis that update dynamics, not only prediction accuracy, govern integration success.

An important operational nuance is that the reconstructed DMAN used here makes use of simplified boundary conditions. This means the gains reported (e.g. $-13.3\%$ vacated slots) are achieved within a simplified system; performance in real operational conditions may differ. The results should therefore not be interpreted as absolute DMAN performance benchmarks but as evidence that filtered PEGT improves relative schedule quality under realistic operational constraints.

### B. Conclusion

This study investigated whether sensor-derived readiness predictions (PEGT) can improve departure management at Schiphol without destabilising the outbound sequence. Using counterfactual replay of 21,152 departures across 230 filter configurations, three principal conclusions emerge.

First, vision-based prediction of ground-handling completion shows clear potential as a complementary readiness signal. The PEGT error profile crosses below that of TOBT at approximately 27 minutes before AEGT, and accuracy continues to improve as departure approaches. This late-window accuracy advantage makes PEGT a valuable supplement to the human-declared TOBT, which is often set hours in advance and sparsely updated.

Second, composite use of TOBT and PEGT outperforms either signal alone. Neither the TOBT-only baseline nor unrestricted PEGT adoption achieves the schedule quality of the filtered composite. The fusion framework extracts PEGT's accuracy benefit while retaining TOBT's early-planning stability, demonstrating that the two signals are complementary rather than substitutive.

Third, selective filtering resolves the stability–accuracy trade-off. Cluster C3 configurations achieve simultaneous improvement on all five metrics, lying outside the Pareto frontier defined by the two baselines. Timing-based suppression of frequent and late-stage updates is the primary mechanism: by preventing updates from entering the DMAN after the sequence has effectively been committed, filtering preserves schedule predictability while still admitting operationally meaningful readiness corrections.

### C. Limitations

Several limitations bound the generalisability of these findings:

- All comparisons are against a *reconstructed* DMAN, not realised operations. While the reconstruction is validated against observed throughput (correlation of 0.885), unmodelled tactical interventions (e.g. runway closures, ATC re-routing) introduce discrepancies.
- The study covers a single month (August 2024) of nominal operations. Behaviour under disrupted conditions (weather, equipment failures) is not characterised.
- PEGT coverage was partial (43% of stands). Whether performance gains scale, diminish, or shift with expanded coverage remains an open question.
- Taxi times are modelled as static per stand–runway pair; dynamic taxi-time variability (or congestion) is not captured.
- The filter thresholds are globally applied; flight-specific or time-of-day-specific tuning was not explored.

### D. Future Work

- Extend temporal scope to multiple months and disrupted operating conditions to assess robustness.
- Investigate dynamic integration approaches (e.g. Kalman filtering of TOBT and PEGT, TSAT-expiry-triggered adoption) as alternatives to static threshold filtering.
- Explore using Deep Turnaround's pushback-truck connection status as a hard constraint for TSAT assignment,

leveraging a high-confidence binary signal to anchor the final sequence commitment.

- Analyse flight-level delay effects: while this study reports aggregate improvements, it remains unclear how benefits distribute across individual flights. Future work should quantify changes in the per-flight delay distribution to assess whether improvements are broad-based or concentrated on specific subsets.

*E. Concluding Remarks*

As Deep Turnaround coverage expands and prediction algorithms mature, the value proposition of sensor-derived readiness estimates will likely strengthen. The interpretable filtering framework developed in this study provides a foundation for incremental deployment and iterative refinement. We recommend that LVNL and Schiphol maintain close monitoring of PEGT integration opportunities as both the DMAN and Deep Turnaround systems continue to evolve.

## References


Alexopoulos, Y. (2024). A-CDM Manual Schiphol Airport. https://www.schiphol.nl/nl/download/b2b/1569488978/7ERl8iHeLELDtgFsnK0mGi.pdf

Aviation Solutions. (2023). *Deep Turnaround*. Retrieved June 12, 2025, from https://www.schiphol.nl/en/aviation-solutions/deep-turnaround/

Bäuerle, N., Engelhardt-Funke, O., & Kolonko, M. (2007). On the waiting time of arriving aircrafts and the capacity of airports with one or two runways. *European Journal of Operational Research*, *177*(2), 1180–1196. https://doi.org/10.1016/j.ejor.2006.01.002

Bikir, A., Idrissi, O., Mansouri, K., & Qbadou, M. (2024). An Optimized Air Traffic Departure Sequence According to the Standard Instrument Departures. *International Journal of Advanced Computer Science and Applications*, *15*(3). https://doi.org/10.14569/IJACSA.2024.01503133

Eurocontrol. (2009). Airport CDM: Steps to boost efficiency. http://www.euro-cdm.org/library/eurocontrol/airport_cdm_steps_to_efficiency.pdf

Eurocontrol. (2025, January 23). *European aviation overview 2024*. Retrieved September 17, 2025, from https://www.eurocontrol.int/sites/default/files/2025-01/eurocontrol-european-aviation-overview-20250123-2024-review.pdf

European Commission. (2025). Directorate-general for mobility and transport. transport modes: Air. https://transport.ec.europa.eu/transport-modes/air_en

Frequentis. (2021). Functional Specification TAPtools CTASNLLVAA: LVNL [Internal Document].

Frequentis. (2022). *Departure Manager* (Product Brief). Retrieved May 16, 2025, from https://www.frequentis.com/sites/default/files/support/2022-06/Frequentis_ATM_product-brief_DMAN.pdf

Furini, F., Kidd, M. P., Persiani, C. A., & Toth, P. (2015). Improved rolling horizon approaches to the aircraft sequencing problem. *Journal of Scheduling*, *18*(5), 435–447. https://doi.org/10.1007/s10951-014-0415-8

LVNL. (2019). *Outbound Planning - Operational Concept Document (OCD)* (tech. rep.) (Internal Document).

Ma, J., Delahaye, D., & Liang, M. (2024). Arrival and Departure Sequencing, Considering Runway Assignment Preferences and Crossings. *Aerospace*, *11*(8), 604. https://doi.org/10.3390/aerospace11080604

Schäfer, J., & Alexopoulos, Y. (2024, October 14). *How Deep Turnaround Enhances A-CDM Efficiency*. Schiphol Aviation Solutions. Retrieved May 19, 2025, from https://www.schiphol.nl/nl/aviation-solutions/blog/how-deep-turnaround-enhances-a-cdm-efficiency/

SESAR-JU. (2011). Basic DMAN Operational Service and Environment Definition (OSED). *SESAR-JU*. Retrieved May 16, 2025, from https://www.sesarju.eu/sites/default/files/documents/solution/Sol106%204%20DMAN%20baseline_OSED.pdf

SESAR-JU. (2012). *DMAN Baseline to be used for Integration of AMAN and DMAN* (Contextual Note) (EXE-06.08.04-VP-470 Validation Exercise at Paris CDG). SESAR Joint Undertaking. Retrieved May 20, 2025, from https://www.sesarju.eu/sites/default/files/documents/solution/Sol106%201%20DMAN%20baseline_Contextual%20note.pdf
EXE-06.08.04-VP-470 Validation Exercise at Paris CDG.

Simaiakis, I., & Balakrishnan, H. (2016). A Queuing Model of the Airport Departure Process. *Transportation Science*, *50*(1), 94–109. https://doi.org/10.1287/trsc.2015.0603

Snijders, D. C. (2024). Investigating the effect on departure capacity of changing Target Off-Block Time uncertainty. Retrieved May 15, 2025, from https://repository.tudelft.nl/record/uuid:8ac0ba80-25b2-4ef1-b1bf-f0aa1b3d8e97

Walker, C. (2025). All-Causes Delays to Air Transport in Europe Annual 2024 - Central Office for Delay Analysis. *Eurocontrol*.

Wang, S.-M., Yang, W.-R., Zhuang, Q.-Y., Lin, W.-H., Tian, M.-Y., Su, T.-J., & Cheng, J.-C. (2025). Application of Three-Dimensional Hierarchical Density-Based Spatial Clustering of Applications with Noise in Ship Automatic Identification System Trajectory-Cluster Analysis. *Applied Sciences*, *15*(5), 2621. https://doi.org/10.3390/app15052621

# Part B

## Appendix

# 2

# PEGT Data Analysis

This appendix extends the PEGT characterisation study presented in Section III of the main paper. It provides a detailed worked example of a single turnaround (Section I), followed by the full-scale statistical analysis of PEGT and TOBT prediction behaviour for August 2024 (Section II). All figures are presented at full resolution to support detailed inspection.

## I. Detailed Turnaround Case Illustrating PEGT Behaviour

To give an intuitive sense of how PEGT evolves throughout a turnaround, Figure 2.1 shows one example flight in detail. The horizontal axis represents *clock time*, while the vertical axis shows the various *CDM scheduled times* associated with the same moment. The plot can be read by selecting any vertical slice: a slice at 12:00, for example, shows the values of TOBT, TSAT, TTOT, and PEGT at that exact time. A diagonal "current time" guideline indicates where the actual clock intersects the scheduler's timeline.

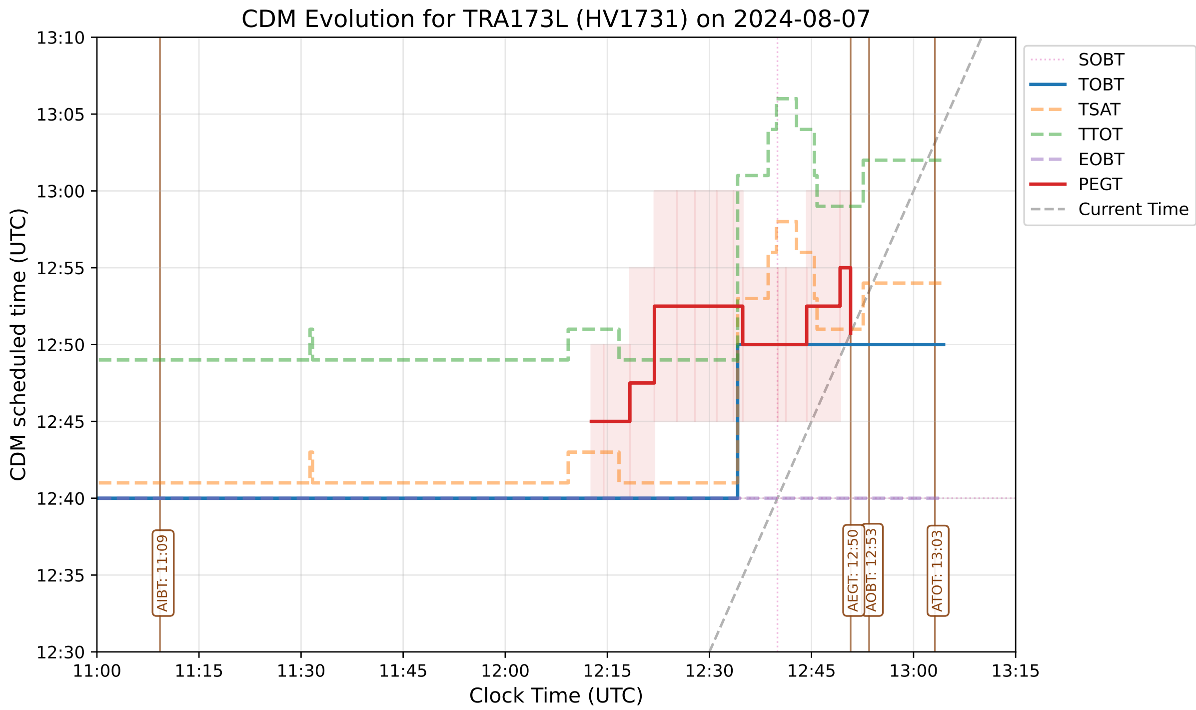



**Figure 2.1:** Evolution of CDM and PEGT times during the turnaround of TRA173L (HV1731) on 7 August 2024 at Amsterdam Airport Schiphol.

For this Transavia flight, the aircraft arrived on stand at 11:09, the Actual In-Block Time (AIBT). From this moment onward the handling agents begin their turnaround process. Early in the turnaround the Target Off-

Block Time (TOBT) remains equal to the Scheduled Off-Block Time (SOBT), which is the departure time printed on the passenger's ticket. This indicates that, at that stage, no delay was anticipated.

Also shown are the TSAT and TTOT, which are computed by the Collaborative Pre-Departure Sequence Planner (CPDSP) used by the tower's outbound planning position. Given a TOBT, the CPDSP determines the runway sequence. The difference between TSAT and TTOT reflects the predicted taxi time.

During this turnaround a manual TOBT update was entered, shifting it from 12:40 to 12:50. However, this update was only submitted at 12:34, six minutes before the originally planned off-block time. Until that very late moment, the scheduler was sequencing this flight using the outdated TOBT.

PEGT, in contrast, signalled a delay much earlier. At 12:12, PEGT predicted an off-block time of 12:45, which is far closer to the actual end of ground-handling at 12:50. Having this information earlier reduces the likelihood of sudden, last-minute changes to the runway sequence. Such late changes are highly disruptive for the outbound planner and may threaten runway throughput.

No prediction system is flawless, however, after this first accurate early estimate, PEGT continued to adjust its prediction. Near the end of the turnaround, at 12:49, PEGT briefly jumped to 12:55 even though the aircraft was close to completion. This behaviour is a known limitation of the Deep Turnaround system. Deep Turnaround relies heavily on binary state inputs, such as whether the passenger bridge is connected, without information on whether boarding is nearly complete. When the bridge remains connected very late in the process, the algorithm interprets this as evidence of ongoing boarding and predicts additional delay, even when only a quick final action remains.

This example illustrates both the potential and the limitations of PEGT: while it can provide valuable early indications of delay, it cannot serve as a wholesale replacement for TOBT. The operational picture is more nuanced. The remainder of this chapter analyses PEGT behaviour at scale, using the full dataset for August 2024, to determine how consistently these patterns occur and where PEGT adds the most value.

# II. PEGT Data Analysis for August 2024

Building on the single-turnaround example in the previous section, this section analyses the behaviour of PEGT at scale and compares its performance to TOBT over a full month of operations. All results presented here are based on turnaround data from August 2024 at Amsterdam Airport Schiphol, provided by Schiphol Aviation Solutions for research purposes. August was selected as a representative high-traffic summer month, ensuring a sufficiently large and operationally relevant sample.

The dataset is predominantly composed of narrow-body operations (approximately 80% of turnarounds with PEGT available). This reflects two factors: more stands serving narrow-body aircraft were equipped with Deep Turnaround cameras as of August 2024, and narrow-body stands have higher daily turnover due to shorter turnaround times.

## A. Prediction error over the course of the turnaround

The prediction error of PEGT and TOBT is defined as the difference between the predicted off-block time and the Actual End of Groundhandling Time (AEGT). To allow comparison across flights, time on the horizontal axis is expressed relative to the AEGT, such that positive values indicate minutes before departure.

Figure 2.2 compares the absolute prediction error of PEGT and TOBT as a function of time before AEGT. Only turnarounds for which both TOBT and PEGT were available at some point during the process are included.

A clear crossover point can be observed. Up to approximately 27–28 minutes before AEGT, TOBT is on average more accurate than PEGT. From this point onward, PEGT consistently outperforms TOBT, both in terms of median error and dispersion. As departure approaches, PEGT exhibits a faster reduction in error dispersion than TOBT, with increasingly narrower interquartile ranges and shorter tails. This indicates that, once available and sufficiently close to departure, PEGT provides a more accurate estimate of the true end-of-ground-handling time. Note that data at horizons beyond approximately 30 minutes before AEGT are sparse due to limited PEGT availability at those lead times (see Section B); statistics in that region should therefore be interpreted with caution.

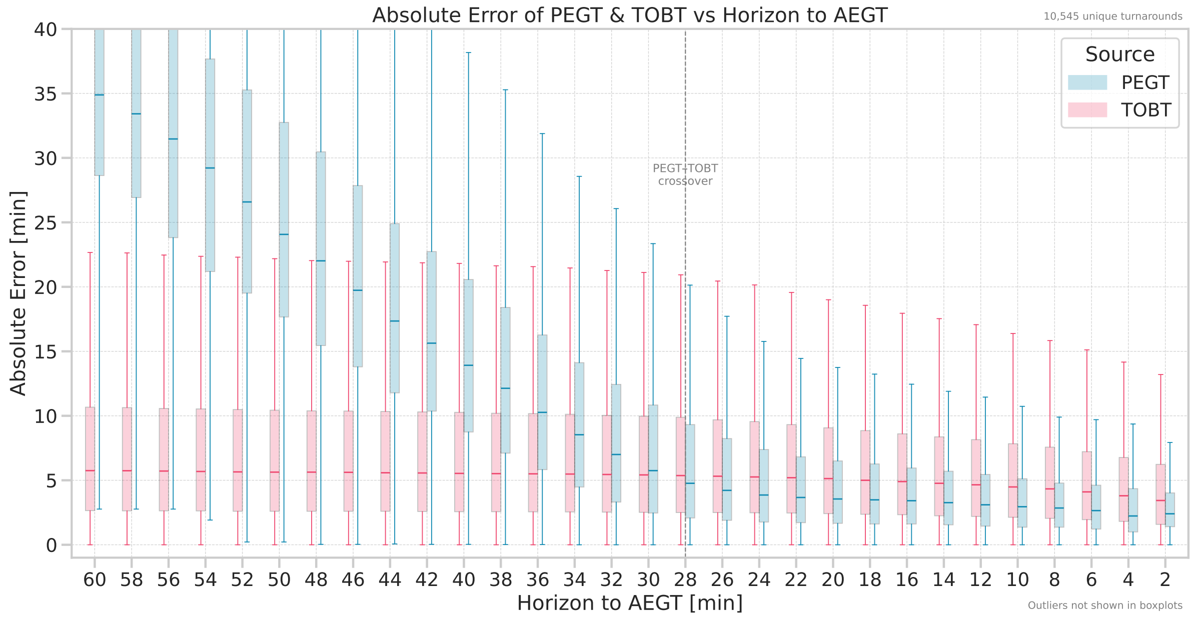



**Figure 2.2:** Absolute prediction error of PEGT and TOBT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

## B. Availability of PEGT during the turnaround

The interpretation of Figure 2.2 requires additional context regarding the availability of PEGT. Early in the turnaround, PEGT predictions are sparse, as the system requires sufficient sensor and process information before producing its first estimate.

Figure 2.3 shows the distribution of the first availability of PEGT and TOBT relative to AEGT. While TOBT is typically available three hours in advance, PEGT becomes available much later. Only 50% of turnarounds receive their first PEGT prediction by approximately 28 minutes before AEGT. At the 25th percentile, the first prediction arrives at 23 minutes before AEGT; at the 10th percentile, PEGT does not appear until fewer than 19 minutes remain.

This spread indicates that not all PEGT predictions are equally useful from an operational perspective. For a substantial fraction of flights, outbound planning must still rely on TOBT for most of the turnaround. In those cases, PEGT only becomes available very late into the turnaround, limiting its potential contribution to earlier sequencing decisions and overall operational efficiency.

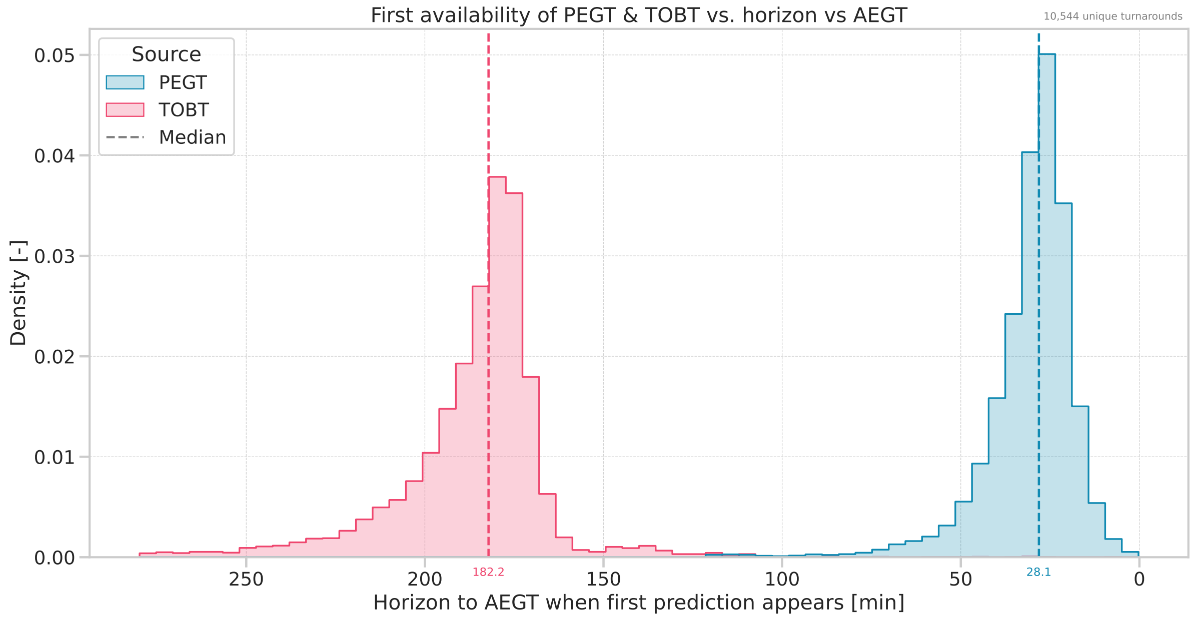



**Figure 2.3:** First availability of TOBT and PEGT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

## C. Bias in readiness estimates

The analysis in Figure 2.2 focuses on absolute prediction error and therefore does not capture systematic over- or underestimation. To complement this, the bias is analysed in Figure 2.4. It is defined as the median of estimate minus AEGT.

At horizons earlier than 30 minutes before AEGT, PEGT exhibits an extremely large negative bias, reflecting the incorrect predictions as shown previously in Figure 2.2 and limited data due to sparse availability of PEGT. In this region, TOBT shows almost no bias. Between roughly 30 and 12 minutes prior to AEGT, PEGT remains systematically optimistic, predicting earlier readiness than observed in reality, while TOBT stays closer to zero bias.

A clear turning point occurs around 12 minutes before AEGT. From this point onward, PEGT becomes pessimistic, overestimating the remaining time to completion. This shift is markedly stronger for PEGT, whose bias increases rapidly in magnitude. This behaviour is consistent with the system characteristics illustrated in Section I, where late-stage updates tend to introduce conservative corrections.

These results indicate that, despite its lower absolute error close to departure, PEGT exhibits a larger and time-varying systematic bias than TOBT. Both the magnitude of this bias and the horizon-dependent sign change should be considered when interpreting accuracy results and assessing the operational use of PEGT updates.

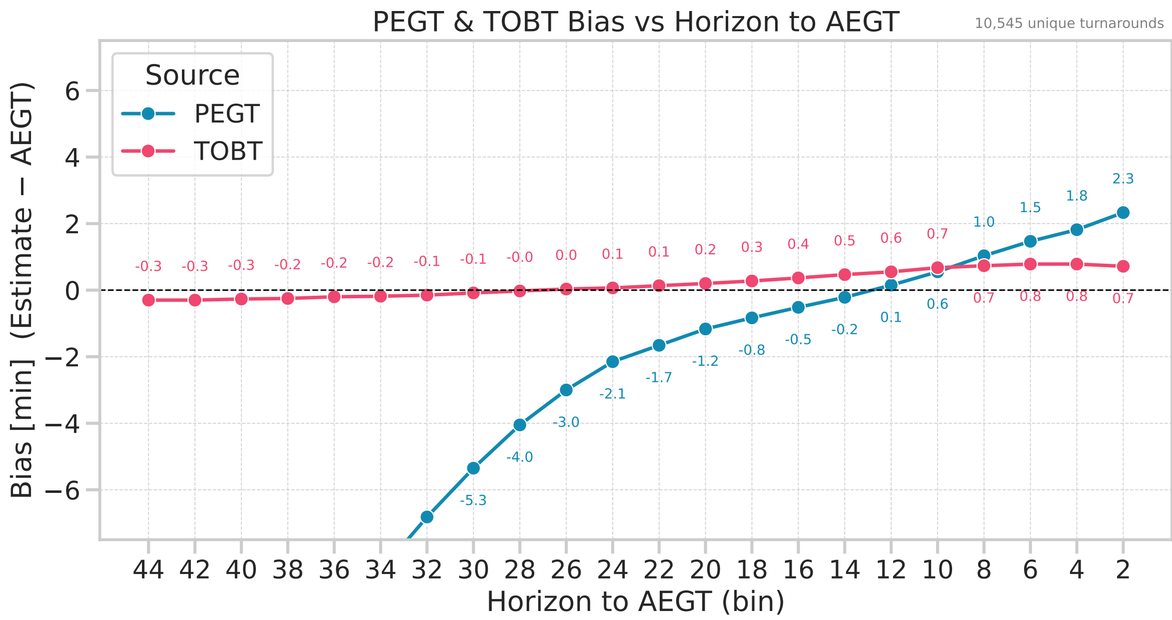



**Figure 2.4:** Bias in prediction error of PEGT and TOBT relative to AEGT for August 2024 at Amsterdam Airport Schiphol.

## D. Operational perspective: error relative to own prediction

The analyses above are based on AEGT, which is only known after the turnaround has completed. While this is suitable for retrospective performance evaluation, it does not reflect the information available during live operations.

To better approximate an operational setting, the same errors are therefore plotted relative to each prediction's own time reference. Figure 2.5 shows the absolute error as a function of time relative to the respective TOBT or PEGT value.

When viewed from this perspective, the crossover point shifts slightly. PEGT begins to outperform TOBT at approximately 32 minutes before the predicted off-block time. Although the difference is modest, the overall conclusion remains consistent: once sufficiently close to departure (~30 min) and once PEGT is available, PEGT provides a more accurate estimate than TOBT. A clear increase in prediction error is visible in the final minutes before AEGT, where PEGT tends to overcorrect and predict additional delay; as discussed in Section I, this is a known limitation related to the coarse, mostly binary input signals and has not yet been resolved operationally. Data at horizons beyond approximately 40 minutes before departure are sparse due to limited PEGT availability and should be interpreted with caution.

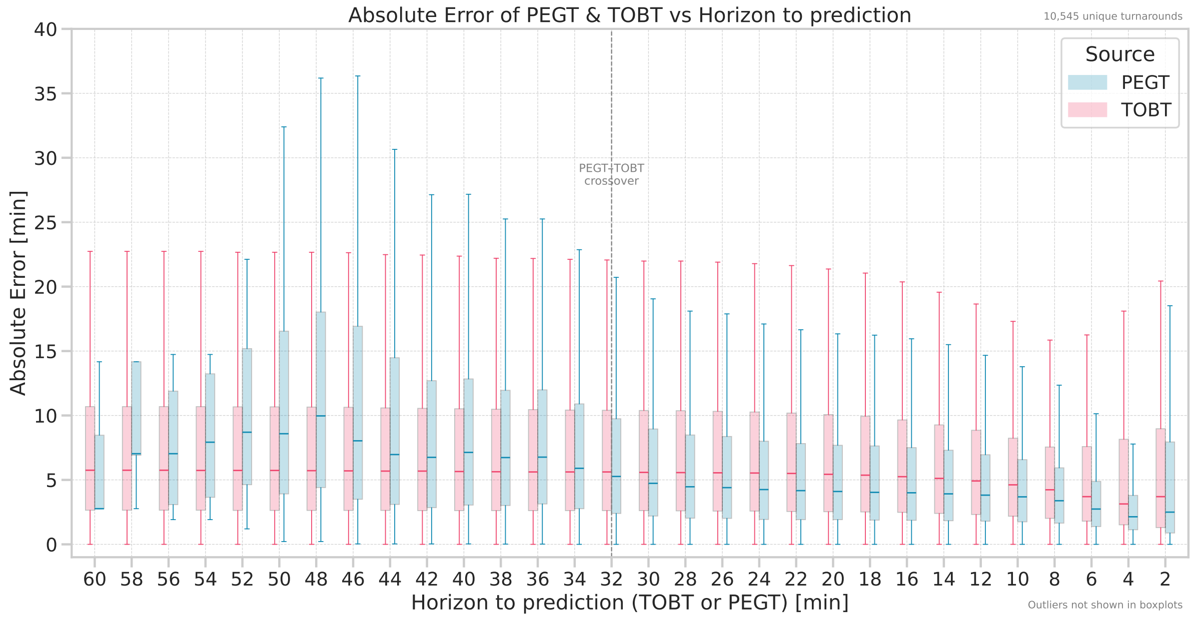



**Figure 2.5:** Absolute prediction error of PEGT and TOBT relative to their own prediction times for August 2024 at Amsterdam Airport Schiphol.

## E. Update frequency during the turnaround

In addition to differences in prediction accuracy, PEGT and TOBT also differ substantially in how often they are updated during a turnaround. Figure 2.6 shows the distribution of the number of predictions per turnaround for both sources.

On average, PEGT produces approximately 1.8 times more updates per turnaround than TOBT This higher update frequency reflects PEGT's data-driven nature: as additional ground-handling information becomes available, the prediction is revised accordingly. In contrast, TOBT updates are largely manual and operationally triggered, resulting in a much lower update rate.

A particularly notable observation is that 47% of turnarounds receive only a single TOBT prediction When combined with the availability analysis in Figure 2.3, this implies that for a large fraction of flights, the TOBT used throughout most of the turnaround was entered three hours before departure. From an accuracy perspective, even one additional update would already provide substantial benefit.

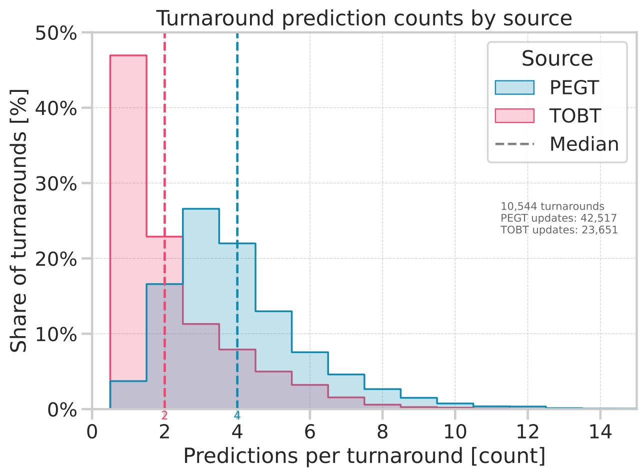



**Figure 2.6:** Distribution of the number of predictions per turnaround by source (TOBT vs PEGT) for August 2024 at Amsterdam Airport Schiphol.

The higher update frequency of PEGT therefore represents an inherent advantage in terms of information fidelity. However, this advantage comes with an important operational trade-off. Each update potentially triggers a resequencing action in the outbound planner, increasing workload and reducing predictability for pilots and ground handlers. Frequent revisions, even if individually more accurate, may therefore be undesirable when viewed from a human-in-the-loop operational perspective.

This tension between prediction fidelity and operational stability motivates the need for explicit policies governing how and when PEGT updates should be allowed to influence outbound planning.

## F. Marginal benefit of successive updates

Beyond absolute accuracy and update frequency, it is also relevant to assess the marginal benefit of each successive prediction update. To this end, the change in prediction error introduced by each new update is analysed as a function of the prediction index within a turnaround. For each update, the marginal benefit is defined as the reduction in absolute error relative to the previous prediction from the same source.

Figure 2.7 shows the distribution of this error reduction for PEGT and TOBT, plotted against the prediction index. Higher indices correspond to later updates within the same turnaround. It should be noted that the number of available samples decreases for higher indices, particularly for TOBT, as already observed in Figure 2.6.

The marginal benefit of successive prediction updates is evaluated by analysing the change in absolute error introduced by each new update relative to the previous one. Across all prediction indices, both PEGT and TOBT show positive mean and median error reductions, indicating that additional updates generally improve accuracy.
While individual updates may increase error, particularly for TOBT, no systematic degradation is observed at higher indices The difference in average marginal error reduction between PEGT and TOBT is primarily driven by update density PEGT issues more frequent updates, resulting in smaller incremental corrections per update, whereas TOBT updates are fewer and therefore induce larger stepwise changes.

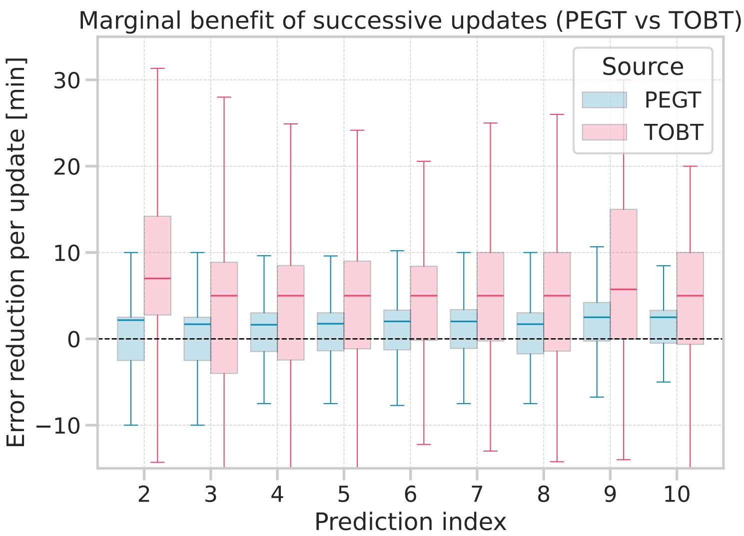



**Figure 2.7:** Marginal reduction in prediction error per successive update for PEGT and TOBT, plotted against the prediction index.

To assess whether single updates can be filtered from an operational perspective, the marginal error reduction is additionally analysed against their own prediction horizon rather than against the index (Figure 2.8). In this representation, no clear horizon or crossover point is observed at which PEGT performance systematically degrades. The mean error reduction remains positive across the operationally relevant range, indicating that later updates do not become predictably harmful.

Only when evaluated against AEGT does a degradation emerge, and this occurs very close to the event. However, this effect is not actionable for operational filtering, as PEGT updates cease approximately three minutes before their own predicted time due to the discrete update structure.

Consequently, no simple a priori filtering rule based on update index or operational horizon can be justified. Although only 60% of PEGT updates reduce absolute error (compared to 87% for TOBT), the absence of a clear degradation regime implies that update acceptance decisions must rely on additional metrics beyond timing alone.

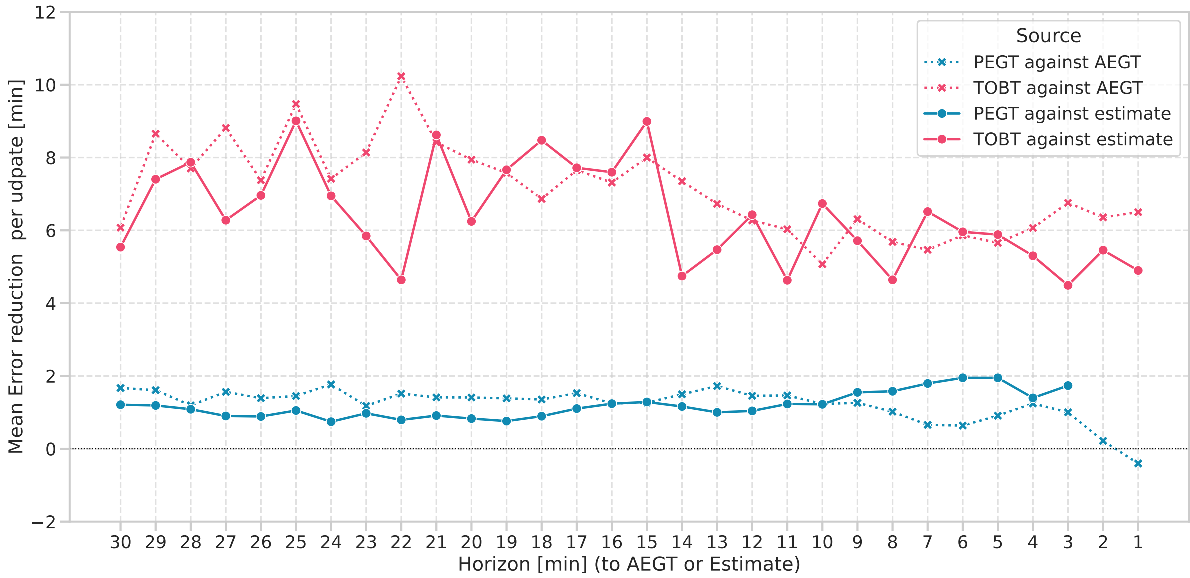



**Figure 2.8:** Mean reduction in absolute prediction error of PEGT and TOBT updates relative to AEGT and their own prediction times for August 2024 at Amsterdam Airport Schiphol.

When the change in error is plotted against the time elapsed since the previous update, no clear pattern emerges. This suggests that, to date, no single metric has been identified that reliably distinguishes beneficial from detrimental PEGT updates.

## G. Relative accuracy of first PEGT value per turnaround

To assess the added value of PEGT at the moment it becomes available, the first PEGT prediction is compared against the most recent TOBT available at that same time. The comparison is based on absolute prediction error with respect to AEGT and is summarised in Table 2.1.
In half of the cases, the first PEGT yields a lower absolute error than the corresponding TOBT, while in 20% of the cases both predictions perform equally. In the remaining 30% of cases, the PEGT performs worse than TOBT. Overall, this indicates that in approximately 70% of the turnarounds, the introduction of the first PEGT does not degrade prediction accuracy and frequently provides an immediate improvement.
This result suggests that, despite individual instances of performance degradation, the first PEGT update provides a net accuracy benefit at the time of its introduction per turnaround.

**Table 2.1:** Relative accuracy of the first PEGT of a turnaround compared to the TOBT.

| Comparisons | Count | Share |
|---|---|---|
| PEGT lower error | 5,209 | 49.4 % |
| TOBT lower error | 3,215 | 30.5 % |
| Equal error | 2,113 | 20.1 % |

## H. Error distribution characteristics

The distribution of signed errors (estimate − AEGT) for PEGT and TOBT is shown in Figure 2.9. Compared with TOBT, the PEGT error distribution is more sharply concentrated around zero, with a narrower central spread and less probability mass in the tails. This indicates lower dispersion (i.e., higher precision/consistency) of PEGT estimates relative to TOBT, and therefore suggests potential for improved timing performance. It does, however, exhibit a slight positive bias, indicating a tendency to overpredict AEGT (i.e., more delay) than were actually observed.

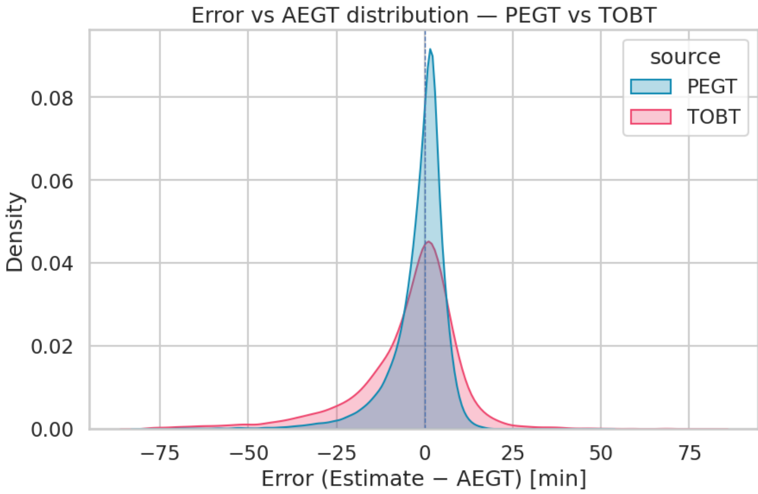



**Figure 2.9:** Kernel density estimates of signed prediction error (estimate − AEGT) for PEGT and TOBT. PEGT is more concentrated around zero, indicating lower dispersion than TOBT.

## I. Normality assessment by prediction horizon

The Q–Q plots in Figure 2.10 assess the distribution of signed forecast errors (PEGT−AEGT) at horizons defined relative to the *predicted* event time (i.e., PEGT-anchored, not AEGT-anchored). From this perspective, the error distribution is clearly non-Normal at most horizons: especially at short horizons (e.g., 5–15 min before PEGT), the Q–Q curves show a pronounced S-shape and tail deviations, indicating skewness and heavy tails (with a stronger negative tail). The central quantiles become more linear for intermediate horizons (roughly 20–30 min), suggesting a better Normal approximation in the center of the distribution, although tail non-Normality remains. The absence of observations at $t \approx 1$ min is structural rather than incidental: PEGT updates are generally not issued within the final few minutes before their own predicted time (minimum lead time of approximately 3 min). Likewise, the limited sample size for horizons beyond ~30 min reflects the operational timing of first predictions, which are typically generated within 30 min of departure; therefore, panels at 45+ min should be interpreted with caution due to sparse data. Because the error distribution is horizon-dependent and departs from Normality, any compensation strategy should be calibrated per prediction horizon and should account for asymmetry and tail risk, rather than relying on a single Gaussian error model. In practical terms, this suggests that bias correction may still be feasible, but probabilistic margins or decision thresholds should preferably be based on empirical error distributions rather than a Normal approximation.

## J. Pier-level error dynamics

Following the distributional analysis, the error progression was examined per pier to assess whether the PEGT error dynamics differ across airport areas (Figure 2.11). Clear pier-level differences are observed. In particular, two broad groups are visible: piers A–D and piers E–H. Relative to E–H, the A–D group shows smaller errors at longer horizons, but transitions earlier from an underestimation regime to an overestimation regime as departure approaches. In addition, the interquartile range remains comparatively wide for several piers, indicating substantial within-pier variability. In the final minutes before departure, all piers tend to move toward positive errors (overestimation), although the magnitude and timing of this transition differ by pier. These findings are descriptive and indicate that spatial/operational stratification may be relevant when modelling or correcting PEGT errors.

### Operational context of pier usage at Schiphol

A plausible operational explanation for the observed pier-level differences is the heterogeneous use of Schiphol piers. Gate allocation is dynamic and depends on factors such as border status (Schengen vs. non-Schengen), aircraft size, turnaround characteristics, and airline preferences, rather than fixed airline–pier assignments. In broad terms, piers B, C, M, and the Schengen section of D are typically used for Schengen traffic, which

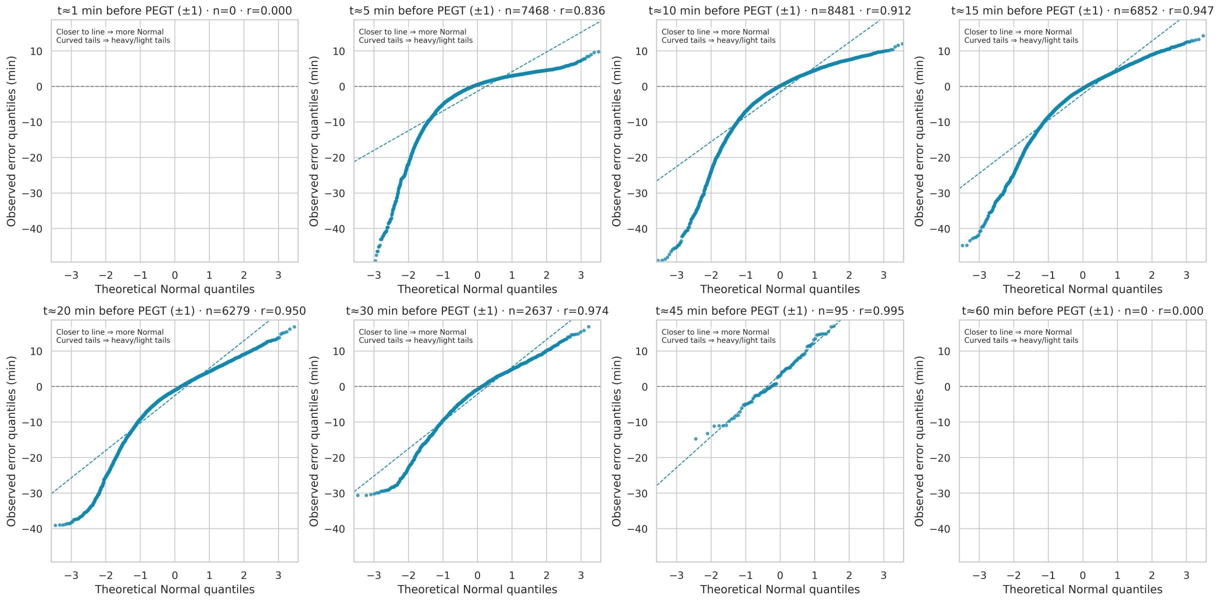



**Figure 2.10:** Q–Q plots of signed forecast error (PEGT−AEGT) by PEGT-anchored prediction horizon. Short horizons show clear deviations from Normality (skewness and heavy tails), while intermediate horizons exhibit a more linear central region with persistent tail deviations.

includes many short-haul European services operated by regional and narrow-body aircraft. By contrast, part of D and piers E, F, G, and H are used for non-Schengen traffic; in operational practice, E/F/G are frequently associated with intercontinental traffic and wide-body aircraft. Pier A is mainly used by KLM Cityhopper for smaller Embraer aircraft at remote stands, where boarding is via bus and stairs rather than a connected bridge. Whereas H/M often accommodate lower-cost operations subject to border-status constraints. Because stand and gate planning are dynamic, these patterns should be interpreted as typical operational tendencies rather than strict assignments. Accordingly, the pier-level error differences in Figure 2.11 should be interpreted as consistent with differences in traffic mix and turnaround processes, rather than as evidence of a single causal factor.

A formal pier-level comparison (e.g., median error and dispersion per prediction horizon with confidence intervals) is left for future work. Such an analysis could reveal systematic, time-varying differences in error behaviour between piers. These differences might support infrastructure-level calibration of the readiness predictor or DMAN inputs. Any operational use would need to be constrained to avoid flight- or airline-dependent prioritisation based solely on gate assignment.

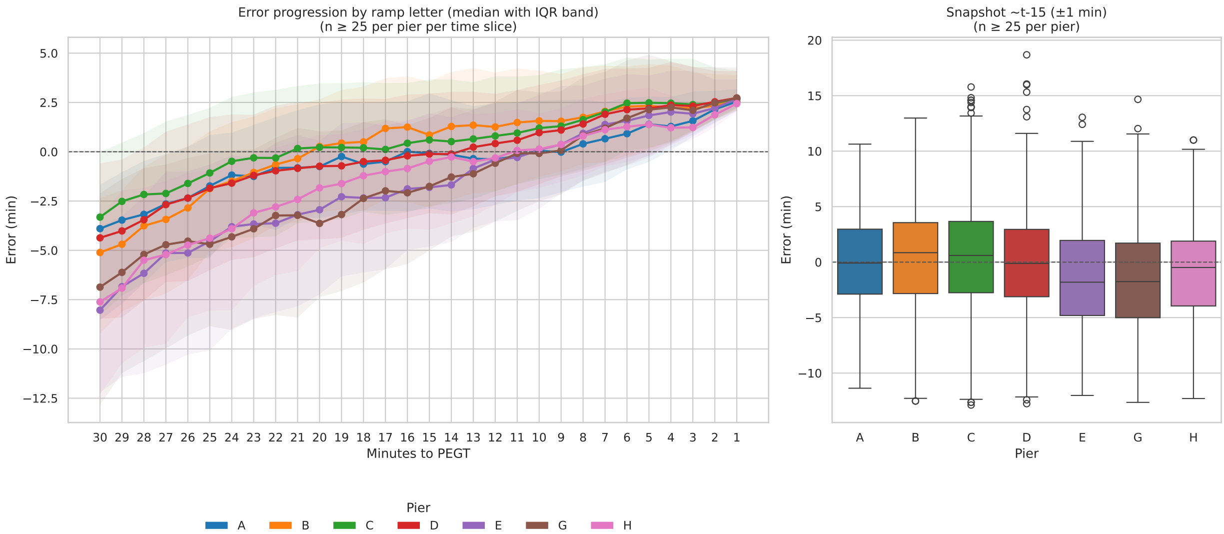



**Figure 2.11:** Progression of signed PEGT error by pier as a function of time-to-departure. Distinct pier-level patterns are visible, including differences in bias transition timing and dispersion, suggesting that error dynamics depend on operational context.

# 3

# Weather Analysis for Nominal Operational Conditions

This appendix documents the meteorological screening performed to verify that August 2024 represents a period of nominal operations at Amsterdam Airport Schiphol (EHAM). The main paper assumes that all 31 operating days are representative of normal traffic flow patterns; this analysis substantiates that assumption. Periods of restricted operations (e.g., due to high crosswind, low visibility, or significant weather phenomena) could bias departure-performance metrics and are therefore identified and assessed for their operational impact.

## A. Data Sources and Methodology

The analysis utilised two data sources from the airport's operational databases:

- **Runway configuration data (RunwayMRI):** Records which runways are designated for landing and takeoff operations at each point in time, with validity periods defined by start and end timestamps. This table is updated whenever the runway configuration changes.
- **Meteorological observations (MeteoActual):** Minute-by-minute records of wind direction and speed, gusts, visibility, temperature, dewpoint, QNH pressure, and weather phenomena codes. In practice, parameter values change at least every 30 minutes.

## B. Nominal Operation Criteria

The criteria for nominal meteorological conditions were based on operational limits for Beperkt Zicht Operaties (BZO; reduced visibility operations) and published Boeing 737 limitations. The thresholds used here are set deliberately below the Boeing 737 limits to ensure a conservative margin, and are intended to be broadly representative of operational constraints for most commercial aircraft operating at Schiphol, excluding general aviation:

**Table 3.1:** Meteorological limits defining nominal operational conditions

| Parameter | Limit | Unit |
|---|---|---|
| Maximum crosswind (steady) | 25.0 | kt |
| Maximum crosswind (gust) | 30.0 | kt |
| Maximum tailwind (steady) | 5.0 | kt |
| Maximum tailwind (gust) | 10.0 | kt |
| Minimum visibility | 2000 | m |
| *Capacity-reducing weather phenomena:* | | |
| TS, CB, FG, FZ, FZRA, FZDZ, FZFG, SN, SG, PL, GR, GS, SQ, DS, SS, +RA | | |

Wind components were calculated relative to each runway's magnetic heading. Crosswind and tailwind components were derived from the reported wind direction and speed using vector decomposition:

$$\text{along-wind} = v_{\text{wind}} \cdot \cos(\theta_{\text{wind}} - \theta_{\text{runway}}) \quad (3.1)$$

$$\text{cross-wind} = v_{\text{wind}} \cdot \sin(\theta_{\text{wind}} - \theta_{\text{runway}}) \quad (3.2)$$

where positive along-wind indicates headwind (favorable) and negative indicates tailwind (restricting).

## C. Analysis Process

The analysis was performed at minute-level resolution for each day in August 2024. For each runway and minute, the following steps were executed:

1. **Runway operational status determination**: Using the RunwayMRI data, boolean time series were constructed indicating whether each runway was open for landing operations, takeoff operations, or both during each minute of the day.
2. **Meteorological condition mapping**: For each runway, minute-level meteorological parameters were extracted from the MeteoActual table. Values were forward-filled to handle brief gaps between meteorological updates.
3. **Non-nominal condition flagging**: Each minute was flagged as non-nominal if any meteorological limit from Table 3.1 was exceeded, specifically: crosswind or tailwind (steady-state or gust), visibility, or the presence of capacity-reducing weather phenomena.
4. **Operational impact assessment**: Non-nominal conditions were only considered significant when they occurred during periods when the affected runway was actively in use according to the MRI.
5. **Daily nominal percentage calculation**: For each day, the percentage of nominal minutes was calculated as the ratio of minutes meeting all criteria to total operational minutes across all active runways.

## D. Results for August 2024

The analysis revealed that August 2024 experienced exceptionally favorable weather conditions. Out of 31 days analyzed:

- 30 days (96.8%) exhibited 100% nominal conditions during all operational periods
- Only 1 day (August 11th) contained a brief period of non-nominal conditions
- The non-nominal period on August 11th lasted less than one hour (approximately 50 minutes)
- This restricted period occurred around 05:00 UTC and affected only the active departure runway
- The restriction was primarily due to crosswind conditions marginally exceeding operational limits

Given that the single non-nominal period represented less than 0.1% of total operational time in August 2024, and only 1 of the 2 active departure runways, the entire dataset for August 2024 was deemed representative of normal operations and suitable for subsequent analysis without requiring exclusion of any days or time periods.

Figure 3.1 illustrates the runway operational timeline for August 11, 2024, the only day with non-nominal conditions. The figure shows:

- Horizontal axis: Time of day (UTC), spanning 00:00 to 24:00
- Vertical axis: Runways, with separate rows for landing (LD) and takeoff (TO) operations
- Blue bars: Periods when runway is designated for takeoff operations
- Sand/beige bars: Periods when runway is designated for landing operations
- Red shading: Overlay indicating periods when meteorological conditions exceeded nominal limits

The visualization clearly shows that the non-nominal conditions (red shading) were confined to a brief morning period and affected only a subset of the active runways, confirming the minimal impact on overall operational representativeness.

## E. Implications for Study Validity

The exceptional meteorological conditions during August 2024 provide an ideal dataset for this study's objectives. The near-complete absence of weather-related operational restrictions ensures that traffic patterns reflect normal procedures rather than weather-imposed constraints, that capacity utilisation measurements represent airport capabilities under favourable conditions, and that statistical analyses are not confounded by meteorological variability. No data exclusions were necessary, maximising the available sample size for subsequent analyses.

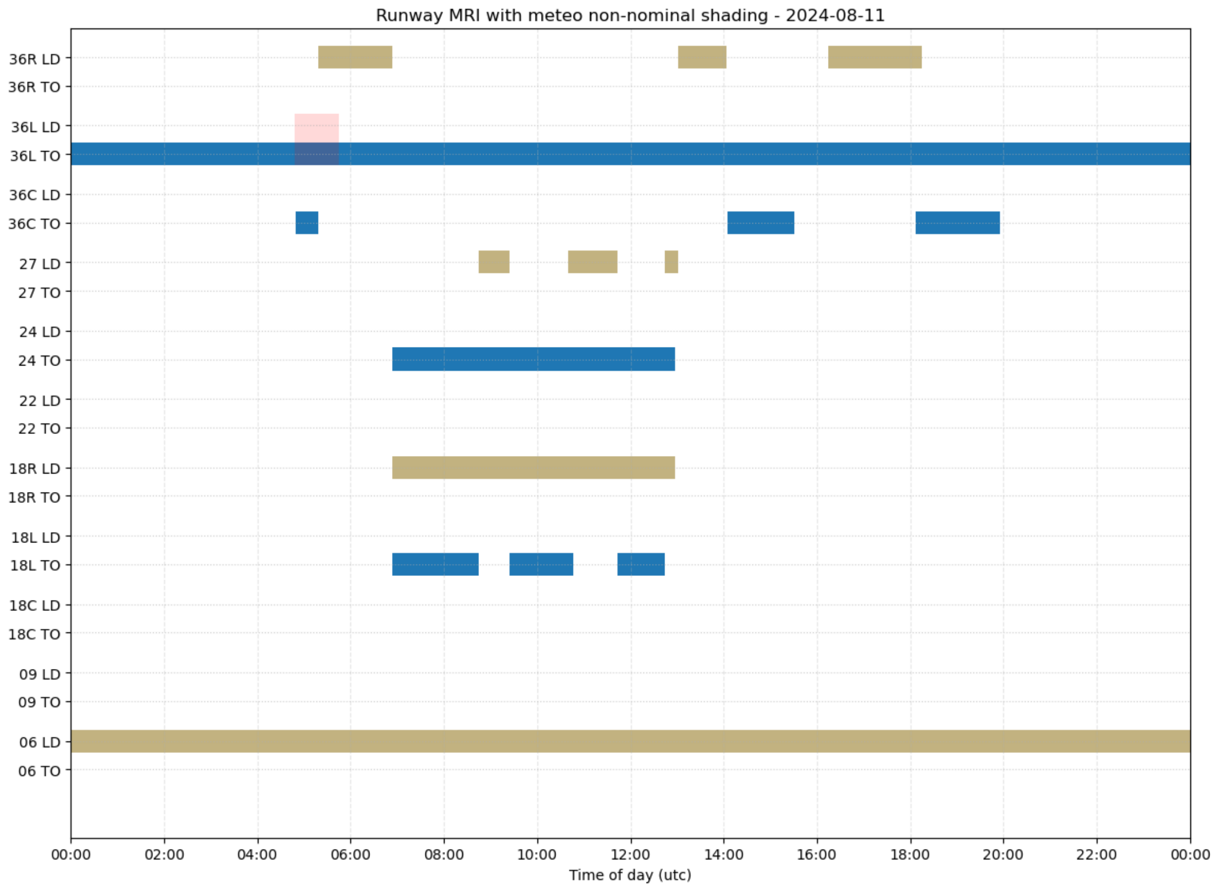



**Figure 3.1:** Runway operational timeline for August 11, 2024, showing runway opening times per operation type (LD = Landing, TO = Takeoff) with red shading indicating periods exceeding nominal meteorological conditions. Blue bars indicate takeoff operations, sand-colored bars indicate landing operations, and red shading marks the brief morning period when crosswind limits were exceeded.

# 4 SID Divergence and Inter-Departure Separations

This appendix presents an exploratory analysis of observed inter-departure separation times at Amsterdam Airport Schiphol, differentiated by Standard Instrument Departure (SID) pair. The analysis complements the main paper's departure-management study by examining a factor not captured by the reconstructed DMAN: the dependence of effective runway throughput on the SID combination of consecutive departures. The results are based on August 2024 operational data for the three most frequently used departure runways (24, 18L, 36L).

## I. Background: SIDs, Wake Turbulence, and Separation

A Standard Instrument Departure (SID) is a predefined departure route that connects the runway to the en-route airway structure. Each SID prescribes an initial climb profile and lateral track. Because different SIDs route aircraft in different lateral directions immediately after takeoff, the track divergence between a consecutive pair of departures affects the minimum separation that ATC must apply.

Minimum time-based wake turbulence separations between consecutive departures are governed by the RECAT-EU (European Wake Turbulence Re-Categorisation) scheme (Dijkstra et al., 2020; van Baren, 2016), which assigns each aircraft to one of six categories based on its wake vortex characteristics. For a given leader–follower category pair, RECAT-EU prescribes a minimum separation. However, when two consecutive departures are assigned to the same or a non-diverging SID, additional radar separation is required to ensure safe spacing for approach control. This may exceed the wake turbulence minimum. Conversely, when SIDs diverge (i.e., the leader and follower bank in opposite directions shortly after takeoff), the required radar separation is achieved more quickly, and the two aircraft can be released with a shorter interval. This mechanism makes SID-pair composition a significant lever for runway throughput.

In operational practice, the final release decision rests with the runway controller in the ATC tower, who accounts for additional factors such as wind conditions, dependent runway interactions, and real-time traffic flow. These factors contribute to the variability in observed separations beyond what SID-pair geometry alone would predict.

## II. Heatmap Construction

### A. Definition of a SID Pair

Each cell in the heatmaps (Figures 4.1–4.3) represents a *consecutive* departure pair on the same runway. The row corresponds to the *preceding* flight's SID (abbreviated to its first three characters), and the column corresponds to the *following* flight's SID. A pair is included only if the two departures occur within 180 s of each other (based on ATOT timestamps).

### B. Computation of Cell Values

For each combination of runway, Wake Turbulence Category (WTC) pair (e.g., M–M, H–M), and SID pair, the *modal* observed separation time is computed from the filtered ATOT differences. The modal value is displayed as the number inside each cell. A minimum sample-size threshold of 50 occurrences is applied to suppress noise from infrequent pairs.

### C. Colour Encoding

The heatmap colours encode the deviation from the RECAT-EU separation standard: $\Delta$ = observed mode – RECAT-EU (seconds).

- **Green (negative $\Delta$):** observed separations are *shorter* than the RECAT-EU minimum (more efficient).
- **White ($\Delta \approx 0$):** aligned with the RECAT-EU standard.
- **Red (positive $\Delta$):** observed separations are *longer* than the RECAT-EU minimum (more constraining).

Cells with large $|\Delta|$ and high sample counts identify the most consistently constraining (dark red) or efficient (dark green) SID pairs.

## III. Key Observations and Relevance for DMAN

Observed separations differ substantially across SID pairs even within the same WTC class, indicating pair-specific operational behaviour driven by track divergence, controller workload, and flow interactions. The most pronounced variation appears in the Medium–Medium (M–M) category, which dominates traffic volume at Schiphol. Under the reconstructed DMAN's fixed 10-minute binning structure, outbound sequencing does not account for SID-pair-dependent separation differences. The heatmaps therefore highlight a potential opportunity: incorporating SID-pair awareness into departure sequencing logic could increase effective runway throughput or reduce delay without modifying formal separation standards.

This analysis is exploratory and does not claim causal attribution; the observed patterns may also reflect correlated factors such as runway configuration interdependencies or time-of-day effects. A quantitative assessment of attainable throughput gains through SID-pair-aware sequencing (particularly for M–M pairs) is left for future work.

## IV. Heatmaps by Runway

The runway-specific heatmaps are provided in Figures 4.1–4.3. Each figure contains multiple panels corresponding to the WTC pairs observed on that runway.

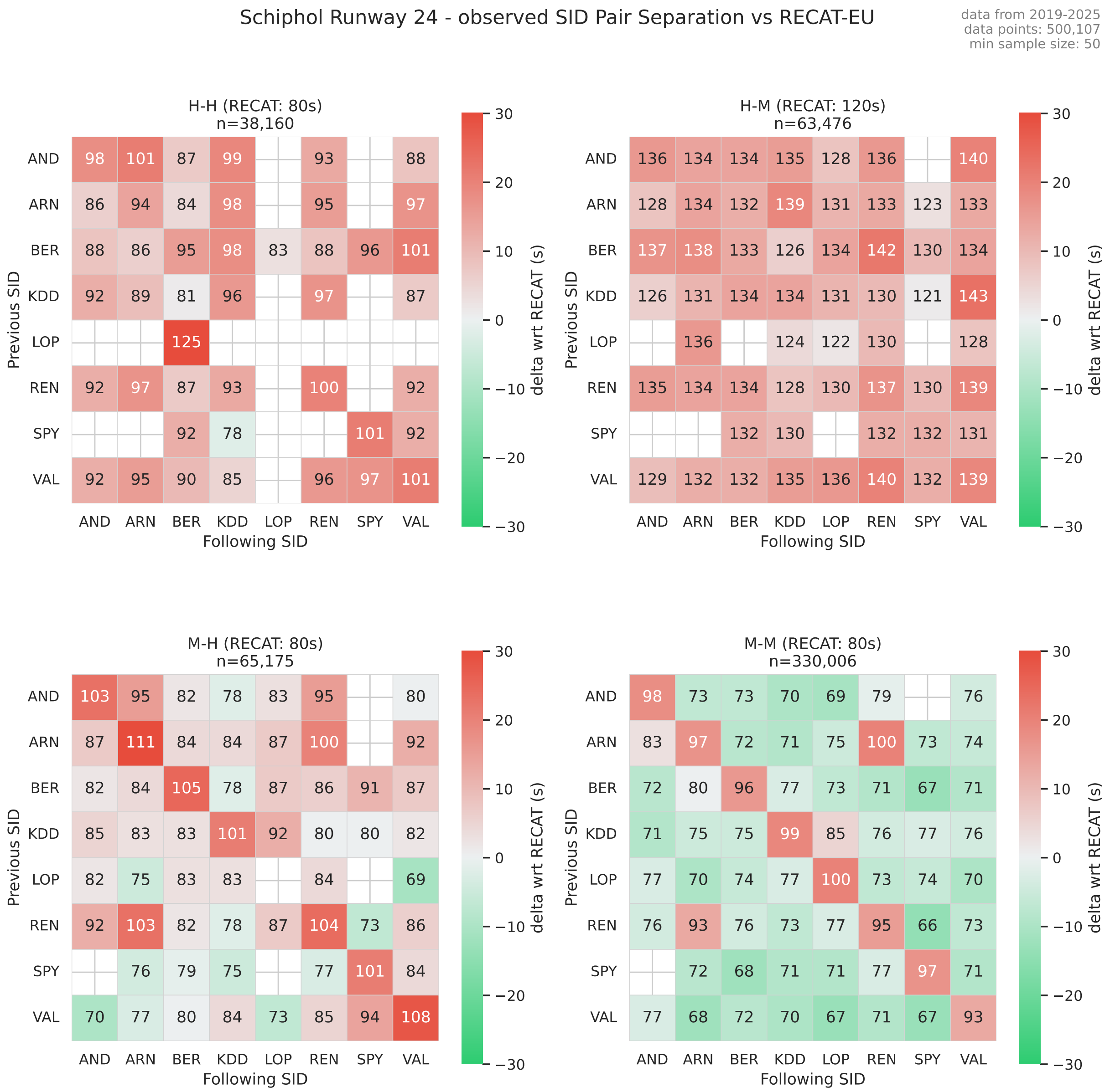



**Figure 4.1:** Runway 24: observed SID-pair separation versus RECAT-EU for the most common WTC pairs. Cell values show modal separations (s); colors show $\Delta$ vs RECAT-EU.

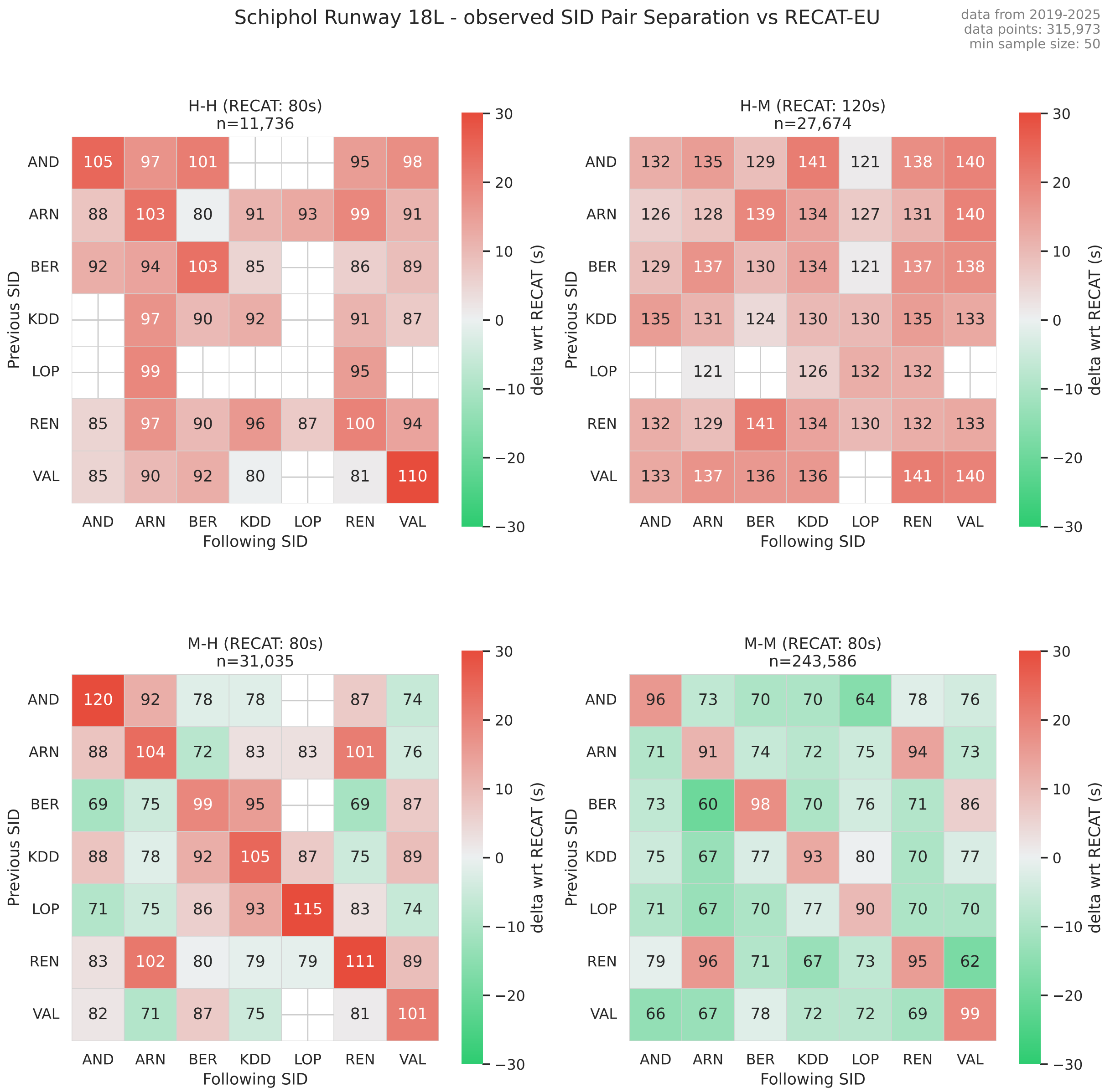



**Figure 4.2:** Runway 18L: observed SID-pair separation versus RECAT-EU for the most common WTC pairs. Cell values show modal separations (s); colors show $\Delta$ vs RECAT-EU.

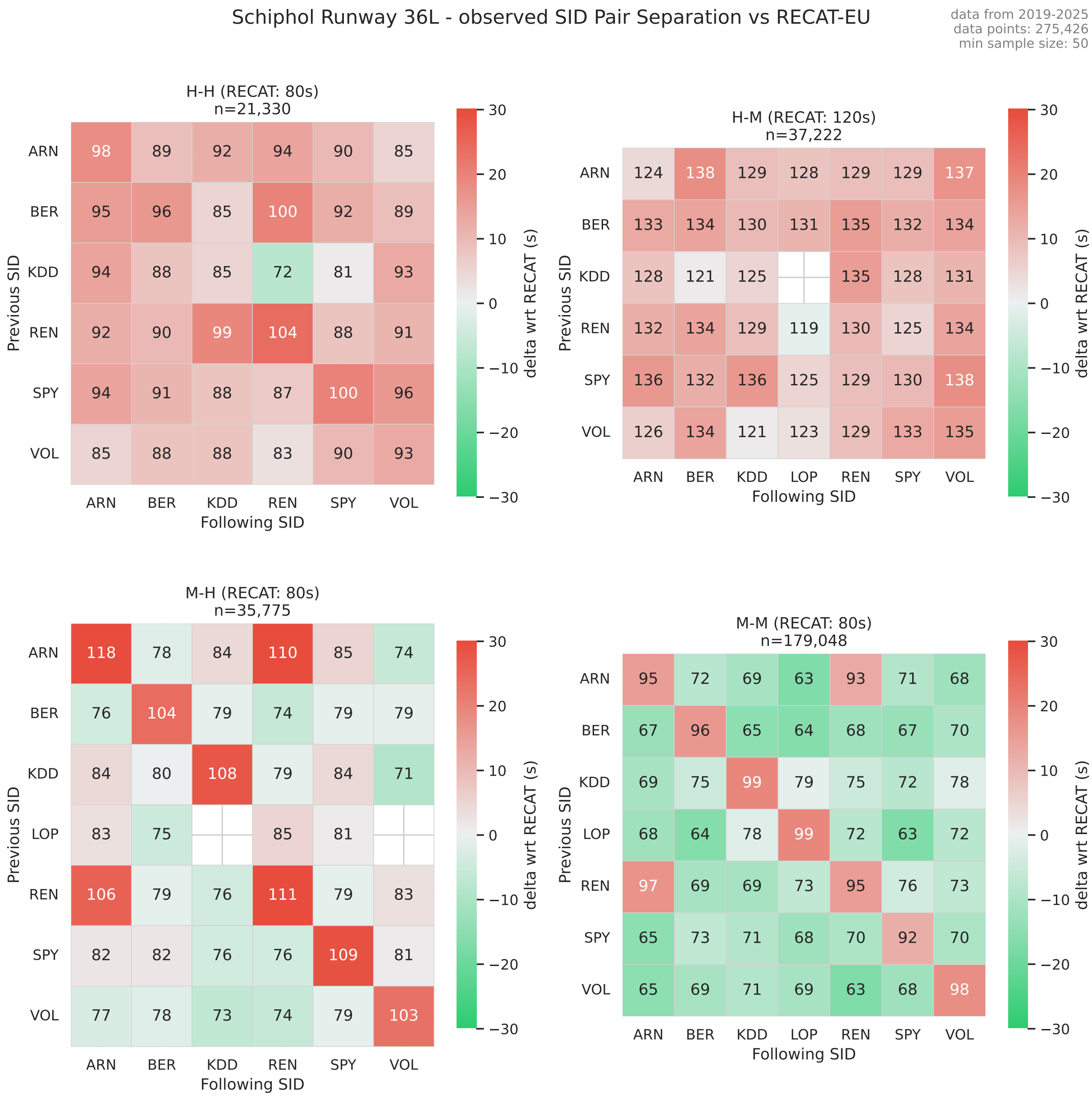



**Figure 4.3:** Runway 36L: observed SID-pair separation versus RECAT-EU for the most common WTC pairs. Cell values show modal separations (s); colors show $\Delta$ vs RECAT-EU.

# Part C

## Literature Review & Research Definition

# 5
# Literature Review

This chapter provides the theoretical foundation for the research and identifies key gaps that motivate the proposed methodology. Its purpose is to consolidate the current state of knowledge relevant to departure sequencing at major airports, particularly Schiphol, and to contextualize how emerging data-driven techniques can enhance predictability within operational constraints. The literature review is structured into five main parts. First, it introduces the concept of the departure manager (DMAN) and outlines the critical constraints that must be considered when sequencing departures, such as wake turbulence, route dependencies, and slot compliance. Second, it explores the aircraft turnaround process at A-CDM airports, emphasizing the role of accurate target off-block time (TOBT) reporting in enabling effective departure planning. Third, the chapter examines Deep Turnaround, an AI-powered system at Schiphol that provides high-confidence predictions (predicted end of ground-handling time (PEGT)) of turnaround completion based on real-time visual monitoring. Fourth, it reviews data fusion methods that can combine the early availability of TOBT with the improved accuracy of PEGT, aiming to produce a more reliable single estimate of actual off-block time (AOBT) suitable for use in the DMAN. Finally, it discusses techniques to classify TOBT update behavior and uncertainty profiles, with the goal of enhancing AOBT prediction accuracy by accounting for flight-specific delay tendencies during the turnaround.

Together, these sections form the conceptual basis for the proposed fusion-based improvement to Schiphol's departure sequencing process.

## 1. Departure Manager

A departure manager (DMAN) is a comprehensive system responsible for strategic departure planning and aircraft sequencing at airports. While often confused with simpler Pre-Departure Sequencers (PDS) that primarily manage push-back timing, a DMAN operates as a sophisticated runway-departure planner that optimizes the entire departure flow up to 2-3 hours ahead of takeoff. These systems determine the optimal departure sequence and timing by considering complex operational constraints including runway capacity, arrival-departure interlacing, standard instrument departure (SID) spacing requirements, CTOT compliance, de-icing capacity, and weather conditions. Modern DMAN implementations typically incorporate PDS functionality while providing strategic optimization that maximizes throughput, minimizes delays, and maintains network flow efficiency.

This allows for a larger prediction horizon and gives everyone in the chain information to work with. For example, to predict capacity constraints, or how many ground personnel are needed to be ready.

For busy airports a DMAN becomes a necessity, since it is not feasible for air traffic controller to continuously solve it in their head. During an outbound peak at Schiphol, it is common to see 3 runways in use simultaneously, 1 arrival and 2 departure runways. During such a peak, around 34 aircraft per hour arrive and 80 aircraft per hour depart.

These peaks require optimal planning so as not to let any runway capacity go to waste. The throughput of the runway is highly dependent upon the exact departure sequence. This is due to several constraints, which are further explored below.

### 1.1. Aircraft Separation Constraints

The air traffic controllers in the tower take quite a lot into account before giving a pilot 'cleared for takeoff'. They need to be certain all aircraft remain (1) horizontally separated by 3 nautical miles or (2) vertically separated


**Figure 5.1:** Runway layout at Schiphol Amsterdam Airport (Schiphol, n.d.)

by 1,000 feet. These are the requirements for the terminal maneuvering area (TMA) in Dutch airspace (Air Traffic Control the Netherlands (LVNL), 2025). Outside the TMA the horizontal requirement becomes more stringent, namely 5 nautical miles.

The most important considerations regarding the separation are discussed in this section together with their impact on the DMAN.

**Dependent Runways**

Amsterdam Airport Schiphol operates a complex runway system comprising of six runways (Figure 5.1), which, due to their spatial configuration, often function in dependent modes. This dependency arises when the operation of one runway influences or restricts the use of another, thereby impacting overall airport capacity (Derks, 2020). Such dependencies are primarily categorized into intersecting and converging runway operations.

Intersecting runways physically cross each other, necessitating careful coordination to prevent conflicts. For instance, the 'Buitenveldertbaan' (Runway 09/27) and 'Aalsmeerbaan' (Runway 18L/36R) intersect, making simultaneous operations infeasible under standard procedures. This physical intersection mandates staggered usage, thereby limiting capacity during peak periods.

Converging runways do not physically intersect but have approach or departure paths that converge, leading to potential conflicts, especially during missed approaches or due to jet blast effects. A notable example is the interaction between arrivals on the 'Zwanenburgbaan' (Runway 36C) and departures from the 'Kaagbaan' (Runway 24). Arrivals from the south on Runway 36C intersect with departures heading southwest on Runway 24, creating a dependency that requires precise timing and coordination to maintain safety and efficiency.

These dependencies necessitate increased separation minima and coordination, often resulting in reduced runway throughput. For example, a departure from the 'Aalsmeerbaan' (Runway 18L) may be delayed until the jet blast from a preceding departure on the 'Kaagbaan' (Runway 24) has dissipated, ensuring safe operations.

The complexity of Schiphol's runway layout, combined with these dependencies, means that certain runway combinations are avoided during normal operations due to their complexity and the potential for reduced capacity. Others remain in use but require meticulous timing and coordination, further limiting efficiency.

**SID divergence**

The concept of 'miles-in-trail' refers to the distance two consecutive aircraft share the same flight path. When aircraft follow different SIDs routes, their paths diverge. SIDs are predefined departure procedures that aircraft must follow after takeoff. The shorter the shared distance between aircraft paths, the sooner separation requirements are satisfied, allowing for reduced separation times and increased runway throughput. Air traffic controllers can strategically leverage this by alternating SIDs in the departure sequence to optimize capacity.

This principle is particularly powerful when applied to parallel runways, where departure traffic can be segregated based on their SID. At Schiphol Airport, this strategy is regularly employed with the 'Polderbaan' and 'Zwanenburgbaan' runways operating simultaneously for departures. By assigning aircraft with similar SIDs to the same runway, immediate divergence after takeoff is guaranteed for aircraft departing from different runways. This optimization not only ensures safe separation but also maximizes operational efficiency, as aircraft on one runway can commence their takeoff roll while the previous aircraft from another runway has just taken off.

**Wake Turbulence (RECAT-EU)**

All aircraft generate wake turbulence (rotating air masses known as wake vortices) that are formed whenever a wing produces lift. These vortices originate at the wingtips due to pressure differences between the upper and lower wing surfaces and trail behind the aircraft in a pair of counter-rotating spirals. In large aircraft, the vortex core can reach speeds of up to 100 m/s and extend over 30 meters in diameter. Wake vortices can persist for several minutes, especially in calm, stable air. The greatest hazard they pose is induced roll and yaw, which can be particularly dangerous during takeoff or landing when there is little altitude for recovery (Anderson Jr., 1999). Lighter aircraft, especially those with short wingspans, are most vulnerable. In extreme cases, encountering a strong vortex can cause rapid loss of control, structural overload, or high descent rates exceeding 1000 feet per minute. Consequently, maintaining safe separation between aircraft is critical to avoiding wake-related incidents.

The (older) International Civil Aviation Organization (ICAO) standard for wake turbulence categories grouped aircraft into just three categories: Heavy, Medium, and Light, based solely on weight. Although safe, this method often resulted in unnecessarily large separation distances, reducing runway throughput and operational efficiency. European wake turbulence re-categorisation (RECAT-EU) refines this approach by introducing six categories (A–F), from 'Super Heavy' to 'Light', allowing for more precise and often reduced separation requirements depending on the leader-follower aircraft pair.

Eurocontrol, the European air navigation organization, developed RECAT-EU, which defines required separations between aircraft based not only on weight but also on aerodynamic characteristics such as wingspan and approach speed. The RECAT-EU distance-based separation requirements are shown in Table 5.1. In conditions with strong headwinds, distance-based separation can become unnecessarily conservative. Time-based separation offers a more optimal solution, though it is more challenging to visualize on radar screens. This approach has only been implemented in recent years, with the time-based RECAT-EU requirements shown in Table 5.2.

By tailoring separation to the actual wake risk posed and the following aircraft's vulnerability, RECAT-EU enables airports to safely increase capacity, reduce delays, and lower fuel burn and emissions.

From the perspective of the departure manager, RECAT-EU enables tighter departure spacing, facilitating more efficient scheduling, and quicker recovery after delays or runway reconfigurations. Runway throughput can increase by up to 5%, particularly at airports with a traffic mix dominated by A320 and B737 aircraft. Importantly, this is achieved without compromising safety: the RECAT-EU separation logic is backed by over 100,000 wake vortex measurements and a rigorous safety case reviewed by European Union Aviation Safety Agency (EASA) and industry stakeholders.

**Restricted Visibility (BZO)**

Restricted Visibility Conditions (beperkt zicht omstandigheden (BZO)) significantly impact airport operations, particularly affecting departure procedures. These conditions are declared when visibility drops below 1,500 meters and the cloud base is lower than 300 feet. During such conditions, pilots and air traffic controllers cannot rely on visual observations for aircraft separation on and around the airport terrain, necessitating stricter operational procedures.

BZO conditions are classified into four phases (A through D), with phase D representing the poorest visibility conditions. When visibility falls below 550 meters, the TMA requires a minimum separation of 5 nautical miles between aircraft, increased from the standard 3 nautical miles. This increased separation ensures safe operations despite the limited visibility.

Several additional safety measures are implemented during BZO conditions at Schiphol:

- Runways are equipped with additional safety measures, including stop bars at entry and exit points

**Table 5.1:** RECAT-EU distance-based separation requirements. Empty fields represent minimum radar separation (which in TMA is 3 NM)

| RECAT-EU scheme<br>Leader / Follower | | Super Heavy<br>A | Upper Heavy<br>B | Lower Heavy<br>C | Upper Medium<br>D | Lower Medium<br>E | Light<br>F |
|---|---|---|---|---|---|---|---|
| Super Heavy | A | 3 NM | 4 NM | 5 NM | 5 NM | 6 NM | 8 NM |
| Upper Heavy | B | | 3 NM | 4 NM | 4 NM | 5 NM | 7 NM |
| Lower Heavy | C | | | 3 NM | 3 NM | 4 NM | 6 NM |
| Upper Medium | D | | | | | | 5 NM |
| Lower Medium | E | | | | | | 4 NM |
| Light | F | | | | | | 3 NM |

**Table 5.2:** RECAT-EU time-based separation requirements. Empty fields concern pairs for which no time separation minima apply, so the distance minima still hold.

| RECAT-EU scheme<br>Leader / Follower | | Super Heavy<br>A | Upper Heavy<br>B | Lower Heavy<br>C | Upper Medium<br>D | Lower Medium<br>E | Light<br>F |
|---|---|---|---|---|---|---|---|
| Super Heavy | A | | 100 s | 120 s | 140 s | 160 s | 180 s |
| Upper Heavy | B | | | | 100 s | 120 s | 140 s |
| Lower Heavy | C | | | | 80 s | 100 s | 120 s |
| Upper Medium | D | | | | | | 120 s |
| Lower Medium | E | | | | | | 100 s |
| Light | F | | | | | | 80 s |

- Aircraft separation distances are increased
- Runway usage and capacity are adjusted based on current visibility values
- Ground activities in the maneuvering area are limited or suspended
- Pilots receive active guidance during taxiing operations

These measures, while essential for safety, significantly impact airport capacity and departure scheduling. The airport has to cancel flights in order to comply with these increased separation requirements and reduced operational capacity.

**Winter Operations**

Winter conditions significantly impact airport departure capacity through multiple mechanisms that affect both infrastructure and aircraft operations (Huet et al., 2016). The combination of these factors often makes de-icing capacity (not runway throughput) the critical constraint during winter operations.

When precipitation occurs at freezing temperatures, aircraft must undergo de-icing procedures before departure to remove accumulated ice and snow that could affect aerodynamic performance and flight safety. The required de-icing method varies with precipitation type and temperature, with more intensive procedures demanding greater time and resources. This creates a bottleneck in the departure process that directly limits the maximum departure rate.

De-icing operations introduce substantial complexity into departure management. Aircraft must first taxi to dedicated de-icing facilities, undergo the de-icing procedure, and then proceed to the departure runway, significantly extending taxi times and potentially causing taxiway congestion. Additionally, de-icing fluids have a limited effectiveness window (holdover time), creating a critical timing constraint; aircraft must take off within this window or undergo repeated de-icing. This time pressure can even necessitate accepting less favorable runway configurations, such as departures with tailwind components, to maintain safe operations within holdover time limitations. Although the latter example is for extreme cases, rarely seen at the latitudes of the Netherlands, it illustrates the extent to which de-icing procedures can constrain departure capacity.

The availability and condition of movement areas (runways, taxiways, and aircraft stands) further influence capacity. Airport operations continuously assess and maintain these surfaces according to coordinated snow and ice removal plans. However, the scope of infrastructure that can be effectively maintained depends on storm severity and available resources.

During winter operations, stakeholders must maintain a delicate balance between arrivals and departures to prevent aircraft congestion at stands, which could further constrain the system. The critical capacity constraint is identified through coordination between airport operations, air traffic control, de-icing organizations, and airlines, with all parties adjusting their processes accordingly (IATA, 2020). Air traffic control typically adapts their sequencing procedures to align with these constraints, emphasizing the importance of integrated departure management during winter conditions.

**Wind**

Among weather phenomena affecting airport operations, wind conditions are the most frequently encountered factor in the Netherlands. Significantly impacting runway selection and departure procedures. Wind effects can be categorized into three main components: headwind/tailwind, crosswind, and wind shear, each presenting unique operational challenges.

Aircraft preferentially take off and land into the wind (headwind) for optimal performance and safety. Headwind reduces the ground speed required for takeoff, consequently reducing the runway length needed. Conversely, tailwind operations require higher ground speeds to achieve the necessary lift, demanding substantially longer runway lengths (Dijkstra et al., 2020). This extended runway requirement must also account for potential aborted takeoff scenarios, ensuring sufficient stopping distance remains available.

In an outbound peak, maintaining runway throughput is critical. However, strong headwinds can significantly reduce aircraft ground speeds, which in turn impacts the spacing between departing aircraft and can diminish departure capacity (van Baren, 2016). In essence, when a departing aircraft faces a significant headwind, it takes longer to cover the same distance over the ground. This means that fixed distance-based separation minima (whether for radar or wake turbulence separation) take more time to be achieved, forcing longer intervals between takeoffs and reducing the number of departures per hour.

**Example** Consider a typical outbound rush where aircraft depart every 90 seconds, achieving a rate of 40 departures per hour. This precisely orchestrated sequence demonstrates how tightly air traffic controllers must manage departures. Even a slight increase in departure time of just 2.3 seconds per aircraft, due to headwind for example, results in one fewer departure per hour. In severe headwind conditions, where each departure easily takes 15 seconds longer than nominal, the departure rate can drop below 35 aircraft per hour. This reduction of 5 aircraft per hour significantly impacts airport capacity, potentially leading to delays and flight cancellations.

Crosswind conditions (wind perpendicular to the runway) present additional operational challenges. Each aircraft type has specific crosswind limitations that must be strictly adhered to for safety considerations. At Schiphol, the configuration of runways can result in scenarios where strong crosswind components render multiple runways inoperable, potentially constraining operations to a single viable runway. While the final decision to operate under such conditions remains at the pilot's discretion (Air Traffic Control the Netherlands (LVNL), 2025), these wind constraints significantly influence runway configuration decisions and, consequently, departure capacity. It should be noted that crosswind has an even more pronounced impact on arrivals than on departures, as landing aircraft are more susceptible to wind-induced control challenges during the critical final approach and touchdown phase.

The ICAO general operational recommendation for runway usage is 15 knots crosswind, including gusts, especially in relation to preferential runway selection Schiphol has a 25 knot crosswind limit, including gusts. In low visibility or low braking conditions, this reduces significantly and is catastrophic for the capacity of the airport.

Wind shear, characterized by sudden localized changes in wind direction and speed, poses particular challenges for aircraft. While modern aircraft are equipped with Enhanced Ground Proximity Warning systems to detect such conditions, confirmed wind shear reports often lead to temporary operational restrictions or runway changes. Air traffic controllers actively relay pilot reports of wind shear to subsequent pilots, potentially requiring adjustment of departure sequences or even temporary suspension of operations.

These wind-related factors directly impact departure capacity through multiple mechanisms: they determine which runways are available for use, may require increased separation between departures, and can necessitate runway configuration changes that disrupt the planned departure sequence. During strong wind conditions, the DMAN must incorporate these operational limitations while maintaining safe and efficient departure operations.

**Aircraft Performance Differences**

Aircraft performance characteristics significantly influence departure sequencing decisions and directly impact runway throughput (Bikir et al., 2024a). These characteristics vary considerably between aircraft types and are affected by factors such as takeoff weight, engine type, and aircraft configuration. The key performance parameters that affect departure sequencing include initial climb rate, acceleration during takeoff roll, and cruise speed capability (Anderson Jr., 1999). Initial climb rate, typically measured in feet per minute, directly affects how quickly aircraft can achieve required separation. Takeoff roll acceleration determines runway occupancy time, with heavier aircraft generally requiring longer distances. Cruise speed differences between consecutive departures impact the required initial spacing to prevent compression during climb. Therefore it is necessary to accurately model these performance characteristics (Sun et al., 2020).

When sequencing departures, consideration of these performance differences becomes crucial for maintaining both safety and efficiency. Slower aircraft followed by faster ones create particular challenges for air traffic control, as the speed differential can lead to compression of separation distances during the climb phase. This situation often requires controllers to implement additional spacing at takeoff or tactical interventions during the climb, both of which reduce operational efficiency.

**Example** Consider a scenario where a Fokker 50 turboprop is followed by a Boeing 737, both proceeding along the same SID. The significant performance difference between these aircraft creates a challenging situation. The Boeing 737, with its superior climb rate and higher airspeed, quickly catches up to the slower Fokker 50. This forces the air traffic controller to either: (1) impose a significant initial departure delay for the Boeing 737, (2) tactically vector the Fokker 50 off its planned SID, allowing the Boeing to overtake, or (3) instruct the Fokker 50 to maintain a lower altitude while the Boeing 737 climbs overhead. Each of these solutions compromises operational efficiency and increases controller workload. Reversing this sequence would largely eliminate these issues.

An effective DMAN should incorporate aircraft performance data to optimize departure sequences. When possible, it should arrange departures in order of decreasing performance (Bikir et al., 2024b). This approach minimizes the need for tactical interventions and helps maintain separation during the initial climb phase. However, this optimization must be balanced against other constraints such as wake turbulence requirements and slot times.

## 1.2. Departure Manager Constraints

The DMAN aims to reduce taxi-out queues, fuel burn, and delays, while improving predictability. Morning outbound peaks stress the airport's capacity and make effective DMAN operation critical. Each airport's unique characteristics (demand level, runway layout, operating modes, airspace structure, noise abatement procedures, ground infrastructure limitations) shape how the DMAN is configured (Karapetyan et al., 2017).

The DMAN must take into account these (local) operational constraints when determining the departure sequence, apart from the separation constraints for aircraft themselves. These constraints are crucial for ensuring safe and efficient operations at the airport. This section explores in detail the key operational constraints faced by major European hub airports (including Amsterdam Schiphol, Frankfurt, London Heathrow, and Paris Charles de Gaulle). The most important constraints are discussed below.

#### Slot Allocation and Airspace Flow Constraints

One of the most foremost constraints for European departures is compliance with air traffic flow management (ATFM) slots allocated by Eurocontrol's network manager. These take the form of calculated take-off time (CTOT) assigned to certain flights to regulate traffic into congested en-route or arrival sectors. A CTOT defines a strict slot tolerance window (CTOT -5 to +10 minutes) within which the aircraft must depart (Karapetyan et al., 2017). Missing this window forces the airline to request a new slot, often incurring additional delay. Consequently, DMAN systems treat CTOT adherence as a hard scheduling constraint. Any sequence plan that would violate a CTOT is considered unacceptable (Karapetyan et al., 2017). In practice, DMAN gives the highest priority to flights under flow control to ensure they depart within their slot (SESAR-JU, 2011). This has measurably improved slot compliance at collaborative decision making (CDM) enabled airports, making it far more likely flights depart on time (Huet et al., 2016).

In addition to ATFM departure slots, DMAN must consider local airspace constraints such as SID or departure fixes capacity. Each SID or airway has a limited throughput to maintain safe separation in the downstream en-route airspace. Major hubs often have specific limits on the number of aircraft that can enter a given sector or fix per time interval. For instance, at London Heathrow certain departure fixes are capped (e.g., no more than 5 aircraft through the CLN (Clacton) fix in any 10-minute period) (Marayat, 2008). Exceeding this rate (also referred to as 'bunching') would overwhelm en-route air traffic control (ATC), so a minimum departure interval (MDI) is imposed between flights in the same route.

DMAN algorithms incorporate these SID/route constraints as hard separation rules in the departure sequence (Marayat, 2008). In fact, there is a high likelihood that any given departure at a European hub will be subject to multiple overlapping MDI restrictions simultaneously. For example, a flight might simultaneously need extra spacing due to wake turbulence, plus a 2-minute gap behind another flight on the same SID, plus compliance with cross-border flow restrictions (Eun et al., 2019). Eun et al. (2019) found that over 90% of departures had to comply with two or more such constraints at Icheon International Airport. These compounding airspace constraints significantly increase the complexity of the departure scheduling problem, often requiring DMAN to find an optimal sequence that satisfies all route, flow, and slot rules whilst still using the runway efficiently (Karapetyan et al., 2017).

#### Turnaround Times and Ground Handling Constraints

Before an aircraft can join the departure queue, it must of course be ready. Hence the timing of the turnaround operations (further discussed in Section 2) feeds directly into DMAN planning. In the airport CDM framework, airlines or ground handlers provide a target off-block time (TOBT) for each flight, representing when they expect to have the aircraft ready for pushback. The accuracy and timeliness of TOBT updates are crucial. If TOBTs are updates late or turn out unrealistic, the carefully optimized DMAN sequence can be disrupted by last-minute unavailability of a flight, causing inefficiencies or gaps.

A recent analysis of the new DMAN at Schiphol confirmed that early and accurate TOBT updates significantly improve departure scheduling, whereas late updates propagate 'like a snowball' through the plan, undermining capacity and predictability (Snijders, 2024). 'One cannot schedule better than on time', when flights are not ready as planned, the whole sequence must be reworked, often resulting in cumulative delays for everyone

(Snijders, 2024).

To mitigate this, DMAN relies on the CDM process to continuously align its planning with the ground operations. If a TOBT is not updated by a certain threshold (e.g., 15 minutes beyond the last known off-block estimate), the system raises alerts prompting stakeholders to provide a new time (SESAR-JU, 2011). Ground handling delays, such as late fueling, last-second passenger issues, or slow boarding translate into TOBT shifts. DMAN must be flexible to re-sequence or delay flights whose TOBT slips, while minimizing impact on others. Some airports allow airlines to indicate priority flights (for instance, a flight with many connecting passengers or a long-haul that must depart on time), such user preferences can be input to the DMAN so that if re-sequencing is necessary, lower-priority flights absorb more delay (Frequentis, 2022).

Often at the busiest hubs, such as Schiphol Airport, there is limited slack in the departure sequence, and trade-offs between efficiency and fairness are sometimes made to ensure optimal runway usage (Buchli et al., 2019). Airlines understand that a small deviation from their planned schedule may be accepted or even necessary if it helps prevent a larger departure gap that would waste valuable runway capacity (IATA, 2020). This efficiency-driven behavior is facilitated through airport collaborative decision making (A-CDM), which operates under principles like 'best planned, best served' and depends heavily on timely, accurate TOBT information from all stakeholders (IATA, 2020).

A collaborative mindset, supported by transparency in A-CDM information sharing, is therefore essential for the CDM to function optimally (Schäfer & Alexopoulos, 2024). If transparency or trust in the system is lacking, airlines may become reluctant to share accurate TOBT updates, fearing these will be used against them in future TSAT assignments or slot allocation negotiations (Jester et al., 2019). For example, if an aircraft is ready ahead of schedule but the updated TOBT leads to a less favorable TSAT, handlers might deliberately delay updating the TOBT to maintain queue position. This behavior introduces inaccuracies in the shared operational picture, leading to inefficiencies and reduced predictability (Jester et al., 2019).

In this way, a vicious cycle can emerge: low trust leads to withheld or inaccurate information, which in turn degrades the performance of the system, further eroding trust. Building mutual confidence, refining the CDM interface, and increasing user awareness are thus crucial steps toward maximizing the performance of Schiphol's DMAN and preserving fairness while still ensuring efficient use of scarce departure capacity (Jester et al., 2019; Schäfer & Alexopoulos, 2024).

Another ground handling constraint comes from push-back conflicts and apron management. Large hub airports have dense gate areas where simultaneous push-backs of adjacent stands can conflict. Modern DMAN implementations take these airport-specific rules into account, effectively de-conflicting push-back schedules as part of the pre-departure sequence optimization (Frequentis, 2022). For example, if two aircraft at neighboring gates are planned to depart close in time, the DMAN may slightly stagger their TSATs to ensure ground controllers aren't faced with an impossible push-back scenario. By embedding such apron resource constraints in the planning, DMAN eases the workload on ground control and prevents downstream taxi delays.

**Taxi-Out Time Estimation and Predictability**

Accurately predicting taxi times (from off-block to takeoff) is crucial for effective departure sequencing. DMAN uses these predictions to determine the TSAT. If taxi durations are misestimated, the entire departure sequence timeline can drift. An overly optimistic taxi time means aircraft might not reach the runway by their planned takeoff slot (creating gaps), while overly padded times lead to unnecessary queuing.

European CDM airports employ standard taxi time tables that vary by stand, runway, visibility and icing conditions. Research shows that using constant taxi-out times for all flights leads to prediction errors that propagate through the departure schedule (SESAR-JU, 2011). Different aircraft types have varying taxi speeds, particularly when cornering. Weather conditions and airline procedures also significantly impact taxi times. Therefore, high-quality taxi time data - accounting for current runway configuration, queuing, and weather effects - is essential for DMAN to reliably compute takeoff times (Herrema et al., 2018).

During winters, one major factor affecting taxi times is the de-icing process. Aircraft must be de-iced before departure, which can add significant time to the taxi-out duration. The DMAN must account for this additional time when scheduling departures, especially during peak winter operations when multiple aircraft may require de-icing simultaneously. During winter operations, de-icing becomes the bottleneck of the departing flow (Huet et al., 2016).

Many airports set a queue limit (often 3-5 aircraft maximum) at the runway holding point for both efficiency and environmental reasons (SESAR-JU, 2011). By keeping the takeoff queue short, taxi-out times become more predictable (less variance due to waiting) and fuel burn and emissions are reduced. Importantly, once an aircraft does push back and taxi out, DMAN will not impose any further planned delay on it (Marayat, 2008). Any needed holding is done at the gate (engines off) rather than out on the taxiway with engines running.

Static buffers are used in schedules to guard against taxi time uncertainty. As Karapetyan et al., 2017 notes, 'startup times and taxi times are hardly ever precisely predictable', so some slack is needed in the schedule to absorb these uncertainties. The art for a DMAN is to include just enough buffer to handle typical variability without sacrificing too much throughput.

**DMAN Strategies**

Studies of peak-hour operations show that airports can be capacity-constrained in fundamentally different ways. Frankfurt's morning hub outbound peak is a clear example of a concentrated wave of departures. These waves lead to periods of temporarily overloaded demand. This makes the airport highly sensitive to short-term reductions in departure capacity. For instance, Elver et al., 2024 showed that even small constraints introduced during Frankfurt's morning bank can trigger extensive queuing and delays extending into the afternoon.

Heathrow, by contrast, rarely sees demand above scheduled capacity because it is heavily slot-controlled. It operates at a state of permanent saturation. This strict slot control ensures scheduled demand matches maximum declared capacity. This leaves no flexibility in the departure sequence, any disruption immediately results in delay due to the lack of unscheduled slack (Odoni, 2020). The use of a single departure runway further amplifies sensitivity to wake turbulence and routing constraints, which cannot be absorbed across multiple departure streams as at Frankfurt or Schiphol. Accordingly, DMAN strategies differ: Heathrow emphasizes precise, minute-level optimization of a single queue, while Frankfurt and Schiphol focuses on dynamic runway assignment and load balancing (SESAR-JU, 2011).

Compounding constraints are particularly evident when something like bad weather occurs during a peak. In winter 2017 at Schiphol, for example, heavy snowfall reduced runway availability and required de-icing, causing large departure delays[1]. Huet et al., 2016 noted that remote de-icing pad capacity can sharply limit departure rates, and integrating de-icing planning with DMAN is essential to avoid gridlock. By metering departures before they even leave the gate, the airport can prevent long queues of idling aircraft and instead sequence them through de-icing and onto the runway as capacity allows. This is essentially the DMAN solving a multistage scheduling problem: each flight has to go through stages (push → taxi → de-ice → taxi → takeoff) with limited resources at each stage, under various constraints. The state of the art is moving toward holistic algorithms that handle such multi-stage sequencing. One NASA study even modeled minimum departure interval (MDI) in a holistic way and found many flights were under multiple simultaneous restrictions, reinforcing that peak departure scheduling is a complex multidimensional puzzle (Eun et al., 2019).

Despite the complexity, European hubs have shown improvements by using CDM and DMAN, especially during peaks. Predictability of off-block and takeoff times has increased. Trials at Paris CDG showed a 7.8% improvement in off-block predictability, with 85% of flights departing within a 5-minute window of their TOBT. Average taxi-out times dropped by 9%. Additionally, 81% of flights adhered to their ATFM slots, up from 75.6% in the previous year (SESAR-JU, 2012). This led to an average reduction of 14.6 kg of fuel per flight.

Thus, while constraints can compound negatively, good CDM practices and DMAN optimization can compound positively, leading to a smooth sailing operation.

## 1.3. Key Performance Indicators

To systematically assess the effectiveness of different departure configurations and enhancements to the DMAN, this study builds upon the key performance indicators (KPIs) proposed by Snijders, 2024, who evaluated the impact of TOBT uncertainty on departure capacity at Schiphol Airport. The selected KPIs are designed to provide measurable and interpretable metrics across multiple configurations, allowing for a clear comparison of performance and trade-offs.

The KPIs serve as dependent variables in this research, and they are chosen to reflect operational goals including throughput maximization, delay minimization, planning accuracy, and temporal stability. Each KPI is defined below with a rationale.

[1] https://nltimes.nl/2017/12/11/code-red-snow-alert-leaves-airline-passengers-stranded

**Runway Throughput (Flights per Hour)**
This KPI measures how closely the departure system approaches the theoretical runway capacity. Snijders modeled multiple configurations, showing that earlier TOBT updates can increase actual throughput, especially under high-demand scenarios. In this research, throughput is evaluated by comparing the number of successful departures per 10-minute slot to the declared capacity.

Higher throughput indicates better capacity utilization and reduced lost slots due to expired or late TOBT updates.

**Departure Delay**
Delay is a central metric, but its interpretation can vary. This study focuses on:

- Slot Delay per Aircraft: The difference between the assigned runway slot and the preferred or ideal slot.
- Cumulative Daily Delay: Summed slot shifts over all flights.
- Slot Adherence Rate: Percentage of flights departing in their initially preferred (or best-fit) slot.

The delay is primarily tracked per aircraft, as this aligns with the DMAN's role in sequencing rather than passenger flow management. However, extensions to delay per passenger may be considered for wide-body prioritization analysis. Especially since these aircraft often carry connecting passengers.

Lower delay reflects better predictability and operational efficiency.

**Sequencing Accuracy**
To capture how well the DMAN anticipates future states, a sequencing accuracy metric is introduced: Sequencer Consistency, measured as the percentage match between the projected sequence (e.g., from 15 minutes prior) and the final executed sequence.

This KPI reflects whether the system's forward planning aligns with eventual execution, particularly under conditions of turnaround uncertainty. Higher accuracy supports the reliability of the DMAN as a predictive tool.

**Temporal Stability**
Airlines report that frequent changes to the TSAT erode trust in the planning system. This study evaluates temporal stability through Slot shifts: Number of substantial (>5 minute) changes in TSAT per flight throughout the planning process.

Evaluating both total and average shifts per flight, and defining thresholds for 'instability'. A stable plan reduces controller workload and improves pilot and airline confidence. Increased stability is indicative of a robust, predictable planning environment.

**Objective of KPI Use**
This research does not merely aim to improve each individual KPI, but to analyze their interaction. For example, a configuration that increases throughput might induce more instability, or a reduction in delay might come at the cost of decreased accuracy. By systematically tracking each KPI, trade-offs can be explicitly evaluated, guiding future improvements to the departure manager.

# 2. Aircraft Turnaround Process in A-CDM Airports

The aircraft turnaround process encompasses all activities from the moment an aircraft arrives at its parking stand to the moment it departs again for the next flight. In other words, a turnaround spans from actual in-block time (AIBT), when the chocks are placed at the wheels, until actual off-block time (AOBT), when the aircraft pushes back to taxi out (Snijders, 2024). During this period, the aircraft is serviced, prepared, and readied for departure on its next leg. Major international hubs like Amsterdam Airport Schiphol have complex turnarounds involving multiple simultaneous services, such as refueling, cleaning, catering, and baggage handling. The turnaround process is a critical component of airport operations and is essential for maintaining punctuality and efficiency in air travel. Strict timing and efficient coordination of these tasks is critical to avoid delays.

Airport collaborative decision making (A-CDM) is an operational concept adopted at many major airports, including Schiphol, to enhance efficiency and predictability during turnarounds. A-CDM ensures that the right partners get the right information at the right time, enabling joint decision making based on accurate data (Eurocontrol, 2009). At large hubs such as Schiphol, the complexity is significantly increased by the presence of multiple service providers;often several companies performing the same function for different airlines;all requiring precise coordination during the turnaround process. It focuses on improving the turnaround and pre-departure phase through information sharing, common situational awareness, and agreed procedures among all stakeholders. (Technology, 2024).

## 2.1. Turnaround Process Steps from Arrival to Takeoff

Turnarounds at a CDM airport involve a series of well-defined steps, from the aircraft's arrival through servicing and finally departure. Below is a detailed sequence of key turnaround activities, many of these occur in parallel or overlap, rather than strictly sequential (Snijders, 2024).

1. **Aircraft arrival and in-block**
   After landing, the aircraft taxis to its assigned gate or stand. The AIBT is recorded when the aircraft comes to a stop and wheel chocks are in place. Once 'on-blocks', engines are shut down and ground power is connected to run onboard systems without engines/APU (Snijders, 2024). Airport staff position the passenger boarding bridge (or stairs for remote stands) while the crew and ground personnel perform initial safety checks, marking the start of the turnaround clock.
2. **Deboarding of passengers and unloading baggage**
   Once the doors are opened, passengers disembark under ground staff supervision while ramp crews simultaneously begin unloading baggage and cargo using belt loaders. If required, crew changes occur during this phase. Flight administration tasks commence, including flight log updates, turnaround reports, and preparation of the next flight plan (Snijders, 2024). These administrative tasks continue throughout the turnaround process.
3. **Aircraft servicing and cabin preparation**
   As soon as passengers are disembarked and the cabin is free, multiple servicing activities occur simultaneously (Snijders, 2024):
   - **Cleaning**
     The cabin crew performs cleaning tasks including tidying seat areas, removing trash, and servicing lavatories. For long-haul flights, more extensive cleaning may be required. Potable water tanks are refilled as necessary.
   - **Catering**
     Catering trucks resupply food and beverages, unloading used galley carts and loading fresh meals according to airline requirements. For short-haul flights, catering might occur every other flight to optimize turnaround time.
   - **Refueling**
     Refueling occurs via fuel truck or hydrant system, typically starting early in the turnaround as it is often on the critical path. Safety regulations restrict fueling during passenger boarding unless specific precautions are in place, such as an attached boarding bridge and crew present for potential evacuation (Snijders, 2024).
4. **Maintenance and checks**
   Ground engineers perform transit checks and address any minor technical issues reported from the inbound flight. This includes exterior walk-around inspections and routine maintenance like tire pressure checks or oil top-ups. The duration of these activities varies based on aircraft size, flight type, and staff availability.
5. **Boarding preparation**
   As servicing concludes, the new flight crew (if applicable) arrives to perform pre-flight checks and cockpit preparation. This includes setting up avionics, inputting flight plans, and obtaining weather and ATC clearance. The loadmaster provides final load sheets and fuel figures. Passenger boarding typically begins 20-40 minutes before scheduled departure, depending on aircraft size (Snijders, 2024). Simultaneously, ramp workers load outbound baggage and cargo, often timed to ensure all connecting bags are included.
6. **Closing and final checks**
   Once boarding completes, crews perform final checks including door closure verification and passenger count confirmation. Ground support equipment is removed, and the aircraft switches to APU power. In winter conditions, de-icing may occur at the gate or a remote pad, coordinated by specialized teams. The flight crew then signals readiness for startup and pushback to ATC.
7. **Pushback and departure**
   After receiving ATC clearance, the aircraft pushes back from the stand, marking the AOBT. Engine

start-up occurs during or after pushback. At CDM airports, pre-departure sequencing optimizes taxi movements to meet the target take-off time (TTOT). The process concludes when the aircraft takes off, recorded as the ATOT.

Each step in the turnaround process must be completed efficiently and in coordination with others, with many activities occurring in parallel to minimize ground time. Strict safety and regulatory procedures constrain the process, for instance, during refueling. Airlines often incorporate schedule buffers for ground time, but minimizing delays in any step is crucial for on-time departure. The critical path (sequence of tasks determining minimum turnaround time) typically includes the longest tasks such as passenger handling and fueling. Figure 5.2 illustrates common turnaround process 'blocks' and highlights a typical critical path.

Turnaround processes can vary by aircraft and flight type. Long-haul flights generally require longer turnarounds with more extensive services like catering and cleaning, while short-haul flights may streamline or skip certain services for quicker turnarounds. However, the fundamental steps outlined above remain common across most operations.

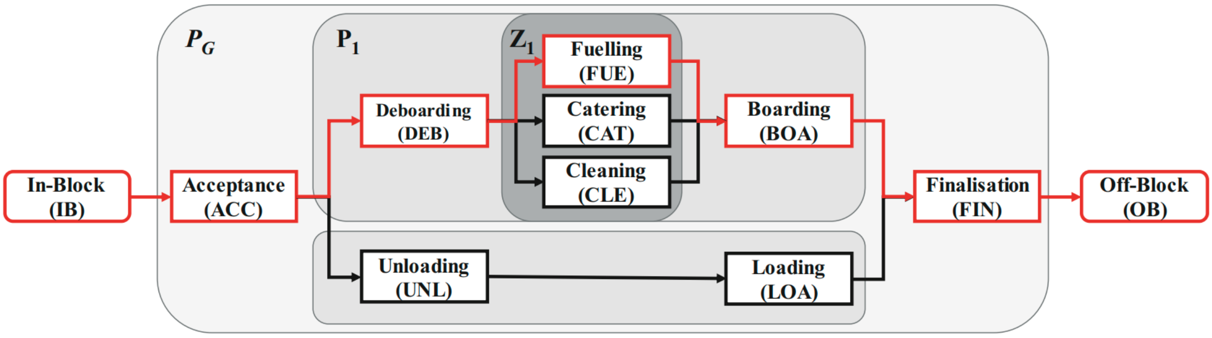



**Figure 5.2:** Overview of aircraft turnaround processes showing typical sequential and parallel service activities. The critical path (highlighted in red) represents the sequence of time-critical activities that determine the minimum possible turnaround duration. Many parallel processes are omitted for clarity. Adapted from Asadi et al. (2020)

## 2.2. Stakeholders Involved and Their Roles

The turnaround involves multiple stakeholders, each with specific responsibilities. At a CDM airport, four primary stakeholder groups collaborate closely: (1) the airline, (2) the ground handling agent, (3) air traffic control, and (4) the airport operator. In addition, there are several indirect stakeholders that provide supporting facilities (fuelling companies catering providers, de-icing contractors, etc.). Effective coordination and communication are essential among these parties. Below is a summary of the key stakeholders and their roles in the turnaround process:

### Airline

The airline bears ultimate responsibility for flight schedules and operational decisions during turnarounds. They provide and coordinate flight and cabin crew, establish standard operating procedures for turnarounds, and appoint and oversee the Main Ground Handling Agent (MGHA). Through their operations center, airlines coordinate with ground staff to determine service requirements including fuel loads, catering needs, and any special handling requirements. They also manage flight planning.

At CDM airports like Schiphol, airlines must agree on target off-block time management through their ground handlers. They ensure crews are aware of TSAT, typically displayed on electronic gate displays, or communicated manually when displays are unavailable (Yiannis Alexopoulos et al., 2019). When operational issues arise, airlines make critical decisions regarding delays or flight cancellations. As the stakeholder accountable for on-time performance, airlines have an interest in efficient coordination among all parties involved in the turnaround process.

### Ground Handling Agent

The ground handling agent (GHA) executes most physical turnaround activities on behalf of airlines. As the Main Ground Handling Agent (MGHA) at CDM airports, they coordinate and perform essential services including passenger handling, baggage and cargo operations, ramp services (aircraft towing, marshalling), cabin cleaning, and catering loading (Alexopoulos, 2024). They also oversee subcontracted services like fueling and maintenance.

A critical responsibility at CDM airports is maintaining accurate turnaround progress information. The MGHA must ensure all ground handling processes are completed by the planned off-block time and continuously update the TOBT to reflect realistic completion estimates. This requires constant assessment of all service components and updating TOBT with $\pm 5$ minute accuracy when delays occur (Yiannis Alexopoulos et al., 2019). By TOBT, the aircraft must be fully ready with all ground equipment removed, doors closed, boarding bridge retracted, and pushback equipment available.

The ground handler serves as the key operational coordinator between flight crews, service providers, and other stakeholders. They conduct departure briefings with pilots, manage ramp resources (staff and equipment), and promptly communicate status updates through the CDM system (Yiannis Alexopoulos et al., 2019). This central coordination role is essential for maintaining schedule integrity and ensuring efficient information flow between all parties involved in the turnaround process.

**Air Traffic Control (ANSP)**

The air navigation service provider (ANSP), which at Schiphol includes tower and ground control provided by Luchtverkeersleiding Nederland (Air Traffic Control the Netherlands) (LVNL), manages aircraft movements both in the air and on the ground. ATC's primary role in turnarounds focuses on the outbound phase, where they receive TOBT from ground handlers and plan startup and pushback sequences accordingly (Snijders, 2024).

At CDM airports, ATC issues a target start-up approval time (TSAT) for each departing flight, indicating when ATC expects to grant start-up clearance based on current traffic conditions. ATC employs pre-departure sequencing systems that consider TOBT, runway capacity, taxi times, and any applicable CTOT to generate optimal departure sequences (Snijders, 2024).

ATC relies heavily on the shared CDM system for coordination with airport operators and ground handlers, particularly requiring timely TOBT updates to accurately assess flight readiness (Alexopoulos, 2024). Strict adherence to TSAT windows (typically TSAT $\pm$ 5 minutes) is enforced - ATC will only grant pushback clearance when aircraft are within their TSAT window and indicate readiness. If flights miss their TSAT or request early starts, ATC generally denies clearance and refers them back to ground handlers for a new TOBT assignment, effectively losing their position in the departure queue (Yiannis Alexopoulos et al., 2019).

Once pushback is approved, ATC ensures flights taxi and depart within their allocated slot windows. When post-pushback delays occur (such as de-icing or queuing at the runway) that risk missing en-route slots, ATC coordinates with the Eurocontrol Network Manager to request slot extensions or new slots (Yiannis Alexopoulos et al., 2019). ATC also provides runway allocation through the CDM system, which feeds into TSAT calculations, and communicates capacity-reducing events to all stakeholders for plan adjustments.

The Network Manager Operations Centre (NMOC) in Brussels oversees European air traffic flow management, receiving CDM updates including Target Take-Off Times and delay information via electronic messages. When en-route or destination constraints exist, NMOC issues CTOTs that must be met by airlines and ATC. At Schiphol, the airport operator maintains real-time CDM information exchange with NMOC, ensuring network awareness of current planning and helping minimize wasted slots while improving overall European network flow (Yiannis Alexopoulos et al., 2019).

**Airport Operator**

The airport operator manages overall airport resources and infrastructure during turnarounds. Their core responsibilities include:

Stand and Gate Management: Schiphol's Airport Planning & Control (APC) department allocates and maintains stable parking stands/gates based on incoming flight data (Estimated In-Block Times) and TOBT/TSAT information. They actively adjust allocations when significant delays occur to preserve connections and prevent conflicts.

A-CDM Information Platform: The airport maintains the critical information-sharing infrastructure underpinning the entire A-CDM process. At Schiphol, this includes the Airport Operations Database (CISS system) and the Schiphol CDM Portal, providing all stakeholders with real-time flight updates. They ensure system reliability and distribute critical CDM alerts for flights requiring immediate attention (Yiannis Alexopoulos et al., 2019).

Operational Coordination: The CDM coordination cell communicates infrastructure disruptions or capacity changes across all stakeholders. During winter operations, they facilitate de-icing coordination by optimizing gate assignments and de-icing bay usage. Schiphol maintains a dedicated CDM contact desk for real-time stakeholder support and to ensure adherence to established CDM procedures (Yiannis Alexopoulos et al., 2019).

This multi-faceted role positions the airport operator as the critical link connecting turnaround operations with both arrival and departure processes while providing the essential infrastructure and coordination mechanisms that enable effective collaborative decision making.

**Service Providers and Supporting Stakeholders**
While the airline, ground handler, ATC, and airport operator form the core stakeholders in turnaround operations, several additional entities provide specialized services critical to successful turnarounds (Yiannis Alexopoulos et al., 2019):

- **Fueling Services**
  Third-party fuel providers receive fueling requests from ground handlers or airlines and must complete refueling within the turnaround window. Though technically part of ground handling, fuel providers operate independently and their performance directly impacts TOBT adherence. Coordination requirements include communicating estimated arrival times, responding promptly to quantity changes, and reporting completion status. Modern CDM implementations increasingly incorporate fuel status tracking to improve visibility of this critical-path activity.
- **Catering Services**
  Catering companies coordinate with ground handlers to access aircraft once passengers have deplaned. They must align their delivery schedules with the turnaround timeline and rapidly respond to last-minute service adjustments. Any delays in catering truck arrival or extended loading times can delay passenger boarding, making timely communication of issues essential for TOBT management.
- **De-icing Contractors**
  During winter operations, de-icing becomes a significant turnaround factor. At Schiphol, a dedicated De-icing Coordinator plans and directs these operations, determining whether de-icing occurs at the gate or a remote pad (ADIP). The coordinator establishes and updates the estimated de-icing start time (EDIT) and de-icing wait time (DIWT), which feed directly into TSAT calculations. Once de-icing completes, the coordinator must promptly clear these parameters from the system to signal ATC that the aircraft is ready for departure.
- **Meteorological Services**
  Weather service providers supply critical information affecting turnaround planning and execution. Forecasts of adverse conditions (thunderstorms, low visibility, strong winds) enable proactive adjustments to ramp operations and influence ATC capacity calculations. At CDM airports, weather updates are typically shared through the common portal, ensuring all stakeholders maintain situational awareness when conditions deteriorate ('German A-CDM Harmonization Group', 2018).
- **Security and Customs**
  Though not directly involved in physical aircraft servicing, security and customs agencies significantly impact passenger flow. Security screenings, aircraft checks, and immigration processing must align with turnaround timelines to prevent boarding or deplaning delays. These agencies coordinate with airline ground staff to manage passenger volumes and adjust staffing levels during peak periods or disruptions.
- **Maintenance Providers**
  Line maintenance teams perform routine checks and address technical issues identified during inbound flights. They coordinate closely with flight crews and ground handlers regarding any findings that might affect departure readiness. When unscheduled maintenance becomes necessary, maintenance providers must promptly communicate revised completion estimates to enable accurate TOBT updates.

These supporting stakeholders typically interface with the turnaround process through the main ground handling agent, who integrates their activities into the overall coordination plan. However, at advanced CDM airports, many now have direct access to portions of the information platform, enabling more transparent and responsive service delivery.

## 2.3. A-CDM Integration with the Turnaround (TOBT, TSAT, TTOT)

Airport collaborative decision making (A-CDM) adds a structured framework of information sharing and timing milestones to the turnaround process. The goal is to improve predictability and optimize the departure flow.

Three of the most critical time-stamps in A-CDM are the target off-block time (TOBT), target start-up approval time (TSAT), and target take-off time (TTOT). Below is explained what each time exactly represents. how they are determined and updated, how they integrate with activities.

**Target off-block time (TOBT)**
TOBT is the time that the aircraft expects to be ready to push back from the gate. In practice, TOBT is a timestamp reflecting when all turnaround activities will be complete and the aircraft will be fully ready for start-up. It is better described as the 'expected end of groundhandling time' ('German A-CDM Harmonization Group', 2018) . Under A-CDM, the responsibility for setting and updating the TOBT lies with the airline/operator or their ground handling agent, since they have the best knowledge of the turnaround progress. The TOBT is initially set based on the scheduled turn time or the latest estimates (for example, after landing the handler might set TOBT = estimated turnaround duration + actual in-block time). As the turnaround progresses, if any delay occurs or if things go faster than expected, the TOBT should be adjusted accordingly.

At Schiphol, ground handlers are required to keep TOBT accurate to within ±5 minutes, continuously updating it if needed so that it always reflects a realistic ready time. This TOBT is a crucial piece of shared information, all stakeholders use it as the authoritative estimate for departure planning. In general, providing an accurate TOBT (and doing so early) helps minimize last-minute surprises and allows ATC to optimize runway usage. If an unexpected issue occurs, stakeholders are advised to delete the TOBT early, which will automatically cancel any TSAT and release the slot for others, then re-enter a new TOBT when the situation is resolved. This practice ensures that a flight that cannot depart does not hold up others.

In practice, groundhandling agents and airlines often update TOBT less frequently than optimal, creating a fundamental tension between individual and collective interests. At the micro level, delaying a TOBT update makes strategic sense for individual flights;operators maintain scheduling flexibility by postponing updates until absolutely necessary, avoiding premature declaration of delays that might still be mitigated. However, this rational individual behavior creates significant inefficiencies at the macro level. When ATC bases sequencing and resource allocation decisions on outdated information, the entire system experiences suboptimal runway utilization, unnecessary queuing, and degraded predictability. This represents a classic collective action problem: actions beneficial to individual stakeholders ultimately reduce system capacity and create congestion that harms all participants. While understandable from a competitive perspective, this dynamic necessitates procedural or incentive-based interventions to align individual behaviors with the broader goal of maximizing airport throughput and operational efficiency.

**Target start-up approval time (TSAT)**
TSAT is the time that ATC plans to give start-up clearance to the flight (effectively the time the flight can expect to push back). TSAT is calculated by the airport's departure manager or by tower controllers for smaller airports. The TSAT is determined by taking the latest TOBT and sequencing it among all other departures, considering various factors: the departure rate (runway capacity and departure interval spacing), any applicable CTOT (network slot) which sets the earliest or latest takeoff time, the expected taxi-out time for that flight (which can vary by gate and destination runway), and other flights' priorities. An important rule is TSAT will never be earlier than TOBT, i.e. an aircraft cannot be assigned a start time before it is ready. Essentially, TOBT + taxi time yields a tentative takeoff time, which then might be adjusted to meet the runway sequencing constraints, resulting in a TSAT (and corresponding TTOT) that fit the overall queue. ATC (or the automated system) continually recalculates TSATs as conditions change (e.g. if one flight updates its TOBT to a later time, other flights might move up in the sequence). At Schiphol, the TSAT for a flight is actually computed as early as 3 hours before TOBT, but it is only made visible 40 minutes before TOBT on the CDM display systems. This prevents crews and handlers from over-reacting to sequence fluctuations far in advance. Within 40 minutes to departure, the TSAT becomes firm and is displayed on stand monitors and the CDM portal. A-CDM procedures dictate that flight crews must abide by the TSAT window: they should be ready for start-up at TSAT (and can call ATC for start clearance at TSAT ±5 minutes) ('German A-CDM Harmonization Group', 2018). If a crew calls ready and it is within TSAT ±5 minutes, ATC will issue clearance when feasible (usually at TSAT itself). If they call earlier than TSAT-5, they will be told to wait. If they call later than TSAT+5 (i.e. they missed their slot), ATC will likely cancel their clearance and require a new TOBT/TSAT to be set. This effectively means missing the TSAT window leads to losing one's place in the departure queue (which can result in further delay). The TSAT is thus both a product of collaborative planning (taking into account for everyone's needs) and a coordination mechanism (it tells the crew and ground handler exactly when pushback is expected so they can be ready).

**Target take-off time (TTOT)**
TTOT represents the predicted takeoff time for a flight. It serves two critical functions: (1) network coordination: by communicating to Eurocontrol's Network Manager when flights will enter the airspace system, allowing comparison with any allocated slots (CTOT), and (2) performance measurement: tracking the accuracy between predicted and actual takeoff times helps assess A-CDM effectiveness. When TTOT falls outside a flight's CTOT tolerance window, either the TSAT is adjusted or a new slot is requested. A-CDM implementations have dramatically improved takeoff time predictability, with Heathrow reducing average prediction error from 7 minutes to just 30 seconds (Atkin et al., 2009, 2013), enabling better downstream planning and resource utilization throughout the European airspace network.

## 2.4. A-CDM milestones

Eurocontrol's A-CDM approach integrates all these times into a larger 'milestone timeline'. This standardized implementation defines 16 milestones from the early planning stage through landing of the next leg (Alexopoulos, 2024). The framework ensures consistent operational procedures and information sharing across European airports, enabling seamless coordination between airports and improved network-wide flow management. By standardizing these milestones, flights can maintain predictable timing and data exchange protocols regardless of which A-CDM airport they operate from or to. Relevant milestones around the turnaround include:

- Milestone 7: In-Block (aircraft arrived at gate).
- Milestone 8: Ground Handling Starts - recorded when ground handling activities commence.
- Milestone 9: TOBT updated/confirmed - this is essentially a continuous milestone, reflecting that the airline/handler provides an accurate TOBT (it can be updated multiple times; milestone 9 is considered achieved when a TOBT is set and kept updated).
- Milestone 10: TSAT Issued - when ATC calculates and issues the TSAT for the flight. At Schiphol, the TSAT is generated in advance and becomes visible at TOBT-40 minutes.
- Milestone 11: Boarding Starts - when passengers actually start boarding the aircraft.
- Milestone 12: Aircraft Ready - usually defined as the moment the aircraft is fully ready to depart, i.e. all doors closed, bridge removed, and possibly the pilot indicating 'ready for start' to the ground crew or via radio. This corresponds closely to the TOBT (ideally, the aircraft ready milestone occurs at the TOBT)
- Milestone 13: Start-Up Request - when the pilot calls ATC for start-up/pushback clearance (or uses datalink to request it).
- Milestone (14 &) 15: (Start-Up Approved &) Off-block - when ATC gives the clearance and the aircraft actually pushes back, i.e. AOBT occurs.
- Milestone 16: Take-Off - the ATOT.

TOBT updates and TSAT recalculation are the core of CDM during turnarounds (Snijders, 2024). If TOBTs were not updated (or kept optimistic even when delays occurred), the aircraft would miss its TSAT and end up waiting, causing knock-on delays and under-utilization of the departure slot ('German A-CDM Harmonization Group', 2018). Schiphol is especially vulnerable to these snowball effects due to its hub-and-spoke operational model, which deliberately concentrates arrival and departure waves to minimize connection times for transfer passengers. This scheduling strategy creates intense peak periods where facility utilization approaches maximum capacity, making efficient turnaround management particularly critical for maintaining the integrity of the entire connection network. With nearly half of Schiphol's passengers being transfer travelers, even minor disruptions during these concentrated operational windows can rapidly cascade throughout the system.

## 2.5. Operational Bottlenecks, Delays, and Areas for Improvement

'In theory, practice and theory are the same. In practice, they are not.'

As with any complex operational process, the reality can deviate quite a bit from the ideal plan. The turnaround process is susceptible to bottlenecks and delays that can disrupt the entire departure sequence. Common issues relevant for the departure manager are discussed below.

Inaccurate or late TOBT updates, which lead to missed TSAT windows and cascading delays. Turnaround handling agents might be optimistic or hesitant to push back a TOBT even when delays are emerging, sometimes due to pressure to meet the schedule or fear of losing sequence priority.

Research at Schiphol has specifically studied the behavior of airlines and handlers in setting/updating TOBT and its effect on departure capacity (Snijders, 2024). If TOBTs are updated too late, the pre-departure sequence can be disrupted, causing a ripple of TSAT changes and lost capacity. A-CDM fundamentally relies on a culture of collaboration and discipline. If even one party frequently deviates, the system's benefits erode. The human factor remains an area for improvement technology provides the data, but people must use it properly.

Infrastructure and Capacity Limitations: Sometimes delays are simply due to airport infrastructure limits. For example, not enough gates, taxiway congestion, or ATC capacity. At Schiphol a capacity bottleneck occurs at times when only one or two runways are available for departures, causing a backlog no matter how efficient turnarounds are. Next to that, the groundcontrollers at Schiphol can quite easily be overloaded. During peak hours in the morning, it can be exceptionally busy on the frequency for the ground controllers to handle all the aircraft in their sector. As became clear during a visit to the ATC tower at Schiphol. This can lead to delays in the turnaround process, as aircraft may have to wait longer than expected for pushback clearance.

Studies show that simply sharing information (the core of CDM) already significantly reduces average delays (Technology, 2024), but the next steps involve using that information smarter. There are currently 2 large projects ongoing, (1) Deep Turnaround, which falls under the category of predictive analytics: machine learning models that predict if a particular flight's turnaround will take longer, so the handler can plan accordingly. And (2) an improved departure manager: one which plans aircraft in bins of 10-minutes to increase the stability of the TSAT. Both projects will be used in this research.

# 3. Deep Turnaround

Amsterdam Schiphol's Deep Turnaround system represents a groundbreaking implementation of computer vision and artificial intelligence in airport operations. Developed by Schiphol Aviation Solutions, this data-driven system transforms the traditionally opaque turnaround process into a transparent, predictable operation by continuously monitoring ground handling activities and predicting their completion times. The system addresses the fundamental challenge of accurately forecasting when aircraft will complete their turnaround activities and be ready for pushback, providing stakeholders with objective, real-time insights that significantly outperform traditional manual TOBT estimates. The information presented in this section is primarily derived from interviews with the developers and operators of the Deep Turnaround system, supplemented by official documentation (Aviation Solutions, 2023).

## 3.1. System Architecture and Technical Implementation

Deep Turnaround's architecture is built around a sophisticated dual-camera monitoring system integrated with cloud-based artificial intelligence processing. Two ultra-wide cameras are strategically positioned at each aircraft stand, approximately 20 meters high, providing comprehensive coverage of all ground handling activities. These cameras capture snapshot images every 5 seconds, which are immediately transmitted to a cloud-based platform for real-time processing. The dashboard interface, shown in Figure 5.3, displays the live camera feeds alongside detected activities and predicted timelines, enabling stakeholders to monitor turnaround progress effectively.

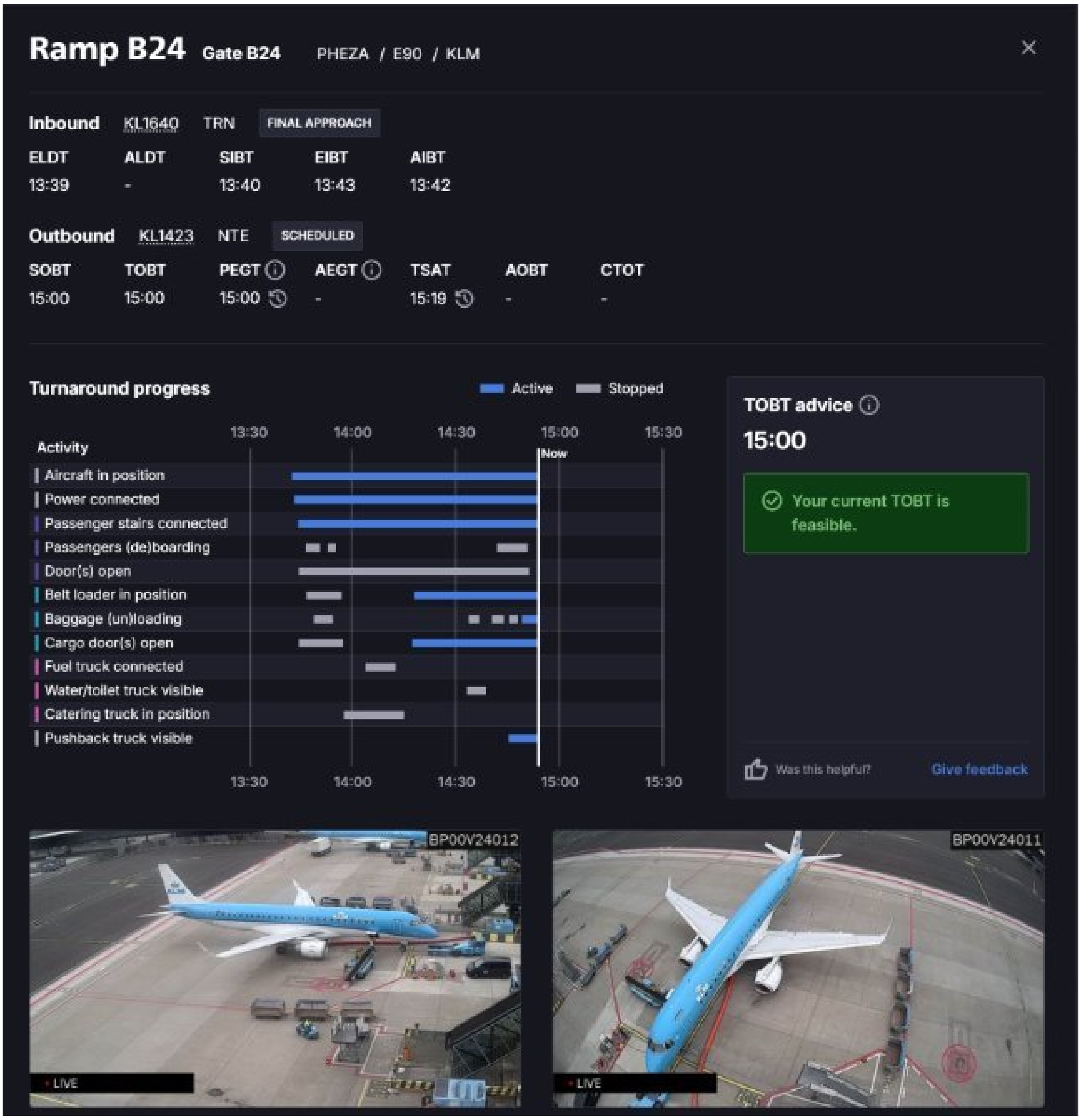



**Figure 5.3:** Deep Turnaround operational dashboard showing the dual camera views, detected ground handling activities, PEGT prediction timeline, and flight-specific information. The interface enables stakeholders to monitor turnaround progress in real-time and anticipate completion times.

The system's AI capabilities are implemented through two distinct deep-learning models operating in sequence:

**Real-Time Event Detection Model** The first AI model processes the continuous stream of images using advanced computer vision techniques to automatically recognize over 70 distinct turnaround events across approximately 30 subprocesses. These include passenger deboarding and boarding, refueling operations, catering services, cabin cleaning, baggage and cargo loading, aircraft cleaning, pushback truck positioning, and crew changes. This model effectively 'opens the black box' of the turnaround process by producing an accurate, real-time timeline of ongoing activities at each stand. The continuous monitoring provides objective visibility into turnaround progress, eliminating the traditional reliance on manual reporting and enabling immediate detection of deviations from planned schedules.

**Predictive Analytics Model** The second AI model leverages the detected events, enriched with flight-specific data, alongside a comprehensive historical dataset to predict how the remainder of the turnaround will unfold. This prediction engine has been trained on over 150,000 past turnarounds, enabling sophisticated clustering and pattern-matching capabilities. By identifying similar historical scenarios and their outcomes, the system can detect potential delays up to 40 minutes before the scheduled off-block time. The model outputs a prediction called Predicted Off-Block Time (POBT), referred to at Schiphol as predicted end of ground-handling time (PEGT) to emphasize that it predicts when ground handling will be complete rather than the actual pushback moment.

## 3.2. Data Processing and Pattern Recognition

The system's predictive capability relies on sophisticated clustering algorithms that match ongoing turnarounds with similar historical patterns. When analyzing a current turnaround, the AI identifies comparable scenarios from its database based on factors such as aircraft type, airline operational procedures, time of day, weather conditions, and the specific sequence of observed events. This pattern-matching approach recognizes that turnaround behaviors follow identifiable patterns that can be learned from historical data.

The clustering methodology enables the system to provide context-aware predictions that account for the specific operational scenario being observed. For example, the system can distinguish between normal boarding delays and those caused by late connecting passengers, or identify patterns associated with specific airlines' operational procedures. This nuanced understanding allows Deep Turnaround to generate predictions that reflect the complexity and variability inherent in real-world turnaround operations.

## 3.3. Known Biases and Timing Characteristics

Analysis of PEGT performance has revealed systematic timing biases that vary depending on the prediction horizon (Aviation Solutions, 2025). The system exhibits a tendency toward:

- **T-15 minute undershoot**: Predictions made 15 minutes before the target TOBT tend to be slightly optimistic, underestimating the actual completion time
- **T-5 and T-1 minute overshoot**: As the prediction horizon shortens to 5 minutes and 1 minute before AEGT, the system tends toward conservative estimates, occasionally overestimating completion times
- **Peak-hour systematic bias**: During high-traffic periods, both PEGT and human-set TOBT may systematically underestimate turnaround durations due to resource constraints and operational pressure

These timing biases reflect the inherent challenges of turnaround prediction, where early estimates must rely more heavily on planned activities and historical patterns, while late-stage predictions can incorporate real-time observations but may be influenced by operational urgency. Importantly, the system operates with a fundamental limitation, it can only detect whether an activity has started or is completely finished. It lacks the nuanced ability to assess intermediate states such as 'nearly complete' or 'proceeding slowly'. This binary detection limitation highlights where human judgment remains crucial, as experienced ground personnel can visually assess subtle progress indicators, estimate remaining durations based on contextual factors, and anticipate completion with nuances the AI cannot perceive. Understanding both these timing biases and detection limitations is crucial for proper integration of PEGT into operational planning systems, emphasizing the complementary roles of AI prediction and human expertise.

## 3.4. Current Advisory Implementation

Deep Turnaround's operational implementation at Schiphol provides sophisticated advisory capabilities that are tightly integrated into the airport's A-CDM process. Rather than operating as a standalone system, PEGT predictions are seamlessly incorporated into existing operational tools and workflows, making the AI-driven insights immediately accessible to all stakeholders.

The system continuously compares its PEGT predictions with the current TOBT set by ground handlers and airlines. When PEGT deviates significantly from TOBT (typically by more than 5 minutes) the system generates automated alerts to relevant stakeholders. These alerts enable ground staff to proactively update the TOBT to a more realistic value, preventing last-minute surprises and expired TSAT slots.

The system generates specific advisory messages tailored to different stakeholders based on detected patterns. For instance, when PEGT indicates a likely delay, ground handlers receive targeted notifications about required resources, while airlines are informed about potential TOBT updates. These actionable recommendations, illustrated in Figure 5.4, help stakeholders take appropriate measures to mitigate delays or optimize the turnaround process.

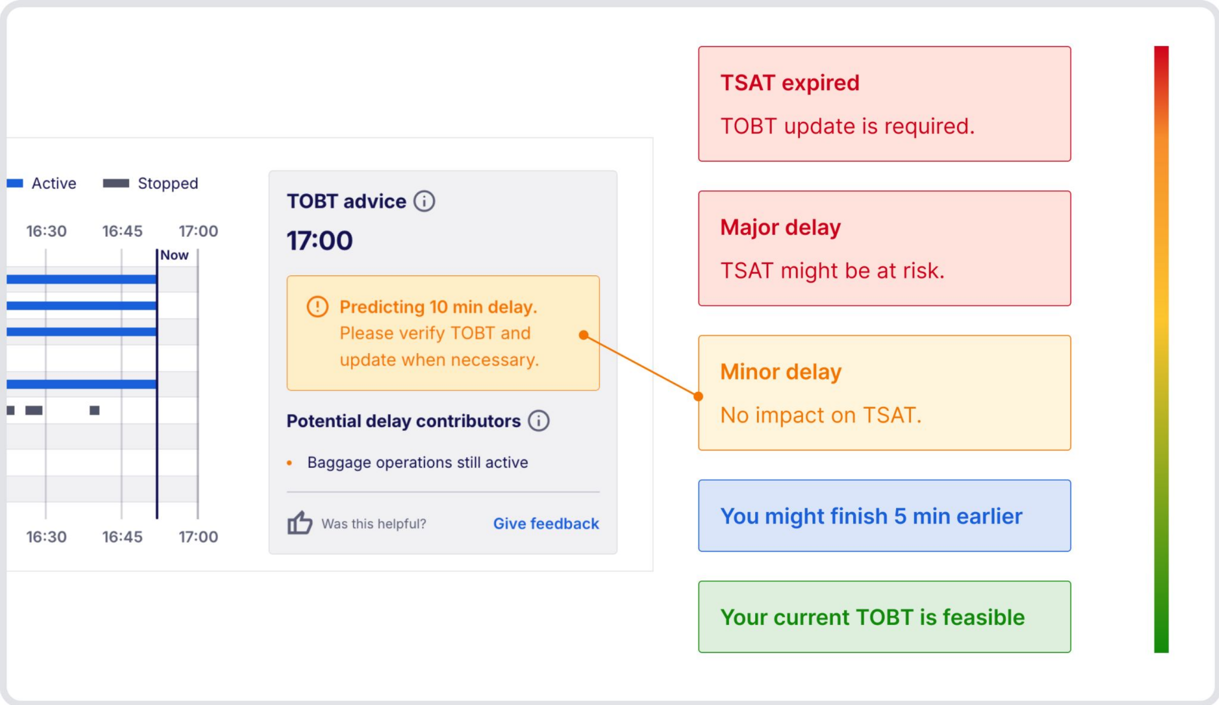



**Figure 5.4:** Deep Turnaround advisory interface showing actionable recommendations to stakeholders. The system generates specific advice such as suggested TOBT updates for airlines, resource allocation recommendations for ground handlers, and operational insights for airport authorities.

All detected events and predictions are accessible through multiple channels:

- **API Integration**: Real-time data feeds enable integration with existing airport operational systems
- **Turnaround Insights Dashboard**: Live visualization of turnaround progress and predictions for operations teams
- **Mobile Applications**: Portable access for ground staff and airline personnel
- **APOC Integration**: Direct integration into Schiphol's Airport Operations Center processes

Importantly, Schiphol has formally adopted PEGT as part of its collaborative decision-making vocabulary, reflecting the airport's confidence in the system and its integration into standard operational procedures (Alexopoulos, 2024). This represents a significant shift from reactive to proactive A-CDM paradigm.

## 3.5. Operational Performance and Benefits

The implementation of Deep Turnaround at Amsterdam Schiphol has delivered measurable improvements in operational performance, transforming what was previously a 'blind spot' into a continuously monitored and optimized process. The system's ability to provide objective, real-time turnaround monitoring has enabled data-driven adjustments that significantly reduce delays and enhance on-time performance.

Deep Turnaround has transformed daily operational practices:

- **Reduced coordination overhead**: Ground handling agents report significantly fewer coordination phone calls, with Menzies Aviation noting 'countless hours' saved on busy days
- **Faster issue identification**: Precise live data eliminates the need to call around to identify delay sources
- **Objective root cause analysis**: Data-driven insights revealed that 11% of missed TSAT slots were due to late 'ready for startup' calls; a previously difficult-to-quantify issue
- **Proactive resource allocation**: Earlier delay detection enables timely deployment of additional staff and resources

The system has also fostered a more collaborative and accountable operational culture. Every delay now has a clear data trail, enabling stakeholders to work together on solutions using objective metrics rather than subjective assessments. This transparency has reduced finger-pointing and encouraged cooperative problem-solving, as all parties can access the same real-time information about turnaround progress.

### 3.6. Integration with A-CDM and Future Development

Deep Turnaround exemplifies successful integration of advanced AI capabilities into established airport collaborative decision-making processes. At Schiphol, the system is an add-on of the A-CDM framework that enhances the accuracy and reliability of critical operational milestones.

The system directly addresses historical challenges in A-CDM operations where TOBT updates relied on manual communication and were often imprecise or delivered too late. Deep Turnaround provides an objective 'second opinion' on each flight's readiness, enabling proactive TOBT corrections that prevent cascading issues such as expired TSAT slots and gate conflicts.

Future development directions focus on direct integration of PEGT into DMAN systems through sophisticated data fusion techniques, as explored in Section 4. This integration aims to combine the early availability of TOBT with the higher accuracy of PEGT to produce optimal timing estimates for departure sequencing.

By combining real-time situational awareness with historical pattern recognition and uncertainty quantification, Deep Turnaround addresses fundamental limitations in traditional turnaround management while maintaining compatibility with established operational procedures. The system demonstrates how advanced machine learning techniques can enhance aviation operations through improved prediction accuracy and explicit uncertainty modeling, ultimately achieving better punctuality and throughput by only adding two cameras per aircraft stand.

## 4. Data Fusion

As explained in the previous sections, the turnaround is a complex process with many uncertainties. To boost the predictability of the turnaround, Schiphol introduced 'Deep Turnaround' which computes the predicted end of ground-handling time (PEGT).

PEGT is high-confidence (data-driven) but becomes available later in the turnaround (as the AI observes events) compared to the early TOBT calculations. Currently, the PEGT implementation only provides advisory information to ground handlers and airlines (e.g., 'Advice to delay TOBT by 5 minutes due to fueling truck'), leaving the human operator to decide whether to act on this information. In practice, this approach helped improve accuracy, however, it still relied on human intervention and did not fully leverage the predictive power of the available data. This research aims to implement this valuable data stream directly into the existing DMAN system. The goal is to fuse the early, low-confidence TOBT with the later, higher-confidence PEGT into one improved prediction (let's call it the fused-TOBT&PEGT) that DMAN can use in place of a single off-block estimate.

Integrating this PEGT into the certified DMAN system must be done modularly with minimal changes. The methods should handle

- **Asynchronous availability** - TOBT is known well in advance, while PEGT arrives closer to departure
- **Unequal confidence** - PEGT is generally more accurate than TOBT
- **Backward compatibility** - if PEGT is missing, the system should revert to using TOBT as before
- **Modularity** - the fusion module should be easily implementable, replaceable, or upgradable without affecting the rest of the DMAN system

The goal is to create a fused time that is more accurate than either input alone, improving the predictability of the turnaround process. This section analyzes candidate methods from control theory, estimation, data fusion, and systems engineering that could produce the desired fused-TOBT&PEGT. Each methodology is examined conceptually with appropriate formalisms to aid understanding, followed by a critical assessment of feasibility within the airport turnaround context, advantages and limitations, and integration considerations as a modular component.

### 4.1. Kalman Filter

A Kalman filter offers a principled way to fuse two streams of time estimates based on their uncertainties. It treats the actual off-block time (or end-of-ground-handling time) as an unknown state to be estimated, and each prediction (TOBT or PEGT) as a noisy measurement. The filter operates in two steps: a prediction (time update) and an update (measurement fusion). In this scenario, the 'process model' can be very simple. It is assumed that the best current estimate of the off-block time stays the same unless corrected (since it's more of a static parameter for each flight, possibly with a slow drift if delays accumulate). When a new prediction comes in, the Kalman update step computes a weighted average of the current estimate and the new input, weighting more heavily the source with lower uncertainty (Grewal & Andrews, 2001). Essentially, the Kalman filter automatically assigns higher weight to PEGT if it has a higher accuracy (lower variance), and a lower weight to the noisier TOBT, in an optimal Bayesian sense. This is why Kalman filtering is widely used in sensor fusion problems, it yields an estimate with minimized mean-square error by blending information based on confidence.

The Kalman filter implementation for the turnaround time prediction can be formulated as follows:

When TOBT is provided, it is treated as an initial measurement:

$$z_{\text{TOBT}} = x + e_{\text{TOBT}} \tag{5.1}$$

where $x$ is the true off-block time (unknown state) and $e_{\text{TOBT}}$ represents measurement noise with a variance reflecting TOBT's reliability. At this point, the Kalman filter would initialize its state estimate using TOBT.

As time progresses without new information, the filter can either:

- Maintain the current estimate (assuming a static process)
- Apply a predictive model to account for how off-block times typically evolve (e.g., a slight random walk to model potential delays)

When new TOBT updates arrive, they are processed like any other measurement:

$$z_{\text{TOBT,new}} = x + e_{\text{TOBT}} \tag{5.2}$$

with the update equations:

$$K_k = \frac{P_{k|k-1}}{P_{k|k-1} + R_{\text{TOBT}}} \tag{5.3}$$

$$\hat{x}_{k|k} = x_{k|k-1} + K_k(z_{\text{TOBT,new}} - x_{k|k-1}) \tag{5.4}$$

The system treats these updated TOBTs with the same uncertainty model as the initial ones, though historical data could be used to adjust $R_{\text{TOBT}}$ dynamically if repeated updates tend to follow patterns.

When PEGT becomes available at time $t_k$, it is treated as a new measurement:

$$z_{\text{PEGT}} = x + e_{\text{PEGT}} \tag{5.5}$$

where $e_{\text{PEGT}}$ has a smaller variance since PEGT is higher-confidence. The Kalman update equations then adjust the estimate as:

$$K_k = \frac{P_{k|k-1}}{P_{k|k-1} + R_{\text{PEGT}}} \tag{5.6}$$

$$\hat{x}_{k|k} = x_{k|k-1} + K_k(z_{\text{PEGT}} - x_{k|k-1}) \tag{5.7}$$

Here, $P_{k|k-1}$ is the prior variance of the estimate and $R_{\text{PEGT}}$ is the PEGT measurement variance. If PEGT is very reliable (small $R_{\text{PEGT}}$), the Kalman gain $K_k$ will be high, causing the estimate to move strongly toward PEGT. Conversely, if PEGT is absent, the filter continues with the TOBT-based prediction.

This asynchronous update mechanism is a natural strength of Kalman filters, updates are performed whenever a measurement arrives, and irregular timing can be handled by adjusting the prediction step accordingly. This makes the Kalman filter particularly well-suited to the airport turnaround scenario, where TOBT and PEGT become available at different times with different confidence levels.

This approach is highly feasible. It would act as a small estimation module receiving TOBT updates and later PEGT updates, outputting a continuously refined fused-TOBT&PEGT. It does not require altering DMAN's core, DMAN would just receive the Kalman filter's current output as the best off-block time. Kalman filters are computationally lightweight and well-understood, even in safety-critical applications (they have been used in aircraft navigation for decades (Grewal & Andrews, 2001)). Tuning would involve setting the noise variances for TOBT and PEGT, based on historical error distributions.

**Advantages:**

- **Optimal data fusion:** For Gaussian noise, the Kalman filter yields the mathematically optimal fusion, essentially performing an uncertainty-weighted merge of the two predictions. It will improve the accuracy of the off-block time prediction by leveraging PEGT's precision while still using TOBT as needed.
- **Handles asynchronicity:** The KF naturally supports irregular measurement updates. It will keep propagating (with or without change) until a new measurement (PEGT) arrives, then incorporate it.
- **Confidence estimation:** It provides a running estimate of uncertainty (error covariance) for the fused-TOBT&PEGT, which could be useful to stakeholders.
- **Minimal integration impact:** This can be a self-contained module feeding an improved time into DMAN, without needing to modify the DMAN algorithm. If PEGT is unavailable (e.g., system outage), the filter's update step just won't run and it will effectively rely on TOBT (backward compatible by design).
- **Extensibility:** If in the future more predictive signals emerge, they can be added as additional measurements in the Kalman framework easily.
- **Single-point output:** The basic Kalman filter will output a single best estimate. DMAN requires a single deterministic time, not a distribution.

**Limitations:**

- **Requires tuning:** Reasonable values for TOBT and PEGT error variances must be obtained (and any process model parameters). Poor tuning can reduce performance – e.g., if TOBT's uncertainty is underestimated, the filter might stick too long to a bad TOBT.
- **Assumes error characteristics:** The standard Kalman filter assumes roughly linear dynamics and Gaussian noise. In this reasearch the 'dynamics' are non-trivial and error distributions of time estimates are most likely non-linear.
- **Correlation:** PEGT and TOBT are not independent (for example, ground handlers might adjust TOBT when seeing PEGT advice), the Kalman formulas could overestimate confidence by double-counting information. In such cases, additional techniques like Covariance Intersection can be used to fuse estimates conservatively without assuming independence. Covariance intersection essentially finds a fused covariance between the two that is guaranteed not to be over-optimistic regardless of correlation – trading a bit of optimality for robustness.

## 4.2. General Bayesian Data Fusion Approach

While the Kalman filter provides an optimal solution for the linear-Gaussian case, a more general Bayesian fusion framework offers additional flexibility. This approach conceptualizes the actual off-block time $x$ as a random variable, with both TOBT and PEGT providing evidence to update our belief about $x$.

In the Bayesian framework, it begins with a prior distribution for $x$, typically centered on the initial TOBT with appropriate uncertainty. When PEGT becomes available, a standard Bayesian update is performed:

$$P(x|\text{PEGT}) \propto P(\text{PEGT}|x)P(x) \tag{5.8}$$

The errors can be modeled as additive Gaussian noise:

$$\text{TOBT} = \text{actual} + e_{\text{TOBT}}, \quad e_{\text{TOBT}} \sim \mathcal{N}(0, \sigma^2_{\text{TOBT}}) \tag{5.9}$$

$$\text{PEGT} = \text{actual} + e_{\text{PEGT}}, \quad e_{\text{PEGT}} \sim \mathcal{N}(0, \sigma^2_{\text{PEGT}}) \tag{5.10}$$

Then the posterior mean (the fused estimate) becomes a confidence-weighted average:

$$\hat{x} = \frac{\sigma^{-2}_{\text{PEGT}} \cdot \text{PEGT} + \sigma^{-2}_{\text{TOBT}} \cdot \text{TOBT}}{\sigma^{-2}_{\text{PEGT}} + \sigma^{-2}_{\text{TOBT}}} \tag{5.11}$$

This result is mathematically equivalent to the Kalman filter solution for this scenario. However, the Bayesian approach extends naturally to non-Gaussian uncertainties or more complex prior knowledge. For example, it could accommodate a bimodal distribution representing that '90% of flights will depart by 15:00, but 10% experience last-minute delays of 30+ minutes.'

In Bayesian terms, each data source provides a likelihood function;PEGT's higher confidence corresponds to a narrower likelihood, while TOBT provides a broader one. The fused result combines these to produce a posterior distribution with reduced uncertainty. When the signals agree, the posterior peaks strongly around that time; when they differ, it may become skewed, providing a natural way to detect conflicts.

For implementation, the simplest case (with Gaussian error assumptions) reduces to the weighted average formula above. For more complex distributions, techniques like particle filtering could represent the full posterior distribution, though for DMAN's needs, providing the posterior mean as a single point estimate is likely sufficient. The framework can also be extended to incorporate methods such as Dempster-Shafer theory (Shafer, 1976) for handling explicit uncertainty and belief functions.

The method is highly feasible for the airport context. It offers a modular approach that can be implemented as a preprocessing step before feeding timing information to DMAN. Calibration would involve analyzing historical data to determine appropriate error distributions for TOBT and PEGT. The output would be a single time estimate (posterior mean or other statistic) that DMAN can use directly.

**Advantages:**

- **Conceptual transparency:** The approach explicitly models uncertainty in each source and combines them rationally, providing interpretable results (e.g., '90% confidence the flight will be off-block by 15:20 $\pm$ 2 min').
- **Flexibility:** Can handle arbitrary probability distributions beyond Gaussian. If TOBT tends to be optimistic by 5 minutes on average, the prior can be adjusted accordingly.
- **Minimal interference with DMAN:** Distributions can be derived from historical data without altering the entire DMAN system.
- **Output alignment:** The approach can provide either a mean estimate or a percentile-based estimate (e.g., 85th percentile latest time) if minimizing missed slots is critical. Or tuned differently for any objective.

**Limitations:**

- **Computational complexity:** Non-linear, non-Gaussian cases require numerical methods for Bayesian updates, potentially adding unnecessary complexity.
- **Prior calibration:** Requires quantitative models for the confidence of both TOBT and PEGT, which must be developed through historical error analysis.
- **Correlation handling:** TOBT and PEGT are not independent (e.g., if handlers adjust TOBT based on PEGT), naive application can overestimate certainty. Solutions include covariance intersection (Julier & LaViola, 2007) or variance inflation to account for correlation.
- **Distribution simplification:** The full posterior distribution must ultimately be reduced to a single point estimate for DMAN, potentially limiting the advantage over simpler methods like the Kalman filter.

## 4.3. Ensemble and Weighted Average Methods

Machine learning ensemble techniques offer a complementary approach to the statistical methods discussed previously. In this framework, TOBT and PEGT are treated as two 'expert predictors' whose outputs can be combined to produce a more accurate off-block time estimate. The simplest implementation is a weighted linear combination:

$$\hat{x}_{\text{fused}} = w_{\text{TOBT}} \cdot \text{TOBT} + w_{\text{PEGT}} \cdot \text{PEGT} \tag{5.12}$$

where $w_{\text{TOBT}} + w_{\text{PEGT}} = 1$. This approach provides considerable flexibility: weights can be static (fixed values) or dynamic (changing based on context). Before PEGT becomes available, $w_{\text{PEGT}} = 0$ by necessity, so $\hat{x}_{\text{fused}}$ = TOBT, ensuring backward compatibility. Once PEGT arrives, $w_{\text{PEGT}}$ could increase, potentially to $w_{\text{PEGT}} \approx 1$ if PEGT is known to be substantially more accurate.

Time-varying weight schemes can further refine this approach. As a flight's scheduled departure time approaches, the weight assigned to PEGT could increase, reflecting its growing reliability relative to TOBT. Additionally, rule-based gating mechanisms can be implemented; for instance, if PEGT differs significantly from TOBT (suggesting TOBT might be outdated), the system could automatically increase PEGT's influence. Conversely, for slight differences, a smoothing function might be applied to prevent oscillations in the output.

A more sophisticated implementation involves training a meta-learner;a small regression model that takes both TOBT and PEGT as inputs and outputs the fused estimate. This approach resembles a stacked ensemble or mixture-of-experts model (Waterhouse et al., 1995):

$$\hat{x}_{\text{fused}} = \alpha \cdot \text{TOBT} + \beta \cdot \text{PEGT} + \gamma \tag{5.13}$$

Using historical data comparing AOBT with predictions, the parameters $\alpha$, $\beta$, and $\gamma$ can be learned. This might reveal systematic biases - consultations with Schiphol data scientists indicate that PEGT may exhibit certain biases in specific conditions (Aviation Solutions, 2025), while TOBT is consistently optimistic by several minutes (reflected in $\alpha < 1$ or $\gamma < 0$).

In reality, this would most likely require a time-evolving regression model, since the optimal weights would shift with the progression of the turnaround process. For example, TOBT might be more reliable at certain phases, while PEGT gains accuracy as specific turnaround milestones are observed. This temporal dimension significantly complicates the implementation, requiring much larger datasets to properly train such dynamic models. Unfortunately, these datasets are not yet available, partly because PEGT is relatively new and partly because data from recent years does not represent normal operations (since it's affected by the COVID-19 pandemic), limiting its value for model training. Once trained, this small model would function as a dedicated fusion module without requiring modifications to DMAN's core algorithms. The model could be periodically updated offline with new data in a controlled manner.

This ensemble approach is analogous to techniques used in weather forecasting (Krishnamurti et al., 1999), where multiple meteorological models are combined to produce more accurate predictions than any individual model alone. Similarly, in predictive maintenance, physics-based models are often fused with data-driven models, with weights assigned based on each model's historical accuracy for specific components (Romero et al., 2010).

**Advantages:**

- **Implementation simplicity:** Weighted averages are straightforward to implement and interpret, potentially as effective as more complex methods for this application.
- **Domain knowledge integration:** Expert rules can be directly encoded e.g., 'rely entirely on PEGT if available >10 minutes before departure; otherwise use a 50/50 blend to avoid disruptive last-minute changes.'
- **No system dynamics required:** Unlike Kalman filtering, no state evolution equations are needed, making this approach more accessible for straightforward implementation.
- **Operational control:** Airport stakeholders can explicitly adjust weights during irregular operations (e.g., reducing dependence on automated predictions during disruptions).
- **Ensemble benefits:** Machine learning research consistently demonstrates that combining models improves accuracy when individual models have complementary strengths (Dietterich, 2000).

**Limitations:**

- **Manual uncertainty encoding:** Unlike Bayesian approaches that naturally handle uncertainty through probability distributions, weighted averages require explicit encoding of confidence levels.
- **Potential discontinuities:** Abrupt weight changes (e.g., suddenly switching from TOBT to PEGT) could cause jumps in the fused-TOBT&PEGT, potentially triggering disruptive queue reshuffling in DMAN.
- **Data dependency:** Learned meta-models require historical datasets with both predictions and actual times, which are currently insufficient due to PEGT's novelty and pandemic-affected data.
- **Temporal complexity:** The time-evolving nature of optimal weights adds significant model complexity and training requirements.
- **Certification challenges:** Machine learning components may face additional scrutiny in aviation certification processes, as their behavior is not fully deterministic, though the proposed regression model is simple enough to potentially mitigate these concerns.

## 4.4. Online Learning and Adaptive Fusion

While the previously discussed methods use static or offline-determined parameters, an online learning approach would adapt the fusion strategy continuously based on performance feedback. This methodology is particularly valuable when systems face unprecedented disruptions or gradual shifts in operational patterns.

**Incremental Weight Adaptation**
Online learning treats TOBT and PEGT as 'expert advisors' whose reliability may change over time. Drawing from the 'learning with expert advice' framework (Cesa-Bianchi & Lugosi, 2006), an adaptive algorithm can continuously adjust the weighting between these experts based on their recent predictive accuracy:

$$\hat{x}_{t+1} = w_{\text{TOBT},t+1} \cdot \text{TOBT}_{t+1} + w_{\text{PEGT},t+1} \cdot \text{PEGT}_{t+1} \tag{5.14}$$
$$w_{i,t+1} = w_{i,t} \cdot \exp\big(-\eta \cdot \ell(p_{i,t}, \text{AOBT}_t)\big) \tag{5.15}$$

where $\ell(p_{i,t}, \text{AOBT}_t)$ measures the prediction error of expert $i$ for the recently completed flight $t$, and $\eta$ is a learning rate controlling adaptation speed. After each update, weights are normalized to sum to 1. This approach, a variant of the Hedge algorithm (Freund & Schapire, 1997), gradually increases the influence of the more accurate predictor, allowing the system to adapt to changing conditions.

A more structured approach involves recursive estimation of regression coefficients:

$$\hat{x}_t = \alpha_t + \beta_t \cdot \text{TOBT}_t + \gamma_t \cdot \text{PEGT}_t \tag{5.16}$$

Using recursive least squares (RLS) (Haykin, 2002), the parameters $\alpha_t$, $\beta_t$, and $\gamma_t$ are updated incrementally as each flight completes, minimizing accumulated prediction error. This captures not only relative importance but also systematic biases in each data source.

**Context-Aware Adaptation**
More sophisticated variants can incorporate contextual information, effectively learning different fusion strategies for different operational scenarios:

$$\hat{x}_t = f_t(\text{TOBT}_t, \text{PEGT}_t, \text{context}_t) \tag{5.17}$$

Here, $f_t$ is a lightweight model that considers additional features such as airline, aircraft type, time of day, or weather conditions. This contextual approach transitions from pure data fusion toward a bandit-style learning problem (Bubeck & Cesa-Bianchi, 2012), where the system learns which fusion strategy works best in each context.

**Implementation Considerations**
For practical implementation in airport operations:

- **Safety boundaries:** Parameters can be constrained within safe limits (e.g., $w_{\text{PEGT}} \in [0.3, 0.9]$ when available) to prevent extreme behavior
- **Learning rate decay:** Using $\eta_t = \frac{\eta_0}{\sqrt{t}}$ or similar schedules prevents overreaction to temporary anomalies
- **Forgetting mechanisms:** Exponential weighting of historical data ensures adaptation to changing conditions rather than the permanent influence of past patterns
- **Anomaly detection:** Before updating weights, outlier flights (e.g., those with exceptional delays) can be excluded from the learning process

The online module operates alongside DMAN, producing the fused estimate without altering core algorithms. After each flight departs, actual off-block times are logged and used to update the model parameters, creating a feedback loop that improves over time. If the module fails or produces suspect results, the system can fall back to static weights or TOBT-only operation.

**Advantages:**

- **Automatic adaptation:** Adjusts to evolving conditions without manual parameter tuning, handling gradual shifts in TOBT or PEGT reliability

- **Anomaly accommodation:** During disruptions gradually recalibrates to new delay patterns without requiring manual intervention
- **Infrastructure adaptation:** Automatically learns how operational changes (e.g., runway configuration changes, terminal maintenance, or ground handling equipment restrictions) affect turnaround times and adjusts fusion weights accordingly
- **Continuous improvement:** Learns from operational experience, effectively implementing a small-scale continuous learning loop
- **Operational insights:** Parameter evolution can reveal changing operational patterns (e.g., if certain airlines' TOBT quality improves)

**Limitations:**

- **Certification challenges:** Adaptive components face additional scrutiny in safety-critical environments, requiring careful bounds and failure modes
- **Delayed feedback:** Learning occurs only after flights depart, creating a lag before adaptations benefit operations
- **Parameter sensitivity:** Learning rate and constraints significantly impact performance;too slow fails to adapt, too fast overreacts to anomalies
- **Start-up data needs:** Initial deployment requires historical data for reasonable starting parameters
- **Reactivity vs. proactivity:** Being inherently reactive, the system 'learns after the fact' rather than anticipating disruptions

The online learning approach offers a compelling middle ground between static fusion and full DMAN replacement. It maintains the modularity and safety of simpler fusion methods while adding a controlled adaptive element. However, its reactive nature means it will always lag behind real-world changes to some degree. For daily operations with moderate variability, this approach provides substantial benefits with manageable complexity. During unprecedented disruptions like pandemic shutdowns or volcanic ash events, it would gradually adapt but couldn't predict such events in advance;a limitation shared by all data-driven systems without explicit modeling of such rare scenarios.

## 4.5. Robust Filtering and Outlier Handling

While the previous sections focus on methods that optimize for mean prediction accuracy, real-world airport operations must also consider the impact of outliers. Outliers, instances where either TOBT or PEGT significantly deviates from reality, can cause disruptive re-sequencing in DMAN if not properly handled.

### Robust Estimation Methods

Standard Kalman and Bayesian methods minimize mean-squared error under assumed noise characteristics, but they can be sensitive to violations of these assumptions. Robust methods sacrifice some average-case performance to protect against worst-case scenarios.

The $H_\infty$ filter represents one principled approach to robust estimation. Unlike the Kalman filter (which is $H_2$-optimal), the $H_\infty$ filter minimizes the worst-case error under bounded uncertainties (Simon, 2006). This approach treats prediction error as an adversarial disturbance and is less sensitive to misspecified noise models. For off-block time prediction, an $H_\infty$ filter would:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k(z_k - \hat{x}_{k|k-1}) \tag{5.18}$$

$$K_k = \frac{P_{k|k-1}}{P_{k|k-1} + \gamma R_k} \tag{5.19}$$

Here, $\gamma > 1$ is a robustness parameter that makes the filter more conservative about measurement updates than a standard Kalman filter. The practical effect is that the $H_\infty$ filter will trust new measurements less aggressively, particularly when the measurement model is uncertain.

Simpler robust methods include M-estimators or direct outlier rejection mechanisms. For example, the Kalman filter update could be modified using a Huber loss function:

$$\hat{x}_{k|k} = \hat{x}_{k|k-1} + K_k \cdot \psi(z_k - \hat{x}_{k|k-1}) \tag{5.20}$$

Where $\psi(\delta)$ equals $\delta$ for small errors but saturates for large deviations. This effectively caps the influence of any single measurement, preventing one erroneous PEGT from completely resetting the fused estimate.

**Fault Detection and Exclusion**
Another approach involves explicit fault detection. If PEGT and TOBT differ beyond a reasonable threshold (e.g., 15 minutes), one could be flagged as potentially faulty:

$$\text{if } |\text{PEGT} - \text{TOBT}| > \text{threshold} \Rightarrow \text{trigger fault logic} \tag{5.21}$$

The fault logic might:

- Use the more credible source exclusively
- Increase the assumed uncertainty of the outlying prediction
- Flag the flight for human operator attention
- Apply a weighted blend but limit the maximum allowed shift

This is conceptually similar to fault-tolerant interval intersection in sensor fusion, where each input provides a range of plausible values rather than a single point estimate (Zhang et al., 2006).

**Operational Considerations**
In the turnaround context, robustness serves several practical purposes:

- **Stability:** Prevents 'jitter' in the DMAN sequence due to questionable updates
- **Fault isolation:** Ensures issues with one data source don't contaminate the entire system
- **Predictable degradation:** Provides controlled fallback when data quality deteriorates

However, robust methods also introduce additional complexity and potential conservatism. For example, an $H_\infty$ filter might react too slowly to genuine updates if tuned too conservatively, while overly aggressive outlier rejection could discard valid but surprising updates (such as a sudden major delay that PEGT correctly identifies).

**Mean Optimization vs. Worst-Case Guarantees**
For this research, the optimal balance likely favors mean prediction accuracy over worst-case robustness for several reasons:

- The airport environment maintains human oversight, outbound planners continuously monitor the system and can override automated predictions when necessary
- Extensive robustness mechanisms may sacrifice too much average-case performance
- External factors like Eurocontrol's Network Manager assigning CTOTs, which can cause predictable 'outliers' as flights finish ground handling and then wait for their flow management slot
- The system already has processes for handling exceptional circumstances (e.g., adverse weather conditions)

Therefore, while robust elements should be incorporated;particularly simple outlier detection with appropriate thresholds;the primary focus will remain on optimization for mean prediction accuracy. The fusion model should aim to be robust to common variations but defer to human judgment for truly exceptional cases. This recognizes the reality that in aviation operations, computers optimize nominal operations while humans manage exceptions; a principle that has served the industry well for decades. Implementation could follow a hybrid approach: using Kalman or Bayesian fusion as the core method, with a lightweight robustness wrapper that watches for inconsistencies exceeding predetermined thresholds. This balances computational simplicity with protection against the most disruptive outliers.

## 4.6. Correlation of TOBT and PEGT

A critical assumption underlying most data fusion methods is the independence of input sources. However, in the airport turnaround context, TOBT and PEGT are not independent measurements of the same underlying phenomenon. They exhibit complex correlations that can significantly impact fusion performance if not properly addressed.

**Sources of Correlation**
The correlation between TOBT and PEGT arises from several operational realities. The most direct source stems from PEGT's advisory role. When Deep Turnaround generates a PEGT estimate that differs significantly

from the current TOBT, it provides advisory messages to ground handlers and airlines (e.g., 'Advice to delay TOBT by 5 minutes due to fueling truck delays'). If operators act on this advice by updating their TOBT, the two estimates become correlated through this feedback mechanism. This creates a scenario where PEGT influences TOBT, violating the independence assumption of standard fusion methods.

Correlation can additionally arise from systematic biases affecting both estimates. For example, during peak hours, both TOBT (set by time-pressured ground handlers) and PEGT (trained on historical data) might systematically underestimate turnaround times, leading to positive correlation in their errors (Aviation Solutions, 2025).

**Asymmetric Information**
While correlation often assumes symmetric information sharing, the TOBT-PEGT relationship exhibits important asymmetries. Ground handlers possess contextual information that PEGT may lack access to. This includes last-minute crew scheduling changes, informal communications between ground teams, visual observations of equipment status or passenger flow, knowledge of airline-specific operational preferences, and awareness of imminent but not yet recorded events (e.g., maintenance issues discovered during inspection). In these cases, TOBT updates may reflect information unavailable to PEGT, creating scenarios where human judgment provides superior predictions despite PEGT's generally higher accuracy.

Conversely, PEGT has access to comprehensive data streams and pattern recognition capabilities that exceed human cognitive capacity. These include real-time processing of multiple simultaneous data sources, pattern recognition across large historical datasets, objective assessment without optimism bias, and consistent application of learned relationships. This asymmetry suggests that correlation patterns may vary depending on the type of operational scenario and the availability of different information sources.

**Correlation Impact on Fusion Methods**
The presence of correlation affects different fusion approaches in distinct ways. Standard Kalman filtering assumes uncorrelated measurement noise. When TOBT and PEGT are correlated, the filter may overestimate the information content of the combined measurements, leading to overconfident estimates with artificially reduced uncertainty bounds.

Similarly, Bayesian fusion methods that treat TOBT and PEGT as independent evidence sources will overestimate the posterior certainty. The Bayesian update:

$$P(x|\text{TOBT}, \text{PEGT}) \propto P(\text{TOBT}|x) \cdot P(\text{PEGT}|x) \cdot P(x) \tag{5.22}$$

assumes conditional independence of the likelihoods given the true state $x$. Violation of this assumption leads to overconfident posterior distributions.

**Handling Correlation in Practice**
Several approaches can address correlation while maintaining the modularity and practicality required for DMAN integration. Covariance Intersection (Julier & LaViola, 2007) provides a conservative fusion method that does not require knowledge of the correlation structure. Instead of assuming independence, it finds the fused estimate that minimizes the worst-case error regardless of the actual correlation. For Gaussian estimates, this reduces to:

$$\hat{x}_{\text{fused}} = \frac{\omega\sigma_{\text{PEGT}}^{-2} \cdot \text{PEGT} + (1-\omega)\sigma_{\text{TOBT}}^{-2} \cdot \text{TOBT}}{\omega\sigma_{\text{PEGT}}^{-2} + (1-\omega)\sigma_{\text{TOBT}}^{-2}} \tag{5.23}$$

$$\sigma_{\text{fused}}^{-2} = \omega\sigma_{\text{PEGT}}^{-2} + (1-\omega)\sigma_{\text{TOBT}}^{-2} \tag{5.24}$$

where $\omega \in [0, 1]$ is chosen to minimize the fused variance. This approach trades some optimality for robustness against unknown correlation.

If correlation patterns can be characterized through historical analysis, they can be explicitly modeled. For example, if analysis reveals that TOBT updates following PEGT advisories exhibit a correlation coefficient $\rho$, the fusion uncertainty can be adjusted accordingly:

$$\sigma^2_{\text{fused,adjusted}} = \sigma^2_{\text{fused,independent}} \cdot (1 + \rho \cdot f(\Delta t)) \tag{5.25}$$

where $f(\Delta t)$ represents how correlation strength varies with the time interval between PEGT generation and TOBT update.

A more sophisticated approach involves real-time correlation monitoring. By tracking the statistical relationship between TOBT and PEGT over recent flights, the fusion algorithm can adapt its assumptions:

$$\rho_t = \alpha \rho_{t-1} + (1 - \alpha) \cdot \text{corr}(\text{TOBT}_{\text{recent}}, \text{PEGT}_{\text{recent}}) \tag{5.26}$$

This running correlation estimate can then inform the fusion process, increasing conservatism when correlation is detected and reverting to optimal fusion when independence is observed.

A practical implementation approach involves temporal restrictions on fusion. If TOBT is updated within a threshold time window after PEGT becomes available (e.g., within 10 minutes), the fusion algorithm could assume correlation and apply conservative weighting. For TOBT updates occurring outside this window, standard fusion methods could be applied under the assumption that correlation is less likely.

The correlation between TOBT and PEGT represents a fundamental challenge for data fusion in operational environments where human operators can act on automated advisories. While this correlation complicates fusion mathematics, it also reflects the intended operation of the advisory system, where human operators should respond to PEGT insights when deemed appropriate. The key is developing fusion methods that harness this operational reality while avoiding the statistical pitfalls of assuming independence where it does not exist.

# 5. Classification of Delay Behavior Using TOBT Update Patterns

To enhance turnaround time prediction and improve operational realism within the PEGT-scenario, this section explores classifying flights based on their TOBT update behaviors. Rather than using static characteristics like aircraft type or airline, we aim to uncover latent structure in how flights respond to operational changes.

## 5.1. TOBT Update Patterns

To capture these patterns, each flight's delay behavior is represented as a structured matrix – a 'TOBT–TSAT behavior matrix' where one axis represents the timing of TOBT updates (how long before the last TOBT each update occurs) and the other axis represents the magnitude of each update (how many minutes the flight's off-block time was adjusted). Each cell in this matrix encodes the average difference between pre-update and post-update startup delays. This matrix is visualized in Figure 5.5.

| | Difference between new and old TOBT | | | | | | |
|---|---|---|---|---|---|---|---|
| Time to previous TOBT | 00-05 | 05-10 | 10-15 | 15-20 | 20-25 | 25-30 | 30+ |
| <-10 | . | . | . | . | . | . | . |
| -10--5 | . | . | . | . | . | . | . |
| 00-05 | . | . | . | . | . | . | . |
| 05-10 | . | . | . | . | . | . | . |
| 10-15 | . | . | . | . | . | . | . |
| 15-20 | . | . | . | . | . | . | . |
| 20-25 | . | . | . | . | . | . | . |
| 25-30 | . | . | . | . | . | . | . |
| 30+ | . | . | . | . | . | . | . |

**Figure 5.5:** Structure of the TOBT–TSAT behavior matrix

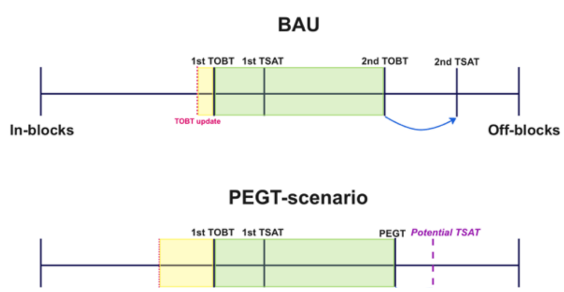



**Figure 5.6:** Example visualization of TSAT adjustments

This approach allows for a realistic approximation of airport congestion conditions that influence TSAT allocation. As shown in Figure 5.6, in the PEGT-scenario, TOBT updates occur earlier than in the Business As Usual (BAU) scenario, where updates often happen shortly before the initial TOBT.

By clustering these matrices, the aim is to group flights with similar latent behavioral patterns in their departure timing adjustments. The ultimate goal is to use these clusters to improve predictive performance – identifying categories of flights with high departure-time uncertainty versus those that are relatively stable, thus informing real-time uncertainty estimation and improving decisions in the DMAN system.

This analysis leverages two primary data sources (Deep Turnaround and CDM messages) to compare standard operations against a PEGT-enhanced scenario, focusing on start-up delay as the central metric. This classification framework will enable more context-aware uncertainty estimates, supporting the core objective of improving departure predictability through data-driven enhancements.

## 5.2. Unsupervised Clustering Methods for Behavioral Matrices

Clustering TOBT–TSAT behavior matrices requires methods that can handle structured matrix data points while discovering latent groupings without supervision. This section analyzes suitable unsupervised methods, examining their characteristics for the airport operations context.

**K-Means Clustering**

K-means assigns data points to $k$ clusters by minimizing within-cluster variance using squared Euclidean distance. This partitioning approach is computationally efficient but requires specifying $k$ in advance and tends to form convex clusters of similar size (Moloko et al., 2023). While widely used for its simplicity, K-means may struggle if behavioral patterns are not clearly separated or if clusters have irregular shapes.

In the airport turnaround context, $k$ could be chosen based on validation metrics or domain judgment;for example, representing categories like 'high uncertainty', 'moderate uncertainty', and 'stable' flights. K-means has been applied in delay prediction studies to cluster similar delay patterns before fitting separate prediction models (Guvercin et al., n.d.).

**Gaussian Mixture Models**

Gaussian Mixture Models (GMMs) employ a model-based approach, assuming data is generated from a mixture of Gaussian distributions. GMM clustering via the Expectation-Maximization algorithm provides probabilistic cluster assignments and can capture clusters with varying shapes and sizes through individual covariance matrices. This offers advantages when latent patterns vary in spread or orientation.

Like K-means, GMMs require specifying the number of components, though this can be determined using information criteria. GMMs excel when data resembles a multimodal Gaussian mixture, but may overfit or produce spurious small clusters if the component count is overestimated. GMMs have proven effective for identifying overlapping behavior patterns and anomalies where hard clustering is overly rigid (Reynolds, 2009), making them potentially valuable for flights with ambiguous delay characteristics.

**Density-Based Clustering**

Density-based methods group points that are closely packed in feature space while marking sparse regions as noise. DBSCAN (Density-Based Spatial Clustering of Applications with Noise) does not require a predefined number of clusters and can identify arbitrarily shaped clusters. It uses two parameters: $\epsilon$ (neighborhood radius) and MinPts (minimum points to form a dense region).

DBSCAN automatically determines the number of clusters and identifies outliers, enhancing robustness for discovering unusual patterns. This capability has been successfully leveraged in flight-data studies to detect abnormal operations without preset cluster counts (Li et al., 2015). However, DBSCAN may struggle with varying cluster densities.

HDBSCAN (Hierarchical DBSCAN) extends DBSCAN by building a hierarchy of clusters and extracting the most stable ones. This approach effectively handles variable densities by analyzing cluster stability across scales and typically outperforms K-means and basic DBSCAN on complex, heterogeneous data (Wang et al., 2025). In maritime research, HDBSCAN achieved superior trajectory clustering compared to K-means in terms of compactness and separation metrics (Wang et al., 2025).

For TOBT-TSAT behavior matrices, HDBSCAN shows particular promise as it can reveal natural groupings, including differentiating a cluster of 'noisy' outlier flights with irregular behavior, without requiring a preset cluster count.

**Mean Shift and Density Peak Clustering**
These algorithms identify clusters by locating modes (peak density points) in feature space. Mean Shift iteratively shifts data points toward nearby high-density regions as estimated by kernel density functions, with points converging to the same mode forming clusters. Like DBSCAN, it automatically determines cluster count and handles non-convex clusters, but can be computationally expensive for large datasets.

Mean Shift is particularly suitable when clusters correspond to density peaks in data distribution. In aviation contexts, similar mode-seeking approaches have been used to cluster aircraft taxiing positions into regions of operational interest' (Schultz et al., 2022), identifying operational hotspots through density estimation. For TOBT-TSAT matrices, Mean Shift could distinguish clusters like 'flights with no updates' from 'flights with multiple large delays' if such distinct modes exist in the data.

**Hierarchical Clustering**
Hierarchical clustering builds a tree (dendrogram) of clusters through successive merges of the closest clusters (agglomerative) or splits of clusters (divisive). This multi-level approach allows cutting the dendrogram at different levels to yield various partitioning schemes. Linkage methods (single-link, complete-link, Ward's) can use any distance metric and don't require specifying cluster count in advance.

This flexibility benefits the analysis of flight-behavior similarities, potentially revealing a natural hierarchy of patterns; broad groups that further subdivide into finer patterns. Though computationally intensive for large datasets, hierarchical clustering can complement other methods by initializing cluster centers for K-means or validating cluster stability. Different linkage criteria (Ward, complete, average) can significantly affect cluster formation, requiring careful selection based on the expected structure of the delay behavior data.

**Self-Organizing Maps**
Self-Organizing Maps (SOMs) project high-dimensional data onto a lower-dimensional (typically 2D) grid of 'neurons' while preserving topological relationships (Nichanian et al., 2024). During training, each neuron learns to represent a prototype pattern, with similar patterns mapping to nearby neurons. The resulting 2D map visualizes the data's manifold, effectively clustering the data spatially.

SOMs are well-suited for structured data like TOBT-TSAT matrices because they capture nonlinear relationships while providing visual insight. Aviation applications include clustering flight approach performance profiles to differentiate optimal versus suboptimal operations (Nichanian et al., 2024). For departure delay behaviors, a SOM could map different patterns spatially; one region representing flights with stable times, another showing flights with last-minute large delays.

Implementation requires choosing the map size and training parameters, typically optimized via grid search using metrics like the silhouette coefficient or quantization error. Modern libraries such as MiniSom in Python facilitate SOM implementation for matrix-structured data points.

**Method Selection Considerations**
For TOBT-TSAT behavior clustering, several factors guide method selection:

- **Data characteristics:** The size and dimensionality of behavior matrices affect computational feasibility
- **Cluster shape assumptions:** K-means and basic GMMs assume convex clusters, while density-based and hierarchical methods handle arbitrary shapes
- **Outlier handling:** Density-based methods naturally identify anomalous flights, while K-means forces all points into clusters
- **Interpretability:** SOMs and hierarchical clustering provide visual representations of cluster relationships
- **Validation:** Internal metrics (silhouette, Davies-Bouldin) and external metrics (when labels are available) should guide parameter tuning and method selection

A practical approach may combine multiple methods; for example, using HDBSCAN to determine cluster count and identify outliers, followed by K-means for the final clustering to enhance interpretability. Alternatively, hierarchical clustering might identify the overall structure, with SOMs providing visual insights into the relationships between different delay behavior patterns.

## 5.3. Cluster Validation Without Ground Truth

Evaluating clustering quality without ground truth labels presents unique challenges in the context of TOBT-TSAT behavioral matrices. This section examines internal validation metrics and techniques appropriate for this domain.

**Internal Validation Metrics**
The Silhouette Coefficient measures how well each data point fits within its assigned cluster versus neighboring clusters. For each flight $i$, we compute $a(i)$ (average distance to other points in the same cluster) and $b(i)$ (average distance to points in the nearest different cluster). The silhouette value $(b(i) - a(i))/\max(a(i), b(i))$ ranges from -1 to 1, with values near 1 indicating well-separated and cohesive clusters. This metric helps optimize cluster count by computing average silhouette scores across different $k$ values. In aviation studies clustering flight performance, silhouette values around 0.34 have been considered 'moderately strong' clustering (Nichanian et al., 2024).

The Davies-Bouldin Index (DBI) evaluates the ratio of within-cluster scatter to between-cluster separation, with lower values indicating superior clustering. Similarly, the Dunn Index, the ratio of minimum inter-cluster distance to maximum intra-cluster diameter, rewards well-separated, compact clusters with higher values. These complementary metrics have effectively validated clustering solutions in transportation studies, confirming HDBSCAN's superior performance through lower DBI scores (Wang et al., 2025). The Calinski-Harabasz Index offers another perspective by comparing between-cluster to within-cluster dispersion, functioning similarly to an F-statistic.

**Stability and Robustness Analysis**
Robust clustering solutions should demonstrate stability under minor perturbations. We can assess this by applying clustering algorithms to bootstrapped data samples and measuring agreement between results using the Adjusted Rand Index (ARI) or Adjusted Mutual Information (AMI) (Taskesen, 2020). These metrics range from 0 (random agreement) to 1 (identical clustering), adjusting for chance.

Temporal stability can be examined by clustering flights from different time periods (e.g., first versus second half of the month) to verify consistent pattern identification. Parameter stability tests how sensitive results are to algorithm configurations; if K-means with different random initializations converges to similar partitions (high ARI between runs), this suggests the identified structure is inherent to the data rather than an algorithmic artifact. Similarly, density-based clustering can be validated by examining which flights consistently form isolated clusters across various parameter settings, as demonstrated in anomaly detection research (Li et al., 2015).

**Domain-Specific Validation Approaches**
Beyond numeric indices, operational context provides crucial validation. Each identified cluster should be profiled according to meaningful characteristics: average TOBT update counts, timing patterns, and magnitude distributions. These profiles must align with recognizable operational scenarios; for example, one cluster might represent flights with minimal early adjustments followed by a single delay near departure (potentially normal operations), while another might show cascading delays of increasing magnitude (suggesting disrupted turnarounds).

Cross-referencing clusters with external variables not used in clustering can further validate the discovered structure. Meaningful associations with airline, aircraft type, weather conditions, or operational outcomes would suggest the clustering has captured operationally relevant distinctions rather than statistical artifacts. While these correlations don't constitute ground truth, they help confirm that the clustering reflects genuine operational patterns.

**Predictive Performance Validation**
Since a primary goal of this classification is improving prediction, cluster quality can be evaluated through predictive utility. This approach involves:

1. Clustering flights using their TOBT-TSAT behavior matrices (unsupervised)
2. Training cluster-specific predictive models for targets like off-block delay or TSAT deviation
3. Comparing performance against a global model without cluster segmentation

If cluster-specific models demonstrate significantly improved accuracy, this validates both the clustering approach and its practical value. This methodology has shown success in other domains where regression models built for specific clusters outperformed global models in both accuracy and reliability (Moloko et al., 2023). The approach directly aligns with the objective of enhancing predictive performance by identifying distinct categories of departure uncertainty.

In practice, implementation will combine multiple validation strategies; using silhouette coefficients and Davies-Bouldin indices to optimize algorithmic parameters, assessing stability through resampling or temporal splits, profiling clusters for operational interpretation, and ultimately evaluating predictive performance improvements.

This multi-faceted validation framework ensures that identified patterns are statistically sound, operationally meaningful, and predictively useful for uncertainty estimation in departure management.

## 5.4. Operational Implications of Delay Profile Clustering

The clustering of delay profiles offers valuable insights from two complementary perspectives, each of which can inform and enhance airport departure management strategies.

First, historical data can be used to derive a set of delay behavior clusters, each representing a distinct pattern of TOBT evolution. These clusters can be interpreted as uncertainty profiles. For instance, there could emerge the following 4 clusters: stable flights, early minor adjustments, incremental delayers, big last-minute delays. Or any other number of clusters that capture the essential characteristics of flight delay behaviors.

In the operation, real-time monitoring of a flight's TOBT update pattern allows for its classification into the most likely cluster, for example, by measuring the distance to predefined cluster centroids. This real-time cluster assignment acts as a proxy for estimating uncertainty. For example, if a flight begins to exhibit characteristics similar to any cluster, such as multiple early TOBT changes, it can be flagged early as high-risk, enabling preemptive measures.

Second, these clusters can be analyzed to uncover common attributes among flights with similar delay patterns. Perhaps significant differences are observed between airline types (e.g., low-cost vs. full-service), but the clustering could also reveal more nuanced, operationally meaningful patterns. Such insights can help stakeholders understand which types of flights are prone to which types of delay behavior, potentially informing strategic decisions or targeted interventions.

Together, these two applications support a better understanding of TOBT uncertainty for any given flight, both historically and in real-time.

Integrating these uncertainty profiles into DMAN logic creates opportunities for more robust scheduling. Typically, DMAN optimizes sequences assuming nominal TOBTs. However, if a flight is identified as belonging to a high-risk cluster, DMAN can adapt its sequencing strategy, delaying the TSAT, building in buffers, or proactively reordering flights to mitigate the risk of a last-minute disruption. This aligns with robust optimization principles: incorporating uncertainty into planning generally leads to more stable throughput compared to deterministic scheduling. As Mori, 2019 emphasizes, accurate pushback time estimation is critical to an efficient throughput. This classification approach does not eliminate TOBT variability, but it quantifies and contextualizes it, providing the DMAN with a probabilistic understanding of likely disruptions.

Moreover, these cluster profiles can directly support human decision-making. For example, if an airline operations manager sees that a flight has been classified as high-risk, they may investigate the turnaround progress and address potential bottlenecks (e.g., catering or maintenance delays). Similarly, air traffic controllers or airport operations centers could use a dashboard that visualizes flight uncertainty based on cluster membership. This acts as an early warning system: rather than reacting after a TSAT is missed, stakeholders receive actionable signals in advance, enabling proactive TOBT updates and smarter sequencing adjustments.

In conclusion, clustering structured delay profiles creates a layer of meta-information that enhances both situational awareness and operational robustness. It enables DMAN to operate with a finer understanding of TOBT reliability, improving punctuality and reducing throughput losses due to late surprises. More broadly, this approach addresses a key gap identified in Collaborative Decision Making (CDM) literature: that milestone tracking alone is insufficient when data inputs are unreliable. By transforming historical delay patterns into predictive, actionable insights, clustering directly strengthens the CDM framework and supports more adaptive, data-driven airport operations.

# References

Air Traffic Control the Netherlands (LVNL). (2025, May). eAIP Netherlands – integrated aeronautical information package. ENR 1.6 ATS surveillance services [AIRAC Amendment 04/2025, effective 12 Jun 2025]. https://eaip.lvnl.nl/web/2025-05-29-AIRAC/html/index-en-GB.html

Alexopoulos, Y. (2024). A-CDM Manual Schiphol Airport. https://www.schiphol.nl/nl/download/b2b/1569488978/7ERl8iHeLELDtgFsnK0mGi.pdf

Anderson Jr., J. D. (1999). *Aircraft performance and design*. McGraw–Hill.

Asadi, E., Evler, J., Preis, H., & Fricke, H. (2020). Coping with Uncertainties in Predicting the Aircraft Turnaround Time at Airports. In *Operations Research Proceedings 2019* (pp. 773–780). Springer, Cham. https://doi.org/10.1007/978-3-030-48439-2_94

Atkin, J. A. D., Burke, E. K., Greenwood, J. S., & Reeson, D. (2009). An examination of take-off scheduling constraints at London Heathrow airport. *Public Transp*, *1*(3), 169–187. https://doi.org/10.1007/s12469-009-0011-z

Atkin, J. A. D., De Maere, G., Burke, E. K., & Greenwood, J. S. (2013). Addressing the Pushback Time Allocation Problem at Heathrow Airport. *Transportation Science*, *47*(4), 584–602. https://doi.org/10.1287/trsc.1120.0446

Aviation Solutions. (2023). Deep Turnaround. Retrieved June 12, 2025, from https://www.schiphol.nl/en/aviation-solutions/deep-turnaround/

Aviation Solutions. (2025). *TSAT Analysis Schiphol Airport* (tech. rep.). Schiphol Group.
Internal document.

Bikir, A., Idrissi, O., Mansouri, K., & Qbadou, M. (2024a). An Optimized Air Traffic Departure Sequence According to the Standard Instrument Departures. *IJACSA*, *15*(3). https://doi.org/10.14569/IJACSA.2024.01503133

Bikir, A., Idrissi, O., Mansouri, K., & Qbadou, M. (2024b). The Impact Of Aircraft Performance Variety In A Departure Sequence Prior to Takeoff Operations Following A Standard Instrument Departure Route. *2024 11th International Conference on Wireless Networks and Mobile Communications (WINCOM)*, 1–6. https://doi.org/10.1109/WINCOM62286.2024.10657711

Bubeck, S., & Cesa-Bianchi, N. (2012). Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems. *MAL*, *5*(1), 1–122. https://doi.org/10.1561/2200000024

Buchli, J., Huttner, D., Bosworth, C., & Stet, N. (2019, August). *Improving slot compliance: Addressing slot scarcity at Schiphol airport* (tech. rep. No. Version: 1.2). PA for Ministry of Infrastructure and Water Management. Retrieved May 19, 2025, from https://zoek.officielebekendmakingen.nl/blg-916200.pdf

Cesa-Bianchi, N., & Lugosi, G. (2006, March). *Prediction, Learning, and Games*. Cambridge University Press.

Derks, J. L. C. (2020). Coordinated Arrival and Departure Management. Retrieved May 15, 2025, from https://repository.tudelft.nl/record/uuid:fe455144-b0d1-459a-babc-34b847966d11

Dietterich, T. G. (2000). Ensemble Methods in Machine Learning. *Multiple Classifier Systems*, 1–15. https://doi.org/10.1007/3-540-45014-9_1

Dijkstra, F., Okina, A., & Jester, N. (2020, October). *RECAT-EU FOR DEPARTURES AT SCHIPHOL* (tech. rep. No. FWY-2020-01). Ferway & MovingDot. Retrieved May 14, 2025, from https://kdc-mainport.nl/wp-content/uploads/2021/04/FWY-2020-01-RECAT-EU-for-departures-v1.0.pdf

Elver, J., Schultz, M., Fricke, H., & Cook, A. (2024). Stochastic Delay Cost Functions to Estimate Delay Propagation Under Uncertainty. *ResearchGate*. https://doi.org/10.1109/ACCESS.2022.3152570

Eun, Y., Jeon, D., Kim, H., Jung, Y., Lee, H., Zhu, Z., & Hosagrahara, V. (2019). A Tactical Scheduler for Surface Metering under Minimum Departure Interval Restrictions. *2019 IEEE/AIAA 38th Digital Avionics Systems Conference (DASC)*, 1–10. https://doi.org/10.1109/DASC43569.2019.9081701

Eurocontrol. (2009). Airport CDM: Stepts to boost efficiency. http://www.euro-cdm.org/library/eurocontrol/airport_cdm_steps_to_efficiency.pdf

Frequentis. (2022). *Departure Manager* (Product Brief). Retrieved May 16, 2025, from https://www.frequentis.com/sites/default/files/support/2022-06/Frequentis_ATM_product-brief_DMAN.pdf

Freund, Y., & Schapire, R. E. (1997). A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/10.1006/jcss.1997.1504

'German A-CDM Harmonization Group'. (2018, April). Facts on Airport Collaborative Decision Making. Retrieved June 4, 2025, from https://www.munich-airport.com/_b/0000000000000007897042bb5dd3c114/flyer-cockpit-crews-en.pdf

Grewal, M., & Andrews, A. (2001, January). *Kalman filtering: Theory and practice using MATLAB*. Retrieved June 10, 2025, from https://www.researchgate.net/publication/243786641_Kalman_filtering_theory_and_practice_using_MATLAB

Guvercin, M., Ferhatosmanoglu, N., & Gedik, B. (n.d.). Forecasting Flight Delays Using Clustered Models Based on Airport Networks. *ResearchGate*. https://doi.org/10.1109/TITS.2020.2990960

Haykin, S. S. (2002). *Adaptive filter theory*. Pearson Education India.

Herrema, F., Curran, R., Visser, H., Huet, D., & Lacote, R. (2018). Taxi-Out Time Prediction Model at Charles de Gaulle Airport. *Journal of Aerospace Information Systems*, *15*(3), 120–130. https://doi.org/10.2514/1.I010502

Huet, D., Booth, D., & Pickup, S. (2016, March). *A-CDM Impact Assessment* (tech. rep.). EUROCONTROL. Retrieved May 16, 2025, from https://www.eurocontrol.int/sites/default/files/2019-04/a-cdm-impact-assessment-2016.pdf

IATA. (2020). *IATA Airport CDM Recommendations* (tech. rep.). International Air Transport Association. Retrieved May 19, 2025, from https://www.iata.org/contentassets/5c1a116a6120415f87f3dadfa38859d2/iata-acdm-recommendations-v1.pdf

Jester, N., van Ravenswaaij, C., The, J., van Dronkelaar, J.-H., & van Muynck, R. (2019). Managing of turnaround priorities. *MovingDot and NLR for KDC*. https://kdc-mainport.nl/wp-content/uploads/2019/09/Managing-of-Turnaround-Priorities-report-MD-NLR-v1.0.pdf

Julier, S. J., & LaViola, J. J. (2007). On Kalman Filtering With Nonlinear Equality Constraints. *IEEE Trans. Signal Process.*, *55*(6), 2774–2784. https://doi.org/10.1109/TSP.2007.893949

Karapetyan, D., Atkin, J. A. D., Parkes, A. J., & Castro-Gutierrez, J. (2017). Lessons from building an automated pre-departure sequencer for airports. *Ann Oper Res*, *252*(2), 435–453. https://doi.org/10.1007/s10479-015-1960-z

Krishnamurti, T. N., Kishtawal, C. M., LaRow, T. E., Bachiochi, D. R., Zhang, Z., Williford, C. E., Gadgil, S., & Surendran, S. (1999). Improved Weather and Seasonal Climate Forecasts from Multimodel Superensemble. *Science*, *285*(5433), 1548–1550. https://doi.org/10.1126/science.285.5433.1548

Li, L., Das, S., John Hansman, R., Palacios, R., & Srivastava, A. N. (2015). Analysis of Flight Data Using Clustering Techniques for Detecting Abnormal Operations. *Journal of Aerospace Information Systems*, *12*(9), 587–598. https://doi.org/10.2514/1.I010329

Marayat, A. (2008, May). Departure Metering. https://www.eurocontrol.int/archive_download/all/node/9726

Moloko, L., Bokov, P., Wu, X., & Ivanov, K. (2023, December). Clustering and Uncertainty Analysis to Improve the Machine Learning-based Predictions of SAFARI-1 Control Follower Assembly Axial Neutron Flux Profiles [Comment: 29 pages, 11 figures]. https://doi.org/10.48550/arXiv.2312.14193

Mori, R. (2019). Evaluation of Departure Pushback Time Assignment Considering Uncertainty Using Real Operational Data. https://www.sesarju.eu/sites/default/files/documents/sid/2019/papers/SIDs_2019_paper_21.pdf

Nichanian, A., Li, W.-C., Korek, W. T., Wang, Y., & Chan, W. T.-K. (2024). Self-organising Maps for Comparing Flying Performance Using Different Inceptors. In D. Harris & W.-C. Li (Eds.), *Engineering Psychology and Cognitive Ergonomics* (pp. 109–122, Vol. 14693). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-60731-8_8

Odoni, A. R. (2020). A review of certain aspects of the slot allocation process at level 3 airports under regulations 95/33. *MIT International Center for Air Transportation (ICAT)*, (ICAT-2020-09). https://dspace.mit.edu/bitstream/handle/1721.1/132655/ICAT-2020-09_Slot%20Allocation%20Process.pdf

Reynolds, D. (2009). Gaussian Mixture Models. In *Encyclopedia of Biometrics* (pp. 659–663). Springer, Boston, MA. https://doi.org/10.1007/978-0-387-73003-5_196

Romero, C., Ventura, S., Pechenizkiy, M., & d Baker, R. S. J. (2010, October). *Handbook of Educational Data Mining*. CRC Press.

Schäfer, J., & Alexopoulos, Y. (2024, October). How Deep Turnaround Enhances A-CDM Efficiency. Retrieved May 19, 2025, from https://www.schiphol.nl/nl/aviation-solutions/blog/how-deep-turnaround-enhances-a-cdm-efficiency/

Schiphol. (n.d.). Vliegroutes en baangebruik. Retrieved June 13, 2025, from https://www.schiphol.nl/nl/schiphol-als-buur/vliegroutes-en-baangebruik/

Schultz, M., Rosenow, J., & Olive, X. (2022). Data-driven airport management enabled by operational milestones derived from ADS-B messages. *Journal of Air Transport Management*, *99*, 102164. https://doi.org/10.1016/j.jairtraman.2021.102164

SESAR-JU. (2011). Basic DMAN Operational Service and Environment Definition (OSED). Retrieved May 16, 2025, from https://www.sesarju.eu/sites/default/files/documents/solution/Sol106%204%20DMAN%20baseline_OSED.pdf

SESAR-JU. (2012). *DMAN Baseline to be used for Integration of AMAN and DMAN* (Contextual Note). SESAR Joint Undertaking. Retrieved May 20, 2025, from https://www.sesarju.eu/sites/default/files/documents/solution/Sol106%201%20DMAN%20baseline_Contextual%20note.pdf
EXE-06.08.04-VP-470 Validation Exercise at Paris CDG.

Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press. https://doi.org/10.2307/j.ctv10vm1qb

Simon, D. (2006, June). *Optimal State Estimation: Kalman, H Infinity, and Nonlinear Approaches*. John Wiley & Sons.

Snijders, D. C. (2024). Investigating the effect on departure capacity of changing Target Off-Block Time uncertainty. Retrieved May 15, 2025, from https://repository.tudelft.nl/record/uuid:8ac0ba80-25b2-4ef1-b1bf-f0aa1b3d8e97

Sun, J., Hoekstra, J. M., & Ellerbroek, J. (2020). OpenAP: An Open-Source Aircraft Performance Model for Air Transportation Studies and Simulations. *Aerospace*, *7*(8), 104. https://doi.org/10.3390/aerospace7080104

Taskesen, E. (2020, September). Clusteval is a python package for unsupervised cluster validation. https://erdogant.github.io/clusteval

Technology, H. K. H. (2024, October). Airport Collaborative Decision Making: The Complete Handbook. Retrieved June 4, 2025, from https://www.haisenglobal.com/blog/airport-collaborative-decision-making-complete-handbook

van Baren, G. (2016, August). *Benefits analysis of RECAT-EU for Schiphol Aiport* (tech. rep. No. NLR-CR-2016-131). NLR. Retrieved May 14, 2025, from https://kdc-mainport.nl/wp-content/uploads/2016/12/NLR-CR-2016-131.pdf

Wang, S.-M., Yang, W.-R., Zhuang, Q.-Y., Lin, W.-H., Tian, M.-Y., Su, T.-J., & Cheng, J.-C. (2025). Application of Three-Dimensional Hierarchical Density-Based Spatial Clustering of Applications with Noise in Ship Automatic Identification System Trajectory-Cluster Analysis. *Applied Sciences*, *15*(5), 2621. https://doi.org/10.3390/app15052621

Waterhouse, S., MacKay, D., & Robinson, A. (1995). Bayesian Methods for Mixtures of Experts. *Advances in Neural Information Processing Systems*, *8*. Retrieved June 11, 2025, from https://papers.nips.cc/paper_files/paper/1995/hash/9da187a7a191431db943a9a5a6fec6f4-Abstract.html

Yiannis Alexopoulos, van Leeuwen, E., Zwaaf, D., & Kelder, h. (2019, September). Schiphol Airport CDM Operations Manual.

Zhang, J., Xiu, J., He, Y., & Xiong, W. (2006). Distributed interacted multisensor joint probabilistic data association algorithm based on D-S theory. *SCI CHINA SER F*, *49*(2), 219–227. https://doi.org/10.1007/s11432-006-0219-3

**J.P.H. Bremer**