Stochastic Optimisation of Tail Assignment and Maintenance Task Scheduling with Health-Aware Models

Melissa Oremans

Delft University of Technology

TUDelft

Stochastic Optimisation of Tail Assignment and Maintenance Task Scheduling with Health-Aware Models

by

Melissa Oremans

to obtain the degree of Master of Science
at Delft University of Technology,
to be defended publicly on 2 August 2024

Date and version document:	July 19, 2024	final version
Student number:	4558391	
Thesis supervisors:	Dr. M. Arias Chao	ZHAW / TU Delft
	Dr. B.F. Lopes Dos Santos	TU Delft (until April 2024)
	Dr. M.J. Ribeiro	TU Delft (from April 2024)
Further members graduation committee:	Dr. J. Sun	TU Delft
	Dr. V. Yaghoubi Nasrabadi	TU Delft

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

Acknowledgements

I lift up my eyes to the mountains [Swiss Alps] - where does my help come from?

My help comes from the Lord, the Maker of heaven and earth.

He will not let your foot slip— He who watches over you will not slumber;

indeed, He who watches over Israel will neither slumber nor sleep.

The Lord watches over you— the Lord is your shade at your right hand;

the sun will not harm you by day, nor the moon by night.

The Lord will keep you from all harm— He will watch over your life;

the Lord will watch over your coming and going both now and forevermore.

Psalm 121, the Bible (NIV translation)

Contents

Part I

Scientific Paper

List of Figures

List of Tables

List of Abbreviations

4M Material, Machinery, Method and Manpower
AOG Aircraft On Ground
ARP Aircraft Recovery Problem
CBM Condition-Based Maintenance
CI Confidence Interval
CM Corrective Maintenance
DES Discrete Event Simulation
EOL End Of Life
FH Flight Hours
GPR Gaussian Process Regression
KPI Key Performance Indicator
MC Monte Carlo
MI(L)P Mixed-Integer (Linear) Programming
MRO Maintenance, Repair and Overhaul
MS Maintenance Scheduling
NN Neural Network
PDF Probability Density Function
PM Predictive Maintenance
RUL Remaining Useful Life
SP Stochastic Programming, 2SP: Two-Stage Stochastic Programming
TA Tail Assignment
TAT Turn Around Time
UQ Uncertainty Quantification
WO Work Order

Stochastic Optimisation of Tail Assignment and Maintenance Task Scheduling with Health-Aware Models

Melissa Oremans
Delft University of Technology, Delft, The Netherlands

Abstract

Efficient maintenance management requires an integrated approach that balances downtime (maintenance scheduling, MS) and uptime (tail assignment, TA). Current methods often use sequential decision-making, which neglects the interdependencies between MS and TA, resulting in sub optimal outcomes. Recent research has started to integrate MS and TA modelling, and separately also to integrate MS with predictive maintenance (PM) strategies. These trends highlight the need for a comprehensive approach that optimises these interdependencies and considers the uncertainties of prognostic models. This research presents a new scheduling framework for optimal maintenance management that integrates PM, MS and TA. Historical sensor data is used with Gaussian Process Regression to predict the end-of-life of aircraft components and quantify prediction uncertainties. These uncertainties are incorporated into the decision-making process using Monte Carlo simulations. The resulting model assigns aircraft and tasks to maintenance slots and aircraft to flight legs under various predictor scenarios, enabling a more adaptive scheduling strategy. A case study with Swiss International Air Lines over a three-day period for five aircraft demonstrates the model's effectiveness, showing improvements in operational efficiency, reduced costs, fewer cancellations, and higher fleet availability. The model's predictions and schedules were validated by expert planners, suggesting that a holistic, adaptive scheduling approach has significant potential for operational improvements in the airline industry.

1. Introduction

In the the highly competitive aviation sector, even minor enhancements in operational efficiency can result in substantial competitive benefits. Maintenance, which represents approximately 11% of total costs [4], is a key area for strategic cost optimisation. Efficient management of maintenance operations is crucial not only for cost reduction but also for sustaining operational availability, fulfilling passenger demand, responding to market changes, and maintaining safety standards.

Successful maintenance management demands an integrated approach to balancing downtime, i.e. maintenance scheduling (MS), and uptime, i.e. tail assignment (TA) for airline operations. Current methods for maintenance optimisation often rely on sequential decision-making, which overlooks the interdependencies between MS and TA, resulting in a sub optimal balance. Therefore, state-of-the-art research works, such as Lagos et al. [5], Iwata et al. [6] or Varenna et al. [7], have started to integrate MS and TA modelling. The complexity involved is highlighted by the frequent choice of heuristic, instead of exact, algorithms even when solely solving for aircraft network operations [8] or solely for aircraft maintenance scheduling [9].

Another emerging trend in maintenance optimisation is the inclusion of predictive maintenance strategies within the traditional framework of preventive and corrective maintenance. Predictive maintenance strategies resort to models to asses the current condition of aircraft components (health indicator models) and forecast their future states (prognostics models), aiming to prevent unexpected failures and optimise maintenance in an operational environment. However, a significant challenge remains in integrating these predictive strategies with conventional practices. Multiple studies, notably those by Tseremoglou et al. [10] and Lee and Mitici [11], have explored the impact of probabilistic prognostic models on MS, but have not extended their analysis to the influence on and interplay with TA.

To the best of the authors' knowledge, no research has yet integrated TA and MS with predictive tasks. Therefore, there is a need for a comprehensive approach that (1) both integrates and optimises the interdependency between MS and TA, and (2) accounts for the inherent uncertainties of prognostic models.

This paper introduces a novel scheduling framework for optimal maintenance management incorporating predictive strategies, that simultaneously solves TA to account for the interdependency, namely aircraft availability and usage. The proposed framework is depicted in Figure 1. The ground time here refers to the aircraft being physically on the ground, either for maintenance, turn-around-time (TAT) or parking, which is a key parameter for both MS and TA. Using historical sensor data, Gaussian Process Regression is applied to predict the end-of-life (EOL) of aircraft components and to quantify prediction uncertainty. This uncertainty in the decision-making process is included through a Monte Carlo sampling approach. The resulting model assigns aircraft and tasks to maintenance slots and aircraft to flight legs under different predictor scenarios, which shows which tail-flight and slot-tail combinations are more robust than others, allowing for a more adaptive scheduling strategy.

A case study with Swiss International Air Lines is performed which analyses the fleet schedule over a three-day time horizon for five aircraft. The results show improved operational efficiency by lower costs, fewer cancellations, and higher fleet availability. Moreover, Monte Carlo solutions show which tail-flight or tail-maintenance slot combinations are robust to to uncertainties in EOL predictions. The model's predictions and schedules are validated by expert planners, showing that the proposed holistic, adaptive scheduling approach holds great potential for operational improvements in airlines.

The structure of this paper is as follows: the problem context and key assumptions are presented in section 2, followed by related work in section 3. Next, the methodology of handling the complexity of all model components is discussed in section 4. The case study set-up is elaborated on in section 5, followed by the results in section 6. Section 7 outlines the discussions and recommendations for future work. Finally, the conclusions are presented in section 8.

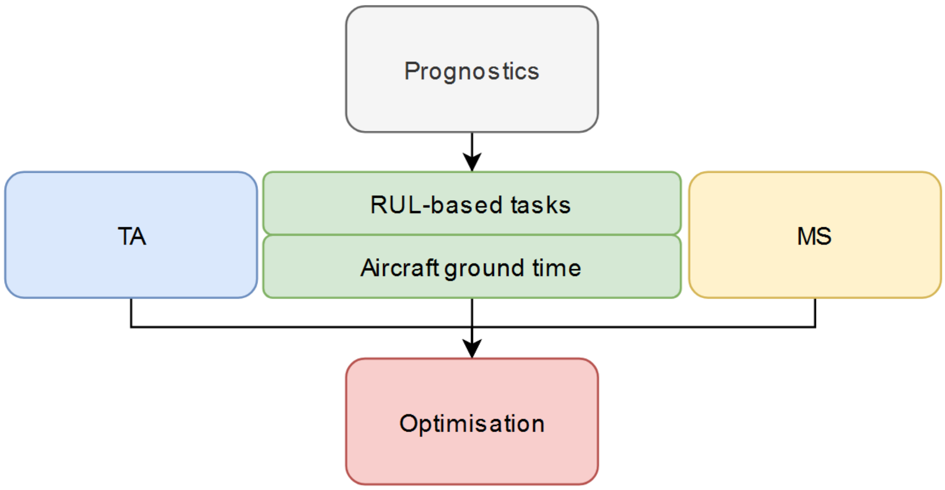

Figure 1: Modular structure of the proposed integrated scheduling framework for TA, MS and RUL-based predictive maintenance.

2. Problem Context

In this section, the problem of maintenance scheduling with tail assignment constraints under a hybrid maintenance strategy comprising reactive, preventive, and predictive interventions is introduced (subsection 2.1). For clarity, the assumptions and conventions upon which the proposed methodology is based are stated in subsection 2.2.

2.1. Problem description

To reduce costs and enhance time efficiency, airline operation and maintenance management are constantly innovating. This complex and dynamic field involves various agents, including crew, staff, facilities, and aircraft that all need to be planned for on a strategic, tactical and operational level [12]. This study focuses on the latter, namely fleet operational planning. The two key processes that create the short-term aircraft schedules are tail assignment (TA) and maintenance task scheduling (MS).

Tail assignment (TA) is the assignment of a specific aircraft (tail) to a flight leg. This step comes after fleet assignment, which means that the flight legs in the schedule are already assigned to subtypes, but not yet to specific aircraft registrations. Key drivers for optimal tail assignment are the fuel efficiency of an aircraft subtype and the amount of seats available versus the passenger forecast on a flight leg [13]. When stochastic disruptions are considered, TA is also referred to as the Aircraft Recovery Problem (ARP).

Maintenance scheduling (MS) in this paper refers to maintenance task allocation, which allocates the open tasks for an aircraft to an available maintenance slot while satisfying all requirements. These requirements can be summarised by the 4M concept, namely the availability of material, machinery, method and manpower. It is common practice for airlines to contract out some maintenance operations to maintenance-repair-overhaul (MRO) companies, reducing the airline's influence in maintenance scheduling. In this study, it is assumed that the airline keeps complete control.

A combination of preventive, corrective and predictive maintenance tasks need to be scheduled. Preventive tasks are routinely scheduled maintenance tasks, that go due based on calendar days, flight hours or flight cycles. Corrective tasks need to be performed when failure occurs and are thus not known in advance.

Predictive maintenance (PM) tasks result from prognostics that predict when failure will occur based on historical sensor data. This means that a corrective measure, which is more expensive, can be prevented and instead the maintenance can be scheduled as if it was a preventive measure.

Traditionally, MS and TA are scheduled sequentially (without PM), ignoring the interdependency between these processes and leading to sub optimal results. The interdependency is seen in aircraft availability and usage. An aircraft can either be scheduled to a maintenance slot, to a flight (including the required TAT and other regulatory time on the ground) or it sits idle (parked). Both the MS and TA do not have the full fleet available at all times, and each schedule depends on the other. In addition, an aircraft that is used more frequent or in operating conditions that degrade a component more, will require more maintenance hours. Thus, the decisions made by TA influence the MS requirements.

Therefore, the need to evaluate the effectiveness of integrated aircraft scheduling arises in both the airline industry and academic work. This requires a schedule optimisation model of tail assignment and maintenance task scheduling, including predictive maintenance with its inherent uncertainty, for commercial airline operations. Therefore, this study proposes a proof-of-concept version of a decision support tool for use by airline short-term planners. Consequently, the model must consistently provide feasible solutions and strive to complete computations in the order of minutes.

Finally, it is good to be aware of the definitions of the following words in this paper. A 'slot' always refers to a maintenance slot, not airport (network) slot. Predictors describe the component, or system, that are included in the prognostic methods and thus in predictive maintenance. Prognostics in this work refers only to predicting the RUL; fault detection and fault diagnostics are outside the scope of this study.

2.2. Assumptions

The problem formulation described above implies certain assumptions about the availability of the information upon which the proposed framework is built. For clarity, these assumptions are explicitly listed below.

Aircraft-based short-term scheduling models, such as the one presented in this study, means that passenger flows and staff allocations are not included. Also, it focuses on short-term scheduling, namely up to three days from day of operations, as is done in practice by commercial airlines. A hub-and-spoke airline model is considered with two hubs. No disruptions are included, neither from network operations (like flight delays) nor from maintenance (for example delays or unknown failures leading to grounding of aircraft). It should be noted that the model is formulated such that these can be included in future development, as explained in section 4.

The main costs for TA, fuel efficiency and seats, are summarised in one parameter, the operating costs, which is specified for every tail/flight combination.

Regarding MS, maintenance can be performed both during the day and at night, meaning that the available slots are defined over 24 hours. It is also assumed that no slots are pre-assigned, which would be the case in actual practice, for example for letter checks. In addition, all maintenance is performed by the airline itself.

The assumptions on the issue and due dates of tasks are as follows. Firstly, this model simplifies the due dates of preventive tasks by converting the deadline of all tasks to calendar days. This is because of incomplete historical input data. Secondly, due to the fact that this model does not include operational disruptions and only includes three days of operations, it is assumed that also corrective tasks are known at the start of the time horizon. Thirdly, the model's input includes the maintenance tasks of predictors as if they were preventive measures, however they are set at non-active. Only when the RUL reaches its threshold these tasks are activated.

Furthermore, only time-based constraints are considered, namely the required ground-time, required amount of man hours and the maximum amount of tasks that can be done in parallel. It is assumed there are always enough facilities, machines and material at hand if there is an available slot. Maintenance tasks in this model can therefore be line maintenance or hanger maintenance in actual practice, but for this model this distinction is irrelevant. Additional limitations, like staff certifications and the clustering of tasks such that a panel should only be opened once, are also not considered. Again, note that the model formulation allows for these additional constraints to easily be added.

3. Related Work

This research touches on four different research directions that have not been solved together before, as explained in the problem context. The following sections will discuss relevant literature for each of these research directions: TA, MS, combined TA and MS scheduling, and PM with uncertainty quantification.

3.1. Tail assignment

TA and ARP models are often modelled as MILP formulations [14], because of its versatile range in variables and constraints and well-developed solvers. Over 80% of recent ARP models rely on heuristic algorithms rather than exact methods, highlighting their complexity [8]. The challenge lies in finding solutions that accurately model a comprehensive scenario, considering recovery actions, a large fleet size, and operational constraints such as crew limitations and passenger itineraries, which all come with a high computational cost.

Therefore, most models choose to consider only a selection of constraints. Deterministic models include the MILP-formulated framework by Barnhart et al. [15], considering sequences of flight legs, which laid the foundation for future TA models. Typically, aircraft availability is limited by taking into account fixed maintenance times, such as Sarac et al. [16]'s model which schedules daily aircraft routings according to set night stops where maintenance is done.

Vink et al.'s state-of-the-art ARP framework effectively reduces computational time to under one minute [1] while still including disruptions. Their model incorporates aircraft maintenance times, passenger itineraries, and indirectly, crew limitations. The integer-linear-program (ILP) problem was formulated using parallel time-space networks, which will be further explained in section 4. It is dynamically solved by limiting the focus to one subset of the aircraft fleet at a time, mirroring current manual airline scheduling practices. Even though a fixed maintenance schedule is modelled as a constraint, the interaction between TA and MS schedules is not accounted for.

3.2. Maintenance task scheduling

Airline maintenance scheduling is addressed in two directions of research, namely task scheduling, which is the assignment of tasks and aircraft to slots, and resource allocation, which considers resource (typically 4M) constraints for each task.

Recent works for resource allocation include Witteman et al. [17] who developed a bin packing algorithm that examines workforce availability and skill-specific requirements for preventive tasks. Qin et al. [18] created a MILP-formulated model focusing on hangar and parking spot capacity and lay-out plans.

The study in 2022 by Van Kessel et al. [19] is the first to address maintenance disruptions in the specific setting of airline operations and to combine different maintenance 4M requirements. They created a MILP model taking into account a stochastic disruption process for the timing and frequency of tasks. The model also includes constraints on resource and manpower availability. Their state-of-the-art framework for MS does not cover any interaction with TA however.

The next subsection discusses the literature of interest for task scheduling, namely the ones that take into account (up to some extent) the interplay with TA.

3.3. Combined tail assignment and maintenance task scheduling

An overview of studies that have integrated MS and TA, to varying extents is shown in Table 1. Only two published studies and one master thesis (Varenna [7]) have formulated a holistic TA and MS optimisation model. The other works mentioned in the table are added for completeness since they include the interplay between TA and MS, however, their models are build from a MS perspective.

Reference	TA / MS based	Application	Methodology	TA considerations	MS considerations	Components introducing uncertainty
Dufuaa et al. 1999 [20]	MS	Commercial airline operations; short-haul / long-haul fleet; time horizon not specified	Qualitative analysis; conceptual modelling	Flight schedule; Flight cancellations; Flight delays	Minimum repair time; Spare parts inventory; Station availability; Staff requirements	-
Sachon et al. 1999 [21]	MS	Commercial airline operations: 1 system analysis (leading edge slats)	Probabilistic risk analysis model	Flight schedule; Flight cancellations; Flight delays	Parts availability; Staff requirements and availability; In-flight safety	All are stochastic variables except: actual maintenance time, task deferral allowed or not, flight delay allowed or not
Öhman et al. 2020 [22]	MS	Commercial airline operations: long-haul fleet	Discrete event simulation; No optimisation; Uncertainty not stochastically modelled	Aircraft arrival time, Aircraft departure time, Aircraft location	Frontlog time buffer	Random variable for: maintenance tasks (volume)
Iwata et al. 2013 [6]	TA & MS	Military aircraft operations: supply chain spare parts	Discrete event simulation, No optimisation; Uncertainty not stochastically modelled	Flight schedule (mission type and length)	Repair time; Parts inventory	Random variables for: mission length, part life time, replacement time, repair time, transit times (spare parts)
Lagos et al. 2020 [5]	TA & MS	Commercial airline operations: short-haul fleet	MIP formulation; Markov design process with approximate dynamic programming; Monte Carlo simulation for expected value	Line of flights schedule; Cancellations	Staff requirements; Space capacity; Repair time; Parts availability	Stochastic process for: task arrival, task criticality, task due date
Varenna 2023 [7]	TA & MS	Commercial airline operations: long-haul fleet	MILP formulation; Discrete event simulation, Optimisation: simplex and barrier solver (Gurobi)	Rotation-based flight schedule; Flight delays; Flight cancellations; Reserve aircraft	Staff availability; Repair time;	Stochastic process for: flight delay; AOG occurence
Proposed model in this study	*TA & MS*	*Commercial airline operations: short-haul fleet*	*MIP formulation; Monte Carlo simulation, Optimisation: simplex and barrier solver (Gurobi)*	*Flight schedule*	*Staff availability; Repair time; Flexible slot schedule; Predictive tasks*	*Stochastic optimisation for: predictor scenarios*

Table 1: Literature overview of aircraft operations models that combine TA and MS. The column 'TA / MS based' indicates what is main focus of the model formulation. In case this is MS, still TA objectives and constraints are considered to a larger extent than conventional, but the model is not build with the idea to optimise both simultaneously.

Firstly, Iwata et al. [6] modelled a military aviation operation using discrete event simulation (DES). Their focus on military aviation operations, where flight schedules represent missions (type and length), differs significantly from commercial airlines' needs to schedule flights according to a fixed flight schedule. Also, not all solution methods are clearly explained in this paper. Secondly, Lagos et al. [5] developed a Markov decision process (MDP) scheduling formulation for the short-haul fleet of a commercial airline, using approximate dynamic programming (ADP) to solve this high-dimensional problem. The flight schedule is simplified by scheduling line of flights and corrective maintenance tasks are included. Similarly, Varenna [7] modelled a holistic TA and MS model using DES, but with a focus on the long-haul fleet and flight schedule disruptions. All these frameworks, however, do not include predictive maintenance or stochastic optimisation.

3.4. Predictive maintenance and uncertainty quantification

In predictive maintenance, tasks are initiated based on models forecasting the future health status, and potential time of failure of the system under consideration. These predictions inherently come with uncertainty. The leading uncertainty quantification (UQ) models for data-driven predictive maintenance include [23]: Gaussian process regression (GPR), Bayesian neural networks (BNNs), neural network (NN) ensembles, and deterministic approaches like spectral-normalized Gaussian process (SNGP) and deep neural network (DNN) GPR. While GPR ranks highly in terms of quality, it is associated with high computational demands and the curse of dimensionality. In contrast, BNNs, NN ensembles, and deterministic methods are highly scalable and provide medium to high-quality solutions [23].

Predictive maintenance in airline planning is a relatively new research topic. De Pater and Mitici [24] effectively implemented an aircraft maintenance planning model that adds predictive maintenance to traditional corrective and preventive approaches. Predictive tasks are included by the Remaining-Useful-Life (RUL) and reliability of multi-component, repairable systems. The stock of spare components and the maintenance slot schedule are considered, but operational constraints are not included in this work. Next, Lee and Mitici [11] included uncertainty quantification in a comparable problem context. Using a convolutional neural network with Monte Carlo dropout, an estimated distribution of the RUL is generated. This distribution is used in the maintenance schedule by applying deep reinforcement learning (DRL). Tseremoglou and Santos [10] developed an airline maintenance scheduling model that also includes RUL predictions. The framework is solved in two stages using POMDP (at component level) and DLR algorithms (at fleet level). Although more capacity and availability constraints are included, the interplay with operations is not considered.

4. Methodology

To address the challenge of MS with TA constraints under a hybrid maintenance strategy comprising reactive, preventive, and predictive interventions, a novel modular framework composed of three main components is proposed, which is visually shown in Figure 2. The sections below describe each module in detail. First, subsection 4.1 explains the middle row, namely the mixed-integer program (MIP) formulation of TA, MS and PM. Next, subsection 4.2 elaborates on using Gaussian Process Regression (GPR) as a prognostic method to define the input for the RUL-constraints. Finally, Monte Carlo simulation for uncertainty analysis in decision-making is described in subsection 4.3.

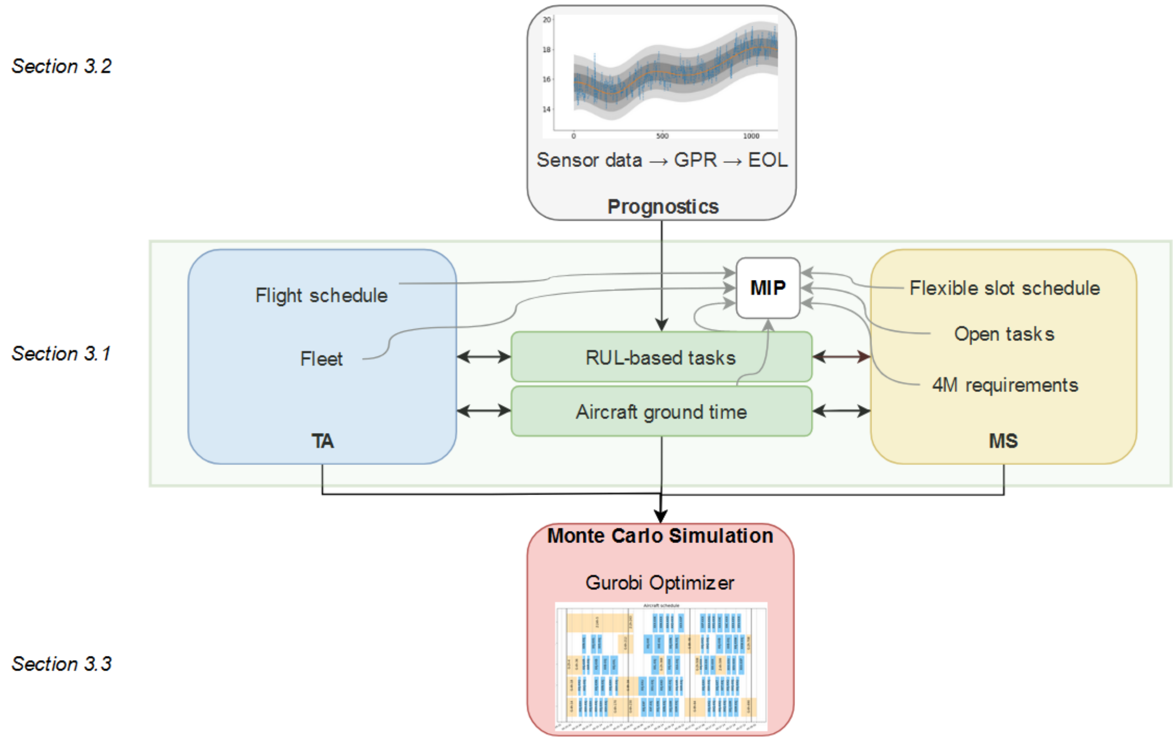

Figure 2: Detailed overview of the modular framework. The MIP formulated model, shown in the middle row, integrates TA, MS, RUL-based tasks and aircraft ground time objectives and constraints. Gaussian Process Regression is used on the sensor data of the predictors and Monte Carlo simulations are used to include the predictor's uncertainty.

The input for this model consists of the flight schedule per aircraft subtype, a list of maintenance tasks per aircraft, all available maintenance slots per aircraft subtype, prognostics for the predictors and the specifications and constraints on aircraft tails, maintenance tasks and slots. All times, so for example departure and arrival time in the flight schedule or the maintenance task duration, are rounded to the nearest discrete time step. The model input is elaborated on in the case study, section 5.

4.1. MIP formulated holistic MS and TA model

The core of model is the MIP formulation, which can be optimised for only MS, only TA or for both at the same time. Also the predictive maintenance tasks can be turned on or off. This is beneficial for benchmark comparisons and validation with historical airline data, since there MS and TA optimisations are solved separate from each other.

Mixed-integer linear programs (MILP) or MIP programs are often used for decision-making problems, because they can handle both continuous and binary variables and a wide variety of constraints. Also, well-developed and widely used solvers are available for this formulation, which allows the focus on the problem formulation and not the solving algorithm.[1]

Before going into the mathematical formulation, it is good to understand the underlying structure. The framework used is a parallel time-space network as previously developed by Vink et al. [1] for the aircraft recovery problem (ARP). Please refer to section 3 for a description of their work. A network of nodes (airports) versus discrete, homogeneous time steps is defined for each aircraft in the fleet. They are connected by flight arcs and ground arcs. The greatest benefit of a separate network per aircraft is that tail specific constraints can be taken into account, both for TA and MS. Each network contains all other flight arcs as well to allow different aircraft to cover different flights, granted that the subtype requirements are fulfilled. The disadvantage of this method is that it requires a large number of variables, which slows down the optimisation.

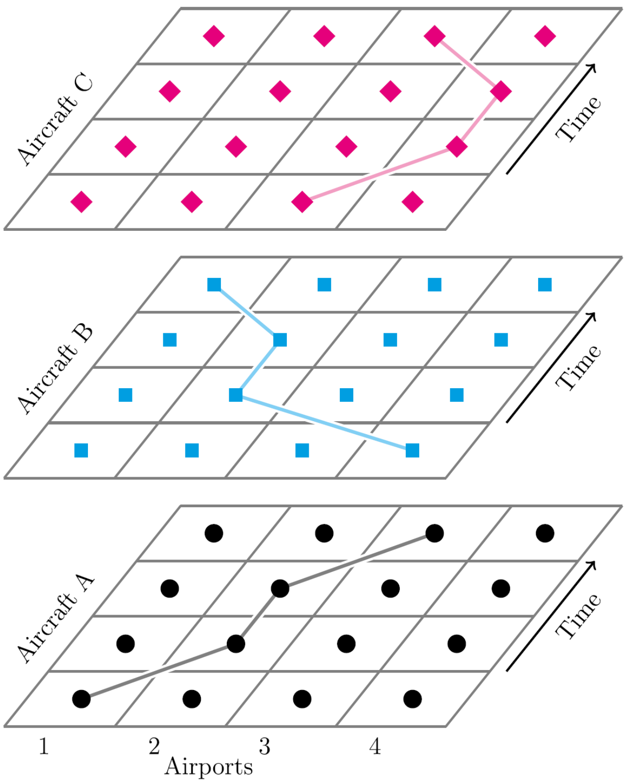

Figure 3: The concept of parallel time-space networks as illustrated by Vink et al [1]. Each aircraft network contains all eligible flight arcs and ground arcs, of which only the activated ones are shown here. Only three aircraft are shown, however any n number of aircraft can be used in the fleet.

[1]The model is solved using Gurobi software [25] which is very efficient in solving high-dimensional and non-linear MIP problems by its advanced branch-and-cut-algorithm. Gurobi allows for multiple objective functions to be solved either as a linear combination or hierarchically. Since the goal is to optimise TA and MS as one integrated problem, Gurobi is set to solve the objectives as a linear combination with equal weights for both objective functions. The optimisation gap is set to 0.8%, allowing for near-optimal results without excessive computation.

Next, the mathematical formulation is stated below. The formulation of the stand-alone MS and stand-alone TA frameworks are derived from state-of-the-art MILP formulated works, namely Van Kessel et al. [19] for MS and Vink et al. [1] for TA. Both Vink et al. and Van Kessel et al. allow for rescheduling and disruptions in their respective schedules, which is the daily reality for airlines and fits the future vision of this project. However, for the current scope they are not activated, because the focus is to formulate a model that allocates for TA, MS, PM and its uncertainty all together.

Compared to the MILP formulations of the two works mentioned before, several new formulations are introduced in this paper. Firstly, the connection between TA and MS, which is seen in both the constraints and the adaption of the objective functions. Additionally, the the RUL-based maintenance tasks and flexible slot schedule are introduced. The RUL-constraint is recursive, breaking the linearity of the model constraints.

Below the model's two objective functions and 17 constraints are presented in four parts: TA, MS, connection between TA and MS and the RUL-based constraints. They all use the same variables, sets and parameters.

Sets and subsets		
P		Aircraft
F		Flights
F_{arr}	$\subseteq F$	Arriving flights
F_{dep}	$\subseteq F$	Departing flights
$F_{bid}(n,p)$	$\subseteq F$	Forbidden aircraft - airport combination
T		Time blocks within the current schedule window
N		Arc nodes in time-space network (airports)
$T(n,p)$	$\subset N$	Arcs terminating at node n (arrival) for aircraft p
$O(n,p)$	$\subset N$	Arcs originating at node n (departure) for aircraft p
S		Maintenance slots
$S_{overlap}$	$\subseteq S$	Maintenance slots
S_{pos}	$\subseteq S$	Maintenance slots that fulfil the requirements to accommodate predictive task t on aircraft p
S_{Fict}	$\subseteq S$	Fictitious maintenance slot that starts after the considered time window
G		Maintenance task (or task group)
G_p	$\subseteq G$	All tasks for aircraft p
G_{due}	$\subseteq G$	Tasks that go due during the time window selected
G_{prev}	$\subseteq G$	Preventive tasks
G_{corr}	$\subset G$	Corrective tasks
G_{defer}	$\subset G$	Tasks that are deferred
PR		Predictors

Decision Variables		
$F_{p,f}$	$\in \{0,1\}$	1 if aircraft p assigned to flight f, 0 otherwise
$G_{p,n,t}$	$\in \{0,1\}$	1 if aircraft p assigned to ground arc n at time t, 0 otherwise
C_f	$\in \{0,1\}$	1 if flight f is cancelled, 0 otherwise
$M_{p,s}$	$\in \{0,1\}$	1 if aircraft p assigned to maintenance slot s, 0 otherwise
$T_{g,s}$	$\in \{0,1\}$	1 if task group g assigned to maintenance slot s, 0 otherwise
$Mpr_{p,t}$	$\in \{0,1\}$	1 if predictor pr on aircraft p assigned to maintenance at time t, 0 otherwise
$Clean_p$	$\in \{0,1\}$	1 if aircraft p has no open tasks in the coming n days after the end of the schedule window, 0 otherwise
$Slack_{p,s}$	$\in \{0,1\}$	1 if aircraft p assigned to slot s, 0 otherwise
$RUL_{pr_{p,t}}$		RUL value (integer) for predictor pr on aircraft p at time t

Parameters (all unitless unless stated otherwise)	
$OC_{p,f}$	Operating costs for aircraft p assigned to flight f
CC_f	Cancellation cost for flight f
GC_n	Ground time cost for ground arc n
W_{DUE}	Costs for task going due before it is assigned
W_{GROUND}	Ground time cost for maintenance slot
$W_{INT_{prev}}$	Preventive task interval costs
$W_{INT_{corr}}$	Corrective task interval costs
W_{DEFER}	Costs to defer a task
W_{CLEAN}	Costs for open tasks after time window
W_{SEQ}	Costs to non-sequential slot selection
$B_{n,p}$	Flow balance at node n for aircraft p. 0 at intermediate nodes, 1 at start node
TAT_p	Minimum turn-around-time for aircraft p
t_{arr}	Arrival time
f_{dur_f}	Flight duration (en route time) for flight f
C_g	Criticality coefficient for task g
$DD_{g,s}$	1 if the start date of slot s is before the due date of task g, 0 otherwise
$GndTime_{g,s}$	1 if the task duration is smaller than slot duration, 0 otherwise
$MaxTasks_s$	Maximum number of tasks allowed in slot s
WH_g	Amount of workforce hours required for task g (h)
$MaxWH_s$	Available workforce hours in slot s (h)
ndd_g	1 if task group g goes due in n days after the scheduling window, 0 otherwise
$RUL_{pr_p}^{TH}$	Threshold RUL value for predictor pr on aircraft p
$RUL_{pr_p}^{MAX}$	Maximum RUL value for predictor pr on aircraft p
s_{dur_s}	Maintenance slot duration for slot s

Tail assignment:

$$\begin{aligned}Min: &\sum_{p\in P}\sum_{f\in F} OC_{p,f}\cdot F_{p,f}\\ &+\sum_{f\in F} CC_f\cdot C_f\\ &+\sum_{p\in P}\sum_{n\in N}\sum_{t\in T} GC_n\cdot G_{p,n,t}\end{aligned} \tag{1}$$

$$\sum_{p\in P} F_{p,f}+C_f=1 \quad \forall f\in F \tag{2}$$

$$\sum_{n\in N\cap O(n,p)} G_{p,n,t}-\sum_{n\in N\cap T(n,p)} G_{p,n,t}+\sum_{f\in F\cap O(n,p)} F_{p,f}-\sum_{f\in F\cap T(n,p)} F_{p,f}=B_{n,p} \quad \forall n\in N \quad \forall p\in P \tag{3}$$

$$\text{if } F_{p,f}=1\rightarrow \sum_{t_{arr}}^{t_{arr}+TAT_p} G_{p,n,t}=TAT_p \quad \forall p\in P,\forall f\in F_{arr_t} \tag{4}$$

$$\sum_{\text{Fbid }(n,p)} G_{p,n,t}=0 \quad \forall t\in T \tag{5}$$

The first term in the objective function of TA makes sure that the tail-flight combination with the minimal operational costs is selected. The second term are the costs for cancelling a flight and the third term minimises the ground time costs.

Constraint (3) is the flow balance at each node in the network, constraint (4) ensures a minimum turn-around-time after every flight and constraint (5) ensures that aircraft cannot be assigned to flights going to forbidden airports, namely aircraft subtype - airport combinations that are non-compatible due to runway length and noise limitations.

Maintenance task scheduling:

$$\begin{aligned}Min: &\sum_{g\in G_{due}} T_{g,s=Fict}\cdot W_{DUE}\cdot C_g\\ &+\sum_{s\in S} M_{p,s}\cdot s_{dur_s}\cdot W_{GROUND}\\ &+\sum_{s\in S}\Big(\sum_{g\in G_{prev}} T_{g,s}\cdot W_{INTprev}+\sum_{g\in G_{corr}} T_{g,s}\cdot W_{INTcorr}\Big)\cdot C_g\\ &+\sum_{g\in G_{defer}} T_{g,s=Fict}\cdot W_{DEFER}\cdot C_g\\ &+\sum_{p\in P}(1-Clean_p)\cdot W_{CLEAN}\\ &+\sum_{p\in P}\sum_{s\in S} Slack_{p,s}\cdot W_{SEQ}\end{aligned} \tag{6}$$

Task - slot scheduling

$$\sum_{s\in S} T_{g,s}=1 \quad \forall g\in G \tag{7}$$

$$\sum_{s\in S}(1-DD_{g,s})\cdot T_{g,s}=0 \quad \forall g\in G_{due} \tag{8}$$

$$\sum_{s \in S} (1 - GndTime_{g,s}) \cdot T_{g,s} = 0 \quad \forall g \in G \tag{9}$$

$$\sum_{g \in G_p} T_{g,s} \leq MaxTasks_s \cdot M_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{10}$$

$$\sum_{g \in G_p} T_{g,s} \cdot WH_g \leq MaxWH_s \cdot M_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{11}$$

$$\sum_{g \in G_p} T_{g,s=Fict} \cdot ndd_g \leq MaxTasks \cdot (1 - Clean_p) \quad \forall p \in P \tag{12}$$

Aircraft - slot scheduling

$$\sum_{p \in P} M_{p,s} \leq 1 \quad \forall s \in S \tag{13}$$

$$\text{if } M_{p,s} = 1 \rightarrow \sum_{s \in S_{overlap}} M_{p,s} = 0 \quad \forall s \in S \quad \forall p \in P \tag{14}$$

$$G_{p,loc_s,t_{s_{end}}+1} \leq 1 - M_{p,s} + Slack_{p,s} \quad \forall p \in P \quad \forall s \in S \tag{15}$$

The first term in the MS objective penalises tasks that are not scheduled before going due, the second term aims to minimise ground time, the third term indicates that the preferred time interval for scheduling preventive and corrective tasks, the fourth term minimises the deferral of tasks, the fifth term penalises aircraft with open tasks at the end of the time window and the last term penalises the scheduling of two slots with idle ground time in between.

Next, the set of constraints that schedule a task to a slot are the following. Constraint (7) makes sure that all tasks are scheduled to a slot, including a fictitious slot if it cannot or should not be assigned to a slot active in the time window. Constraint (8) ensures that all tasks are scheduled ahead of their due date, and (9) dictates that the required ground time for all tasks in a slot does not exceed the slot duration. The total amount of (parallel) tasks in a slot is limited by constraint (10) and in the same way (11) ensures the man hours required for the tasks are within the available man hours. To avoid the model pushing tasks to the end of the time horizon resulting in infeasible solutions long-term, constraint (12) checks if there are open tasks left in the n next days after the time window considered.

In addition, there are three constraints which schedule the right aircraft to a slot. Constraint (13) dictates that each slot can only have one aircraft assigned to it. Constraint (14) ensures that no overlapping slots (in time) can be allocated to the same aircraft. Constraint (15) is set in place to give a preference to sequential slots, when possible, instead of idle ground-time between slots.

Connection tail assignment and maintenance task scheduling (groundtime):

$$\text{if } M_{p,s} = 1 \rightarrow \sum_{t_{s_{start}}}^{t_{s_{end}}} G_{p,loc,t} = s_{dur_s} \quad \forall s \in S, \forall p \in P \tag{16}$$

Constraint (16) dictates TA and MS compatibility by ensuring ground time for the duration of the maintenance slot when a slot is assigned.

RUL-based maintenance task scheduling:

$$RUL_{pr_{p,t}} = RUL_{pr_{p,t-1}} - \sum_{f \in F_{dep}} F_{p,f} \cdot f_{dur_f} + Mpr_{p,t} \cdot (RUL_{pr_p}^{MAX} - RUL_{pr_{p,t-1}}) \quad \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \tag{17}$$

$$RUL_{pr_p}^{TH} \leq RUL_{pr_{p,t}} \quad \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \tag{18}$$

$$\text{if } Mpr_{p,t} = 1 \rightarrow \sum_{s \in S_{pos}} T_{pr,s} = 1 \quad \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \tag{19}$$

Finally, the RUL-based constraints are part of both MS and TA. Recursive constraint (17) calculates the RUL for every predictor at every time step: the flight hours flown are subtracted and the RUL is reset to its maximum value when maintenance is performed on this predictor (Mpr). Maintenance on a predictor is activated when the RUL is below a set threshold, as described in constraint (18). When maintenance on a predictor is triggered, constraint (19) ensures that the predictor task is activated, meaning this task is also assigned to a slot. Note that the model can also decide to assign less or no flights to an aircraft with low RUL-values. This will postpone the RUL reaching its threshold and thus postpones maintenance on the predictors.

The RUL for each predictor at $t = 0$ in constraint (17) is set as $RUL_{pr_{p,t=0}} = EOL_{pr_p} - AGE_{pr_{p,t=0}}$, all expressed in flight hours (FH). The component's predicted end-of-life (EOL) value is derived from sensor data, which will be explained next.

4.2. RUL-based predictors

The five steps to translate historical sensor data into mean predictions of EOL values with their corresponding uncertainty quantification are visualised in Figure 4. The first step involves data collection, in this case sensor readings supplied by the airline, and will not be further expanded on here. The other steps are explained in more detail.

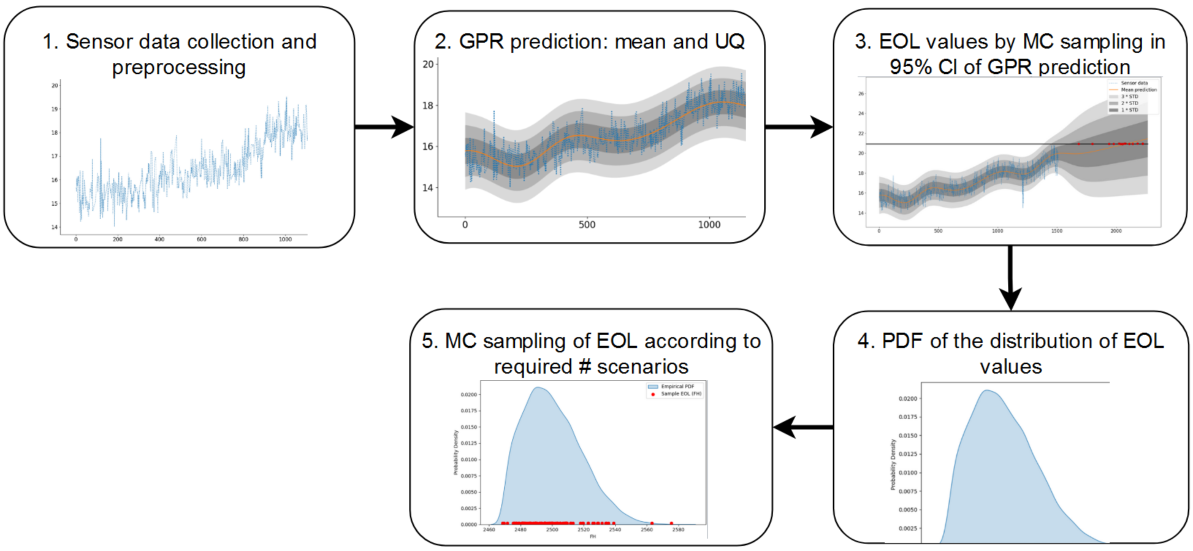

Figure 4: The five steps to translate historical sensor data into mean predictions of EOL values with their corresponding uncertainty quantification.

Gaussian Process Regression (step 2)

In the second step a mean prediction curve and its uncertainty should be quantified for the observed data. Since the model is build modular, in principle any prognostics method can be selected that fits the available data and component, as long as the output is the EOL with the prediction's uncertainty. For the predictors in this model it is assumed that the functional form of the prediction curve is unknown. Therefore, a Gaussian Process Regression (GPR) is applied, which is non parametric and comes with a high quality of uncertainty quantification [23]. It defines a prior over functions, which is then updated with observed data to form a posterior distribution. This posterior can be used to make predictions and quantify uncertainty for unknown, future sensor readings.

A prior is defined by a mean and covariance function (kernel) that encodes the data behaviour. Given observed data points, GPR then updates this prior to obtain a posterior distribution. Predictions made by GPR are not single point estimates but distributions, providing a mean prediction and an uncertainty measure, both required for the scheduling model.

The combined kernel for this study is set to:

$$k(x, x') = \theta_0(x \cdot x') + \exp\left(-\frac{\|x - x'\|^2}{2\ell^2}\right) + \sigma^2\delta(x, x')$$

This kernel combines multiple components to capture different aspects of the sensor data's underlying structure and can be applied to any predictor. Even though the function is not known, it is known that the sensor data increases over time, contains noise and its mean, or starting value in the time window considered, can vary. Therefore, the kernel is a sum of a constant kernel, a dot product kernel, an RBF kernel, and a white noise kernel. The constant kernel, θ_0, represents a bias term, ensuring that the mean function of the Gaussian Process can shift vertically. The general trend of increasing values over time is included by the dot product kernel, $x \cdot x'$, that indicates a linear trend. The RBF kernel, $\exp\left(-\frac{\|x-x'\|^2}{2\ell^2}\right)$, is needed for modelling smooth and non-linear variations, as it measures similarity based on the distance between points in the input space, with the length scale parameter ℓ controlling the smoothness of the function. Finally, the white noise kernel, $\sigma^2\delta(x, x')$, accounts for the noise in the observed data points.

End-Of-Life predictions sampling (step 3,4,5)

Next, the GPR prediction should be translated to a distribution of EOL values, expressed in flight hours (FH). This is done by drawing many GPR curves within the 95% confidence interval of the mean prediction. The FH at which each curve first crosses the threshold is recorded as an observation. These sample curves are drawn by assigning random Z-values to the prediction curve. The random Z-values are drawn from a normal distribution.

Afterwards, the probability density function is found for the distribution of these observations. Finally, a sample is drawn from this distribution, representing the EOL of the predictor, which is then fed into the holistic model. Only one sample per predictor is needed to solve the holistic model deterministically. To include the uncertainty of the EOL, however, many samples are drawn from the distribution, namely one sample per scenario. When running a sufficient amount of simulations, the probability distribution of the scenarios corresponds to the probability distribution of the EOL.

4.3. Stochastic optimisation

The uncertainty of the predictor's EOL, and therefore the uncertainty of the RUL value in the MIP parameters, should be included in the optimisation to understand its influence in decision-making. This is done using Monte Carlo sampling. Monte Carlo sampling is a simple-to-implement yet effective method, because of its assumption that a smaller sample size adequately approximates the exact solution.

The only input parameter changing in each simulation is the predicted EOL (for each predictor, for each aircraft), which has a known probability. This results in N optimal solutions, one for each of the N simulations. These N solutions with known probability can then be analysed for the features of interest, which will be explained at the end of the next section.

5. Case Study

The proposed method is demonstrated and evaluated for a case study with Swiss International Air Lines involving a fleet of five short-haul aircraft over a period of three days. The case study consists of three variations, namely sequential versus combined scheduling and including or excluding PM and its uncertainty. By examining these diverse scenarios, the advantages and applicability of the proposed method is shown.

Subsection 5.1 presents the operational input covering fleet characteristics, flight schedules and maintenance tasks. Subsection 5.2 introduces the slot schedule policies. In subsection 5.3, the predicted models are considered. Subsection 5.4 introduces the weights considered in the optimisation objective. The model settings for the three variations of the case study are presented in subsection 5.5. Finally, the evaluation metrics are discussed in subsection 5.6.

5.1. Fleet, flight schedule, maintenance tasks

Input data

The general model input is supplied by pre-pandemic historical airline data. It is expected that this data is representative for (near) future traffic volumes. The input data's parameters are summarised in Table 2.

Fleet	Flight Schedule	Maintenance work orders	Predictors
Aircraft registration	Departure time	Due date	Sensor data
Aircraft subtype	Departure airport	Criticality coefficient	Operational conditions
Forbidden destinations	Arrival time	Required ground time	(flight schedule)
Minimum TAT	Arrival airport	Required manhours	
Fuel efficiency	Required aircraft subtype		

Table 2: An overview of the input data parameters, clustered in four parts: fleet, flights, maintenance and predictors.

Firstly, the fleet needs to be described by aircraft registrations and subtypes. For each subtype the forbidden destinations, minimum TAT and fuel efficiency are defined. The fuel efficiency is used to calculate the operational costs in the TA objective function.

Secondly, the model is given a list of flights that need to be assigned. This input specifies the departure time and airport, destination time and airport and required aircraft subtype. Note that airline scheduled times and airports are used, not the actual flown schedule which may deviate from the original plan due to delays, cancellations or closed airports for example.

Next, the maintenance input consists of the task list and predictor's data per tail. Every aircraft's open maintenance tasks are specified by a list of work order (WO) numbers with a specified due date, criticality coefficient, required ground time and required amount of work hours. Again, it is assumed all due dates are calendar based, not cycles or hours, except for the RUL-based tasks. The RUL-based tasks are also included in the list of work orders per aircraft with the above parameters specified, however they are set at non-active. As explained in the methodology, they are only scheduled when the RUL-constraint activates them.

Finally, sensor data over a period of one year or more is provided for each predictor, which can be linked to the operational conditions when required. The next section will explain how this sensor data is used for prognostics.

Hypotheses driven by input data

The open task list does not contain all actually performed task during maintenance performed on those days, because of limitations on the data that could be provided by the airline. Therefore, a setting is created in the model that can artificially change the work orders. The core input is treated as valid, but can be scaled up or down. In this way the different tails always have different amounts of tasks and a different total duration, as in the original input. The parameters that can be scaled up or down are: amount of tasks, task duration and the due date can be moved ahead or pushed back.

In this experiment this is used to build a scenario where there is an abundant number of work, which would lead to flight cancellations. The settings of the work order input are done in consultation with the airline. Note that the resulting flight cancellations will be more than in reality under normal circumstances, which is done to push the model's limits. This situation namely necessitates a robust optimisation strategy to effectively manage both MS and TA and thus shows the model's capabilities.

5.2. Slot schedule

In order to simulate a continuous time in which maintenance can be done, a slot schedule is made with available slots over 24 hours, each starting at different times and for different durations. The settings of the slot duration, frequency and starting time can be changed, according to the constraints set by the airline. To model all options, so initiating a slot at every discrete time step, would blow up the computational load, especially considering different slot durations.

For this case study, every 24 hours the following slots are created, which are in line with the required maintenance load per aircraft and staff availability:

- Every two hours, for every aircraft subtype, the following slots start: two 2 hour slots, two 4 hour slots and one 6 hour slot.
- At 20:00h, 22:00h and 0:00h an 8 hour night slot is starting, for every aircraft subtype.
- Two 24 hour slots per aircraft subtype.

5.3. GPR on sensor data and EOL sampling

One single component system is selected as predictor; it is on every aircraft and extensive historical data is available by the airline. It is considered independent from other systems. The feature selected can be measured by sensors and is recorded at least every flight. The sensor values increase over time and are dependent on different operational conditions. When the sensor reading reaches a predefined threshold, maintenance should be performed on the component.

Please refer to subsection 4.2 for the full explanation of translating the sensor data to an EOL distribution that can be used as model input. Below the results of this methodology for the selected predictor (for one aircraft) are shown.

First, a prediction is made when the component's critical values will reach the set threshold using GPR, as visualised in Figure 5. The historical sensor data is shown with the blue dots and the set threshold, 21 in this case, is shown by the black horizontal line. The prediction's mean is shown by the orange line and the grey shaded areas show its uncertainty. Logically, the uncertainty converges in future, when the actual behaviour is unknown. The red dots give an example of the sampling procedure, by drawing 10,000 GPR curve samples (normally distributed) in the 95% confidence interval (CI) and recording when the threshold is first exceeded. This thus results in a distribution of EOL values.

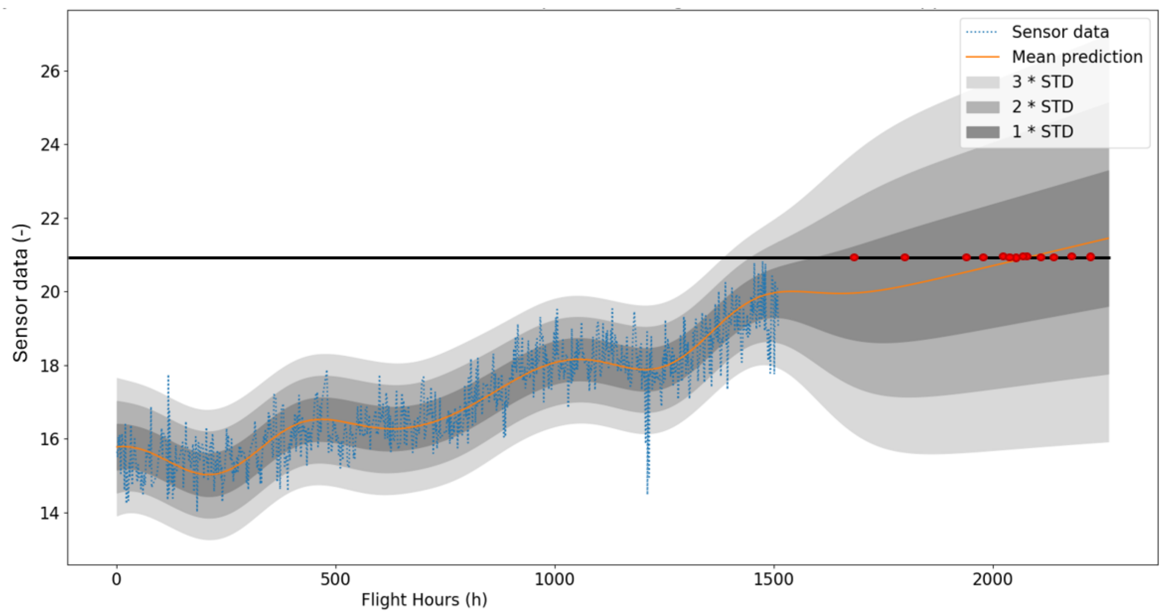

Figure 5: The predictor's sensor data (blue dots) with GPR prediction (orange line) and its uncertainty (shaded grey area). The black horizontal line at 21 indicates the replacement threshold. The red dots illustrate the FH sampling in the 95% CI when the threshold is first exceeded. To keep the image readable, not all samples are shown in this plot.

Figure 6 shows two calibration metrics of the GPR curve. It shows that the curve for the section where observed data is available is well calibrated.

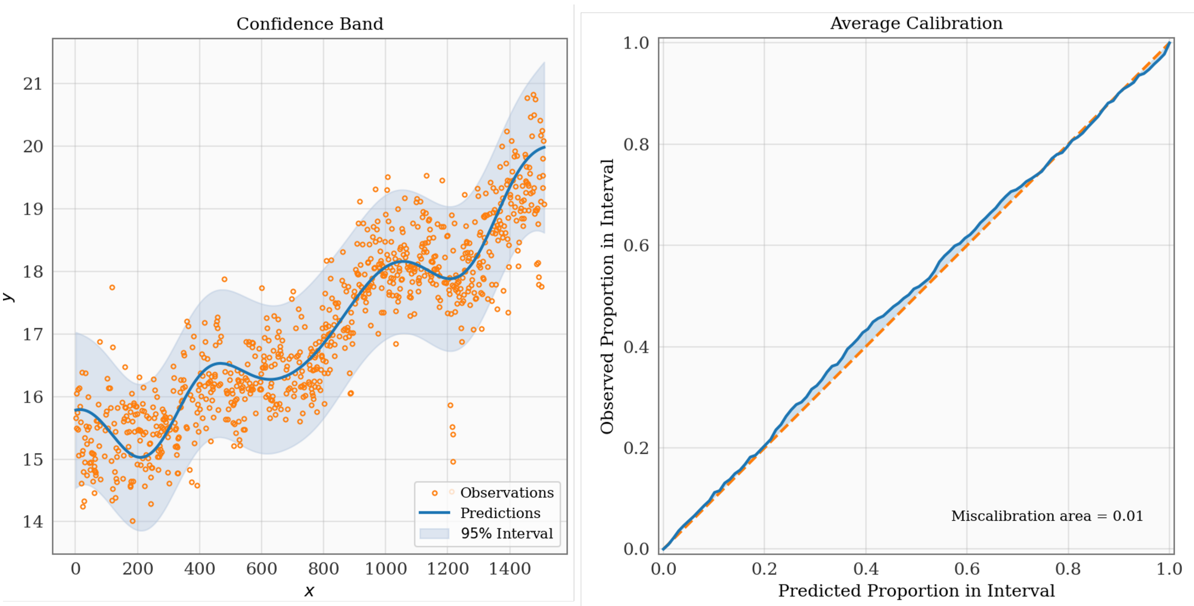

Figure 6: Uncertainty metrics for GPR results show a well-calibrated curve (note: only for the FH where observed data is available). The uncertainty metrics plots are based on the Uncertainty Toolbox by Chung et al. [2]

Next, the probability density function for the distribution of GPR curve samples can be found in Figure 7. Again, random sampling (according to a normal distribution) is applied, which is visualised with the red dots. As a reminder, these represent the EOL values at $t = 0$ that are fed into the model's MIP formulation. The number of samples drawn is thus dependent on the amount of scenarios run in a simulation.

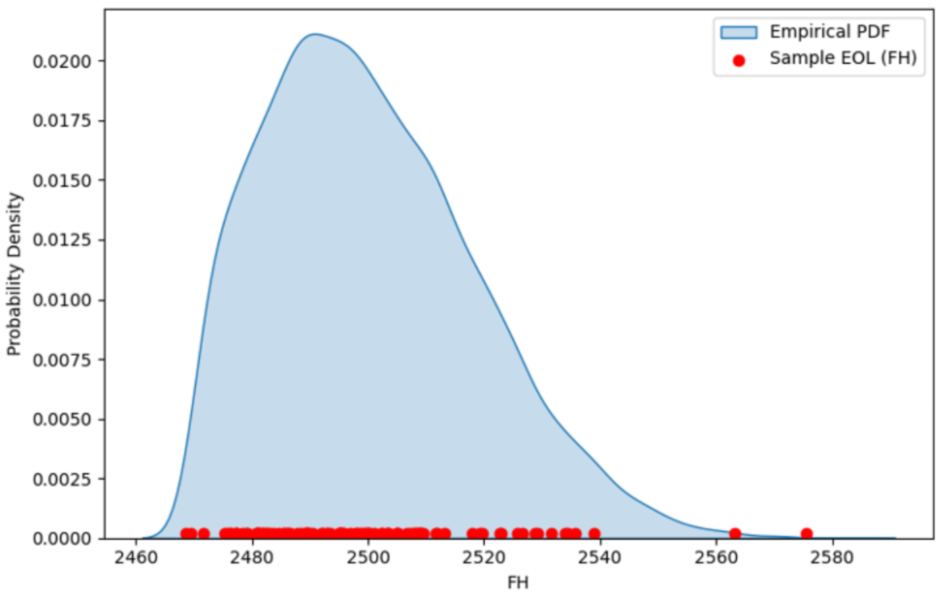

Figure 7: Probability density function for the EOL of the predictor. Samples are drawn (randomly, according to a normal distribution) that are fed into the model formulation.

5.4. Weights in objective functions

The order of magnitude used for the weights in the objective functions can be found in Table 3. These values are tuned based on airline practice, which determines the hierarchy, and comparable literature, specifically the work of Vink et al. [1], Van Kessel et al. [19] and Varenna [7], all discussed in section 3.

The main driver for operational costs is fuel consumption, which is determined for each flight by the en-route time estimate and fuel efficiency of the aircraft subtype assigned. Other drivers, such as seats management, are not considered.

The weights for task interval are simplified to two values, one for preventive tasks and one for corrective tasks. The distinction is made, because of the different nature of the tasks. This is made more clear when considering the weights of the task interval drawn from linear functions. For all tasks, first a buffer needs to be set before the due date. Any tasks, preventive or corrective, scheduled between the buffer time and the due date should be penalised very heavily, increasing more the closer it is scheduled to the due date. Next, for preventive maintenance scheduled before the buffer time point, the costs should decrease the closer it is scheduled to this point to minimise the wasted interval. For corrective tasks it is the opposite, they should be scheduled as soon as possible, and thus has increasing costs the closer it is scheduled to the buffer time point.

Weight	CC	W_{due} OC	W_{ground} GC	W_{clean} W_{seq} W_{slack}	$W_{int_{prev}}$ $W_{int_{corr}}$	W_{defer}
Value	10^5	10^2	10^1	10^1	0.5-1	0.5

Table 3: Weights used in the objective functions for MS and TA

5.5. Model settings

Due to the nature of this study being an applied model approach and not a general mathematical problem formulation, it only makes sense to analyse a specific case, namely specific calendar days for a certain (sub)fleet. The model is tested for a combination of different time windows and aircraft registrations in the fleet; one such combination is presented in this paper.

The case study variations are run for the same time window of three days with a fleet size of five aircraft, consisting of two different subtypes. The time discretisation is set to 15 minutes. The model size of the cases (5 aircraft, 3 calendar days) is in the order of magnitude of 10,000 variables and 100,000 constraints. The model is solved using Gurobi Optimizer version 10.0.3 [25] on a computer with four 2.60GHz cores and 8 GB memory. In order to keep the computational time feasible, Gurobi's solution finding time is limited to four minutes per scenario if before that time no solution has resulted in a MIP gap less than 1%. By running experiments similar in size and complexity, this limit is found to result in a MIP gap of less than 20%.

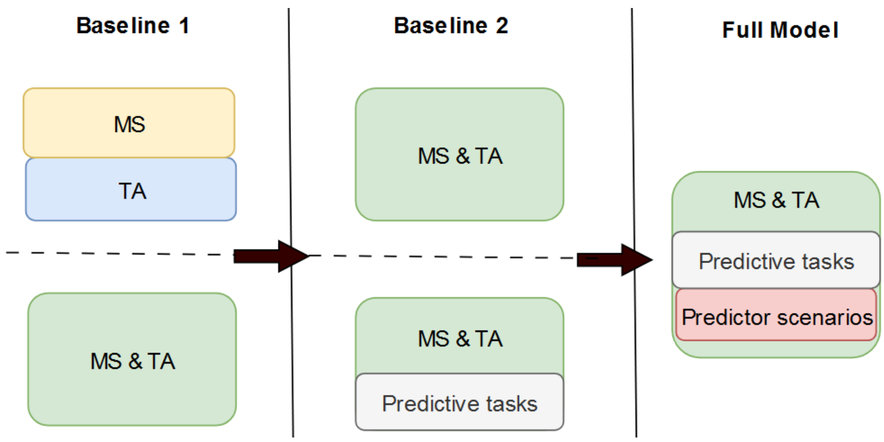

Figure 8: The model settings for the three cases, each extending the previous: baseline 1 is a comparison between sequential and combined scheduling, baseline 2 a comparison between the inclusion and exclusion of predictive tasks. The full model accounts for uncertainty of predictors by analysing multiple scenarios together.

The different case study variations, as visualised in Figure 8, are:

1. In baseline comparison 1, two different optimisation approaches are explored: sequential and combined. In the sequential approach, MS and TA are optimised one after the other, while in the combined approach, both MS and TA are optimised simultaneously. The goal is to evaluate the impact of these strategies on operational efficiency and to understand how well each approach can handle the high volume of WOs while minimising significant disruptions, such as flight cancellations.
2. Next, predictive maintenance is introduced in the combined scheduling framework (baseline 2). PM is used in order to prevent unexpectedly needing to ground an aircraft, called an AOG. In this second model setting the scenario of a unexpected failure occurring is analysed for two cases: without PM or with the inclusion of PM.
3. Lastly, the uncertainty that comes with predictive tasks is included in the model settings. The same case (3 calendar days, 5 aircraft) as the previous two model settings is considered for many different RUL scenarios. The aim is to understand the expected solution and analyse the schedule's robustness in order to find the most optimal schedule and to be best prepared when a certain RUL scenario occurs.

5.6. Evaluation metrics

The airline's interests drive the evaluation metrics for this aircraft-based scheduling tool. Firstly, the three macro parameters are total costs, number of flight cancellations and fleet (un)availability. Secondly, the two sensitivity features are TA balance and slot-tail balance. The last metrics is computational time, which is not a schedule performance indicator, but still relevant for the model's intended use. All these metrics are briefly explained below.

- Schedule costs: MS costs plus TA costs should be minimised, meaning tail-flight combinations should be found that have the lowest operational costs, maintenance ground time should be minimised by efficiently filling the slots with the right tasks and slots should be planned such that they do not interfere with the flight schedule.
- Number of cancellations: flight cancellations should be minimised, because they induce high direct and soft (such as customer satisfaction) costs, and cause consecutive operational disruptions.
- Fleet (un)availability: the goal is to minimise ground time and to maximise aircraft availability for flights. Fleet unavailability is determined by maintenance slots planned during day time, since there is a flight night ban on the two hub airports in this case study.
- TA balance: knowing which tail-flight combinations are fixed or flexible under different RUL scenarios allows for effective planning in a dynamic environment, namely a certain RUL scenario that realises. A heatmap is chosen to show this robustness, namely the density of the allocation of aircraft x to flight y.
- Slot balance: similar to TA robustness, knowing which aircraft-slot combinations are fixed or flexible allow for optimal planning when a certain RUL scenario presents itself. Also in this case a heatmap represents this robustness, namely the density of the allocation of slot x to aircraft y.
- Computational time: since the vision for this model is to be used as a decision support-tool, the model should produce feasible solutions as soon as possible, the aim is in the order of minutes.

6. Results

The results of the three case study variations, as explained in the previous section, are presented here and analysed according to the evaluation metrics. Lastly, validation is discussed.

6.1. Deterministic model with sequential or combined MS and TA optimisation

Please refer to subsection 5.5 for the description this case study variation: baseline 1.

The schedules in Figure 9 clearly show the benefits of combined versus sequential scheduling. The schedule for every aircraft (y-axis) is shown over three days, starting at midnight on day one. The blue time slots represent flights and the yellow ones maintenance slots, their durations are corresponding to flight durations and slot durations. The flights' information shown are two letters, corresponding to the departure and arrival airport. The slots' information consists of the duration, followed by a unique identifier.

As can be seen, the sequential scheduling approach, maintenance slots during the day are avoided, because it is unknown where there will be gaps in the schedule based on the flight schedule alone. This reduces the flexibility of the planning and leads to sub-optimal solutions. In the combined scheduling, however, these opportunities are found during the optimisation, resulting in a greater flexibility and less cancellations.

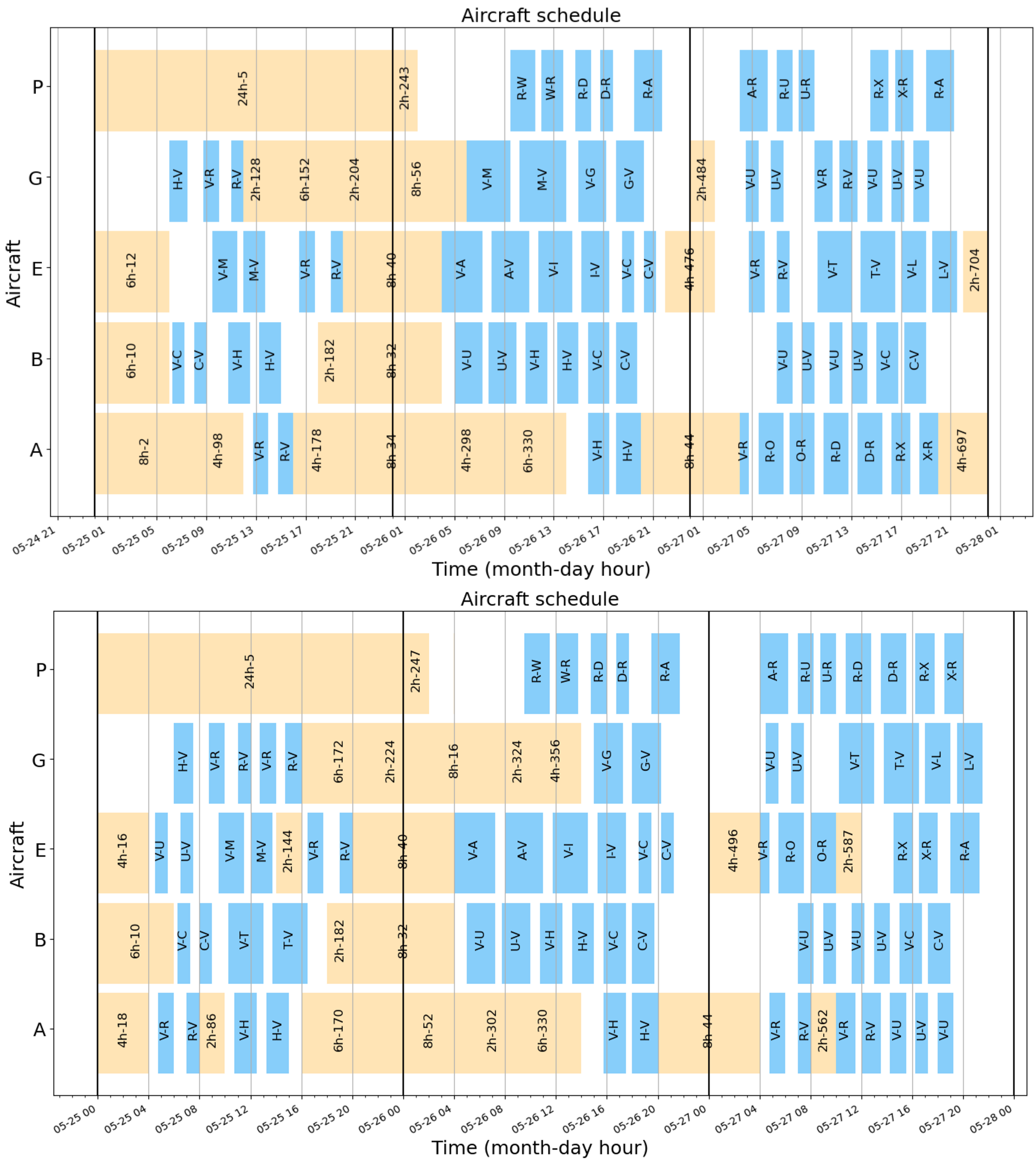

Figure 9: Aircraft schedules for sequential scheduling (top) and combined scheduling (below) for the same three calendar days, fleet, maintenance load and flight schedule input. The blue time slots represent flights and the yellow ones maintenance slots, their durations are corresponding to flight durations and slot durations.

Table 4 shows the effect in numbers, namely the reduction in cancellations, which has the largest influence on TA costs, and also on total schedule costs. Still, the maintenance costs are also reduced in the combined scheduling approach. The amount and nature of the tasks to be scheduled is the same, and in both approaches all tasks are scheduled on time. The difference is thus in the total ground time occupied by slots, which is longer for the sequential approach. Note that this is the average slot usage factor over all slots; more than half of the individual slots have a slot usage factor above 95%, which is consistent with the model design. Slots are filled up to their maximum capacity (which is already accounted for by a safety margin) and only then additional slots are added.

	Total schedule cost [-]	TA cost [-]	MS cost [-]	Flight cancellations	Slot usage factor (average)
Sequential optimisation	1421110	1341621	79489	12	57%
Combined optimisation	1022212 (39% ↓)	944039 (42% ↓)	78173 (1.7% ↓)	8 (50% ↓)	63% (10.4% ↑)

Table 4: Costs, total amount of cancellations and average slot usage factor (the higher the better) are compared for sequential and combined deterministic MS and TA scheduling.

6.2. Deterministic model with or without predictive tasks

Please refer to subsection 5.5 for the description this case study variation: baseline 2.

The schedule solution is illustrated in Figure 10. After the failure occurs (AOG, dashed line), immediate maintenance is required which means that one rotation of flights has to be cancelled. However, as can be seen in this schedule, there was availability before in this aircraft's schedule for additional maintenance. Either by scheduling extra slots (shown in the example in the lower schedule), or by adding a task to an already scheduled slot which is not completely full. When using predictive tasks, this situation can be prevented because the failure is expected, and thus the availability in the maintenance slots can be optimally used. In this case, the overall costs will be lower, because flight cancellations are more costly (by a factor of 10^4) than scheduling extra maintenance for the predictors.

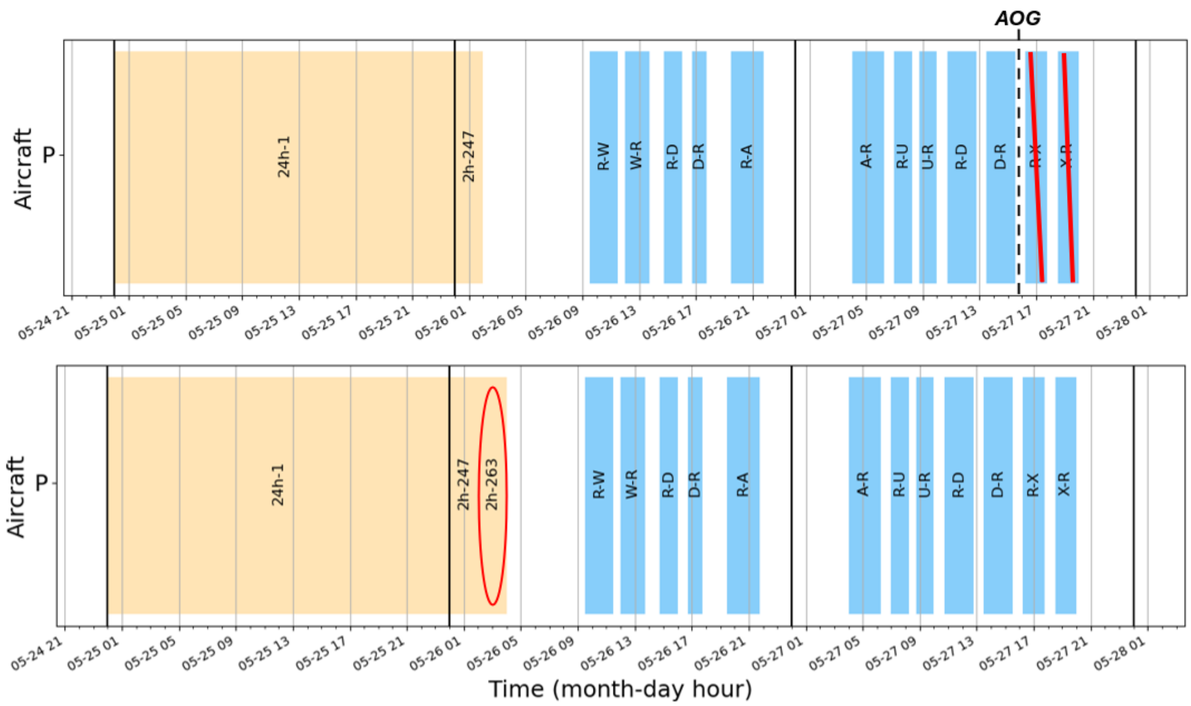

Figure 10: Schedule example when a failure leads to an AOG and corrective maintenance needs to be done, resulting in cancellations (top). Predictive maintenance would have prevented the AOG by scheduling the task in advance (below).

However, scheduling predictive tasks too early (when the component will still be healthy for a considerable time) would result in higher costs, because over time this requires more frequent maintenance on the predictors, i.e. the wasted interval is increased. The model prevents this from happening, since maintenance on the predictors is only activated when the RUL threshold is exceeded. The RUL threshold is set in a comparable way to the task utilisation interval for preventive tasks. In addition, as shown in the example in Figure 10 as well, predictor tasks are added to slots when there is space or scheduled sequential to other slots when possible, in order to avoid additional maintenance costs.

The advantage of using prognostics to avoid corrective interventions is clearly shown by the model. Still, wasting healthy component time is likely to occur in a deterministic approach. Predictions come with an uncertainty, but when this is not quantified, a choice needs to be made whether to account for this. For example, by adding safety factors (conservative approach) the risk of failure will be lower, but the wasted interval higher. This trade-off will add costs due to the lack of information of the predictions' uncertainty quantification.

Thus, this deterministic approach means that human planners would treat the predictive tasks similar to preventive tasks, just the methodology to determine the due date is different. To deal with the uncertainty that comes with predictive tasks' due date, that is not present for preventive tasks, does require a new scheduling strategy. This will be explained in the next section.

6.3. Stochastic model with predictive tasks

Please refer to subsection 5.5 for the description this case study variation: the full model.

The Monte Carlo simulation consists of a 100 different RUL scenarios. Each scenario has different EOL samples for each predictor. Thus, in the case of five aircraft with one predictor per aircraft, this results in 500 EOL samples.

Figure 11 shows the probability distribution of total costs for the MS and TA schedule, number of cancellations and fleet unavailability. This simulation shows that there are essentially four different solutions in terms of cancellations, either 6, 8, 10 or 11, where 8 is most likely and 11 has very small probability. The flight numbers cancelled are the same for these scenarios. That means that the planners can know in advance that six flights have to be cancelled anyway, most likely two more, or potentially three to four more. It also shows which flights are not compatible with the required maintenance tasks due dates of the fleet in this simulation. Since the most influential factor in the total costs is the cancellation costs, it shows a direct correlation with the amount of cancellations.

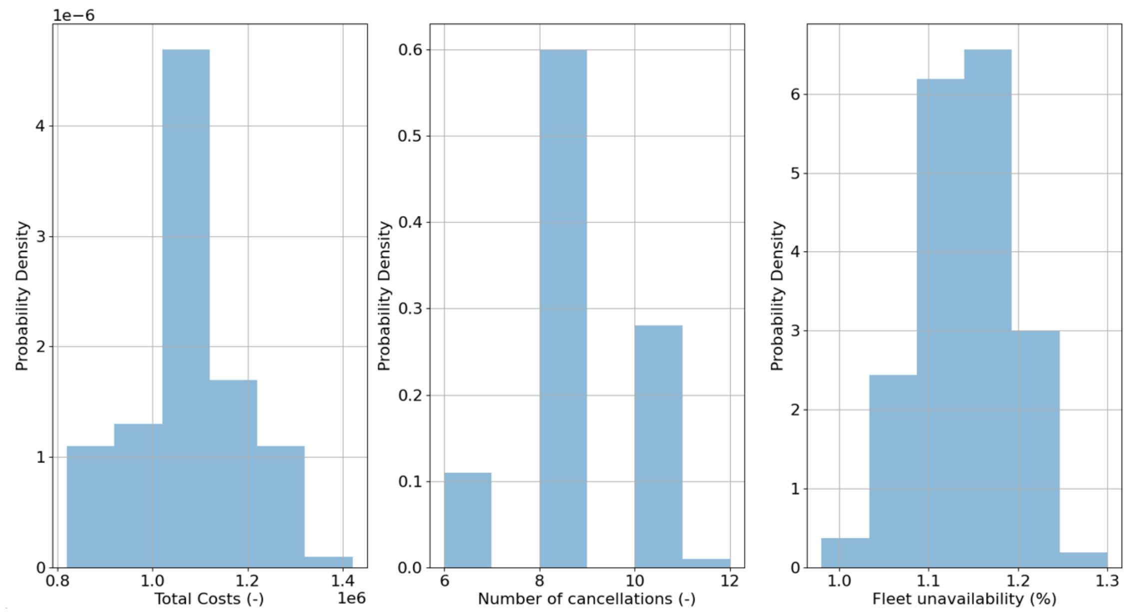

Figure 11: Probability density for three KPIs: total schedule costs, number of cancellations and fleet unavailability.

The fleet unavailability variation for this case is insignificant, a difference of less than 0.5%. The small variation shows that the different combination of slots that are possible are most likely of the same slot duration, and thus do not, or very limited, influence fleet unavailability. This is logical, given that only one predictor per aircraft was considered with a required maintenance ground time of only one hour.

The density distribution of the use of aircraft x for flight y, Figure 12, is one metric to show robustness in the schedule. It can be seen that some flights are always assigned to the same aircraft, such as F5 and F14. Other flights, such as F3, can be assigned to any of four aircraft, with the highest probability for aircraft G. These metrics can be used by planners to effectively deal with uncertainty of RUL-based tasks, by knowing in advance for a set of likely scenarios which flights and aircraft combinations are locked (only one option possible) or flexible, i.e. several combinations can be made while maintaining full schedule feasibility.

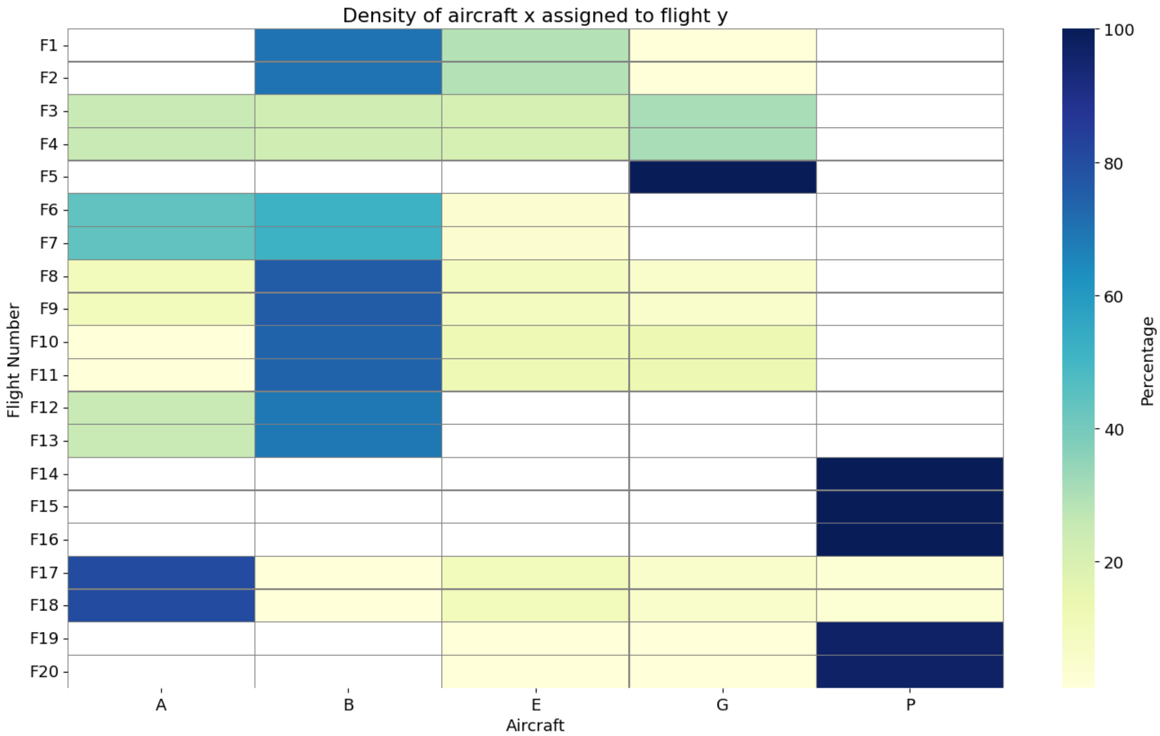

Figure 12: Density distribution of the use of aircraft x for flight y. For clear illustrative purpose, only the first twenty flights in the schedule are shown. Note that rotations are clearly visible by the same aircraft assignment and subsequent flight numbers. Appendix A shows the full figure, including percentages.

An example to better interpret this heatmap is shown in Figure 13. This figure zooms in on the schedule of one day and three aircraft for two different predictor scenarios. The same flights are scheduled in both scenarios, but aircraft E and A have a higher maintenance load (due to the RUL-based tasks being activated) in scenario 2 than scenario 1. Therefore more or longer maintenance slots need to be selected. This triggers a different TA and MS solution for all three aircraft, as different flights are assigned to different aircraft and different maintenance slots are selected, not just to allocate more maintenance time but also choosing slots at different times to avoid extra cancellations. Note the TA costs are nearly the same in both scenarios, namely a difference of 0.01%, which is due to a slightly different fuel cost efficiency factor. The flight rotation that is allocated the same way is encircled in red. This is directly linked to the slot before, which is also the same in both scenarios.

Thus, this example again shows that model solution interpretation should be case specific, but that certain consistencies across different predictors scenarios do appear. When planners use the model as a decision-support tool in live operations, this information will help to mitigate the uncertainty of the required maintenance slots (due to the predictive tasks) by knowing in advance: 1) what is the most optimum solution for a given predictor scenario realisation and 2) when tail swaps need to be made: which tail-flight or tail-slot combinations are best selected to maintain schedule feasibility at low costs. Note that this model is not designed to make the tail swaps itself, but the solutions give valuable information for this procedure.

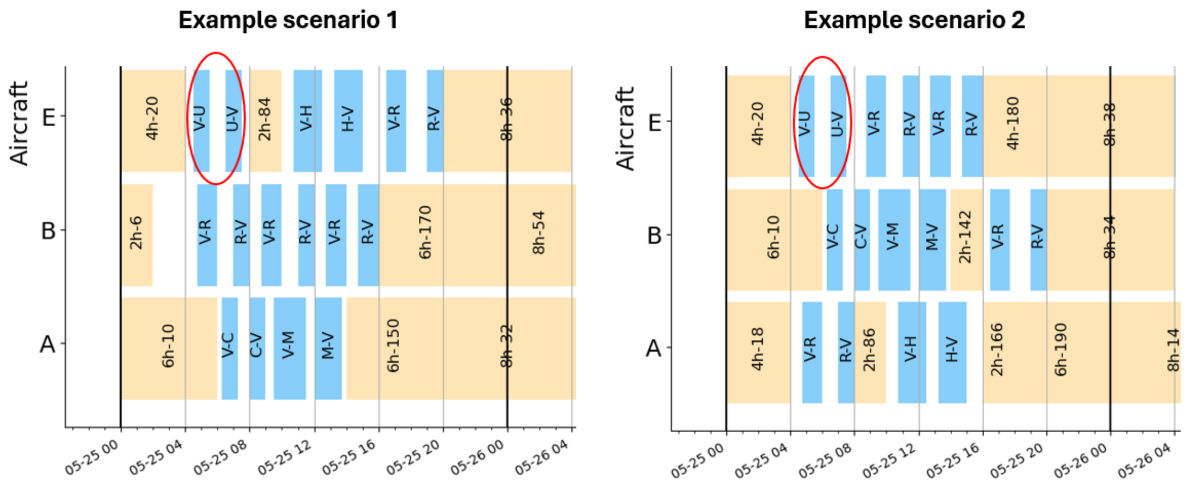

Figure 13: Two example scenarios (zoomed in to one day and 3 aircraft) to clarify the heatmap in Figure 12. Aircraft E and A have a higher maintenance load in scenario 2 (right) than scenario 1 (left), therefore more or longer maintenance slots are selected. This triggers a different TA solution. The flights that are allocated the same way are encircled in red.

Similarly, a metric for MS is the density distribution of the use of slot x by aircraft y, shown in Figure 14. It shows that for some aircraft a certain slot, especially the ones with longer durations, are always scheduled, whereas other slots have much more flexibility, especially for the shorter durations. In the same way as for TA, this is a practical way for planners to deal with uncertainty of the predictors.

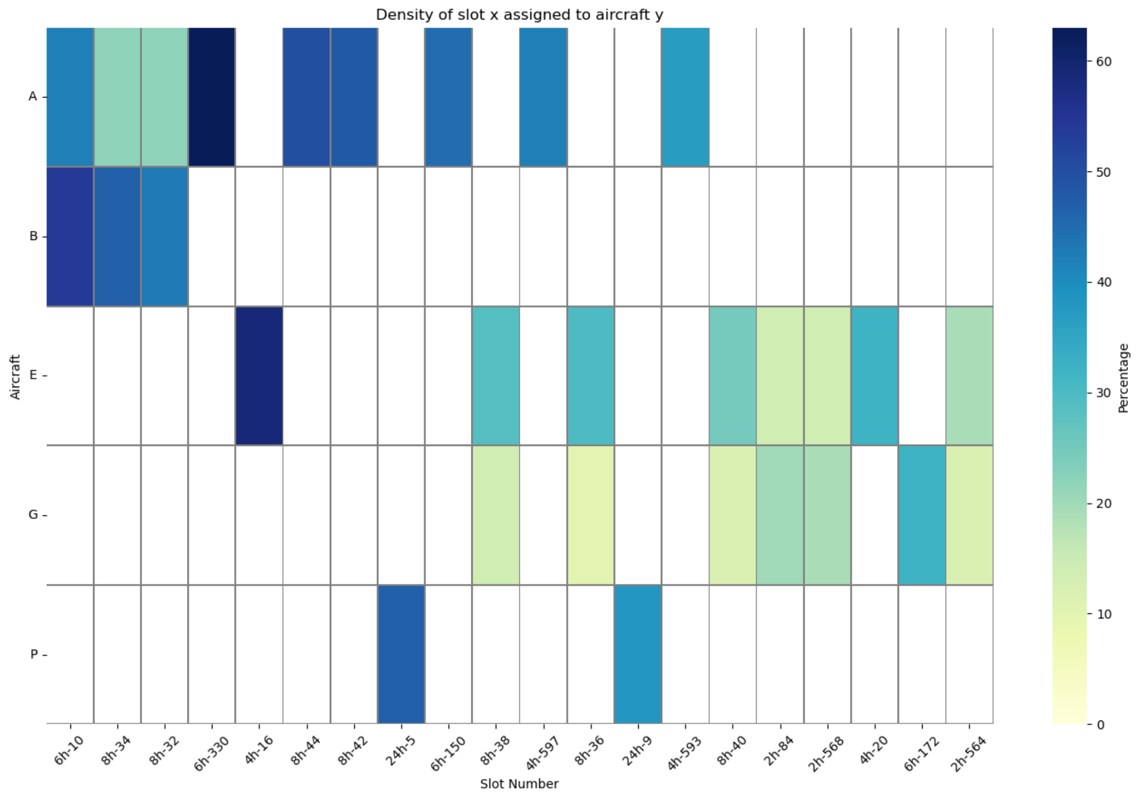

Figure 14: Density distribution of the use of slot x by aircraft y.For clear illustrative purpose, only the twenty most frequently allocated slots are shown. Appendix A shows the full figure.

6.4. Computational time

The computational time is determined by the Gurobi Optimizer time, because the remaining parts take only between 10-20 seconds. Since Gurobi uses a branch-and-cut algorithm, the complexity of the problem drives the computational time. The so-called difficult scenarios that require many cancellations due to a large overload of the WOs input may take up to 10 minutes, and so-called easy scenarios (no cancellations required) can be solved under one minute to achieve a MIP gap tolerance of less than 1%.

As seen in Figure 15, 78% of the Monte Carlo simulation scenarios took just over four minutes, the set limit of Gurobi for this case study, which typically reached a MIP gap between 8-25%. The remaining scenarios reached a MIP gap of <0.8% faster than four minutes. This means that the complete MC simulation with 100 scenarios takes about 7.5 to 8 hours to run.[2] There is no clear explanation why certain scenarios could be solved faster; the RUL-based tasks are scheduled for both the simulations that take longer and less than the set time limit, the schedule solution is feasible and the costs and number of cancellations are no different than the average range. It does show the solution time is highly dependent on the search tree by Gurobi's algorithm.

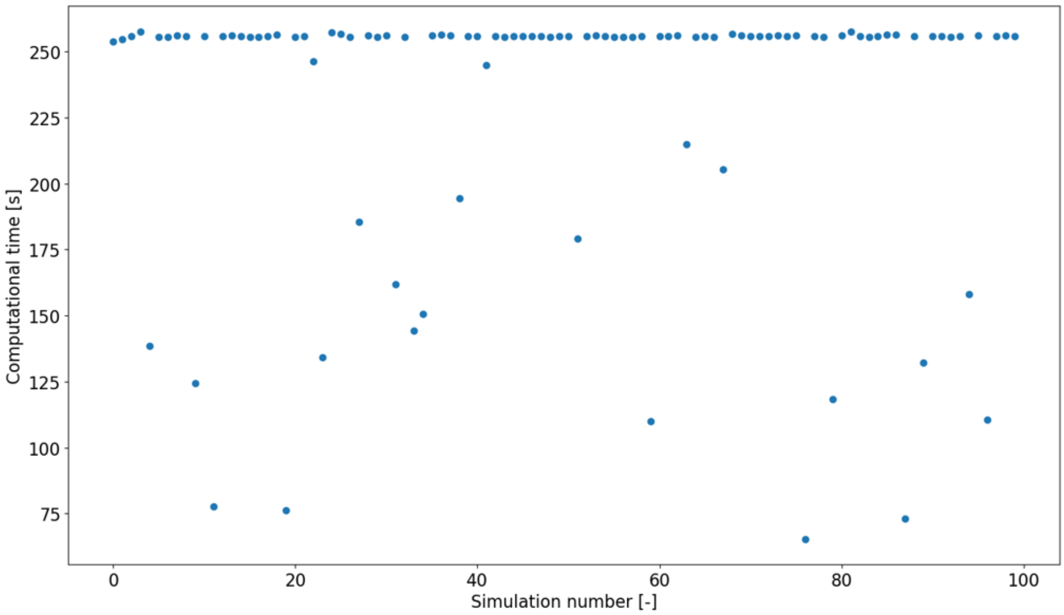

Figure 15: The computational time for each of the 100 Monte Carlo simulations.

6.5. Validation

The model is validated with a qualitative analysis of the results by expert planners, from both TA and short-term MS. A quantitative analysis comparing the model results of a historic scenario with the actual schedule operated during those days is not possible because of the model hypotheses. The two main assumptions that make it too different from the historic dataset are the exclusion of network disruptions and artificial maintenance tasks input.

Set-up

The validation with the expert planners of the two departments consisted of two parts: a comparison study and model analysis. The comparison study was done first, to minimise biasing the planners with example schedules and model solutions. The expert planners were given the same input and assumptions as the model and asked to create a feasible schedule. Note that they were asked to make an aircraft schedule for both TA and MS in one step, and not sequential which would be there normal workflow.

[2]For results analysis all model parameters and solution parameters were saved, and this extra output writing time for each scenario accounts for the missing time when simply multiplying 250 or 260 seconds by 100.

For the model analysis part the planners were shown different solutions that the model found to the case they solved themselves and asked to give feedback. In addition, more cases were presented, also showing results of the stochastic optimisation to highlight the more or less robust parts of the schedule for example. They were asked to give feedback on the reasonability of the model solutions, if they could understand why the model made certain decisions and what they would change and why and how the uncertainty of the RUL-based tasks should be dealt with given the model solutions. Moreover, they were asked to comment on the outlook: which parts of this model development are most/ least promising and why, and what they would like to see in future model development.

Results

Compared to the model results, the expert planners' schedule had the same amount of cancellations and total maintenance slot duration, resulting in the same total costs. This shows that the model results are appropriate. Also, it was noted that it took the planners much longer than the model to make a feasible schedule, because of its complexity, namely the many inputs and the need to go back and forth between MS and TA.

Using a computer-aided combined TA and MS scheduling approach would require a fundamental change in strategy by the planners. The key points of this discussion are as follows. Firstly, the introduction of deterministic predictive tasks could be dealt with similarly to existing strategies, but it is difficult to optimise the schedule manually given the increased complexity. Predictive tasks can be scheduled as preventive tasks, which is consistent with the model. The model also changes TA based on the RUL, which would be a similar strategy to a balanced use of aircraft (in flight cycles and FH) that TA planners already do. The uncertainty might indicate buffer time, tail swaps or different slot allocation, for which it is useful to have the model's results showing which parts of the schedule are most or least robust. However, to apply these strategies also to predictive maintenance does require a change in overall scheduling strategy and is much harder to implement effectively. Optimising TA and MS holistically only further increases this complexity, for which a computerised support-tool is needed.

In addition, the planners pointed out that the current model formulation implicitly favours MS more than TA. Namely, maintenance slots are allowed in-between rotations, adding flexibility for MS, whereas this takes away buffer time for TA in case of flight delays. This is indeed consistent with the model hypotheses, because no network disruptions are considered.

7. Discussion and Recommendations

The results show that combined TA and MS scheduling is promising to reduce overall schedule cost, number of cancellations and shows which schedule elements are more / less stable under different predictor scenarios. For future use as a decision-support tool for planners during live operations, it is essential to have visualisations that provide a quick understanding of the schedule solutions, as well as improved model accuracy. Both of these aspects are elaborated upon in the following sections.

7.1. User interface and model analysis tool

Next to the results already presented, additional visualisations are needed to effectively understand the different schedule solutions, which could eventually be developed into an user interface as well. One example is the parallel coordinates graph. It is an effective tool when used interactively to find patterns or robustness for specific aircraft, slots or flights, but is more difficult to understand as a snapshot. An example that can still be understood without interaction, is shown in Figure 16. It compares two scenarios of different likelihood of EOL predictions, shown by the most right axis and colourbar. The RUL axis shows the value of the RUL over time for the aircraft selected. When maintenance is done on the predictor it is reset to its maximum value of 250 FH. The threshold that triggers maintenance action is set to 20 FH. The middle axis indicates to which flight or maintenance slot an aircraft is assigned. All of this is shown over time, shown on the most left axis, and for each aircraft, shown on the second left axis. Any combination of axes can be selected interactively, to allow a full analysis.

The two cases highlighted in this image show the difference in the schedule for aircraft C for two different RUL scenarios. The complete time window is shown. The top image shows that RUL-based tasks need to be schedule in the time horizon, whereas the bottom image shows that no RUL threshold is

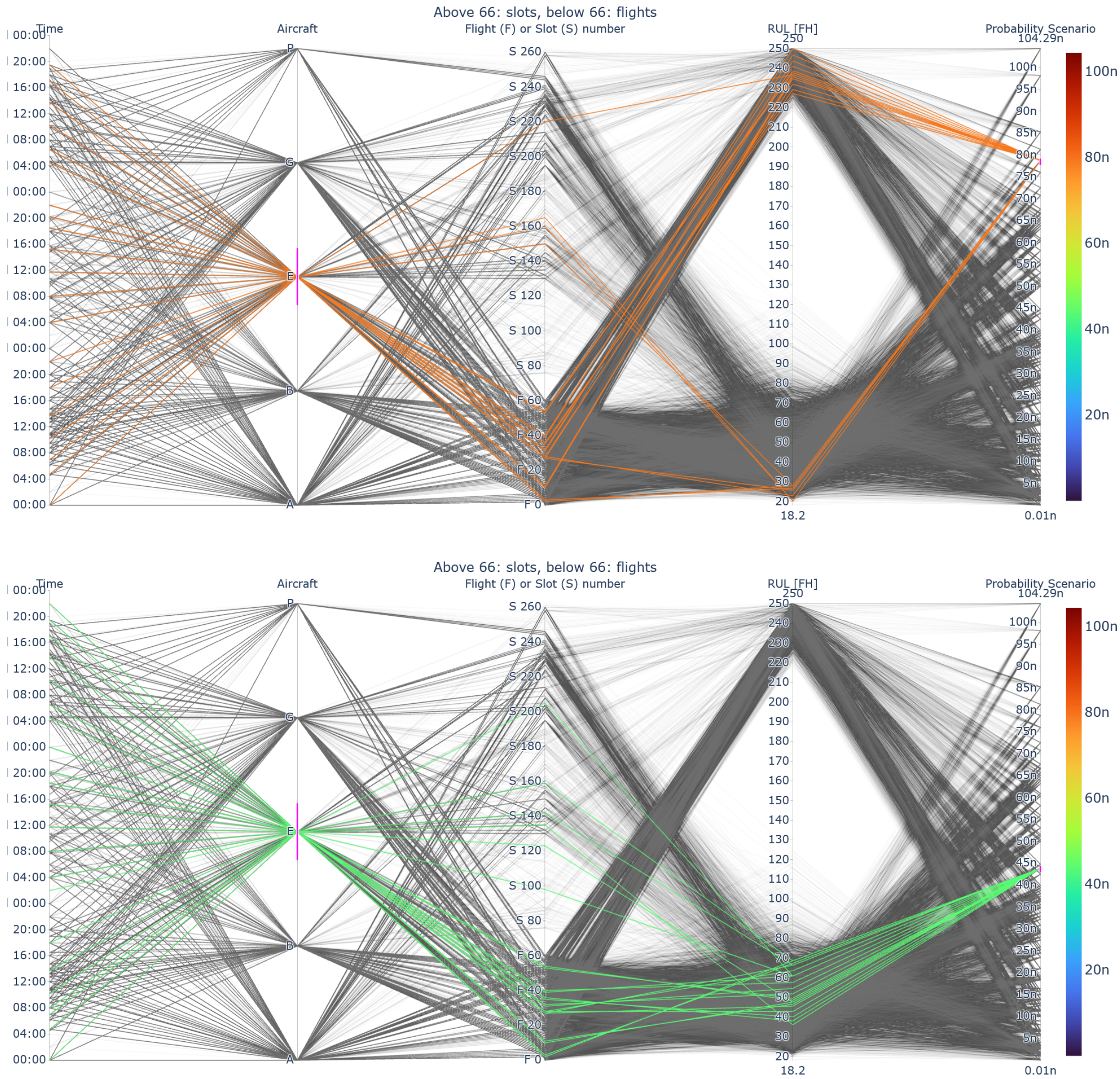

Figure 16: Parallel coordinates graph showing the schedule for the full fleet, 5 aircraft, over 3 days (time increasing from bottom to top on the left axis). The middle axis indicates to which flight or maintenance slot an aircraft is assigned. This snapshot shows the difference in the schedule for aircraft C for two scenarios: RUL-based tasks need to be scheduled in the time horizon (top) versus no threshold reached in the time horizon (bottom).

reached in the time horizon. Based on these two scenarios, there is a higher probability that RUL-based maintenance need to be scheduled. Please keep in mind that this is cannot be concluded for the full simulation based on these two scenarios alone.

7.2. Future work

Further development is needed to model reality more closely, to achieve more accurate results and to be effectively used as a decision-support tool. The modular framework of the model allows for easy adaption and extension, because the future vision was already kept in mind.

Firstly, the current computational load is too expensive for a decision-support tool for expert planners in a live operation environment. MC simulation is computationally expensive, because the complete problem needs to be solved for every simulation. Therefore, it is suggested to investigate methods that lower the computational load, or takes the running of different simulations offline. In practice this could mean to run the program overnight to find all possibilities for the coming days. During live operations, planners could then set the model to the scenario that realises and the model can achieve a result in the order of minutes. This will allow to include the full fleet of the airline, and also to do a feasibility study of the short-term schedule with the long-term schedule, up to several months ahead.

One of the methods that was already investigated, but could not be implemented in this paper due to time constraints, is two stage stochastic programming in combination with a neural network. This method was chosen because it proved effective for similar MIP formulated problems comparable in size [3]. A suggestion for adapting the model in a two-stage formulated problem is presented in Appendix B.

Moreover, a better prognostics model needs to be developed to reduce the uncertainty of the EOL, especially considering that this uncertainty propagates through the model. Also, further accuracy can be achieved by extensive sensitivity studies on the weights and the number of clean days (MS). Additionally, to simulate reality more closely, more accurate maintenance tasks input is required.

Lastly, the model can be extended by adding a more detailed version of the 4M requirements. In the same way, the model can be extended by including disruptions and the necessary countermeasures, both in network operations (flight delays, externally triggered cancellations) and in maintenance delays or failures. At a later stage, the user interface for the planners should be made such that the model gives a proposed fleet schedule solution, with the corresponding KPI values, for a scenario selected by the planner. In addition, the model should highlight where tails or slots can be swapped with the least disturbance and cost.

By implementing these recommendations, the model can become a more powerful and accurate decision-support tool, providing significant benefits in optimising maintenance scheduling and tail assignment under uncertain conditions.

8. Conclusion

This study is the first to demonstrate a successful implementation of combined TA, MS and PM scheduling. The proposed MIP-formulated problem finds an optimal aircraft schedule for TA and preventive, corrective and RUL-based MS. The predictor's RUL values are determined by GPR-based EOL predictions and TA choices, the latter is in turn dependent on the RUL. The prediction's uncertainty results in a range of possible scenarios which are Monte Carlo simulated.

A case study with Swiss International Air Lines shows the effectiveness of the proposed approach. Combined scheduling, compared to a sequential approach, finds a better optimum by allowing maintenance slots between flights. In addition, PM avoids AOGs which would lead to flight cancellations. The MC simulations show to which degree tail-flight and slot-tail combinations can be rotated, as well as the probability of non-compatible flight schedules and the fleet's maintenance requirements, but come with a high computational cost.

Therefore, it is recommended to develop a optimisation method that lowers the computational load. This will also allow to scale-up, both in fleet and time horizon, and to include more 4M constraints and operational disruptions. Moreover, an advanced prognostics model will reduce the predictors' uncertainties. Likewise, more accurate work order input and extensive parameter tuning will model reality even more closely.

Thus, the developed framework highlights the potential of a computerised tool for combined scheduling with uncertain scenarios as a new strategy for planners, resulting in less operational and maintenance costs.

Bibliography

[1] Vink, J., Santos, B., Verhagen, W., Medeiros, I., and Filho, R., "Dynamic aircraft recovery problem - An operational decision support framework," *Computers Operations Research*, Vol. 117, 2020, pp. 104892.

[2] Chung, Y., Char, I., Guo, H., Schneider, J., and Neiswanger, W., "Uncertainty Toolbox: an Open-Source Library for Assessing, Visualizing, and Improving Uncertainty Quantification," *arXiv preprint arXiv:2109.10254*, 2021.

[3] Dumouchelle, J., Patel, R., Khalil, E. B., and Bodur, M., "Neur2sp: Neural two-stage stochastic programming," *arXiv preprint arXiv:2205.12006*, 2022.

[4] IATA, "Airline maintenance cost executive commentary | fy2022 data," 2024, Available: `https://www.iata.org/en/programs/ops-infra/techops/mctg/#tab-2`.

[5] Lagos, C., Delgado, F., , and Klapp, M. A., "Dynamic optimization for airline maintenance operations," *Transportation Science*, Vol. 54(4), 2020, pp. 998–1015.

[6] Iwata, C. and Mavris, D., "Object-Oriented Discrete Event Simulation Modeling Environment for Aerospace Vehicle Maintenance and Logistics Process," *Procedia Computer Science*, Vol. 16, 12 2013, pp. 187–196.

[7] Varenna, S., "A Stochastic Discrete Event Simulation of Airline Network and Maintenance Operations," 2023, MSc Thesis at Delft University of Technology under supervision of Dr. ir. B.F. Santos.

[8] Hassan, L., Santos, B., and Vink, J., "Airline disruption management: A literature review and practical challenges," *Computers Operations Research*, Vol. 127, 2021, pp. 105137.

[9] Shaukat, S., Katscher, M., Wu, C.-L., Delgado, F., and Larrain, H., "Aircraft line maintenance scheduling and optimisation," *Journal of Air Transport Management*, Vol. 89, 2020, pp. 101914.

[10] Tseremoglou, I. and Santos, B., "Condition-Based Maintenance scheduling of an aircraft fleet under partial observability: A Deep Reinforcement Learning approach," *Reliability Engineering System Safety*, 08 2023, pp. 109582.

[11] Lee, J. and Mitici, M., "Deep reinforcement learning for predictive aircraft maintenance using probabilistic remaining-useful-life prognostics," *Reliability Engineering & System Safety*, Vol. 230, 2023, pp. 108908.

[12] Barnhart, C. and Cohn, A., "Airline schedule planning: Accomplishments and opportunities," *Manufacturing & service operations management*, Vol. 6, No. 1, 2004, pp. 3–22.

[13] Zhou, L., Liang, Z., Chou, C.-A., and Chaovalitwongse, W. A., "Airline planning and scheduling: Models and solution methodologies," *Frontiers of Engineering Management*, Vol. 7, No. 1, 2020, pp. 1–26.

[14] Ma, H.-L., Sun, Y., Chung, S.-H., and Chan, H. K., "Tackling uncertainties in aircraft maintenance routing: A review of emerging technologies," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 164, 2022, pp. 102805.

[15] Barnhart, C., Boland, N. L., Clarke, L. W., Johnson, E. L., Nemhauser, G. L., and Shenoi, R. G., "Flight string models for aircraft fleeting and routing," *Transportation science*, Vol. 32, No. 3, 1998, pp. 208–220.

[16] Sarac, A., Batta, R., and Rump, C. M., "A branch-and-price approach for operational aircraft maintenance routing," *European Journal of Operational Research*, Vol. 175, No. 3, 2006, pp. 1850–1869.

[17] Witteman, M., Deng, Q., and Santos, B. F., "A bin packing approach to solve the aircraft maintenance task allocation problem," *European Journal of Operational Research*, Vol. 294(1):365–376, 10, 2021.

[18] Qin, Y., Wang, Z., Chan, F. T., Chung, S., and Qu, T., "A mathematical model and algorithms for the aircraft hangar maintenance scheduling problem," *Applied Mathematical Modelling*, Vol. 67, 2019, pp. 491–509.

[19] van Kessel, P. J., Freeman, F. C., and Santos, B. F., "Airline maintenance task rescheduling in a disruptive environment," *European Journal of Operational Research*, Vol. 308, No. 2, 2022, pp. 605–621.

[20] Duffuaa, S. O. and Andijani, A. A., “An integrated simulation model for effective planning of maintenance operations for Saudi Arabian Airlines (SAUDIA),” *Production Planning & Control*, Vol. 10, No. 6, 1999, pp. 579–584.

[21] Sachon, M. and Paté-Cornell, M.-E., “Delays and Safety in Airline Maintenance,” *Reliability Engineering System Safety*, Vol. 67, 03 1999, pp. 301–309.

[22] Öhman, M., Hiltunen, M., Virtanen, K., and Holmström, J., “Frontlog scheduling in aircraft line maintenance: From explorative solution design to theoretical insight into buffer management,” *Journal of Operations Management*, Vol. 67, 03 2020, pp. 120–151.

[23] Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., Wang, Y., Du, X., Zhang, X., and Hu, C., “Uncertainty Quantification in Machine Learning for Engineering Design and Health Prognostics: A Tutorial,” *Mechanical Systems and Signal Processing*, 2023.

[24] de Pater, I. and Mitici, M., “Predictive maintenance for multi-component systems of repairables with Remaining-Useful-Life prognostics and a limited stock of spare components,” *Reliability Engineering & System Safety*, Vol. 214, 2021, pp. 107761.

[25] Gurobi Optimization, LLC, “Gurobi Optimizer Reference Manual,” 2024.

A. Additional Results

Density distribution flights - fleet

Figure 17 shows the full version of Figure 12 in the Results.

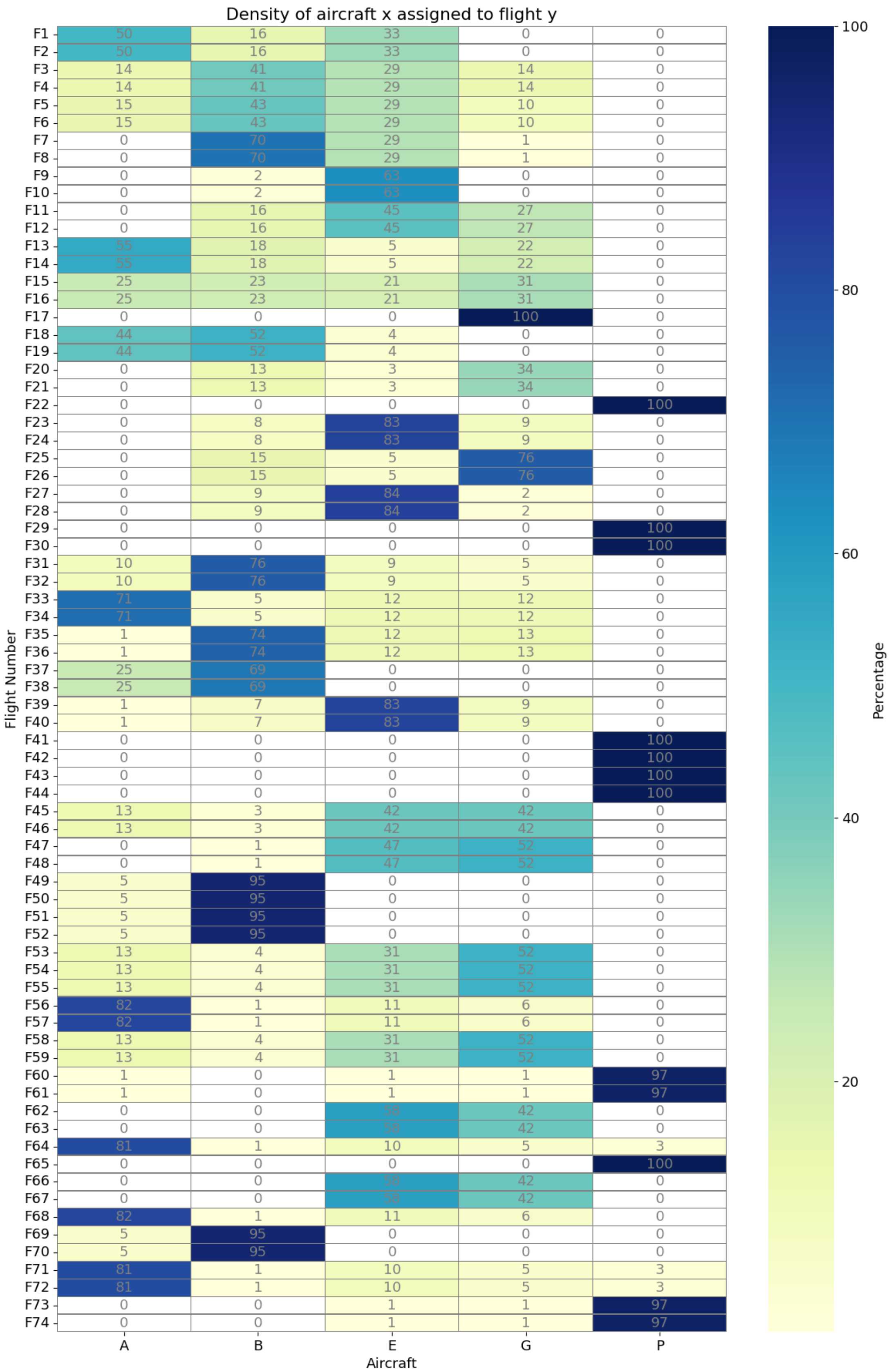

Figure 17: Density distribution of the use of aircraft x for flight y. Note that rotations are clearly visible by the same aircraft assignment and subsequent flight numbers.

Density distribution fleet- slots

Figure 18 shows the full version of Figure 14 in the Results.

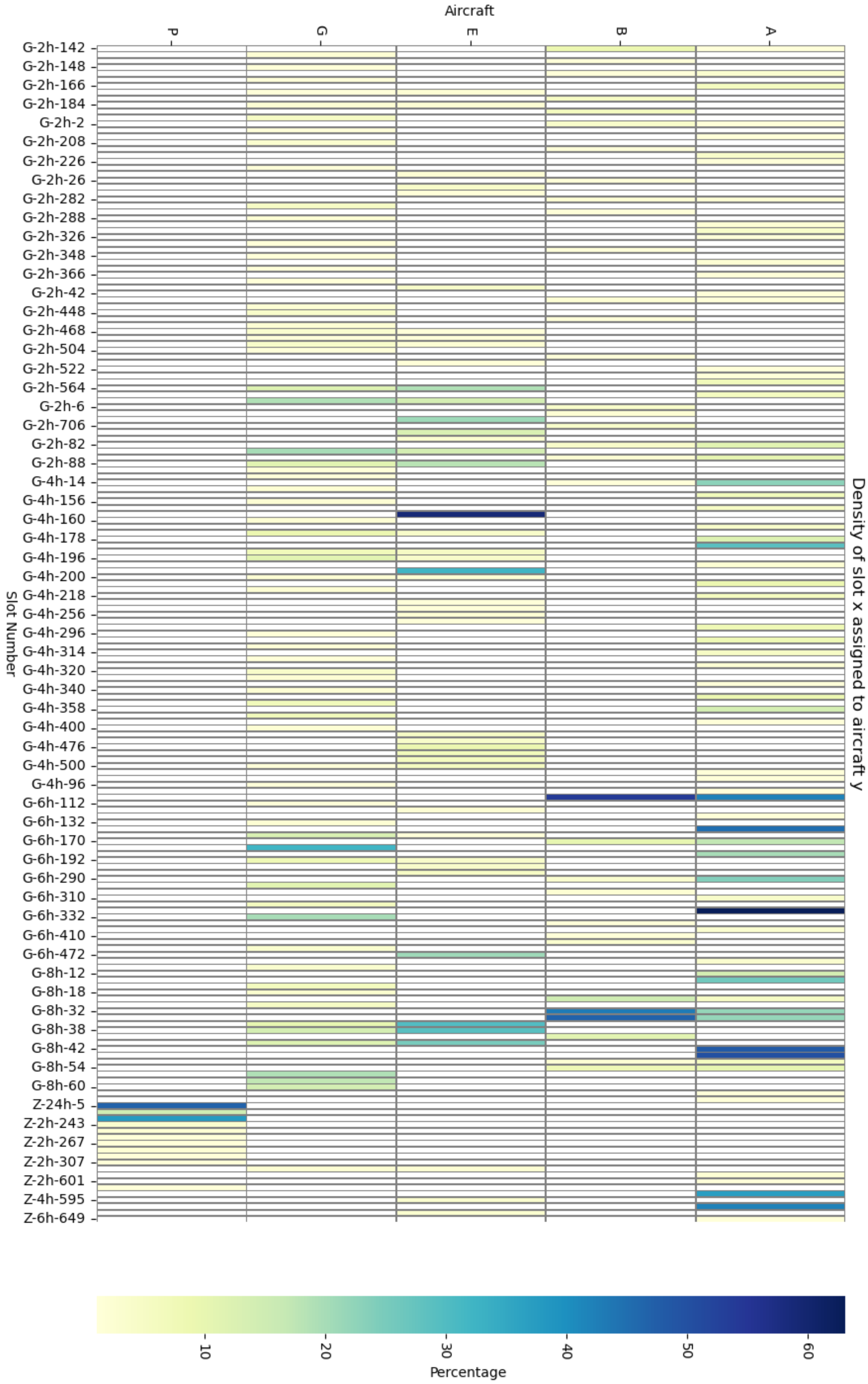

Figure 18: Density distribution of the use of slot x by aircraft y.

B. Two-Stage Stochastic Programming with Neural Network

A recommended solution framework to significantly lower the computational speed for the model developed in this paper is NEUR2SP created by Dumouchelle et al. [3]. Due to time constraints it could not be tested, but the suggested adaptation of this paper's MIP formulation into a two-stage stochastic version is given below.

General principle of the methodology

Two-stage stochastic programming (2SP) is a mathematical approach used to make optimal decisions under uncertainty. In the first stage, decisions are made before the uncertainty is revealed, optimising for the best deterministic outcome. In the second stage, once the uncertain parameters become known, additional decisions (recourse actions) are made to mitigate the impact of the uncertainty. The objective is to minimise the total expected cost, considering both the deterministic first-stage decisions and the expected costs of stochastic second-stage decisions.

Dumouchelle et al.'s NEUR2SP framework [3] is visualised in Figure 19. The framework creates an easier-to-solve, substitute MIP by splitting the original 2SP MIP problem into first-stage decision tuples (x) and a scenario set with corresponding expected second-stage objective values. The second-stage costs are then predicted using supervised deep learning, namely the Rectified Linear Unit Neural Network (ReLU NN). This is done for a given scenario and a known first-stage decision. The resulting trained model is then converted into an approximate MIP which can be solved by any solver of choice.

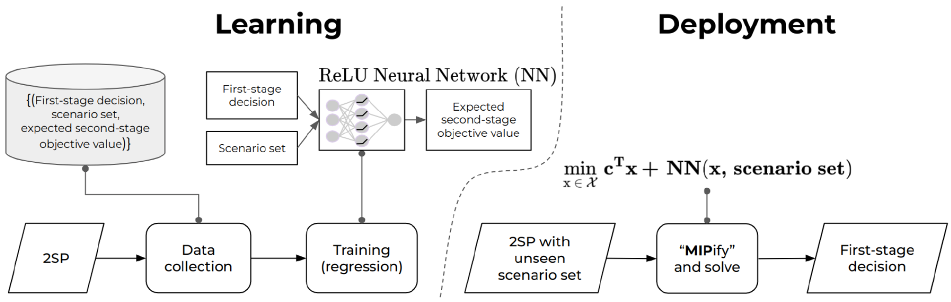

Figure 19: The Neur2SP framework visually explained by Dumouchelle et al. [3]. The deterministic first-stage is represented by x and the second-stage by the 'scenario set'.

Neur2SP's strength lies in its general applicability, supporting formulations with both linear and nonlinear objectives and constraints, as well as binary and continuous decision variables. It also provides high-quality solutions faster than generic methods like the extensive form (EF). The case studies in this paper, based on classic MIP formulations, are solved within seconds. The training phase, which can be done offline, lasts up to three hours, similar to other comparable learning algorithms [3]. Furthermore, the Neur2SP algorithm's code is open-source, facilitating adaptation for different research projects. However, a major limitation is the 'black box' nature of some algorithm steps, such as scenario embedding.

2SP applied to combined TA, MS, PM scheduling

For the problem description in this work the first stage consists of all decisions not influenced by the uncertain parameters: the predictors' EOL, so the preventive and corrective MS. This formulation is the same as for the MS part in section 4. The second stage adds the RUL-based MS and TA under different EOL scenarios (Ω) of the predictors. The objective function then is the deterministic cost of the first stage (X) plus the sum of the probability (p_ω) times cost for each scenario of the second stage (Y_ω), see Equation 0.

Below is the mathematical formulation for the suggested 2SP. The same sets, decision variables and parameters as in the MIP formulation presented in section 4 are used; only the set Ω for the different scenarios is added. Also, the same equation numbering is used as section 4, for easy reference.

Objective function

$$Min : X + \sum_{\omega \in \Omega} p_\omega \cdot Y_\omega \tag{0}$$

First stage: preventive and corrective MS

$$\begin{aligned} X = & \sum_{g \in G_{due}} T_{g,s=Fict} \cdot W_{DUE} \cdot C_g \\ & + \sum_{s \in S} M_{p,s} \cdot s_{dur_s} \cdot W_{GROUND} \\ & + \sum_{s \in S} (\sum_{g \in G_{prev}} T_{g,s} \cdot W_{INTprev} + \sum_{g \in G_{corr}} T_{g,s} \cdot W_{INTcorr}) \cdot C_g \\ & + \sum_{g \in G_{defer}} T_{g,s=Fict} \cdot W_{DEFER} \cdot C_g \\ & + \sum_{p \in P} (1 - Clean_p) \cdot W_{CLEAN} \\ & + \sum_{p \in P} \sum_{s \in S} Slack_{p,s} \cdot W_{SEQ} \end{aligned} \tag{6}$$

Task - slot scheduling

$$\sum_{s \in S} T_{g,s} = 1 \quad \forall g \in G \tag{7}$$

$$\sum_{s \in S} (1 - DD_{g,s}) \cdot T_{g,s} = 0 \quad \forall g \in G_{due} \tag{8}$$

$$\sum_{s \in S} (1 - GndTime_{g,s}) \cdot T_{g,s} = 0 \quad \forall g \in G \tag{9}$$

$$\sum_{g \in G_p} T_{g,s} \leq MaxTasks_s \cdot M_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{10}$$

$$\sum_{g \in G_p} T_{g,s} \cdot WH_g \leq MaxWH_s \cdot M_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{11}$$

$$\sum_{g \in G_p} T_{g,s=Fict} \cdot ndd_g \leq MaxTasks \cdot (1 - Clean_p) \quad \forall p \in P \tag{12}$$

Aircraft - slot scheduling

$$\sum_{p \in P} M_{p,s} \leq 1 \quad \forall s \in S \tag{13}$$

$$\text{if } M_{p,s} = 1 \rightarrow \sum_{s \in S_{overlap}} M_{p,s} = 0 \quad \forall s \in S \quad \forall p \in P \tag{14}$$

$$G_{p,loc_s,t_{s_{end}}+1} \leq 1 - M_{p,s} + Slack_{p,s} \quad \forall p \in P \quad \forall s \in S \tag{15}$$

Connection TA and MS (ground time)

$$\text{if } M_{p,s} = 1 \rightarrow \sum_{t_{s_{start}}}^{t_{s_{end}}} G_{p,loc,t} = s_{dur_s} \quad \forall s \in S, \forall p \in P \tag{16}$$

Second stage: predictive MS and TA

$$\begin{aligned} Y = &\sum_{p \in P} \sum_{f \in F} OC_{p,f} \cdot F^{\omega}_{p,f} \\ &+ \sum_{f \in F} CC_f \cdot C^{\omega}_f \\ &+ \sum_{p \in P} \sum_{n \in N} \sum_{t \in T} GC_n \cdot G^{\omega}_{p,n,t} \end{aligned} \tag{1}$$

Tail Assignment

$$\sum_{p \in P} F^{\omega}_{p,f} + C^{\omega}_f = 1 \quad \forall f \in F \tag{2}$$

$$\sum_{n \in N \cap O(n,p)} G^{\omega}_{p,n,t} - \sum_{n \in N \cap T(n,p)} G^{\omega}_{p,n,t} + \sum_{f \in F \cap O(n,p)} F^{\omega}_{p,f} - \sum_{f \in F \cap T(n,p)} F^{\omega}_{p,f} = B_{n,p} \quad \forall n \in N \quad \forall p \in P \tag{3}$$

$$\text{if } F^{\omega}_{p,f} = 1 \rightarrow \sum_{t_{arr}}^{t_{arr}+TAT_p} G^{\omega}_{p,n,t} = TAT_p \quad \forall p \in P, \forall f \in F_{arr_t} \tag{4}$$

$$\sum_{\text{Fbid } (n,p)} G^{\omega}_{p,n,t} = 0 \quad \forall t \in T \tag{5}$$

Connection TA and MS (ground time)

$$\text{if } M^{\omega}_{p,s} = 1 \rightarrow \sum_{t_{s_{start}}}^{t_{s_{end}}} G^{\omega}_{p,loc,t} = s_{dur_s} \quad \forall s \in S, \forall p \in P \tag{16}$$

RUL-based maintenance task scheduling:

$$\begin{aligned} RUL^{\omega}_{pr_{p,t}} = RUL^{\omega}_{pr_{p,t-1}} - \sum_{f \in F_{dep}} F^{\omega}_{p,f} \cdot f_{dur_f} + Mpr^{\omega}_{p,t} \cdot (RUL^{MAX}_{pr_p} - RUL^{\omega}_{pr_{p,t-1}}) \\ \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \end{aligned} \tag{17}$$

$$RUL^{TH}_{pr_p} \leq RUL^{\omega}_{pr_{p,t}} \quad \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \tag{18}$$

$$\text{if } Mpr^{\omega}_{p,t} = 1 \rightarrow \sum_{s \in S_{pos}} T^{\omega}_{pr,s} = 1 \quad \forall p \in P \quad \forall t \in T \quad \forall pr \in PR \tag{19}$$

Task - slot scheduling

$$\sum_{s \in S} (1 - GndTime_{g,s}) \cdot T^{\omega}_{g,s} = 0 \quad \forall g \in G \tag{9}$$

$$\sum_{g \in G_p} T^{\omega}_{g,s} \leq MaxTasks_s \cdot M^{\omega}_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{10}$$

$$\sum_{g \in G_p} T^{\omega}_{g,s} \cdot WH_g \leq MaxWH_s \cdot M^{\omega}_{p,s} \quad \forall s \in S \quad \forall p \in P \tag{11}$$

Part II

Literature Study

Executive Summary

The complex, dynamic and large-scale environment of an airline's operational planning makes it challenging to model, especially considering the interdependencies of all the various aspects. In this study only aircraft-related operations, namely tail assignment (TA) and maintenance task scheduling (MS), are considered. In addition to preventive and corrective maintenance tasks, predictive tasks and their inherent uncertainty are considered. The main goals of optimising the aircraft flight- and maintenance schedules is to create more fleet availability and lower the costs [12].

The objective of this survey is to review literature in the combined research field of airline operations, condition-based maintenance and stochastic optimisation in order to understand the historical research trend, the current state-of-the-art and, following from these, the research gap.

Both deterministic aircraft network operations and its stochastic form (called the aircraft recovery problem, ARP) with disruptions and recovery strategies have been extensively researched. They are often formulated as mixed integer linear programs (MILP) [1]. More than 80% of the recent ARP models are solved with heuristic, instead of exact, algorithms, showing the complexity involved [20]. ARP models have to make the trade-off between computational speed and completeness of recovery actions, completeness of fleet types and size and inclusion of all maintenance constraints [20].

Disruptions in maintenance task scheduling is not well researched yet in the context of airline operations. The most complete and recent work is by Van Kessel et al. [34]. They developed a MILP formulation with a rolling time-window for scheduling an airline's maintenance activities with a stochastic disruption process of the arrival and frequency of tasks.

Most studies concentrate solely on maintenance or network planning due to the increased challenges associated with merging both. Only two published studies developed an integrated TA and MS approach. Lagos et al. [39] created a Markov decision process (MDP) scheduling formulation for a commercial airline. Approximate dynamic programming (ADP) allowed to solve this high-dimensional problem. Iwata et al. [38] modelled a military aviation operation, focussing on supply lines, using discrete event simulation (DES). Not all solution methods are clearly explained in this paper. Both these frameworks are simplified model formulations and do not include predictive maintenance tasks.

In condition-based maintenance tasks are activated by models predicting the future health state (and moment of failure) of the system of interest. This inherently comes with an uncertainty. The main state-of-the-art uncertainty quantification (UQ) models for data-driven predictive models are [4]: Gaussian process regression (GPR), Bayesian neural network (BNN), neural network (NN) ensemble and deterministic methods spectral-normalized Gaussian process (SNGP) and deep neural network (DNN) GPR. GPR scores very high on quality, but does come with a high computational cost and curse of dimensionality. BNN, and especially NN ensemble and deterministic methods scale very well to higher dimensions while delivering a medium to high solution quality [4].

The most relevant stochastic optimisation methods to introduce uncertainty into high-dimensional scheduling models are Monte Carlo sampling, two-stage stochastic programming (2SP), (partially observable) Markov decision process (POMDP), (deep) reinforcement learning (DRL) and risk-based optimisation [50]. In practice, these are often combined, depending on the nature of the problem formulation.

A research gap is found in the integration of condition-based maintenance task scheduling with the tail assignment, particularly when addressing uncertainties. A stochastic optimisation model that combines these two scheduling approaches as an operational decision-support tool for airline schedulers is thus the suggested research direction. It will provide valuable insights for the formulation of new airline policies that may embrace predictive maintenance next to preventive and corrective practices. In addition, it will contribute to academic knowledge by demonstrating how uncertainties can be integrated into the comprehensive stochastic optimisation model.

Contents

List of Figures

List of Tables

List of Acronyms

AMR Aircraft Maintenance Routing

ARP Aircraft Recovery Problem

AOG Aircraft On Ground

BD Big Data

BNN Bayesian Neural Network

CBM Condition-Based Maintenance

CM Corrective Maintenance

CVaR Conditional Value at Risk

GPR Gaussian Process Regression

HI Health Index

LOF Line Of Flights

MC Monte Carlo

MCMC Monte Carlo Markov Chain

MI(L)P Mixed-Integer (Linear) Programming

ML Machine Learning

MRO Maintenance, Repair and Overhaul

MS Maintenance Scheduling

NN Neural Network

NSRE Non Safety Related Equipment

OOD Out Of Distribution

PM Preventive Maintenance

POMDP Partial Observable Markov Decision Process

(D)RL (Deep) Reinforcement Learning

RUL Remaining Useful Life

SP Stochastic Programming, 2SP: Two-Stage Stochastic Programming

TA Tail Assignment

TAT Turn Around Time

UQ Uncertainty Quantification

1 Introduction

Airline operations are complex, large-scale environments and highly dynamic. Many different stakeholders, facilities and assets all need to come together at the right moment and the right place to avoid flight disruptions. This includes, for example, resource management of staff, hangars and spare parts and the coordination of aircraft flight schedules and maintenance schedules. The inherent interdependencies of all these aspects present challenges in effectively optimising an airline's operations planning.

The scope of this research focusses on network planning and maintenance planning of aircraft. Both in academic research and industry practice these two planning steps have been optimised separately and sequentially. Due to the reciprocal impact the two steps have on each other, it makes sense to optimise them simultaneously. However, when combined the already complex problem formulations quickly turn intractable. Only recently a few research works started integrating tail assignment and maintenance task scheduling to find a more efficient aircraft schedule.

Another research trend is the development of predictive health models and condition-based maintenance strategies. Similarly to network planning optimisation, this trend aims to more effectively schedule maintenance to maximize asset availability and minimise costs. Condition-based maintenance, and the uncertainty that is inherent to predictions, has been researched across many different industries. Also for this research field most works focus on improving one aspect, such as the health state predictive model, but not on bringing these different elements together in a larger scheduling problem.

The objective of this survey is to review literature in the combined research field of airline operations, condition-based maintenance and stochastic optimisation in order to understand the historical research trend, the current state-of-the-art and, following from these, the research gap. Both literature tailored to aviation and to other fields where operational maintenance is researched were taken into account to understand the broader context. Due to the already large scope, most research works referenced in this report are aircraft and / or airline focussed.

First, an overview of the way of working and modelling techniques for airline operations is presented in Chapter 2. This chapter is split in four sections: the broader context, network planning, maintenance planning, and the interaction between these two. Next, Chapter 3 zooms in on predictive models for condition-based maintenance, including uncertainty in these prognostics. This is followed by an overview of stochastic optimisation approaches in Chapter 4. Finally, a recommendation for follow-up research is suggested in Chapter 5.

2 Airline Operations

This chapter will first give a general description of airline operations and the key concepts in Section 2.1. Next, the two planning processes in airline operations that are in the scope of this research are expanded: tail assignment in Section 2.2 and maintenance task scheduling in Section 2.3. Finally, Section 2.4 gives an overview of previous work modelling the interaction, up to a certain extent, between these two processes and an indication for future research.

2.1 Definitions and context

The operations department of a commercial airline involves a diverse range of tasks related to personnel and resources, such as aircraft, facilities and spare parts. The terminology used for the timeline of operations is strategic, tactical and operational planning. Strategic planning typically occurs approximately ten to one year before actual operations, tactical planning spans from one year to one month ahead, and operational planning is carried out from one month up to the day-to-day management of operations [7]. The level of uncertainty is highest during strategic planning and gradually decreases as more information becomes available for accurate forecasting.

The four sequential stages in airline planning process are [8]:

1. Schedule design: strategic process which determines the number of flights serving a specific route, their operational days and times, effectively determining the airline's market share [7].
2. Fleet assignment: aircraft subtypes are allocated to specific flight routes within the established network at a tactical level.
3. Aircraft maintenance routing (AMR): aircraft are allocated to specific flight legs while adhering to maintenance constraints and aviation authorities' regulations. This step is also called tail assignment (TA). Maintenance task scheduling and the scheduling of larger letter checks define the aircraft's maintenance constraints.
4. Crew scheduling: consists of crew pairing and rostering in a tactical / operational time frame.

The four steps are briefly discussed below. It should be noted that in the majority of airlines, these four steps are carried out in a sequential manner where each step is optimized separately due to their intricate nature. However, these planning processes are closely interconnected, with each step required to give a valid input for the subsequent ones. In academic literature, certain studies combine multiple phases into an integrated model, especially for the first two steps. This is done by Lohatepanont and Barnhart [9], for example, who heuristically solved a mixed integer linear program (MILP) to optimize for the flight schedule and fleet assignment.

The key integration of interest is between tail assignment and maintenance task scheduling, which can be found in Section 2.4. Since the other integrations are not in the scope of this study, they are not elaborated upon here.

Schedule design
Schedule design encompasses two primary subproblems: frequency planning, which involves deciding the number of flights on a route, and timetable development, which focuses on determining the specific times for the flights. Both subproblems involve many stakeholders, requiring consideration of factors at the Origin-Destination (OD) market level, network level, fleet availability, and competitive landscape. Automating this process using operations research models is further complicated due to dependencies on data such as competitors' schedules and anticipated market share, which are influenced by the airline's operating schedule [7] .

Due to these complexities and the want of consistency (to keep loyal customers on board), schedule planning often involves partial modifications to previous years' schedules rather than creating entirely new schedules from scratch [7]. This could be done by adjusting flight timings within a specified window around the initially scheduled time [10], for example. An alternative is to pre-determine flights eligible for cancellation or expansion, from which a specific set is chosen [9].

Fleet assignment
Fleet assignment's primary goal is to align supply with demand, aiming to minimize spilled passengers (more bookings than capacity on a certain flight) and spoiled seats (unsold seats). This analysis can be expanded by including the recapture of spilled capacity through similar flights or itineraries [11]. Additionally, minimizing operating costs is a key consideration in fleet assignment, emphasizing the fuel-efficient allocation of aircraft to longer routes [7], for example.

Several constraints come into play in the fleet assignment problem, including fleet-route compatibility and routing feasibility. For instance, not all aircraft types can land at every airport due to runway length constraints.

AMR
Aircraft maintenance routing, or tail assignment, is the focus of this research and therefore this step is elaborated on in the next section, Section 2.2. The necessary scheduling step that provides the constraints for TA, is creating the maintenance task schedule, shortened to MS in this report. This scheduling step is separately discussed in Section 2.3. Afterwards, the interaction and combined modelling of these two problems (TA and MS) is discussed in Section 2.4.

Crew scheduling
The crew scheduling problem is the final step in the planning framework. This particular challenge revolves around assigning cabin and cockpit crew members to flights while ensuring adherence to all aviation regulations. Also this step is a complex process in itself and is often solved in two steps. The first step, known as the crew pairing problem, focuses on generating feasible pairings, usually sequences of flights spanning one to five days. Subsequently, these pairings are assembled into more extended sequences during the crew rostering problem, which results in the creation of rosters for each crew member [8].

2.2 Tail assignment

Tail assignment is the assignment of specific aircraft ('tail'), not just a subtype but the distinct registration, to a flight leg. Hub-and-spoke airlines usually schedule rotations instead of single flight legs. Rotations are combinations of two (or more) flights that start and end at the hub. To further simplify scheduling, these rotations can be combined in a line of flights (LoF). The TA controllers at commercial airlines often consider a time window in the order of days, up to one week. Most research models hence also apply a daily or weekly time horizon [12].

The goal is to most efficiently assign tails to flight legs according to the priorities set by the airline. The performance metrics for TA are selected accordingly, and typically include: flight plan stability (minimum cancellations), punctuality (minimum delays), schedule stability (minimum late changes), fleet availability and total costs [12].

2.2.1 Deterministic tail assignment models

Barnhart et al. [13] have proposed an integrated model addressing fleet assignment and aircraft maintenance routing issues. Their approach utilizes strings, representing sequences of flight legs starting and ending at maintenance stations while adhering to maintenance feasibility constraints related to maximum flight hours and flight time. The proposed model, formulated as a MILP, aims to minimize the overall cost by assigning fleets specific sets of strings that cover the entire flight schedule while satisfying continuity constraints. To overcome the challenge of a large number of feasible strings, the authors employ a branch and price algorithm, demonstrating the model's applicability to real-life scenarios involving over

1000 flights, 89 aircraft, and 9 fleet types. This string-based formulation has been widely adopted in the literature by various authors, including Ageeva [14], Rosenberger et al. [15], and Sarac et al. [16].

In contrast, Sriram and Haghani [17] focus exclusively on short and medium-haul flights pre-assembled into daily routes. These routes are assigned to aircraft in such a way to meet all maintenance constraints within a one-week timeframe. Their MILP model, solved heuristically, is tested on smaller instances, demonstrating a 5% gap from the optimal solution. However, the comparison with Barnhart et al. [13] is challenging due to Sriram and Haghani's reliance on pre-defined sequences of daily flights, limiting flexibility. While Barnhart et al.'s model incorporates flight hour constraints, Sriram and Haghani struggle to solve a version of their model that includes flight hour considerations. On the other hand, Sriram and Haghani's model includes more maintenance constraints, considering availability constraints at each maintenance station and various types of maintenance stops at different intervals.

Sarac et al. [16] also introduce an operational model that addresses daily aircraft routings, satisfying aircraft-level maintenance constraints. Their MILP model assigns strings to aircraft, ensuring compliance with remaining flight hours and guaranteeing a night stop at a suitable maintenance station for all aircraft with tasks due the following day. The model, developed using a connection network structure, is tested on a fleet of 32 aircraft, covering 19 stations with 175 daily flights, and proves capable of providing optimal solutions.

A framework used by several algorithms is the time-space network [18]. A parallel time-space network creates an individual network for each aircraft consisting of nodes at every location (airports) and for every timestep. Those are then integrated using constraints for total network coverage. This approach allows the modelling for specific tails and their requirements regarding maintenance and flight destinations. In order to allow for another aircraft taking over a certain route in case of delay or cancellation, each aircraft network contains the information of all other aircrafts' flight arcs.

2.2.2 Disruptions in tail assignment

Disruption sources and recovery strategies
Within TA, disruptions encompass a wide range of factors, resulting in delays and/or cancellations. They can stem from a variety of sources, like weather changes, airspace changes, airports closing or limiting flights, aircraft health status and disruptions during turn-around-time (TAT) [1].

The disruptions mentioned above lead to uncertainty in parameters such as scheduled departure time (which is then referred to as estimated departure time), scheduled arrival time (becoming estimated arrival time), and the assignment of a specific tail to a flight leg or maintenance slot. They should be dealt with quickly and adequately to allow nominal operations as much as possible.

Disruptions can be dealt with in a reactive or proactive manner. Reactive approaches are the actions taken by operation controllers when a disruption occurs. There are several measures that can be adopted to address these disruptions, including delaying flights, cancelling flights, change aircraft, utilization of reserve aircraft, adjusting the flight speed and alterations in TAT processes [19]. Proactive approaches include the capacity allowance of reserve aircraft, buffer time and resource flexibility (namely short cycles that allow for cancellations and more opportunities for swapping tails).

The aircraft recovery problem (ARP) has been modelled in several ways. The biggest challenge is the computational time, since operations recovery happens live and thus a feasible solution is required within minutes. For a complete overview of many different academic works for aircraft recovery the reader is referred to the literature review by Hassan et al. [20].

One framework that manages to reduce the computational time to under one minue is the work by Vink et al. [18]. The model considers aircraft maintenance schedules and passenger itineraries, and indirectly also crew limitations. The model iterates with each disruption that is introduced, and the latest updated schedule is used as an input for the next iteration. An ILP problem was formulated using parallel time-

space networks and solved dynamically. The ILP problem was reduced by considering only one subset of the aircraft fleet at the time. This is also done in current practices by manual scheduling in airlines. A selection algorithm defines which aircraft are part of the considered subset.
Interestingly, the model also includes a human-in-the-loop by giving feedback about violated requests when an aircraft cannot be scheduled to its final destination.

The fact that more than 80% of the recent ARP models use heuristic algorithms, instead of exact algorithms, only proves ARP's complexity [20]. ARP models have to make the trade-off between computational speed and completeness of recovery actions, completeness of fleet types and size and inclusion of all maintenance constraints. In addition, Hassan et al. [20] found that many models underestimate the disruption costs due to the use of constants for delay and cancellation costs. In reality, the relation between passenger satisfaction and delay times is non-linear [21].

Summary research trend
Hoi-Lam et al. [1] made an extensive literature review covering uncertainty in AMR. They clustered the latest (2012-2020) research papers in four categories: uncertainty and robust solutions, big data and machine learning (BD and ML), smart technologies and integrated information support systems. This is visualised in Figure 2.1.

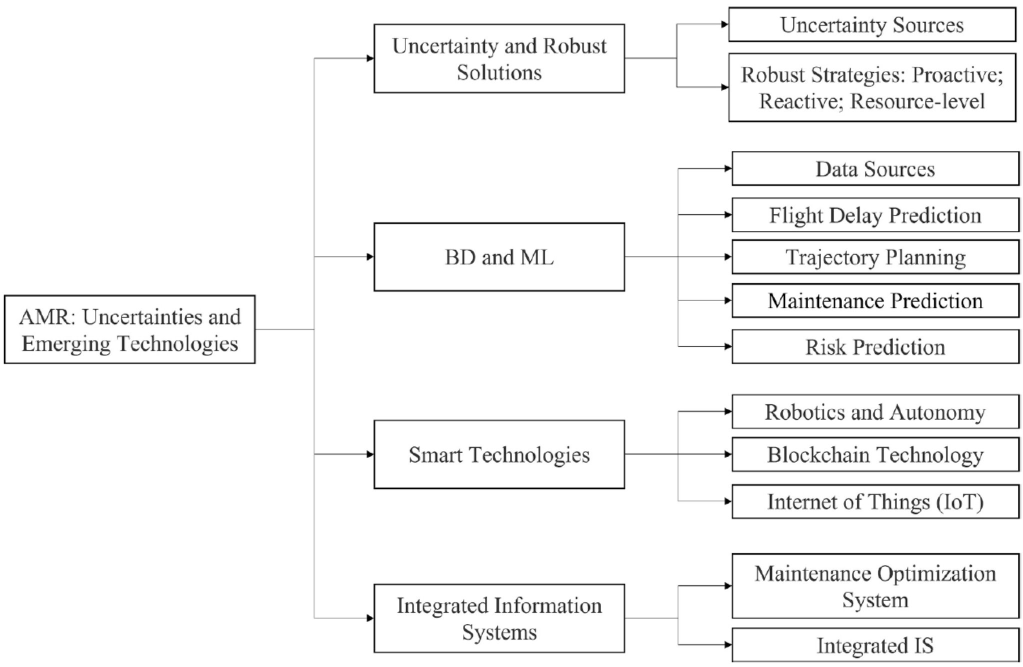

Figure 2.1: Classification of research trends between 2012-2020 in AMR uncertainty, as visualised by Hoi-Lam et al.[1]. Big data (BD) and machine learning (ML) are clustered together.

The research trend of uncertainty in AMR is summarised in Figure 2.2. Most research papers focus on uncertainty sources stemming from the airline, and a few include airport and maintenance sources. Stochastic and reactive strategies (recovery, rescheduling and maintenance responses) are preferred over robustness planning (buffer / flexibility creation). BD and ML show great potential, since their strength is good performance for large datasets and predictive scheduling, which is applicable to TA. This research direction is still in an earlier stage compared to conventional methods [1].

The recommended future directions for uncertainty in AMR research by Hoi-Lam et al. [1] are: collaboration with blockchain (airlines, airport, MROs), recovery and proactive adjustment by internet of things (IoT) for dynamic and proactive schedule adjustments, ML and optimization for uncertainty prediction when sufficient high-fidelity data is available and autonomous systems to enhance automatic optimisation and schedule adjustments.

Author	Year	Uncertainty Source			Strategies			Maintenance Considerations	Optimization model	Solution
		Airline	Airport	Maintenance	Buffer & Flexibility	Stochastic	Reactive			
Lapp and Cohn	2012	√					√	Opportunity	Mathematical programming	CPLEX
Sherali et al.	2013	√			√			Opportunity	MIP	Benders decomposition
Başdere and Bilge	2014	√					√	Opportunity Capacity	IP (Connection Network)	B&B/metaheuristic
Dunbar et al.	2014	√				√		/	Mathematical programming	Heuristic
Maher et al.	2014	√					√	Opportunity	MIP	Benders decomposition/ Column generation
Rosales et al.	2014			√			√	Resources	System dynamics model	Simulation
Liang et al.	2015	√			√			Opportunity Capacity	MIP (LOF network)	Column generation Heuristic
Hu et al.	2015	√					√		MILP	CPLEX
De Bruecker et al.	2015			√		√		Crew Rosters	MILP	Heuristic
Zhu et al.	2015	√		√		√		Staff availability	Two-stage stochastic recovery model	Heuristic and retiming strategy
Bosson and Sun	2016		√			√		/	Job-shop scheduling	Stochastic programming
Jamili	2017	√			√			/	MILP	Heuristic
Ahmed et al.	2017a	√				√	√	Opportunities	MINP	Optimization-simulation
Zhang and Mahadevan	2017	√					√	/	Re-routing optimization model	Simulation
Ahmed et al.	2017b	√			√			/	MIQP	Simulation-optimization
Yan and Kung	2018	√				√		Opportunities	Robust IP	Column/row generation
Kenan et al.	2018	√				√		Opportunities	Two-stage stochastic MIP model	Column generation
Ahmed et al.	2018	√			√			Opportunities	MIP	CPLEX
Kammoun and Rezg	2018	√					√	/	Mathematical programming	Heuristic
Maher et al.	2018	√					√	Opportunities	MIP	Branch and price; Heuristic
Liang et al.	2018	√	√				√	Swappable maintenances	MIP	Column generation heuristic
Scala et al.	2019	√	√			√		/	IP	Optimization & Simulation
Lagos et al.	2020	√		√	√		√	Resources	IP	Heuristic dynamic algorithm
Lee et al.	2020	√	√		√	√	√	/	Stochastic integer programming	Approximate solution algorithm
Qin et al.	2020			√		√		Resources	Stochastic integer programming	Benders decomposition

Figure 2.2: Summary by Hoi-Lam et al.[1] of papers covering uncertainty and robust solutions in AMR.

2.3 Maintenance task scheduling

Maintenance task scheduling is the process of assigning open tasks (also called actions) and the corresponding aircraft to a maintenance slot. This process' performance is typically measured in the following parameters: schedule efficiency (minimum deferred workorders), schedule stability (minimum late changes), fleet availability, maintenance slot capacity utilization and costs [22].

2.3.1 Maintenance definitions and strategies

There are three main maintenance strategies in industrial industry for which different names are used, all referring to a specific nuance. The high-level division is [23]:

- Corrective / reactive / run to failure / incident-based maintenance: maintenance is performed when failure occurs. This is the most expensive approach, because a spare parts inventory, (extra) labour availability and system downtime are required.
- Preventive / scheduled / time-based maintenance: periodically scheduled maintenance when probability of failure thresholds are exceeded, based on operational hours, cycles or calendar days. The advantage is that this can be scheduled and planned for in advance, The downside is that higher costs are unnecessarily involved when corrective actions are taken too early.
- Predictive / condition-based / health-aware maintenance: uses data to model future performance of the considered systems and schedules maintenance based on this expected performance. The goal is to optimise maintenance planning and hereby increase fleet availability and reduce maintenance costs for example [2] .

Airlines practice corrective and preventive maintenance, with a rising interest in CBM practices [22].

The different approaches are visualised in Figure 2.3. Here the degradation of a certain system is shown over time (black line: actual condition state, black dotted line: estimated condition state). For both preventive and corrective maintenance it is assumed that the degradation follows an inverse exponential trend. For predictive maintenance a prognostic model predicts the decay, which results in a trend closer to the real degradation. The horizontal red dotted line shows the failure threshold for this component. As can be seen, preventive maintenance action (blue vertical dotted line) is executed at a set time interval, well before expected failure. Corrective maintenance action (orange vertical dotted line) is taken after the component has failed earlier than expected. Predictive maintenance (green vertical dotted line) action takes place just before expected failure, and not necessarily at a set interval. The buffer between expected and actual failure should be smaller and more accurate compared to predictive maintenance, since the degradation is modelled more precisely.

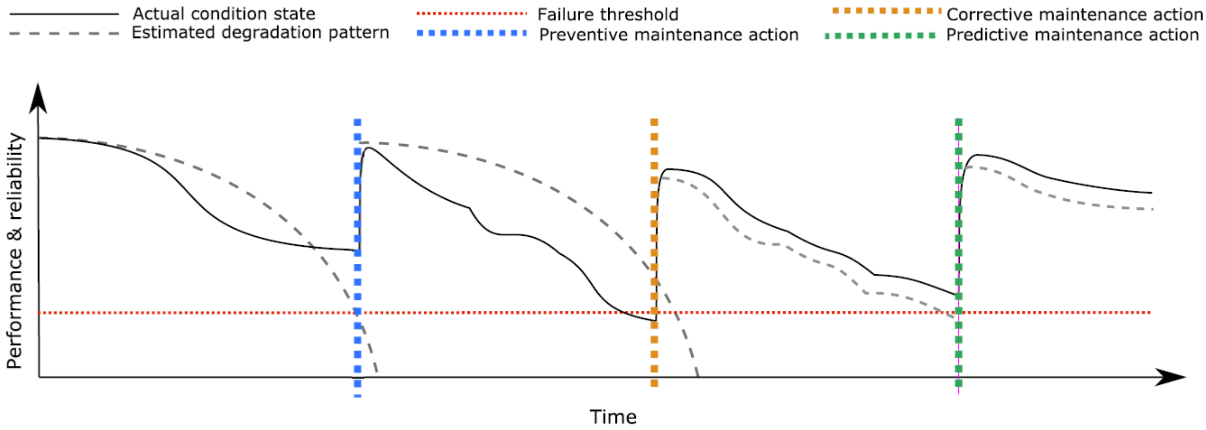

Figure 2.3: Graphical explanation of the three maintenance strategies: corrective, preventive and predictive by Esteban et al. [2].

Note that the focus in the remainder of this chapter is on the current airline practices of preventive and corrective maintenance. CBM is further expanded on in Chapter 3.

Aircraft maintenance is primarily dictated by the rules set by the aircraft manufacturer and aviation authorities, like EASA in Europe [24]. When the aircraft fails to meet these stipulations, it is deemed unfit to fly and referred to as an 'aircraft on the ground' (AOG). AOGs should be prevented due to the substantial expenses it incurs for the operator.

The scheduling of maintenance (MS) can be categorized into planned and unplanned maintenance. Planned maintenance for aircraft is a routine conducted daily, involving either line maintenance (during operations, usually when the aircraft is stationed at the gate / stand) or maintenance carried out in hangars. In order to maximize fleet availability during the day, most (longer) maintenance activities are carried out during the night. Maintenance tasks are organized into work packages, and more extensive and repetitive sets of tasks are referred to as letter checks. These are A checks (scheduled every 2-3 months), C checks (scheduled every 18-24 months), and D checks (scheduled every 6-10 years) [25]. These tasks are scheduled close to their due dates, to avoid waste of resources, and include a reasonable buffer time determined by the airline itself.

Unscheduled maintenance arises when a problem is identified by mechanics, reported by the crew, or a maintenance program change is mandated by engineers or authorities. If such unscheduled maintenance issues cause the aircraft to be grounded, they are categorized as non-routine (NR) maintenance and call for immediate attention. If the issue can be deferred, it is designated as either a 'minimum equipment list' item (MEL) when it concerns regulatory demands or a 'non-safety-related equipment' (NSRE) item when it concerns an airline's own criteria.

2.3.2 Deterministic maintenance scheduling

Letter checks scheduling
The scheduling models for letter checks initially began with a non-stochastic, heuristic method as outlined in Boere et al. [26] in the late 1970s. It evolved into several different frameworks, such as a dynamic programming framework coupled with Monte Carlo simulations by Deng et al. in 2020 [25]. A com-

mon practice is to differentiate between long-term (C, D checks) and short-term (A checks) scheduling, primarily due to their considerably different timeframes. Longterm maintenance scheduling is scarcely researched and in most airlines schedulers plan longterm letter checks manually based on experience [25]. Also, they are dependent on external MRO contracts, which put many constraints on the schedule. Here the goal is feasibility, not so much optimisation.

Maintenance task scheduling
The general problem of maintenance task scheduling involves assigning tasks to specific maintenance slots, and researchers have approached this challenge with varying perspectives on the planning horizon. The range of perspectives includes a four-year planning horizon by Witteman et al. [27] and a decision support model for operational use by Callewaert et al. [28].

Several frameworks are experimented for the assignment of tasks to a maintenance slot, such as a bin-packing problem with a heuristic solving method [29], a MIP formulation [30], and a Monte Carlo simulation [31] [28]. Recent examples are elaborated on below.

Witteman et al. [27] address the optimal assignment of recurring and deferred defects to pre-defined A and C-check maintenance slots, framing the problem as a bin-packing problem. They propose a constructive heuristic that efficiently assigns tasks to checks based on intervals, followed by a prioritized allocation of remaining tasks. Testing the algorithm on a fleet of 45 aircraft over a four-year plan, the heuristic demonstrates the ability to solve instances in less than fifteen minutes, with results within 5% of an exact optimization method.

Shaukat et al. [30] adopt a more tactical perspective, presenting a task scheduling model for line maintenance with a weekly scheduling horizon. The model, formulated as a MIP, considers fixed aircraft rotations and schedules tasks within line maintenance slots while accounting for workforce constraints. Two solution methods are proposed, with the exact solution algorithm providing solutions approximately 3.5% better than the sequantial heuristic approach. However, the optimal approach's run time limits its real-life application, prompting consideration of the heuristic's practical adequacy.

A notable observation across these models is their tendency to schedule tasks as close as possible to their due dates, aiming to reduce wasted intervals. Shaukat et al. [30] stands out as an exception, introducing a preferred buffer to address this issue.

2.3.3 Disruptions in maintenance task scheduling

The main disruptions in MS are: unscheduled work orders (WO), AOG, unavailable or delayed manpower / machines / material / facilities (4M), delays in arriving aircraft and aircraft change in operations [20]. Up to 40-60% of maintenance tasks are non-routine, unplanned or have no prior knowledge [32].

Disruptions in maintenance task scheduling is not well researched yet in the context of airline operations. Most works with applications in airline operations focus on disruptions in the flight network, but keep the maintenance slots and constraints fixed [20]. A little more recent exception is the work of Liang et al. [33]. They model the ARP using a column generation-based heuristic. The model allows for swapping of maintenance slots, given the task can still be scheduled before going due.

Historically, the focus of optimisation was on the combined scheduling of letter checks and maintenance tasks. With the increased knowledge and method developments in prognostics, condition-based maintenance becomes of interest [34]. This asks for a different approach to maintenance task scheduling, where more flexibility, unknowns and late changes are required to avoid disruptions.

Callewaert et al. [28] introduce an operational decision support framework that suggests task postponement or addition during maintenance checks facing delays or advancements. The framework assesses the status of operations and ranks critical and non-critical tasks based on cost, operational risk, and associated delay using Monte Carlo simulation. While savings are predicted in a simulated operational setting, information about run time is absent.

The main limitation is the consideration of only currently scheduled tasks when delaying maintenance to future opportunities, potentially causing over-scheduling and later delays. They also require task-specific data on the probability of failure, posing challenges in obtaining such data for all maintenance tasks. The scalability of these models is limited due to the numerous generated alternatives.

Van Kessel et al. [34] are the first to model maintenance disruptions in the unique environment of airline operations. They developed a MILP formulation for scheduling an airline's maintenance activities with a stochastic disruption process of the arrival and frequency of tasks. Also, constraints on the resources' availabilities are in place. A rolling time horizon is used, in which new data and a new current schedule (output of the previous iteration) is available for every iteration. Each iteration deals with one schedule window, and the total schedule horizon is covered by several overlapping schedule windows, separated one day apart. Two time horizons are chosen: one of 10 days to match that of current practices by airlines, where human schedulers manual execute the rescheduling process, and one to 120 days to analyse the full extent of the model. The main limitations of this framework are the exclusion of line maintenance and no possibility for slot flexibility. The strength of this framework is its demonstration of increased schedule stability and potential revenue when using a decision support tool for maintenance scheduling.

2.4 Interaction network and maintenance operations

The majority of studies concentrate solely on maintenance or network planning due to the increased challenges associated with merging both. Furthermore, a significant number of airlines contract out their maintenance operations to MRO companies, reducing the airline's influence in maintenance scheduling. In this study, it is presumed that the airline retains complete control. This section examines frameworks in existing literature where there is a degree of integration between TA and MS. A summary of the different academic works can be found in Table 2.1.

Including the relation with TA in MS-focussed models
The following papers focus on developing a model for MS, but do include the input of or consequences on TA in their model.

Duffuaa et al. [35] developed a maintenance simulation which also analyses its performance in network operations for Saudi Arabian Airlines. It should be noted that the analysis is only qualitative. The objectives are to minimize flight cancellations, delays and repair time and to maximize effective use of maintenance resources. This is tested for several maintenance policies. The maintenance module includes spare parts inventory, station availability and staff requirements. The interaction of the modules is not elaborated on.

In Sachon et al.'s work [36] the influence of a maintenance policy on flight delays, cancellations and in-flight safety are quantified. The model can be used as a decision-support tool for management to identify which parts of the maintenance process (such as spare parts and skills) should be upgraded to avoid flight delays. Sachon et al. developed a probabilistic risk analysis model and the main focus is on the trade-off between delays and safety. The main limitation is the model only analysing one system at the time (the leading edge of a commercial jet in the presented case study) and the need for data estimation, both airline data and expert knowledge.

Öhman et al. [37] developed a new maintenance approach by introducing a frontlog in work packages. The anticipated tasks in a work packages could then be postponed when necessary without going due. This creates flexibility in maintenance scheduling and hence an improved slot reliability, which reduces delays in the flight network. This connection to network operations is not quantified.

Integrated MS and TA simulation and optimisation
An integrated MS and TA simulation and optimisation framework is developed by the following four research projects.

Reference	TA / MS based	Application	Methodology	TA considerations	MS considerations	Components introducing uncertainty
Dufuaa et al. 1999 [35]	MS	Commercial airline operations; short-haul / long-haul fleet; time horizon not specified	Qualitative analysis; conceptual modelling	Flight schedule; Flight cancellations; Flight delays	Minimum repair time; Spare parts inventory; Station availability; Staff requirements	-
Sachon et al. 1999 [36]	MS	Commercial airline operations: 1 system analysis (leading edge slats)	Probabilistic risk analysis model	Flight schedule; Flight cancellations; Flight delays	Parts availability; Staff requirements and availability; In-flight safety	All are stochastic variables except: actual maintenance time, task deferral allowed or not, flight delay allowed or not
Öhman et al. 2020 [37]	MS	Commercial airline operations: long-haul fleet	Discrete event simulation; No optimisation; Uncertainty not stochastically modelled	Aircraft arrival time, Aircraft departure time, Aircraft location	Frontlog time buffer	Random variable for: maintenance tasks (volume)
Iwata et al. 2013 [38]	TA & MS	Military aircraft operations: supply chain spare parts	Discrete event simulation, No optimisation; Uncertainty not stochastically modelled	Flight schedule (mission type and length)	Repair time; Parts inventory	Random variables for: mission length, part life time, replacement time, repair time, transit times (spare parts)
Lagos et al. 2020 [39]	TA & MS	Commercial airline operations: short-haul fleet	MIP formulation; Markov design process with approximate dynamic programming; Monte Carlo simulation for expected value	Line of flights schedule; Cancellations	Staff requirements; Space capacity; Repair time; Parts availability	Stochastic process for: task arrival, task criticality, task due date
Varenna 2023 [40]	TA & MS	Commercial airline operations: long-haul fleet	MILP formulation; Discrete event simulation, Optimisation: simplex and barrier solver (Gurobi)	Rotation-based flight schedule; Flight delays; Flight cancellations; Reserve aircraft	Staff availability; Repair time;	Stochastic process for: flight delay; AOG occurence
Den Hollander 2023 [41]	TA & MS	Commercial airline operations: short-haul fleet	MIP formulation; Time-space network; Optimisation: simplex and barrier solver (Gurobi)	Flight schedule; Flight cancellations	Repair time; Staff availability; Space capacity Predictive maintenance tasks	-

Table 2.1: Literature overview of aircraft operations models that combine TA and MS. The column 'TA / MS based' indicates what is main focus of the model formulation. In case this is MS, still TA objectives and constraints are considered to a larger extent than conventional, but the model is not build with the idea to optimise both simultaneously.

Firstly, the work of Iwata et al. Iwata et al. [38] modelled the maintenance and logistics, including flight missions, of military aviation operation. The focus is on parts inventory at different local military outposts and vehicle mission capability.

The simulation is performed by a discrete event simulation (DES). Aircraft are assigned to maintenance slots or missions, based on the mission requirements (maintenance could be postponed to fly part of a mission), the aircraft state and geographical locations. Uncertainty is introduced by defining parameters as random variables. These are parameters such as: mission length, part life time, part replacement time, repair time and vehicle transit times within the maintenance processes. The performance metrics considered are mission capable rate versus delivery time from depot to local outpost and the local inventory level of parts. Only four systems are considered, thus the amount of maintenance tasks is very limited.

Returning to the user case of commercial airlines, Lagos et al. [39] developed a dynamic and stochastic maintenance scheduling problem, which schedules tasks on a daily basis for a shorthaul fleet. The decision maker module first assigns a line of flights (LOF) to each aircraft, then assigns aircraft to the maintenance slot and at last selects a subset of open maintenance tasks for this specific aircraft. Uncertainty is introduced by the unknown time a maintenance task is disclosed. The decision maker reacts to newly issued maintenance tasks from a random set and anticipates future scenarios. Hangar space and available manhours are taken into account and it is assumed that maintenance is always done during a fixed night slot. Aircraft itineraries can be changed to facilitate for scheduling maintenance tasks, for example by swapping tails on a certain LOF. Future consequences of the decisions made on the day are included as costs in the objective function.

The total system is modelled as a Markov decision process (MDP). The curse of dimensionality is solved by using an approximate dynamic programming (ADP) approach. Different approaches are tested to solve the MIP problem: a myopic policy, a value function approximation and a rolling horizon. The latter two approaches are combined in a hybrid policy, which gave the best results. A deterministic version of the model solving the whole time windown in one run is used for the underestimation of the expected costs. Also a model mimicking the manual approach of solving TA and MS sequentially based on airline decision rules is tested.

The case study for LATAM airlines analyses 30 days of operations for a shorthaul fleet of 30 aircraft. Five different scenarios are tested in the case study for LATAM airlines: reduced task execution flexibility, modified congestion levels, MS without TA integration, decentralized maintenance plans and the inclusion of periodic tasks. In order to increase the use of residual maintenance capacity, Lagos et al. recommend to schedule tasks requiring less resources first. Also, they found that sharing maintenance resources between airlines gives economic benefits.

Next, a work that is also for a commercial airline but instead focusses on the longhaul fleet. Varenna [40] developed a modular, simulation and optimization model integrating both MS and TA. There are four modules: the scheduler, split into an MS submodule and TA submodule, operations manager, recovery controller and recovery planner. The modules are called sequentially and provide input to each other. Both fixed and flexible maintenance slots are included. The optimal moment in time to do maintenance (at the preferred anticipation time before the task due date) is included by variable costs for assigning tasks to a slot. The TA planner is rotation based and includes reserve aircraft. Primary and propagated departure delays and AOGs are introduced as operational disruptions. The recovery planner can decide to delay or cancel a rotation or maintenance slot, swapping tails for a certain rotation or maintenance slot, use of reserve aircraft and postponing maintenance slots.

The whole system is modelled stochastically as a discrete event simulation (DES). The different modules are MILP formulations. The propagated delays are accounted for by considering the hub disruption state, ie sampling from an exponential distribution to determine the delay times.

The case study includes a longhaul fleet of 50 aircraft and a time window of 180 days. Two scenarios are investigated, namely the effect of using an extra reserve aircraft and the effect of different AOG durations. The use of additional reserve aircraft significantly improved the cancellation factor, but is an expensive measure.

Finally, Den Hollander [41] is the first, to the author's knowledge, to include health-aware models of aircraft parts in the integrated MS and TA model. Hence a condition-based maintenance strategy is explored in the broader context of operations. The objectives are to minimize operating / cancellation / maintenance costs, minimize maintenance ground time, limit rescheduling actions and improve schedule robustness. Disruptions are included in the form of AOGs.

The MS problem is approached as a MILP adapted from Van Kessel et al. [34] and only fixed maintenance slots are considered. RUL predictions are made for two systems only, meaning the tasks related to those parts are scheduled based on their health. The other tasks are scheduled using the conventional preventive maintenance strategy. The TA problem is modelled as a layered time-space-network (TSN) approach, based on the work of Vink et al. [18], with added constraints defining the health-state of parts. The linear RUL prediction model is based on the health-index of a part which is compared to a predefined threshold. For every flight the RUL decreases with one unit and for every maintenance activity the RUL is reset to its maximum value. The entire time window is solved in one run and the MS and TA modules are solved in parallel (both are assigned equal weights with respect to each other).

The case study consists of five shorthaul aircraft, a time window of 14 days and two AOGs. Two models are compared, one with health-aware models and one without. Since the cancellation factor and aircraft availability are comparable between the two models, the benefits of including condition-based maintenance cannot conclusively be shown.

Conclusion
All the papers discussed in this chapter conclude that an integrated TA and MS approach shows promising results for reducing total costs and increasing schedule efficiency. All of the integrated models discussed above consider the classic preventive and corrective maintenance strategies, except for the work by Den Hollander [41]. Due to the limitations in his work, clear benefits or drawbacks of including condition-based maintenance cannot be concluded. Hence advancing the framework he started is one of the interest for future research.

To give a better understanding of the full problem, the next chapter will dive into predictive models for CBM and uncertainty quantification.

3 Predictive Modelling under Uncertainty

In aviation, airlines use a combination of corrective and preventive maintenance, with CBM becoming a new area of interest. In this chapter first some definitions and context are given for the key concepts in CBM, followed by a brief outline of different prognostic models. Next, uncertainty in predictive models and CBM decision-making is discussed.

3.1 Definitions and context

For a more elaborate description of the different maintenance approaches, please refer to Section 2.3. For the sake of clarity and overview hereby first a brief recap of the CBM definition is given followed by more details on how to model this strategy.

Condition-based maintenance
Condition-based maintenance (CBM) is a maintenance strategy that predicts the time until a system reaches its failure threshold. Maintenance actions are only activated when there is a high enough risk of failure. The goal is to reduce the amount of unnecessary maintenance and prevent unexpected failures. This should reduce maintenance costs without giving up safety constraints and increases operational availability.

CBM approaches appear in literature already since the late 1970s [42] and have since been developed across a wide range of industries, such as wind farms, railways, maritime, (nuclear) plants, construction and aviation. An extensive overview of the use of CBM in different industries and systems can be found in the work of Quatrini et al. [23] and Zonta et al. [43].

For this report the terms CBM and predictive maintenance are used interchangeably, as it is defined in Section 2.3. Note: predictive maintenance can be performed on an entire system, a subsystem or a single part. When this chapter refers to a 'system', also a subsystem or a part could be considered for the same principle. In cases where multicomponent systems are considered, possible interdependencies should be taken into account.

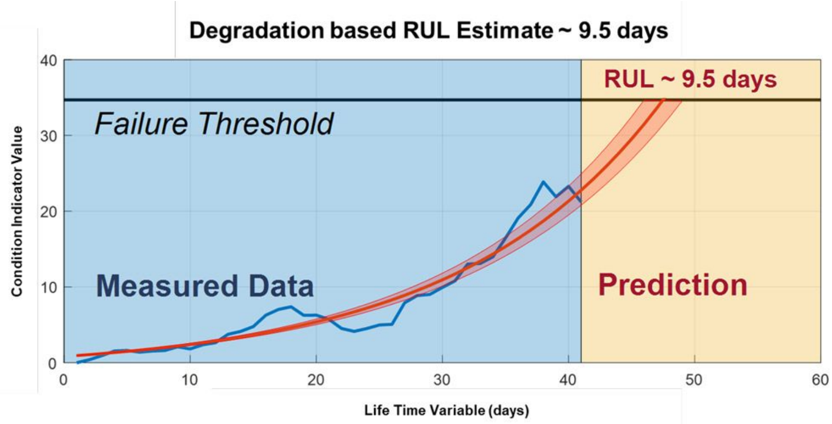

Figure 3.1: Visualisation by Mathworks [3] of a high-speed bearing's predictive degradation model.

Health index and remaining useful life
Predictive models often characterizes a system using either its remaining-useful-life (RUL) or health-index (HI). HI is a specific indicator that shows the system's present level of degradation. Comparing this with a predefined threshold determines whether the system is considered healthy or not.

The RUL quantifies the amount of time or number of cycles that can still be completed before the system necessitates replacement. In Figure 3.1 data of a high-speed bearing is shown as an example to visualise the RUL computation. The bearing's real-time condition is shown by the blue line (measured data). The red line shows the exponential degradation model fitted to the condition data, with the prediction uncertainty shown in the red shaded area. The remaining useful life (RUL) is computed from the current day (vertical line, just over 40 days) till the estimated time when the monitored condition reaches its failure threshold. Hence, this graph shows an estimated RUL of 9.5 days.

Note that prognostics in this report is defined as predicting the RUL. The broader field of prognostics and health management (PHM) also includes fault detection and fault diagnostics, which is outside the scope for this study.

3.2 Predictive models

Below a brief overview of the different steps in predictive models is given.

Data acquisition
In order to build a predictive model, data is required as an input. It drives the next steps in creating the model, since the problem formulation and the use of a certain ML algorithm are determined by the way a system is monitored and the nature of the data [2]. The four main methods described by Esteban et al. [2] for data mining are: (1) sensor connection measuring for example vibration frequency, accelerations, temperatures or electrical signals, (2) external capture by for example image analysis or geopositioning, (3) inspection logs like failure and maintenance logs and (4) simulation to obtain training data that mimic the real system behaviour. These methods can also used in a hybrid form.

After collecting data that ensures that the final dataset used in the model is representative for the operating conditions of the system analysed [44], feature extraction and selection is required. This means the bulk of sensor data is filtered to find the measurement variables that are directly related to the underlying fault. This step has two big advantages: increased predictive outcomes accuracy and reduced computational complexity. The data can be filtered using statistical methods, subspace-aided monitoring and engineering knowledge based methods. Which method will perform the best is depended on the data characteristics and should be chosen accordingly [44].

Model strategies
Predictive models can be broadly categorized into three strategies: physics-model-based, knowledge-based, and data-driven [2]. In the first approach, models are constructed based on the fundamental principles of physics describing the degradation and failure of a system. In contrast, the data-driven approach relies on historical data and real-time sensor outputs, often incorporating machine learning (ML) methods, to predict remaining-useful-life (RUL). In the knowledge-based approach, human expertise, typically acquired through extensive on-the-job experience, takes precedence. It is worth noting that hybrid approaches are increasingly prevalent, where data-driven models are merged with physical principles and existing expert knowledge, particularly given the surge in data availability and the growing use of machine learning in recent years [2].

Variables describing the damage state of a system can be direct or indirect state variables. The most common RUL modelling techniques for direct state variables are regression-based, Wiener processes, gamma processes and Markovian-based [23]. For modelling indirect state variables stochastic-filtering, covariate hazard and hidden (semi-)Markov models are frequently used [23].

Model assessment
Saxena et al. [45] explain that the three main categories to assess predictive models are metrics on: algorithm performance, computational performance and cost-benefit. The parameters most frequently used are: accuracy, precision, mean square error and mean absolute percentage error, and, when applicable, the influence of human factors [46].

3.3 Uncertainty in predictive models

In this section, first an overview is given of uncertainty types, their sources and interpretation. This is followed by methods to quantify uncertainty and integrate this in a CBM approach.

3.3.1 Uncertainty types, sources and interpretation

In continuation of the uncertainty distinction made in engineering risk and reliability analysis, uncertainty in ML is generally classified in two categories: aleatory and epistemic [47]. Aleatory, derived from the Latin word meaning game of chance, is nonsystematic noise inherent to the physical nature of the system. In predictive maintenance models, this shown by the stochastic input / output data. In contrast, epistemic uncertainty is systematic and could in theory be known, but is not in practice. There are two subtypes of epistemic uncertainty: model-form uncertainty, due to simplification of the real problem into a computational model, and parameter uncertainty due to assumptions and simplifications in the model calibration and training [47]. It should be noted that in ML models often both types of uncertainty occur and cannot always be analysed separately [4]. An overview of different sources of uncertainty and their causes is shown in Figure 3.2.

Types, sources, and causes of uncertainty in ML.

Type	Source	Cause(s)
Aleatory uncertainty	Observational uncertainty (model input and output)	Measurement noise (e.g., sensor noise in measuring inputs/outputs of ML models)
	Natural variability (model input)	Variability in material properties, manufacturing tolerance, variability in loading and environmental conditions, etc.
	Lack of predictive power (model input)	Dimension reduction, non-separable classes in input space (classification), etc.
Epistemic uncertainty	Parameter uncertainty	Limited training data, local optima of ML model parameters, low-fidelity training data[a], etc.
	Model-form uncertainty	Choices of neural network architectures and activation and other functions, missing input features, etc.

[a] Data fidelity is the accuracy with which data quantifies and embodies the characteristics of the source [66].

Figure 3.2: Overview by Nemani et al. [4] on the sources of aleatory and epistemic uncertainty in ML. Reference (66) in the original table refers to the work of Sanjay and Sriram, which is reference [5] in this report.

It should be noted that although the uncertainty source of a certain physical component can be both aleatory or epistemic, the interpretation of the uncertainty in condition-based prognostics should be Bayesian (which is in line with epistemic approaches). A system's state is considered, which is deterministic. The system can only be in one state at the time, so there is a true state value, but this value is unknown. The probability distribution represent the state of the system that the analyst believes it is in. This is in contrast to physical probabilities that relate to the variability across several samples of the same system [48].

3.3.2 Uncertainty quantification

Predictions always come with a measure of uncertainty, as discussed above. Quantifying this uncertainty helps the user to understand to what extent the predictions can be trusted and thus also when, or how much, extra caution is required for decision-making based on these predictions. A prediction with high uncertainty may result in a high risk of violating safety constraints or financial loss, and these critical risks in turn trigger different decisions down the line. An additional advantage of uncertainty quantification is a better understanding of complex ML models, which 'black boxes' are difficult to grasp, because the UQ shows the confidence level of the model [4].

Well-known examples of predictive uncertainty are the probability of a result given as a percentage in classification problems and the use of confidence intervals for regression problems. A narrow confidence interval shows low uncertainty.

The main state-of-the-art UQ models for data-driven ML models are [4]: Gaussian process regression (GPR), Bayesian neural network (BNN), neural network ensemble and deterministic methods spectral-normalized Gaussian process (SNGP) and deep neural network (DNN) GPR. Nemani et al. [4] analysed these methods based on a simplified 2D regression problem with 800 samples, which resulted in a qualitative comparison as shown in Figure 3.3. Below the main methods, as summarised in the table, are elaborated on.

Note that deep learning (DL) models are not considered for this study, since they are outside of the scope due to time constraints and very limited prior experience.

A qualitative comparison of state-of-the-art UQ approaches covered in this tutorial.

Quantity of interest	Gaussian process regression	Bayesian neural network			Neural network ensemble	Deterministic method	
		MCMC	Variational inference	MC dropout		DNN-GPR	SNGP
Quality of UQ (e.g., measured by calibration curve)	High	High-medium[a]	Medium	Medium-low	High	Medium	High
Computational cost (training)	High[b]	High	High-medium	Low	Low	High	High
Computational efficiency (test)	High[b]	Low	High-medium	Medium	Medium-low	Low	Low
Ability to detect OOD samples	Strong	Weak	Weak	Weak	Moderate	Strong-moderate	Strong
Scalability to high dimensions	Low	Low	Medium	High	High	High	High
Effort to convert a deterministic to a probabilistic model	Not applicable	High	High-medium	Low	Medium	High-medium	High-medium
Ability to distinguish aleatory and epistemic uncertainty	Yes	Yes	Yes	No	Yes	No	No
Basis of UQ	Analytical	Sampling	Sampling	Sampling	Hybrid	Analytical	Analytical
Stability of quantified uncertainty to parameter initialization	High	High	High	Low	Medium	High	High

[a] Accuracy is largely affected by the quality of the assumed prior.
[b] Efficient only for problems of low dimensions (typically < 10) and small training data (typically < 5000 points).

Figure 3.3: Qualitative comparison of UQ approaches by Nemani et al. [4].

In the explanations below a distinction is made between testing data and training data. Training data are observed values and can be considered as 'evidence' of the real values. Testing data means the data points of interest that are to be tested, namely analysed using the algorithm.

GPR

Methodology
GPR is a Bayesian approach, meaning that a probability distribution is constructed over all possible values (in contrast to ML methods that learn the exact values for every parameter in a function). It assumes that the function to be constructed is drawn from a Gaussian process. As an example, consider the linear function $y = ax + b$, with a the unknown parameter of interest. The Bayesian approach first assigns a prior distribution, $p(a)$ and then redetermines probabilities based on the training data. This results in the posterior distribution $p(a|y,X)$. Next, the predictive distribution is determined by weighting all possible predictions with their calculated posterior distribution [49]. This prediction follows a Gaussian distribution, since the prior distribution and likelihood are assumed to be Gaussian also. The mean represents a point prediction and the variance its uncertainty.

GPR uses Bayes' rule as illustrated in Figure 3.4. The observations are the training data points of which the value is known. The prior is the blue dotted line for which the function (in this case $f(x)$) is specified. The GPR combines the prior, the likelihood function and the observations to make the prediction distribution, showing its mean and covariance (confidence interval). In the figure on the right the noise on the observation is specified, which influences the prediction distribution accordingly.

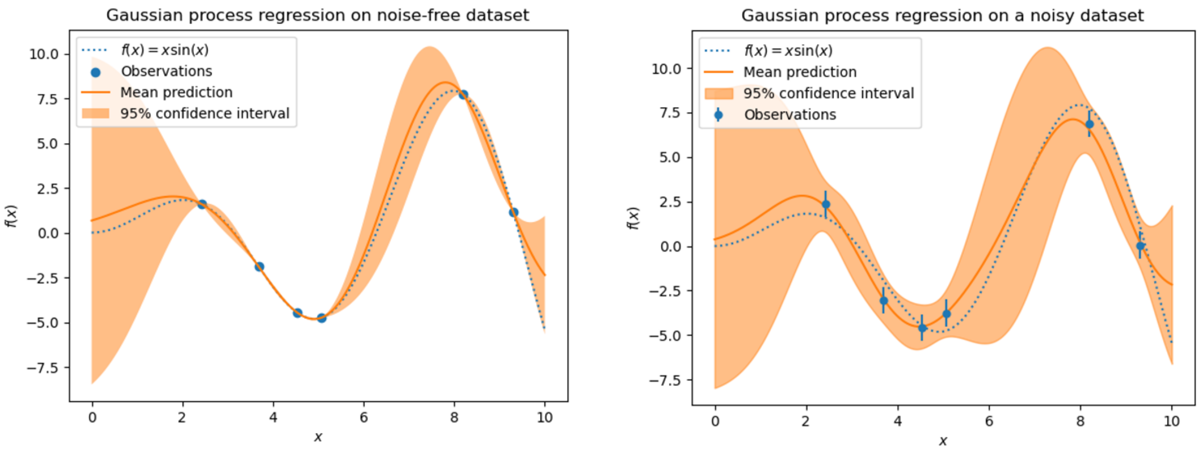

Figure 3.4: Illustrative example of GPR by Scikit Learn [1]. Left: dataset without noise. Right: dataset with noise.

The steps in GPR can be described in more mathematical terms as follows. First a GP prior is assumed: $f(x) \sim GP(m(x), k(x,x'))$. In the GPR prior, which is a multivariate Gaussian distribution independently, identically distributed Gaussian noise can be incorporated. The GP prior is tuned using model selection, which determines the form of the mean function and the covariance kernel. Generally, the mean is zero or the mean of the training dataset, thus a constant value. The covariance kernel function can have many forms, such as constant, linear, square exponential or a composition of several kernels. The radial basis function (RBF) is a commonly used function: $k(x,x') = \sigma_f^2 \exp\left(-\frac{\|x-x'\|^2}{2l^2}\right)$. The two hyperparameters to be tuned are the singal variance, σ^2 and the lengthscale, l. Now the GP posterior, i.e. the predictive distribution, can be computed.

Strengths
The strength of UQ using GPR is its ability to model both aleatory and epistemic uncertainty. The total predictive uncertainty is represented by the posterior (regression) variance for a test point. The aleatory uncertainty is determined by the variance of the additive white noise. The epistemic uncertainty is measured by the covariance vector between the training and testing point. When using such a distance-based covariance function, the GPR model results in low uncertainty for tests points close to data points used for training and results in high uncertainty at test points far away from training data points. Therefore, GPR is very suitable for fault detection in low-dimensional problems and small training data sizes. In addition, GPR is nonparametric, meaning not limited by a specific functional form. This makes it applicable for various datasets with different (or unknown) underlying distributions.

Limitations
Still, the relability of GPR results is not a given but depends on several factors, for example the test point location, the behaviour of the function to be fitted and the kernel and hyperparameter choices. When any of these factors are not correctly tuned, the prediction results will be less accurate. Moreover, 'extreme' test points, namely those far from the training data points, will be less accurate. This is because the

[1] `https://scikit-learn.org/stable/modules/gaussian_process.html`. Retrieved on 18-01-2024.

prediction error will continue to increase due to extrapolation, but the posterior variance will level at its peak value. A detailed explanation of this can be found in the work of Nemani et al. [4].

Also, GPR has a scalability issue for larger datasets and for high input dimensions. It does not scale well for larger training datasets, because of its complexity. A high-dimensional input space does not scale well, because the number of hyperparameters scale linearly with the number of input variables. This in turn requires a large number of training data, leading to a large covariance matrix. Generally, GPR is considered not efficient for a set of 30 features or more [49].

Neural networks

A neural network, inspired by the human brain, is a ML model composed of interconnected nodes arranged in layers. It learns to execute tasks by iteratively adjusting the weights assigned to the connections between nodes, based on the training data. The three main methods to solve the posterior in a Bayesian neural network (NN) are Markov Chain Monte Carlo (MCMC), variational interference and MC dropout. Another way of using neural networks, is NN ensemble. These different methods, their strengths and limitations are expanded upon below.

Markov Chain Monte Carlo
MCMC is used to sample the posterior distribution. Using ergodicity theorems, the average of the MCMC samples converges to the posterior expectation with high certainty, which is its strength. However, they are not suitable for high dimensions, because the high-probability regions of the posterior (called a set) becomes more singular for larger dimensions [4].

Variational interference
For variational interference (VI) methods the posterior is not sampled, but approximated with a parametric family of distributions, often Gaussians [4]. VI scales better and has a faster computational time than MCMC, because it redefines the interference of posterior distributions as an optimisation problem. However, it doubles the parameters the neural network has to approximate and is still computationally expensive [4].

Markov Chain dropout
MC dropout further stretches the scalability of Bayesian NN. The biggest advantages of MC dropout are its straightforward implementation in different neural networks, low computational costs and significant scalability ability. Its disadvantages are in the strong dependency between the uncertainty and the finetuning of hyperparameters and its instability to dropout rate and training parameters. Also, no prior knowledge can be included, which makes MC dropout applicable to only a few problems [4].

Neural network ensemble
Ensemble models consists of several individual models that are trained independently which are merged together to compute the final prediction. The separate models should be very diverse to achieve a high quality final solution. This diversity can stem from randomization approaches, meaning different subsets of the original data, or boosting approaches, which learns each iteration from the errors in the previous iteration. NN ensembles produce a high quality uncertainty prediction. However, its accuracy is significantly lower for test samples outside the training data distribution [4].

Deterministic methods

Deterministic techniques exhibit computational efficiency in UQ by executing a single forward pass, making them computationally less costly than BNN and NN ensembles. These methods perform very well for out-of-distribution (OOD) detection due to their distance awareness property. Yet, they often lack the ability to distinguish between aleatory and epistemic uncertainty. This limitation can be addressed by changing the network architecture by bringing in more hyperparameters. Also, deterministic methods have been reported to yield lower-accuracy UQ, compared to more established methods like MC dropout and NN ensemble. Recent benchmarking studies emphasize the need for further investigation into the calibration performance of deterministic approaches [4].

The predictive models described here result in a probability distribution of the desired parameter. This is just one variable in a maintenance scheduling problem. The question then is how to bring this stochastic variable into the full model. This is the focus for the next chapter.

4 Stochastic Optimisation

Optimisation challenges can be divided into deterministic problems and problems where uncertainty is present. Deterministic problems can be approached with traditional math programming (linear, integer and non-linear). Since the scope of this research includes uncertainty, the focus of this chapter will be on stochastic programming instead. In stochastic optimisation parameters that are uncertain are characterized with known probability distributions.

For an extensive analysis of 15 different stochastic optimisation approaches across different research and/or industry communities, the reader is referred to the survey by Powell [50]. This work also highlights that the consistent, main issue with stochastic optimisation is computational cost. The best way to tackle this challenge depends on the user case and model formulation.

The most relevant approaches for airline scheduling (so a problem formulation that involves a large number of variables) are explained below. In each case the method is briefly described and some recent researches that apply this in airline scheduling problems. Note that a combination of these methods can be used in solving the full optimisation problem. This is also shown in the two works discussed as illustrative examples in the last section.

4.1 Monte Carlo sampling / sample average approximation

The simplest, yet effective, approach to approximate the solution for a stochastic problem is by sampling. The expectation function, $\bar{F}$, is then approximated by the summation of a N amount of samples drawn from the probability distribution of the uncertain variabe, $p(x)$, defined by function $f(x)$, resulting in the full equation: $\bar{F} = \sum_{i=1}^{N} f(x_i) \cdot p(x_i)$. This method is also known as the Sample Average Approximation (SAA). To test how many samples should be taken for the expected value to converge, the expected value should be compared to that resulting from a very large sample size, M (so $M >> N$). The idea is that a smaller sample size sufficiently approximates the exact solution, thus reducing the computational load.

Note that this method does not provide a solution yet; a numerical algorithm is still required after this step to solve the complete objective function. To put it in the context of the MIP examples in the previous chapters, this numerical algorithm could also be an external solver such as Gurobi for example.

An example showing how SAA can be used in airline scheduling is presented in the last section, Section 4.6.

4.2 Two stage stochastic programming

Two-stage Stochastic Programming (2SP) makes two decisions: the first-stage before the uncertainty is realised, and the second-stage after. The random event that is realised in the second-stage influences the first-stage decision. Often a single first-stage decision is made (where no uncertainty is included yet) and several possible scenarios are included in the second-stage decision [51]. The net objective is to minimise the costs of the first decision ($c^T x$, with x the first-stage constraints) and the expected costs of the second decision ($\mathbb{E}[Q(x,\xi)]$, with ξ the uncertain data of the second-stage parameters) [51]:

$$\min_{x \in X} g(x) := c^T x + \mathbb{E}[Q(x,\xi)] \tag{4.1}$$

The numerical solution of this problem is solved by assuming that the random vector has a finite number of actual realizations. These are called scenarios (K) and can be expressed by multiplying their expected value with their probability of realization: $\mathbb{E}[Q(x,\xi)] = \sum_{k=1}^{K} p_k Q(x,\xi_k)$. Effectively, this leads to one large MIP formulation in which all possible scenarios are included with their own decision variables.

In practice, solving this numerically will proof too computational expensive [51]. One way to reduce the total amount of scenarios is by using Monte Carlo sampling for the expectation function, as already discussed before.

It should be noted that two stage stochastic programming involves making a decision at one point in time. When several, sequential decisions need to be made the formulation discussed above can be extended into what is called multi-stage stochastic programming. However, the number of possible scenarios grows exponentially with the number of stages, so only specific multi-stage problems are tractable [51].

Two stage stochastic programming combined with supervised learning
A state-of-the-art example of 2SP that succesfully manages to allow for scaling under low computational cost is the work of Dumouchelle et al. [6]. They developed a framework called *Neur2SP* that creates an easier-to-solve substitute MIP. It splits the original problem into first-stage decision tuples (x) and a scenario set with corresponding expected second-stage objective values. It then uses supervised deep learning, namely the Rectified Linear Unit Neural Network (ReLU NN), to predict the second-stage costs, given a scenario and first-stage decision. This trained model is then converted into an approximate MIP. Finally, the approximate MIP can be solved by any MIP solver. The full algorithm is visualised in Figure 4.1. One of the main achievements is the embedding of multiple scenarios into one single scenario that is a good average of the total set of scenarios. This helps to limit the computational time.

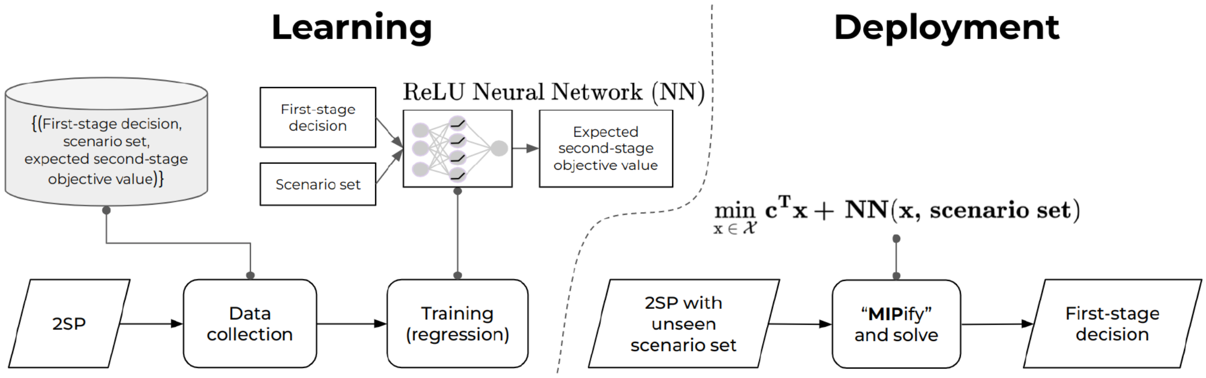

Figure 4.1: Visual explanation of Neur2SP by Dumouchelle et al. [6]. The first-stage, where no uncertainty is present, is represent by x and the second-stage by the 'scenario set'.

One of the advantages of Neur2SP is its generality, because it is applicable for 2SPs with both linear and nonlinear objectives and constraints, and both binary and continuos decision variables. Most other SP models are highly dependent on a specific case study [6]. In addition, it provides high-quality solutions in less computational time than generic methods such as the extensive form (EF). The case studies in the paper, which are classic MIP formulations, are solved in the order of magnitude of seconds. The training phase can be done offline, and can take up to three hours, which is similar to comparable learning algorithms [6]. Also, the Neur2SP algorithm's code is publicly shared and can hence more easily be adapted in different research projects. The main limitation is the 'black box' factor of some steps the algorithm takes, such as the embedding of scenarios.

4.3 (Partially observable) markov decision process

A markov Decision Process (MDP) makes sequantial decisions over discrete time-steps. At each time-step an action must be chosen from the defined set of actions, based on the current state. This decision then results in a reward (every action results in a known reward). The goal is to maximise the rewards over time [50]. The probability to be in a certain state should be known.

Most predictive degradation models are based on only partial information about the state of the system [50]. Therefore, many research studies focus on Partially Observable Markov Decision Process (POMDP). The probability vector of being in a certain state s is then referred to as the belief states. A known observation or indirect measurement gives an indication about the true unobservable state.

They are often solved using point-based solvers. For more information about this technique, the reader is referred to the survey by Shani et al. [52]. For an example how POMDP's can be used in a maintenance scheduling problem, please read the last section of this chapter, Section 4.6

4.4 Risk based optimisation

Another perspective to optimisation with uncertain parameters is a risk-based approach, often used for revenue management and (financial) risk planning [51]. In the airline industry, Cahill et al. [53] developed a model for pre-flight flight planning and briefing tasks and Leon et al. [54] created a value-at-risk framework to assign seat miles. Liu et al. [55] developed a risk-averse stochastic fleet assignment algorithm that is discussed in more detail in the last section, Section 4.6

Three common risk measures are: probability of failure (PoF), conditional value-at-risk (CVaR) and buffered probability of failure (bPoF) [56]. The latter two are conservative measures, because they add a buffer to the threshold. Traditionally these are implemented by adding safety factors, but promising results show the more efficient use of probabilistic approaches [56]. PoF only considers the frequency of hard-threshold failure events. CVaR characterizes failure risk as the expected magnitude of the failure and near-failure events and bPoF accounts for both the magnitude and frequency of failures and near-failures. A benefit of utilizing bPoF and CVaR risk measures is their ability to incorporate additional data regarding the distribution of the limit state function which primarily describes the magnitude of failure [56]. In a MIP formulation, these variables can be introduced as chance-constraints [51] where the variables need to be larger than a certain threshold.

4.5 Reinforcement learning

Reinforcement learning (RL) is very applicable to airline scheduling, because of its strength to be able to handle high-dimensional problem formulations in a computationally efficient way. It can be well formulated as a MDP. It learns the optimal behaviour by maximising the long-term reward. The data is not labelled and the algorithm learns the reward by observing the result of an action on the environment by trial and error, it receives feedback on the decisions made so to speak. The reward function defines the reward of a chosen action with respect to the environment and should be well defined since it drives the solution quality. This should be in line with the objective function. A disadvantage of RL is the need of large datasets and the difficulty to interpret the algorithm.

When the state space or action space are too large to be known or clearly defined, RL can be combined with deep learning. Deep reinforcement learning (DRL) often use a neural network as a foundation, because it allows for large state-action spaces. The neural network learns to map states to values, or state-action combinations to Q values (the reward). It has been used to model condition-based maintenance task scheduling for turbofan engines by Lee and Mitici [57] and for several multi-component systems on an aircraft fleet by De Pater and Mitici [58].

Again, this method is illustrated in the airline scheduling example given in the next section.

4.6 State-of-the-art examples in airline scheduling

Many works focus on just one aspect of all the different concepts and methods discussed in this and previous chapters, for example only on building a predictive model or only on improving an optimisation method. For a better understanding on how these integrate, two researches are explained in more detail that combines all the different 'building blocks' discussed before. These specific works are selected, because they are recent models for commercial airlines and combine several of the stochastic optimisation techniques. The first example is on maintenance task scheduling which is part of the scope of this research, whereas the second example about fleet assignment is not. However, it serves the purpose of showing how different stochastic optimisation methods can be applied very well. In addition, fleet assignment is closely related to tail assignment.

Maintenance task scheduling: POMDP and DRL

Tseremoglou et al. [59] proposed an approach that schedules all maintenance tasks (preventive, corrective and condition-based) for a commercial airline and also includes capacity and availability constraints. The RUL predictions of the predictive components are updated daily and its uncertainty follows a normal distribution.The framework consists of two stages: first the best maintenance policy at component level is determined using a POMDP algorithm, next the maintenance schedule at aircraft fleet level is created with a DRL algorithm. A case study on an international airline showed a maintenance cost reduction of 46% for the prognostics-tasks when compared to a corrective maintenance approach, with 96% of the condition-based components scheduled on time.

The underlying research from Tseremoglou et al. [60] for the framework in the previous paragraph evaluated the effect of prognostics uncertainty in fleet maintenance scheduling. Using a support vector regression (SVR) the RUL is predicted for several components. A Partially Observable Monte Carlo (POMC) algorithm with a rolling time windown is then applied to solve the maintenance scheduling problem. The uncertainty in the RUL means that there is not a deterministic task due date, but a probability distribution for the end-of-life of the component in question. This EOL probability distribution is translated into a beneficial replacement window. Preventive maintenance, repairing the component before it fails, results in less maintenance costs than corrective actions (repair after failure). However, scheduling repair too early, will ultimately result in more maintenance actions during the operational use of the component, which in turn drives up the maintenance costs.

These best maintenance action for each time window, based on the health degradation of the system, are included in the POMC algorithm as belief states. A deep Q-learning algorithm is then applied to solve the full schedule problem. The authors conclude that maintenance tasks for predictors (the components for which a RUL is computed) with an uncertain above 29% are considered high-risk. The impact of these high-risk maintenance actions could be minimised by adding more flexible maintenance slots in the specific time window.

Fleet assignment: risk-averse SP and SAA

An example of risk-based stochastic optimisation for airline fleet scheduling is the work by Liu et al. [55]. A risk-averse two-stage stochastic mixed-integer programming model is made based on a MILP formulation. In the first stage aircraft families are assigned to flight legs and in the second stage aircraft types are assigned to these flight legs (adhering to the predefined families). It is assumed that the passenger demands for individual itineraries are stochastic, and the variability in actual scenarios can be characterized using a conventional probability distribution, such as a truncated normal distribution.

A mean-risk model is applied which minimises the total expected costs and the risk value. The risk value is defined as CVaR of the first-stage decision variables. This risk-averse 2SP is then rewritten as a deterministic equivalent. For a large number of scenarios, this MILP becomes too large to solve for commercial solvers such as CPLEX. Therefore, a SAA algorithm is introduced. After drawing N samples, the upper and lower bound of the objective value is calculated along with the respective variances of the estimator, and finally the optimal solution is computed.

The model is tested on a case study with 72 flight legs, 159 itineraries, 3 aircaft families and 6 aircraft types. The findings suggest that in practical scenarios, a risk-averse model performs better than a risk-neutral model when considering both the robustness of solutions and the decision-makers' risk preferences. It is recommended to dynamically vary the sample size N for the SAA to find the balance the computational costs and estimation quality (in this work sample size N was fixed).

5 Follow-up Research Recommendation

First the research gap concluding this literature review is presented, followed by a suggestion for follow-up research.

5.1 Research gap

As the previous chapters have shown, it is clear that there exists a research gap concerning the integration of condition-based maintenance with tail assignment, particularly when addressing uncertainties. The outcomes of a study in this research direction will provide valuable insights for the formulation of new airline policies that may embrace predictive maintenance. In addition, it will contribute to academic knowledge by demonstrating how uncertainties can be integrated into an comprehensive stochastic optimisation model for this user case.

The relevant research areas are illustrated in Figure 5.1. The research gap is thus in combining the middle and most right circle.

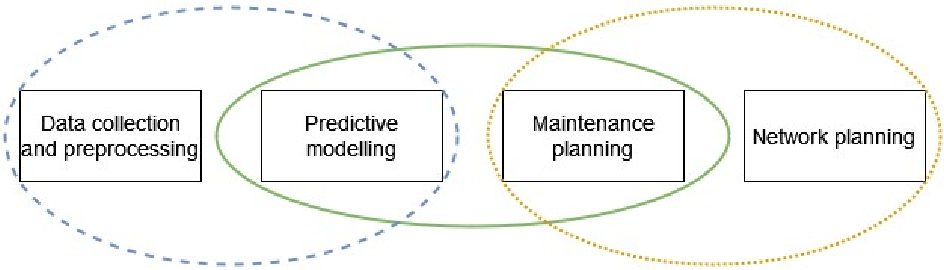

Figure 5.1: The four main research areas in the scope of this study are: data collection and preprocessing, predictive modelling, maintenance planning, and network planning. Typically in existing literature, these aspects are explored independently, occasionally merging two blocks into a integrated model. The three circles in the visualization indicate which blocks are commonly combined in such models.

5.2 Objective and scope

A suitable research objective could be to evaluate the business value of a stochastic optimization model that combines condition-based maintenance scheduling with the tail assignment problem and the associated uncertainties.[1]

With this goal in mind, a decision support tool could be designed for use within the operational context of a commercial airline. Consequently, the model must consistently provide feasible solutions, strive to complete computations in the order of minutes, introduce uncertainty multiple times a day, and adhere to a time window of several days, while also ensuring schedule feasibility assessments for a period of several months.

Long-term schedule design, passenger-related and personnel-related planning, and strategic maintenance scheduling (C, D letter checks) are not in the scope of this project.

Hence, the following research question is suggested:
How effective is the integration of condition-based maintenance scheduling with the tail assignment including uncertainty in an optimisation model for airline operations? [2]

[1] The research objective and research question are the same phrasing as stated in the earlier submitted research methodologies report by the same author as this literature report. These two sentences can thus be considered as a quote.

[2] See footnote 1.

Bibliography

[1] Ma, H.-L., Sun, Y., Chung, S.-H., and Chan, H. K., "Tackling uncertainties in aircraft maintenance routing: A review of emerging technologies," *Transportation Research Part E: Logistics and Transportation Review*, Vol. 164, 2022, pp. 102805.

[2] Esteban, A., Zafra, A., and Ventura, S., "Data mining in predictive maintenance systems: A taxonomy and systematic review," *WIREs Data Mining and Knowledge Discovery*, Vol. 12, No. 5, jun 2022.

[3] Baru, A. and Johnson, R., "Three Ways to Estimate Remaining Useful Life for Predictive Maintenance - Mathworks," 2023.

[4] Nemani, V., Biggio, L., Huan, X., Hu, Z., Fink, O., Tran, A., Wang, Y., Du, X., Zhang, X., and Hu, C., "Uncertainty Quantification in Machine Learning for Engineering Design and Health Prognostics: A Tutorial," *Mechanical Systems and Signal Processing*, 2023.

[5] Sanjay, R. and Sriram, R., *Data fidelity and latency: All things clinical*, Innovaccer 1, 2022, `https://innovaccer.com/resources/blogs/data--fidelity--and--latency--all--things--clinical`.

[6] Dumouchelle, J., Patel, R., Khalil, E. B., and Bodur, M., "Neur2sp: Neural two-stage stochastic programming," *arXiv preprint arXiv:2205.12006*, 2022.

[7] Barnhart, C. and Cohn, A., "Airline schedule planning: Accomplishments and opportunities," *Manufacturing & service operations management*, Vol. 6, No. 1, 2004, pp. 3–22.

[8] Barnhart, C. and Cohn, A., "Airline schedule planning: Accomplishments and opportunities," *Manufacturing and Service Operations Management 6(1):3–22*, 2004.

[9] Lohatepanont, M. and Barnhart, C., "Airline schedule planning: Integrated models and algorithms for schedule design and fleet assignment," *Transportation science*, Vol. 38, No. 1, 2004, pp. 19–32.

[10] Levin, A., "Scheduling and fleet routing models for transportation systems," *Transportation Science*, Vol. 5, No. 3, 1971, pp. 232–255.

[11] Sherali, H. D., Bish, E. K., and Zhu, X., "Airline fleet assignment concepts, models, and algorithms," *European Journal of Operational Research*, Vol. 172, No. 1, 2006, pp. 1–30.

[12] Zhou, L., Liang, Z., Chou, C.-A., and Chaovalitwongse, W. A., "Airline planning and scheduling: Models and solution methodologies," *Frontiers of Engineering Management*, Vol. 7, No. 1, 2020, pp. 1–26.

[13] Barnhart, C., Boland, N. L., Clarke, L. W., Johnson, E. L., Nemhauser, G. L., and Shenoi, R. G., "Flight string models for aircraft fleeting and routing," *Transportation science*, Vol. 32, No. 3, 1998, pp. 208–220.

[14] Ageeva, Y., *Approaches to incorporating robustness into airline scheduling*, Ph.D. thesis, Massachusetts Institute of Technology, 2000.

[15] Rosenberger, J. M., Johnson, E. L., and Nemhauser, G. L., "A robust fleet-assignment model with hub isolation and short cycles," *Transportation science*, Vol. 38, No. 3, 2004, pp. 357–368.

[16] Sarac, A., Batta, R., and Rump, C. M., "A branch-and-price approach for operational aircraft maintenance routing," *European Journal of Operational Research*, Vol. 175, No. 3, 2006, pp. 1850–1869.

[17] Sriram, C. and Haghani, A., "An optimization model for aircraft maintenance scheduling and reassignment," *Transportation Research Part A: Policy and Practice*, Vol. 37, No. 1, 2003, pp. 29–48.

[18] Vink, J., Santos, B., Verhagen, W., Medeiros, I., and Filho, R., "Dynamic aircraft recovery problem - An operational decision support framework," *Computers Operations Research*, Vol. 117, 2020, pp. 104892.

[19] Su, Y., Xie, K., Wang, H., Liang, Z., Art Chaovalitwongse, W., and Pardalos, P. M., "Airline Disruption Management: A Review of Models and Solution Methods," *Engineering*, Vol. 7, No. 4, 2021, pp. 435–447.

[20] Hassan, L., Santos, B., and Vink, J., "Airline disruption management: A literature review and practical challenges," *Computers Operations Research*, Vol. 127, 2021, pp. 105137.

[21] Arıkan, U., Gürel, S., and Aktürk, M. S., "Integrated aircraft and passenger recovery with cruise time controllability," *Annals of Operations Research*, Vol. 236, 2016, pp. 295–317.

[22] Kinnison, H. A. and Siddiqui, T., *Aviation maintenance management*, McGraw Hill, 2013.

[23] Quatrini, E., Costantino, F., Di Gravio, G., and Patriarca, R., "Condition-Based Maintenance—An Extensive Literature Review," *Machines*, Vol. 8, No. 2, 2020.

[24] Regattieri, A., Giazzi, A., Gamberi, M., and Gamberini, R., "An innovative method to optimize the maintenance policies in an aircraft: General framework and case study," *Journal of Air Transport Management, 44-45:8–20*, 2015.

[25] Deng, Q., Santos, B. F., and Curran, R., "A practical dynamic programming based methodology for aircraft maintenance check scheduling optimization," *European Journal of Operational Research*, Vol. 281(2):256– 273, 2020.

[26] Boere, N. J., "Air Canada Saves with AircraftMaintenance Scheduling," *Interfaces*, Vol. 7(3):1–13, 1977.

[27] Witteman, M., Deng, Q., and Santos, B. F., "A bin packing approach to solve the aircraft maintenance task allocation problem," *European Journal of Operational Research*, Vol. 294, No. 1, 2021, pp. 365–376.

[28] Callewaert, P., Verhagen, W., and Curran, R., "Integrating maintenance work progress monitoring into aircraft maintenance planning decision support," *Transportation Research Procedia*, Vol. 29, 01 2018, pp. 58–69.

[29] Witteman, M., Deng, Q., and Santos, B. F., "A bin packing approach to solve the aircraft maintenance task allocation problem," *European Journal of Operational Research*, Vol. 294(1):365–376, 10, 2021.

[30] Shaukat, S., Katscher, M., Wu, C.-L., Delgado, F., and Larrain, H., "Aircraft line maintenance scheduling and optimisation," *Journal of Air Transport Management*, Vol. 89, 2020, pp. 101914.

[31] N. Papakostas, P. Papachatzakis, V. X. D. and Chryssolouris, G., "An approach to operational aircraft maintenance planning," *Decision Support Systems*, Vol. 48(4):604–612, 2010.

[32] Rosales, L. J. S., Yang, J.-B., and Chen, Y.-W., "Analysing delays and disruptions in Aircraft Heavy Maintenance," *The 32nd International Conference of the System Dynamics Society. Delft, Netherlands*, 2014.

[33] Liang, Z., Xiao, F., Qian, X., Zhou, L., Jin, X., Lu, X., and Karichery, S., "A column generation-based heuristic for aircraft recovery problem with airport capacity constraints and maintenance flexibility," *Transportation Research Part B: Methodological*, Vol. 113, 2018, pp. 70–90.

[34] van Kessel, P. J., Freeman, F. C., and Santos, B. F., "Airline maintenance task rescheduling in a disruptive environment," *European Journal of Operational Research*, Vol. 308, No. 2, 2022, pp. 605–621.

[35] Duffuaa, S. O. and Andijani, A. A., "An integrated simulation model for effective planning of maintenance operations for Saudi Arabian Airlines (SAUDIA)," *Production Planning & Control*, Vol. 10, No. 6, 1999, pp. 579–584.

[36] Sachon, M. and Paté-Cornell, M.-E., "Delays and Safety in Airline Maintenance," *Reliability Engineering System Safety*, Vol. 67, 03 1999, pp. 301–309.

[37] Öhman, M., Hiltunen, M., Virtanen, K., and Holmström, J., "Frontlog scheduling in aircraft line maintenance: From explorative solution design to theoretical insight into buffer management," *Journal of Operations Management*, Vol. 67, 03 2020, pp. 120–151.

[38] Iwata, C. and Mavris, D., "Object-Oriented Discrete Event Simulation Modeling Environment for Aerospace Vehicle Maintenance and Logistics Process," *Procedia Computer Science*, Vol. 16, 12 2013, pp. 187–196.

[39] Lagos, C., Delgado, F., , and Klapp, M. A., "Dynamic optimization for airline maintenance operations," *Transportation Science*, Vol. 54(4), 2020, pp. 998–1015.

[40] Varenna, S., "A Stochastic Discrete Event Simulation of Airline Network and Maintenance Operations," 2023, MSc Thesis at Delft University of Technology under supervision of Dr. ir. B.F. Santos.

[41] den Hollander, F., "Aircraft Tail Assignment with Health-Aware Maintenance Scheduling," 2023, BSc Thesis at Zurich University of Applied Sciences under supervision of Dr. Manuel Arias Chao.

[42] Baldin, A. E., "Condition based maintenance: A powerful tool for modern plant management," *Terotechnica*, Vol. 1, No. 2, 1979, pp. 119–129.

[43] Zonta, T., Da Costa, C. A., da Rosa Righi, R., de Lima, M. J., da Trindade, E. S., and Li, G. P., "Predictive maintenance in the Industry 4.0: A systematic literature review," *Computers & Industrial Engineering*, Vol. 150, 2020, pp. 106889.

[44] Zhong, K., Han, M., and Han, B., "Data-driven based fault prognosis for industrial systems: A concise overview," *IEEE/CAA Journal of Automatica Sinica*, Vol. 7, No. 2, 2019, pp. 330–345.

[45] Saxena, A., Celaya, J., Balaban, E., Goebel, K., Saha, B., Saha, S., and Schwabacher, M., "Metrics for evaluating performance of prognostic techniques," *2008 international conference on prognostics and health management*, IEEE, 2008, pp. 1–17.

[46] Saha, S., Goebel, K., Saxena, A., Celaya, J., and Saha, B., "Evaluating Algorithm Performance Metrics Tailored for Prognostics," 2009.

[47] Gal, Y., Hron, J., and Kendall, A., "Concrete dropout," *Advances in neural information processing systems*, Vol. 30, 2017.

[48] Sankararaman, S., "Significance, interpretation, and quantification of uncertainty in prognostics and remaining useful life prediction," *Mechanical Systems and Signal Processing*, Vol. 52-53, 2015, pp. 228–247.

[49] Rasmussen, C. E., . W. C. K. I., *Gaussian processes for machine learning*, The MIT Press, 2016.

[50] Powell, W. B., "A unified framework for stochastic optimization," *European Journal of Operational Research*, Vol. 275, No. 3, 2019, pp. 795–821.

[51] Shapiro, A. and Philpott, A., "A Tutorial on Stochastic Programming," *Stochastic Programming Society*, 2007.

[52] Shani, G., Pineau, J., and Kaplow, R., "A survey of point-based POMDP solvers," *Autonomous Agents and Multi-Agent Systems*, Vol. 27, 2013, pp. 1–51.

[53] Cahill, J., Donald, N. M., and Losa, C. G., "Intelligent planning and the design of a new risk-based, intelligent flight plan," *Cognition, Technology & Work*, Vol. 13, 2011, pp. 43–66.

[54] Leon, S., Szmerekovsky, J., and Tolliver, D., "Using VAR for strategic capacity allocation: an airline perspective," *International Journal of Services and Operations Management*, Vol. 21, No. 2, 2015, pp. 127–149.

[55] Liu, M., Liang, B., Zheng, F., and Chu, F., "Stochastic airline fleet assignment with risk aversion," *IEEE Transactions on Intelligent Transportation Systems*, Vol. 20, No. 8, 2018, pp. 3081–3090.

[56] Chaudhuri, A., Norton, M., and Krämer, B., "Risk-Based Design Optimization Via Probability of Failure, Conditional Value-at-Risk, and Buffered Probability of Failure," 01 2020.

[57] Lee, J. and Mitici, M., "Deep reinforcement learning for predictive aircraft maintenance using probabilistic remaining-useful-life prognostics," *Reliability Engineering & System Safety*, Vol. 230, 2023, pp. 108908.

[58] de Pater, I. and Mitici, M., "Predictive maintenance for multi-component systems of repairables with Remaining-Useful-Life prognostics and a limited stock of spare components," *Reliability Engineering & System Safety*, Vol. 214, 2021, pp. 107761.

[59] Tseremoglou, I. and Santos, B., "Condition-Based Maintenance scheduling of an aircraft fleet under partial observability: A Deep Reinforcement Learning approach," *Reliability Engineering System Safety*, 08 2023, pp. 109582.

[60] Tseremoglou, I., Bieber, M., Santos, B. F., Wim, V., Freeman, F. C., and van Kessel, P., "The Impact of Prognostic Uncertainty on Condition-Based Maintenance Scheduling: An Integrated Approach," *AIAA AVIATION 2022 Forum*, 2022, p. 3967.