FACULTY OF TRANSPORTATION SCIENCES
CTU IN PRAGUE

Department of Air Transport

Data driven and machine learning playbook supporting identification of block time deviations and the assessment of their operational impact

DISSERTATION THESIS

Doctoral Study Programme: Technology in Transportation and Telecommunication
Study Field: Air Traffic Control and Management

Ing. Miroslav Špák

Prague, November 2023

Prague, November 2023

This dissertation thesis was completed as part of the doctoral studies at the Faculty of Transportation Sciences, Czech Technical University in Prague.

Candidate:	Ing. Miroslav Špák Faculty of Transportation Sciences, Czech Technical University in Prague Horská 3, 128 03 Prague 2
Supervisor:	doc. Ing. Bc. Vladimír Socha, Ph.D. Faculty of Transportation Sciences, Czech Technical University in Prague Horská 3, 128 03 Prague 2

Abstract

Nowadays, with the growth of air traffic, airports are congested, and air traffic operations are disrupted by the formation of various bottlenecks on the surface. Hence, the flight predictability and efficiency on the ground play an important role in keeping the whole Air Traffic Management (ATM) business sustainable. Flight efficiency is directly linked to the ability of flight to adhere to airport arrival and departure slots, thus to minimize presence of primary delays further generating reactionary delays. In 2019, overall departure punctuality improved, with 37.6% of flights departing within the 5minute threshold before or after the scheduled departure time. Flights delayed >30 minutes from all causes decreased by 1.6 percentage points to 12.1% compared to 2018. Airline arrival punctuality also improved, with 77.6% of flights arriving within 15 minutes or earlier than their scheduled arrival time, compared to 75.8% in 2018. Flights arriving >15 minutes ahead of schedule saw an increase to 10.3%. Whilst being good for the passenger experience, this high share may affect operations and ground resource allocation process. Excessively late or early arrivals have direct and negative impact on airport operations including air traffic flow management operations as Air Traffic Flow Management (ATFM) regulations are often implemented as a result of demand shifts. It is the variation of traffic in regards with a type of operating fleet, airline's business rules, or administrative requirements, which hinder any flexibility in tactical resource re-allocation process. The focus of a presented proposal of the dissertation thesis therefore lies in improved flight predictability process based on data driven and machine learning model supporting development of a predictive model enabling identification of in-block and landing time deviations.

Keywords: Airport, flight, delay, model, prediction, predictability

Abstrakt

V súčasnosti, s rastom leteckej dopravy, sú letiská preplnené, a letecké prevádzky sú narušované tvorbou rôznych úzkych hrdiel. Preto predvídateľnosť letu a efektívnosť na zemi a teda na letiskách, zohrávajú dôležitú úlohu pri udržiavaní udržateľnosti celého riadenia letovej prevádzky. Efektívnosť letu je priamo spojená s schopnosťou letu dodržiavať príletové a odletové časy na letisku, a tak minimalizovať prítomnosť primárnych meškaní, ktoré ďalej generujú reakčné meškania. V roku 2019 sa celková presnosť odletov zlepšila, pričom 37,6% letov odletelo v priebehu 5 minút pred plánovaným odletom alebo po ňom. Lety meškané viac ako 30 minút zo všetkých príčin sa znížili o 1,6 percentuálnych bodov na 12,1% v porovnaní s rokom 2018. Aj presnosť príletov leteckých spoločností sa zlepšila, pričom 77,6% letov pristálo do 15 minút pred alebo skôr ako plánovaný prílet, v porovnaní s 75,8% v roku 2018. Lety prichádzajúce o viac ako 15 minút skôr než plánovaný čas príletu zaznamenali nárast na 10,3%. Hoci je to pozitívne pre skúsenosti cestujúcich, táto vysoká hodnota môže ovplyvniť prevádzku a proces pridelenia zdrojov na zemi. Prílišné meškanie alebo prílišné skoré prílety majú priamy a negatívny vplyv na leteckú prevádzku ale taktiež na chod samotného letiska. V dobe prevádzkových odchyliek, aspekty ako variabilita dopravy v súvislosti s typom lietadlového parku, či obchodnými pravidlami leteckej spoločnosti alebo administratívnymi požiadavkami, predstavujú limitácie v procese prerozdelenia zdrojov na zemi. Hlavným zámerom predloženého návrhu dizertačnej práce je preto zlepšenie procesu predvídateľnosti letu na základe dátami riadeného modelu strojového učenia, ktorý podporuje vývoj prediktívneho modelu umožňujúceho identifikáciu odchýlok v čase blokovania a pristávania.

Kľúčové slová: Letisko, let, meškanie, model, predpoveď, predvídateľnosť

Acknowledgment

First, I would like to express my sincere gratitude to organizations providing data sources, EUROCONTROL, Prague International Airport, and Czech Hydrometeorological Institute. Further, I would like to thank members of Faculty of Transportation Sciences, doc. Ing. Jakub Kraus, Ing. Peter Olexa Ph.D., Ing. Peter Had, and Ing. Jiří Volt for elaborative work, thoughtful comments, and active support on this dissertation. I also wish to thank doc. Ing. Peter Vittek for the recommendations, enthusiasm, and time invested during my dissertation journey. I am extremely thankful to Ing. Lenka Hanáková, who masterfully guided me through the long nights with a blend of hands-on assistance and never-ending laughter. The work could not be done without the supervisor of this dissertation, and the "Chief Critic Officer" doc. Ing. Bc. Vladimír Socha, Ph.D., who always armed with constructive scepticism, yet miraculously steering me towards the results presented. His invaluable knowledge and support proved essential, and despite occasional differences, I appreciate his patience and look forward to future collaboration. To conclude, I cannot forget to thank my family and friends for all the unconditional support in these very intense academic years.

Declaration on honour

I declare that I have independently prepared the doctoral thesis entitled "Data driven and machine learning playbook supporting identification of block time deviations and the assessment of their operational impact" and that I have used a complete list of citations from the sources I have referenced, which are listed in the attached bibliography of the dissertation thesis.

I have no significant objection to the use of this academic work under the terms of Section 60 of Act No. 121/2000 Coll., on Copyright, on Rights Related to Copyright, and on Amendment to Some Acts (Copyright Act).

In Prague, 30th November 2023 ....

Signature

Contents

List of Figures

List of Tables

List of Acronyms

ADEP	Airport of Departure
ADEP	Airport of Destination
ATC	Air Traffic Control
ATFCM	Air Traffic Flow and Capacity Management
ATM	Air Traffic Management
AOC	Aircraft Operator
A-CDM	Airport - Collaborative Decision Making
APOC	Airport Operations Centre
AO	Airport Operator
AOP	Airport Operations Plan
AODB	Airport Operations Database
BiLSTM	Bidirectional Long Short-Term Memory
CTFM	Current Tactical Flight Model
DC	Demand-Capacity
DCB	Demand-Capacity Balancing
ECAC	European Civil Aviation Conference
EGHF	Estimated Ground Handling Finish
EIBT	Estimated In-block Time
ELDT	Estimated Landing Time
EOBT	Estimated Off-block Time
ETFMS	Enhanced Tactical Flow and Management System
FTFM	Filed Tactical Flight Model
GH	Ground Handler
IFPS	Initial Flight Plan Processing System
KPIs	Key Performance Indicators
LSTM	Long Short-Term Memory
MAE	Mean Absolute Error
MTTT	Minimum Turnaround Time
NM	Network Manager
NMOC	Network Manager Operations Centre
PDS	Pre-departure Sequencing
RTFM	Regulated Tactical Flight Model
SIBT	Schedule In-block time
SOBT	Schedule Off-block Time
RMS	Resource Management System
RMSE	Root Mean Square Error

SESAR	Signel European Sky ATM Research
TOBT	Target Off-block Time
TTOT	Target Take-off Time
TAM	Total Airport Management
TTA	Target Time of Arrival
UDPP	User Driven Prioritization Process

Introduction

Because this thesis has been developed during the COVID-19 period, the reference year for data used in this study dates back to 2019, a year in which the aviation industry was operating at historic maximums prior to the unprecedented disruption caused by the pandemic. According to the report published by Walker (2022) from Eurocontrol's Central office for delay analysis (CODA), delays in European airspace in 2019 improved when compared to 2018. Despite this fact, in the context of long-term performance the level of delay was the third worst in the last 10 years, behind 2010 and 2018. Following on these negative records, the airlines made considerable efforts and investments such as scheduling improvements by increasing schedule buffers (Arikan, 2011) and using hot spare aircraft to improve on-time performance. This all led to boosting of passenger experience as reactionary delays were reduced. The reactionary delays as advised by Walker (2022); IATA (2018) are results of primary delays. On-time performance is not only benefiting passengers but also airports that often operate close to their capacity ceiling and an early or delayed arrival may cause considerable amount of issues linked to resource unavailability.

Non-adherent traffic obviously limits airports in utilizing full potential of their capacity as buffer times in resource allocation plans are included to cater for unplanned deviations in airport slot utilization. However, today we know that enhancing the airport capacity, improving the airport throughput, and building robust operational plans resilient to traffic deviations may be achieved through improved predictability process using big data techniques. The goal is to have accurate information related to traffic prognosis so that airport processes can adequately be planned and managed. Simply said, airports need to know when the traffic is expected to land, to turn around and to depart from their premises in order to maximize utilization of surface resources. Currently, there are solutions such as Advanced Tower or Airport Collaborative Decision Making (A-CDM), which help enhance operational and situational awareness using information sharing platform, where flight updates are provided (for more information see Eurocontrol (2005)). This allows for predictability of events and allocation of resources as flight progresses through its lifecycle. A-CDM process was introduced to improve the efficiency and resilience of airport operations by optimising the use of resources and improving the predictability of air traffic. It achieves this by encouraging the airport stakeholders (Airport operator (AO), Aircraft operator (AOC), Ground handler (GH), Air traffic control (ATC), etc.)) and the Network Manager (NM) of Eurocontrol to work more transparently and collaboratively, exchanging relevant accurate and timely information. It focuses especially on aircraft turn-round and pre-departure processes.

It also allows the exchange of more accurate departure information, particularly target take-off times known as TTOT, with the European Air Traffic Flow and Capacity Management (ATFCM) network, leading to improved en-route and sectoral planning. Despite all benefits A-CDM process brings, its limitations are found in ability of all concerned to anticipate performance deviations beyond its operational horizon. According to Eurocontrol (2005) this typically starts (short haul European flights) at the time of the flight plan activation and that is done three hours prior to departure. This is often too late for effective predictability and coordination of changes related to resource allocation plans. Furthermore, the spatial scope of A-CDM is limited to airside operations and as stated above only covering the day of operations in terms of both time and process management.

Therefore, advanced, and performance-based airport operations are required for a future performance-based Air Traffic Management (ATM) system as prescribed by Single European Sky ATM Research (SESAR). The aim is to build on A-CDM principles and eliminate its limitations described above. This shall be enabled through a new Total Airport Management (TAM) concept (Spies et al., 2008; Pick & Rawlik, 2011), which constitutes a solution for more efficient resource utilisation and a holistic airport management system for airside and landside processes. The efficiency is seen as an ultimate result of a mechanism, which is based on predictability of events related to passenger and aircraft flows. The anticipation of events linked to potentially distorted operations is particularly important for flights of which duration exceeds 5 and more hours. In this context we speak predominantely about interncontinental flights also known as long-haul traffic. The latter presents a specific group of operations, which typically become visible to ATM systems within the area of European Civil Aviation Conference (ECAC) member states hours after the flight activiation. In this context, predictability of events impacting flights along the route beyond the ECAC region also known as Air Traffic Flow and Management (AFTM) area, becomes paramount as deviations from planned performances may occur.

As such, the performance linked to a duration of these flights does not only impact subsequent outbound flow but severely burdens all airport stakeholders as ground resources must be re-allocated. Airport Operations Centre (APOC) was therefore designed to closely monitor the evolution of the traffic and adjust the resource allocation plan when traffic deviations occur. That applies also for the traffic arriving ahead of its scheduled time, thus likewise disrupting original resource allocation plan. Both instances may in fact cause excessive and unexpected demand for ground resources, which do not have to be available at time of the request as they are utilized elsewhere.

It is worth noting, that the impact assessment of both the early arrivals and delayed traffic on the airport performance and resource optimization is currently underestimated and overshadowed by focus on addressing the ATFM delays without due diligence placed on the surface issues. Airports with limited capabilities to improve predictability of non-adherence to scheduled in-block time (SIBT) and scheduled off-block time (SOBT) may not perform efficiently. Operational efficiency in this context consists in timely predictability of SIBT and SOBT deviations, which in return improves impact assessment and situational awareness about potential changes to resource allocation planning. Both late and early arrivals present major issues at airports operating under fully coordinated slot policy, where minimum room for manoeuvring with the capacity is available. All resources on the surface are typically allocated to accommodate a minimum turnaround time of a given flight, which is defined by initial SIBT and SOBT values.

These are determined typically more than six months in advance, and airports operating with limited resources and capacity may suffer from any significant deviations from these scheduled values on the operational day. The presented proposal of the thesis therefore aims at development of an automated APOC mechanism using data driven and machine-learning playbook linked to the long-haul traffic enabling predictability of deviations from SIBT values. The predictability is built around the information and data linked to the operation of a flight including meteorological parameters known prior to and during the phase of a flight execution. In general, the anticipation of the deviations identified may pre-tactically and tactically be used to build a shadow operational plan enabling early impact assessment linked to the resource allocation planning and management process. As such better planning and execution of airport and network operations may result in an improvement in utilisation of resources, airspace and airport infrastructure and in a reduction in reactionary delays. Information sharing between airport operations and network operations could aso assure the best overall system outcome while addressing the needs of airport actors, the ATM network, individual aircraft operators and the passengers who depend on their services. While, this work primarily aims to reveal means for improving operations locally, a network wide aspect could also be accounted for.

1. Current state of the art

As indicated by Barnhart; Bacena et al. (2019), flight delays have negative economic and operational impact on airlines, passengers, network but airports nonetheless. Therefore, any reduction of these consequences can make flying more effective and efficient. As indicated in the introduction, the way forward lies in improved flight predictability of a particular group of flights arriving from beyond ATFM area, therefore moving towards reinforced pre-tactical and tactical planning through principles governed by TAM concept with APOC and Airport Operations Plan (AOP) in place (Spies et al., 2008).

A good quantum of research attention has been devoted to the study of flight predictability and delays as predicting and analysing delays' root causes have long been active subjects in order to support operational planning and management for ATC, airlines, airport operators, ground handlers, but NM nonetheless. The authors such as Reynolds-Feighan & Button (1999); Klein (2010); Glover & Ball (2013) have used different prediction techniques such as statistical method, probability method as conducted by Tu et al. (2005); Abdel-Aty et al. (2007); Evans et al. (2008), or network-based method, which was elaborated on by Mueller & Chatterji (2002); AhmadBeygi et al. (2008); Wan & Roy (2008). Liu & Ma (2008); Hansen (2002) made efforts to use operational methods, while (Abdelghany et al., 2004; Venkatesh et al., 2017; Zanin et al., 2020; Al-Tabbakh et al., 2018) opted for machine learning method to design a model for predictions of flight delays. Part of the machine learning studies as conducted by Kim et al. (2016) developed deep learning models to investigate prediction of air traffic delay including decision trees, or random forest. According to Esmaeilzadeh & Mokhtarimousavi (2020) delay can be propagated and affect subsequent flights at departure and arrival airports. Having said that, the ability to better predict the delay and control all factors affecting the delays is an important objective. Their study has employed a support vector machine (SVM) model to explore the non-linear relationship between flight delay outcomes. Flight delays are not happening for no reason and can actually be attributed to all sort of them. As part of the most known are weather conditions, origin/destination airport or airspace congestion, aircraft mechanical problems, and airline flight scheduling (Arikan, 2011). The latter is considered a critical matter when it comes to dealing with traffic arriving ahead of or after its scheduled airport slot. Last but not least, there is also a category of delays, which may be brought about by an unexpected event such as diversion. Although this category of delays will not be subject of the presented thesis, the author elaborated on diversion airport selection methodology under umbrella of Student Grant Competition (SGS). As function of the efforts, two publications have been released.

Appendix A

Špák, M., Olexa, P., Enhancement of the diversion airport selection methodology, Transportation Research Procedia. Linz: Elsevier BV, 2020. p. 232-242. ISSN 2352-1465.

Appendix B

Olexa, P., **Špák, M.**, Stojić, S., Lán, S., Hamza, M., Static Validation of the Enhanced Diversion Airport Selection Methodology, 2022 New Trends in Civil Aviation (NTCA). Praha: České vysoké učení technické v Praze, 2022. p. 141-146. ISBN 978-80-01-06985-1. ISSN 2694-7854.

One of the most interesting studies using supervised machine learning and data mining technique on published by Deepudev et al. (2020) addressed a need to improve predictability of actual landing times of scheduled flights. They have rightly identified that historical data may help predict delays on arrival. The predictive modelling was based on multi linear regression (MLR) model to predict the arrival variations based on information retrieved at the time of departure. The study has also revealed root cause for early arrival of the aircraft. The flying time constituted one of the building element of MLR equation.

There has also been some research done in respect to evaluation of delays on the turnaround process. Fricke & Schultz (2009) concluded that the scheduled turnaround process is always disturbed if the airplane does not arrive at the allocated gate or apron position on time. They propose to continue integrating time buffers during the gate allocation planning phase as those allow a higher system reliability. They investigated that there is no systematic buffer concept applied and only empirical experiences seem to trigger this process. Therefore, they attempted to develop a model allowing to optimize the time buffer size with regard to the expected average delay.

Another solution led by the author of the presented thesis, aimed to improve the operational performance through the development of a predictive tool providing pre-tactical and tactical predictions of in-block and off-block time deviations supported by predictions on passenger demand (load factors). The project has developed a generic model based, customised for GVA airport based on Random Forest Regression algorithm, for off-block time and turn-around duration predictions. The performance of the predictions is slightly below the performance obtained during the testing phase during algorithms development. The average error when predicting off-block times is 12.47 minutes and 8.45 minutes for turn-around durations. The accuracy of the predicted values reaches 73% of the off-block predicted times with less than 15 minutes of error and 85% of the turn-around predicted durations. Despite the acceptable level of performance and error the algorithms provided, there are a set of drawbacks identified during the trial related to both the prediction model and the designed dashboard. The airport feedback and improvement opportunities have been collected and detailed under a dedicated section. It includes remarks on the understanding of the airport and airline planning and operational processes that should be better reflected into the predictions cases to count with predictions for all GVA flights and not only partially. This involves further development of the algorithms and looking for different ways of linking arrivals and flights, without counting on aircraft registration, and development of a dedicated algorithm for in-block times.

Appendix C

Dalmau R., De Falco P., **Spak M.**, Rodriguez Varela J. D. Probabilistic pre-tactical arrival and departure flight delay prediction with quantile regression. 15th USA/Europe Air Traffic Management Research and Development Seminar. 2023.

As a by-product of the solution described above a Passenger Demand Support Service was developed developed using machine learning principles through analysis of Network Manager data and through monitoring of commercial flights i.e seats offered in the market. Complementary features such as day of the week, time of the day, airline ID, or airport ID are also used to further support the passenger demand prediction accuracy. A current set up of the solution provides solid prediction performance as mean absolute error (MAE) between predicted load factor and real load factor (per flight) reaches on average 8%. This prediction performance is achievable 7 days ahead of the operational day. The overall prediction outlook extends out to 56 days, where prediction error reaches 15% on average ahead of the operational day.

As such The scientific approach used may provide good benchmark in using machine learning techniques within the presented thesis too.

1.1 Total airport management

Moving from complex flight delays prediction analysis, Eurocontrol and other industry partners have also effortlessly been working on TAM concept as part of SESAR Wave 1, where focus lied on developing and validating a number of key elements of the high performing airport operations. As explained TAM takes a 'holistic' view of all key airport operations processes through APOC platform – air and landside, aircraft, passengers, baggage etc. – and importantly, the interaction between them (Spies et al., 2008). If some operations are delayed such as flight arrivals or departures, it may easily affect subsequent chain of processes. The degree of synchronisation between different processes associated with aircraft, passenger and baggage handling constitutes a significant contributory factor to punctual and predictable operations and ultimately passenger satisfaction (Bacena et al., 2019).

As indicated on the Figure 1.1, a core element of TAM is APOC, which provides a common platform to airport stakeholders to jointly organise and coordinate their activities (Reyna et al., 2020) which includes operational exchange with Network Manager Operations Centre (NMOC) in the European airspace (Leeman et al., 2018). The figure further indicates that the final product of the entire concept is called AOP (Leeman et al., 2018), which is designed months away from the operational day until it is finally executed. The AOP is a robust plan composed of various data elements flowing in from ample number of sources. The purpose is to have full situational awareness of impacts of joint decisions on others' operational plans. The APOC can be either a centralised physical command and control room or a distributed solution, connecting stakeholder representatives by existing and new means of supporting tools for arbitrated collaborative decision making. Very first prerequisite to functional TAM stems in availability of flight data stored in systems and databases inluding the airport operations database (AODB). Then it is a role of technical layers and presence of adequate tools and applications catering for different stages of planning, and real-time management.

The AODB may effectively be used to tackle various operational and traffic deviations identified through a Demand – Capacity and Balancing (DCB) process which effectively feeds the AOP (Holoda et al., 2019). The greatest movement forward with TAM lies in the ability of all stakeholders involved to make harmonized decisions using performance monitoring platform and predictive mechanisms enabling operational look-ahead view. To generalize and as seen on Figure 1.2, the principal goals are to extend the planning horizon beyond three hours ahead of operations, and to extend the spatial scope,

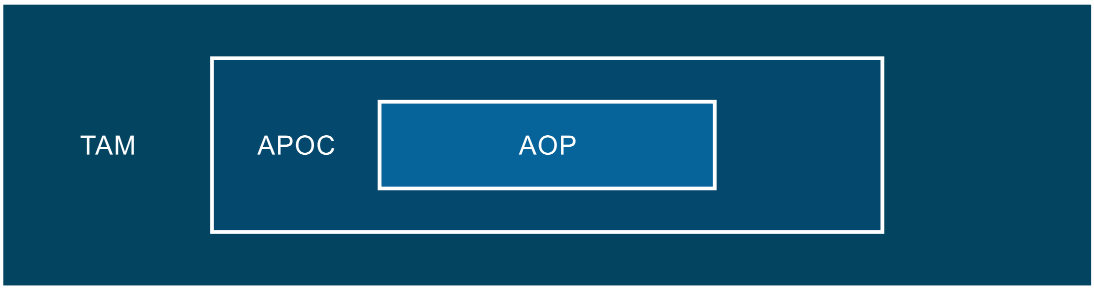


Figure 1.1: Logical relationship between Total Airport Management (TAM), Airport Operations Centre (APOC) and Airport Operations Plan (AOP)

meaning that landside operations would also get a due attention. This should result in identification of potential imbalances between projected demand and available capacity across the airport infrastructure. A demand-capacity (DC) imbalance is an output of the DCB process, and it usually has operational consequences when no preparatory measures are taken. Having knowledge of potential DC imbalances will help implementation of best pre-tactical operational mitigations to improve operational predictability, efficiency, and resource utilization.

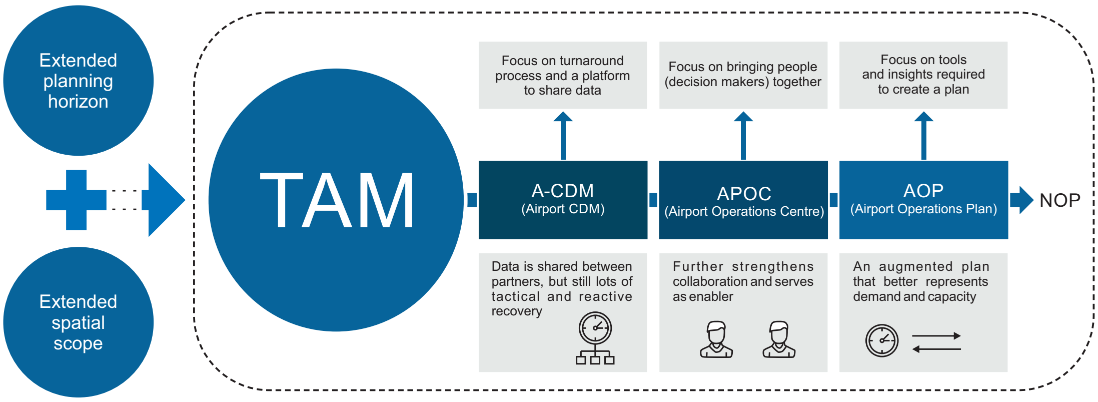


Figure 1.2: Diagram depicting build-up of technical and procedural elements of TAM upon A-CDM

TAM project according to. Günther et al. (2006) aims at increasing predictability, flexibility and efficiency of airport operations as well as resilience through shorter recovery to normal operations. Moreover, it is exactly the predictability element, which calls for attention for the purpose of this thesis proposal, where better knowledge on expected flight performance linked to its arrival time adherence, will increase chances for timely impact assessment. TAM concentrates on strategic, and pre-tactical planning, tactical management on day of operations, and last but not least on the post-ops analysis (Pick & Rawlik, 2011).

When it comes to using big data analysis, Eurocontrol at the end of SESAR Wave 1 initiated two feasibility studies with the airports of London Heathrow and Paris Charles de Gaulle focussing respectively on passenger transfer times and aircraft taxi-out times. Each study also showed the great potential for machine-learning techniques to be applied, based on the use of historic data to determine the underlying drivers of those specific parameters of interest, the relative weight of those drivers, and then enabling predictions of the parameter in question. SESAR Wave 1 also served to perform a shadow mode exercise at Paris Orly airport validating performance dashboards with aim to create airport learning environment to assess landside and airside predictions based on machine-learning technics to support decision-making.

TAM performed more research in this area as SESAR Wave 2 was launched, and some of the simulation exercises further demonstrated operational benefits. SESAR Wave 2 offered significant steps forward in relation to PJ.04 TAM project. Operational improvements focused on airport-centric or airport vs network issues. The work packages covered several solutions focusing on airport integration with the Network (Network Connected Airports) and on rather local airport solutions supporting decision-making process (Digital Smart Airports). Research under SESAR solution PJ.04 and as explained by (Günther et al., 2006) proves that one of the ways forward lies in application of machine-learning techniques based on historical experience stored in operational playbooks. As indicated above, one area that is soliciting a lot of interest at both the academic and the industrial research consists in the exploitation of considerable amounts of data recorded by airports and their partners in order to perform predictions of the behaviour of certain indicators. Furthermore, flight predictability in the context of estimation of flight delays has given opportunity to think out of the box and therefore numerous mitigating solutions to reduce impact of predicted SIBT and SOBT deviations have been developed.

One of the concepts developed under the umbrella of PJ.04 project also elaborated on how to deliver more automated solution to rather regional airports, where APOC would be set up as a virtual arrangement. The idea remained the same, and as such targetted collaborative decision making based on the situational awareness linked to information available. Automated Target Off-block Time (TOBT) updates provided as result of the aircraft and passenger flow monitoring ensured good visibility on potential deviations from Estimated Off-block Time (EOBT) contained in the initial Flight plan (FPL) filed for a given flight. The logic of the automated TOBT included identification of Estimated Landing Time (ELDT) and Estimated In-block Time (EIBT) estimates which triggered TOBT changes. The concept has been co-led by the author and was validated at LKPR in two-step exercise executed in autumn 2022.

Spak M., Rafidison J.P., Marsden A., Vlacil T., Kuran V., Zember M., Sosnovec J., Collaborative Management at Regional Airports - Lite Airport Operations Centre (APOC) concept, SESAR project, 2020 – 2022.

It should be noted that further development of TAM entailed proactive handling of DC imbalances related to slot non-adherence through application certain measures or enablers such as User Driven Prioritization Process (UDPP) or Target Time of Arrival (TTA) (Pilon et al., 2021; Gatsinzi et al., 2018). Potentially, this thesis may also try to take these solutions into account as part of proposed set of measures generated by a new APOC mechanism. However, it has to be underlined, that airlines' disruption recovery and mitigation shall not limited to proposed solutions above.

The UDPP concept gives more flexibility to airspace users to reschedule their flights to keep their business-driven schedule priorities on track when facing capacity constraints and delays. For example, they can reorder the flights in the congested airspace or airports where delays are accumulated. This candidate solution sees the extension of airspace user capabilities, through the UDPP, allowing them to recommend a priority order request to NM, with other ATM stakeholders and appropriate airport authorities, for flights affected by delays on departure, arrival and en-route in capacity-constrained situations. In other words, UDPP allows airlines to change the priority order of unregulated flights among themselves and in collaboration with the airport authorities. Airlines are given this flexibility in the pre-departure sequencing (PDS) for last-minute disruptions, which usually lead to departure delays or cancelled flights. The solution creates more opportunities for departure flexibility within a group of airlines, with benefits increasing as more airlines join. It requires a pre-departure planning process to function, for example using information already shared between operators about planned push-back, start-up and target take-off times. It is especially beneficial in case of disruption with significant financial benefits for the airlines. This solution is available for industrialisation and is now implemented at Paris Charles de Gaulle and Frankfurt airports, and is planned for implementation in Austria and Poland.

Simultaneously to address improvements in arrival slot adherence when traffic is predicted to arrive early, the introduction of TTA came into game. TTA is progressively refined planning time used to coordinate between arrival and departure management and is designed to propose the traffic entry into the

Terminal Manoeuvring Area (TMA) at the right time. UDPP and TTA minimise impact of flow measures on airlines' costs, allow for introduction of more optimum trajectories planned, ensure better use of spare capacity and help decrease demand instability. TTA appears to be one of the solutions to be deployed in short timeframe as it helps facilitate optimization of arrival sequence. However, one has to say, that this solution is rather a reactionary engine, that doesn't necessarily tackle issues linked to predictability of late or early arrivals.

1.2 Linking network and airport operations

Flight predictability is very much linked to what information is available. Working exclusively with isolated information that is restricted to a single user doesn't provide full picture about events, which happened or are about to occur during the execution of a flight. Therefore, it has to be recognized that exchange of information amongst all partners involved in planning, and managament of a flight must be ensured. As explained, APOC connected with NMOC proved to be a right arrangamenet collecting all the vital informaiton linked to a flight lifecycle. NM and its NMOC has a crucial role in european aviation and concretely in ATFM area as it establishes Air Traffic Flow and Capacity Management (ATFCM) that is a complementary to ATC at regional level. The objective of ATFCM is to optimise traffic flows over the entire ECAC region according to air traffic control capacity while enabling airlines to operate safe and efficient flights. It is an arrangement managed from NMOC and is used to organize the flow of air traffic in order to optimize capacity, reduce delays, and enhance safety. The main goal of ATFCM is to ensure that the available airspace and airport capacity is used efficiently and effectively. ATFCM involves a number of different processes and procedures, including traffic demand prediction, flow management, capacity planning, and resource allocation. These processes are supported by a range of different tools and technologies, including air traffic control systems, data analysis software, and communication networks. The primary objective of ATFCM is to ensure that the demand for air traffic is balanced with the available capacity in the airspace and at airports. This involves monitoring the current and predicted traffic levels, and taking action to adjust the flow of traffic when necessary.

As a core element of ATFCM, EUROCONTROL developed ETFMS (Enhanced Tactical Flow Management System), which is a key component of the European ATM system (Eurocontrol, 2023). It is a computerized system that provides a centralized view of air traffic flows in real-time, enabling air traffic controllers and other stakeholders to monitor and manage air traffic in a more efficient and effective way. The system receives and processes data from various sources, such as FPL, radar data, and other air traffic management systems. The ETFMS uses advanced algorithms and decision support tools to provide controllers and other stakeholders with accurate and timely information on traffic flows, airspace capacity, and potential congestion. It helps to optimize the use of available airspace and resources, and provides a common platform for collaborative decision-making between different air traffic control units and stakeholders. The ETFMS supports a range of functions, including tactical flow management, airspace management, and collaborative decision-making. In summary, the ETFMS is a critical component of the European ATM system. It provides real-time information on traffic flows and airspace capacity, helping to optimize the use of available resources and support collaborative decision-making between different stakeholders.

THe ETFMS processes FPL generated by airlines and processed by the Initial Flight Plan Processing System (IFPS) in order to project flight trajectories and to ensure smoothless ATFCM. As part of this process, several flight profiles for each signle flight are generated in order to identify potential conflicts with other trajectories along the route. A Filed Tactical Flight Model (FTFM), which is a mathematical model containing a point and airspace volume profile is created for a flight. The FTFM is an important very first depiction of the intented trajectory for ensuring the safety and efficiency of flights, as it allows airlines and controllers to plan and coordinate their actions in advance, and make adjustments as necessary to ensure that the flight remains on schedule and within safe operational limits. In addition to the FTFM, pilots also use a variety of other tools and procedures to ensure the safe and efficient operation of their flights. These may include pre-flight planning tools, in-flight navigation systems, weather and traffic monitoring systems, and communication and coordination protocols with air traffic controllers and other pilots.

FTFM may change to a Regulated Tactical Flight Model (RTFM), which is typically created shortly before the flight takes place given that a ATFCM measeure in form of a flight regulation along the projected FTFM profile is neccessary. It is designed to ensure that the flight is operated safely and efficiently, while also minimizing delays and optimizing the use of available airspace and airport capacity. The RTFM is based on the same principles as the FTFM, including the planned route, altitude, speed, and other critical flight parameters.

Once a flight is airborne, a Current Tactical Flight Model (CTFM) is generated, which is continuously updated throughout the duration of the flight. The CTFM is based on the FTFM or RTFM and takes into account real-time changes in weather conditions, air traffic congestion, airspace restrictions, and other operational considerations. Therefore, this profile is presents best known and current trajectory guiding the flight crew in making real-time decisions about the flight, including changes to the route, altitude, speed, and other critical flight parameters. The CTFM is also continuously monitored by ATC, who can provide guidance and assistance to the flight crew as necessary. ATC may make adjustments to the CTFM to optimize the flow of traffic, minimize delays, and ensure the safe and efficient operation of the airspace. Overall, the CTFM is an essential part of the air traffic management system, and plays a critical role in ensuring the safety and efficiency of air travel. The Figure 1.3 illustrates differences between respective profiles generated by ETFMS. In comparision to the FTFM and RTFM, the CTFM is more adaptable to changing conditions and provides greater flexibility for the flight crew to make real-time decisions based on the latest information available.

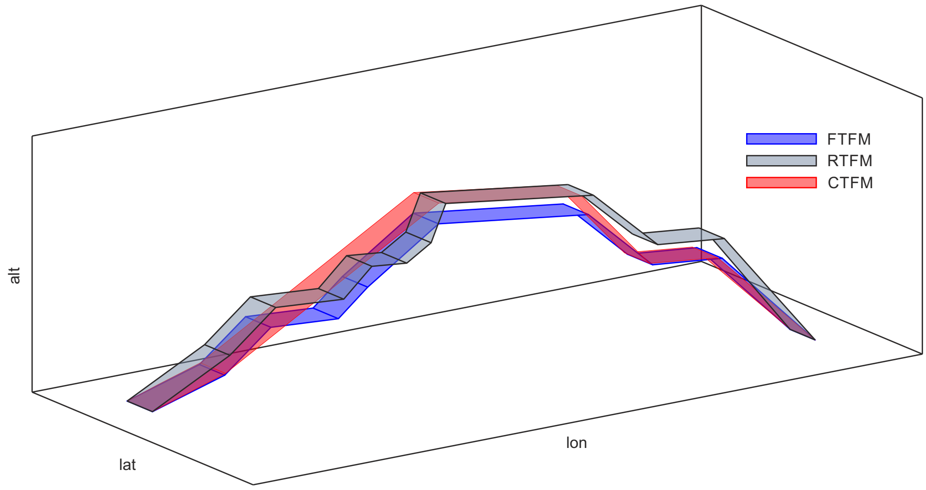


Figure 1.3: Comparision of profiles generated by Enhanced Tactical Flow and Management System (ETFMS)

When relating back to a notion of ECAC region, in our terminology we also call it the ATFCM area. If flight departs from within this area, it may be subject to aforementioned ATFCM measures, which would be translated into RTFM and CTFM profiles accordingly. That means that some of the traffic volumes along the routes exceeeded their capacity in terms of counts at certain time, and that leads to penalization of one or more flights in order to deconflict the traffic. At the same time, this serves to release the workload of respective air traffic controllers in their areas of responsibility. It is imperative to mention that flights departing from outside the ATFCM area and ATFCM adjacent area are exempted from such measures.

The profiles may be modified subject to evolution of airborne or almost airborne flights. For this purpose Koolen & Coliban (2020) was established, which documents messages from and to systems external to the NM called as Flight Progress Messages. These messages are exchanged between ATC, AO, AOC and NM systems and can be classified into the following categories:

- Flight plan filing related messages
- Flight plan status reporting messages
- Flight plan progress messages

For this purpose of this study the flight plan progress messages are the most relevant, as in reality they provide best picture about the flight progress, which may effectively be used to update inbound traffic information at airport of destination. These messages can either be generated by AOC/Air Traffic Services or by NMOC and the overall objective is to improve coordination between all stakeholders. Currently we recognize the following type of messages:

- Arrival planning information
- Aircraft (operator) Position Report
- Departure Planning Information
- First System Activation
- Correlated Position Report
- ETFMS Flight Data message
- Flight Update Message

In principle, Aircraft Position Report (APR) and Correlated Position Report (CPR) are one of those messages, which may effectively be used to provide update on the flight progress for long-haul traffic inbound to Europe. The APR is a message that is sent by AOC and its purpose is to inform the NMOC about the progress of an airborne long haul flight. The APR message informs the NMOC with an accurate update of the Estimated Arrival Time or with an Actual Time Over the aircraft's current position. The APR message will be received and processed by ETFMS. It will be used to update ETFMS flight data, in order to get a more accurate prediction of the sector counts. The main benefit of the APR message is to enable ETFMS to take the accurate flight profile into account for the allocation of ATFM slots to other flights. This will prevent overdeliveries to ATC and reduce holdings for AOC. APR messages are expected to be sent approximately 2 to 3 hours before the flight enters the ATFM area. This gives the NMOC sufficient time to optimise the slot allocation and to prevent overloads of air-spaces where long haul flights form a significant percentage of the traffic. This moment of transmission will also prevent potential conflicts with flight update messages coming from ATC such as FSA messages and radar position reports. Although, the NMOC prefers to receive APR messages approximately 2 to 3 hours before the flight enters the NMOC area, if the AOC wishes to send an APR earlier than that, this will be accepted as well. The APR can be sent on any event that improves the times in the flight data of the AOC. It could e.g. be based upon ACARS messages.

Similarly, the purpose of the CPR as another surveillance data update, is to inform the NMOC about the actual 4D position of the flight when airborne. This message will be received and processed by ETFMS. It will be used to update ETFMS flight data, in order to get a more accurate prediction of the sector counts. The CPR message shall be sent by ATC to the NMOC.

It is also important to say, that all this data exchange is done for the purpose of distributing this further to APOC arrangements. This ultimately, serves to improve predictability linked to inbound traffic, and as a result of it to optimize resource allocation. Good planning presents higher odds for on-time performance, which in return facilitates road towards reduction or elimination of downstream delays in the network.

1.3 Meteorological aspects

Following on the flight trajectories, it should be noted that airlines and their operational peresonnel invest efforts into the most optimal planning of routes. These are then converted into the content of a FPL, which is subsequently submitted to NM via aforementioned IFPS. Ultimatelly, IFPS communicates with ETFMS, where respective mathematical models are used to generate respective profiles. When talking about the FPL driving initial FTFM profile generation, the weather en-route should not be ommitted as this may positively, but also negatively affect the performance and ultimate duration of the flight.

Several weather factors have significant impacts on flying conditions, and below are examples of the associations between meteorology and aircraft operations. Air density, and consequently engine efficiency and lift generated by wings, is influenced by temperature and pressure. Low temperatures below the freezing point are especially significant for icing, and an aircraft flying through clouds or precipitation in such temperatures may be susceptible to icing. Furthermore, the occurrence of condensation is dependent on humidity and dew point, and it plays a significant role in the weather patterns we encounter. Low visibility can result from the formation of fog and clouds. The condition of runways can also be affected by precipitation, in the form of hail, snow, or rain.

However, flying is impacted predominantly by wind and in numerous ways. The lift generated by the airflow over the wings of an aircraft can be impacted by alterations in wind speed. Wind shear, a sudden variation in wind speed or direction with altitude, can be hazardous during take-off and landing. Turbulence caused by wind can produce sudden fluctuations in updrafts and downdrafts. Jet streams are high-speed currents of air that can significantly affect the ground speed of aircraft. Located between 30,000 and 45,000 feet, jet streams only occur at specific latitudes, but they can have a substantial impact on the ground speed of aircraft. With wind speeds exceeding 160 knots, a jet stream is typically 160 kilometres wide and only 2-3 kilometres in vertical extent. Not only jet streams, but also moderate winds along the flown route may cause aircraft ground speed to rise or to drop, ultimatelly leading to different flying times in comparision to scheduled ones.

The flying time may also be impacted by local airport weather, as holding an aircraft above the destination airport might lead to substantial delays. One of the studies conducted by Thiagarajan et al. (2017) dealt with flight predictability with focus on the arrival delays. In their paper a two-stage predictive model was built, where the feature set used consisted of 12 airline data features and 24 weather features. The pertinent weather data used in this study contained only local meteorological reports capturing situation in the vicinity of the airport. Different algorithms such as gradient boosting, random forest, extra-trees, or adaboost classifiers were used for the classification stage in the model. The results proved that gradient boosting classifier performed the best with 94,35 % accuracy. The results of the regression stage in the model were achieved best with the extra-trees regression that can predict the delay value with an error of approximately 8 minutes. The predictive model was further used to develop a decision support tool to visualize prediction of flight delays.

The author was also involved in the development of the airport meteorological application, which was developed in order to evaluate impact of local aerodrome weather on the ground operations. The model was built to parse terminal area forecast, meteorological aerodrome reports, or special reports of meteorological conditions. This further enabled development of the application, which based on the meteorogolical conditions predicted, alerts all local stakeholders about a potential disruption of ground handling services, a change of runway in use, or overall interruption of the turnaround activities.

Spak M., Vlacil T., Kuran V., Zember M., Airport Operations Centre (APOC) - Meteorological application assessing real-time impact on ground operations, Prague, 2018 – 2020.

1.4 Current state of the art limitations

Due to deviations between planned and actual operations, airports are forced to apply tactical adjustments to their resource allocation plans. This creates greatest impact mainly during peak operations when buffer capacity is seldom available. Nowadays, airports have very limited means for projecting pre-tactical and tactical block time (gate slots) deviations for long-haul routes, which would help them optimize their planning. This is mainly due to reason that greatest visibility on flight progress happens only once FPL (flight plan) is activated. However – as explained – this gives a very short notice to airports taking into consideration uncertainty linked to execution of the FTFM profile. Today we know, that due to a limited radar coverage, a great portion of the trajectory for long-haul routes presents a greay area. F.e. for some of the inbound traffic approachin LKPR from far east, the two thirds of the trajectory remains with no real-time positional reporting. If deviations along that portion of the route happen as result of velocity conditions, this may leave the GH and AO in a difficult position. Consequently, the level of service provided to AOCs and their passengers deteriorates too. Therefore, it is proposed to improve operational performance through development of a service helping provide predictions on in-block time deviations. The predictions as part of the service could potentially be visible to APOC or similar operational arrangement and provide primarily pre-tactical, but also tactical support at times of a flight being airborne. The predictions might be built around machine learning or equivalent techniques based on exploitation of massive amount of data. The main limitations of today's system not allowing for such a process are as follows:

- Limited visibility on a flight progress along the trajectory beyond the ATFCM region.
- Existing predictability models in regards with in-block time predictions taking into account predominantly local airport meteorological reports and not full scale en-route data along computed FTFM profiles.

- Substantial amount of experimental work done on off-block predictability through turnaround duration predictability and missing emphasise on in-block time predictability.
- No presence of en-route forecasts including velocity information in AODB enabling timely identification of potential increments or decrements of flying times
- Limited information sharing between AO and AOC on the flight progress, which would allow for updates on estimated landing, in-block timestamps.

There is a crucial need to for AO to possess own means for long-haul flight predictability. Reasons vary, but of them also lies in AOCs behaviour linked to flight planning, where schedule buffering plays a substantial role. Having said that, although meteorological conditions along the projected route are taken into account when submitting the route to NM, economical and business aspects force AOCs to perform on time. The author presumes from his own experience and practice, that initially estimated time over a destination aerodrome may not be realistic. To this effect, the engine installed by the AO allowing him to compute its own estimates based on the flight, trajectory and meteorogolical data along the route, may make resource allocation planning more resilient. AODB is expected to be one of the main data sources as all pertinent information would sit in there. In support of this, a currently missing evaluation of the operational impact of early or delayed long-hauld traffic on airport resources may also be developed.

2. Thesis objective and hypothesis

Considering the results of the research that has been carried out by the author and broader scientific community over the past years, it has been concluded that big data analytics and machine-learning techniques will greatly be used to improve flight predictability within the presented thesis.

2.1 Thesis objective

The main objective of the thesis lies in further exploration on how the current research achievements can be used in broader perspective. This shall be done through development of a new data driven and machine learning model supporting predictability of flight arrival time deviations.

2.1.1 Prediction of in-block time deviations

As some studies revealed there is a number of potential constants and variables also called precursors, or predictors that should be taken into account when designing predictive models helping identify delayed or early arrivals. It is presumed that a data driven and machine learning model using multiple data sources will be developed to anticipate in-block time delay.

This step shall enable development of an engine, which may ultimately be used to support local airport solutions allowing for a decision-making process when allocating ground resources such as parking positions and gates. To maximize efficiency of ground resource utilization and minimize tactical changes to resource allocation plans, timely identification of potential conflicts in resource distribution caused by delayed or early arrivals has to be ensured. This, although not necessarily covered by this thesis, could be done through development of a shadow resource allocation plan, which encompasses arrival time deviations enabling resource allocation assessment. Furthermore, delayed or early traffic impacts more than just resources and so particular attention may also be given to airport standard operating procedures, or other operational nodes, which may also suffer given that unexpected demand kicks in. Obviously, each arrival time deviation may bear certain hallmarks in terms of predictors involved and predicted deviation value itself. Some may generated low, but others moderate or high severity of impact. Based on the thesis objectives, below are the hypotheses, which shall form part of the thesis structure.

2.2 Hypotheses

1. Airline ID, flight ID, origin airport ID, time of day, wind speed, wind direction may be considered as predictors with the highest importance when predicting in-block time deviations.

2. Reduction of tactical stand allocation changes using data driven and machine-learning playbook is proved statistically significant.

3. Delayed and early arrival by more than 15 minutes present the same severity of impact on stand availability.

3. Methodology

To fulfil the objectives of the proposed thesis, a multiple approach may be used. In essence and looking at the nature of expected outcomes while taking into consideration subject matter of the thesis, a predictive model needs to be developed. The model is a pre-requisite for hypotheses to be proven true or false.

The execution of the research entails data acquisition and mathematical approach for data analytics. Ultimately, forecasting techniques will have to be used to enable development of data driven and machine-learning playbook.

The process of literature review settled the background knowledge for this project, involving all relevant information and lessons learned to be reflected in the work to be done. Existing predictive models having as objectives same or similar objective indicators as the ones defined for this project have been identified and further understood to take advantage of the work already done.

The main goal of the model is to improve airport operational performance by providing pre-tactical and potentially tactical predictions of in-block time deviations taking into consideration predicted time increment or decrement along the flight trajectory. For this purpose access to historical data for the model-building process is essential to provide accurate forecasts and predict flight behaviour on the trajectory. Stemming from the objectives an evolving process for the predictive modelling development has been defined.

As such the study is structured on three main technical tasks, starting with the description of the methodology, continuing with development and validation of the predictive model and finally concluding with visualization of prediction outputs.

3.1 Data characteristics, availability and preprocessing

The data inputs described below were provided by Eurocontrol, the Czech Hydrometeorological Institute, Prague International Airport, and Air Navigation Services of the Czech Republic. These data inputs were chosen based on their availability and therefore, some of the elements potentially contributing to delay predictability might be missing. Instead, this is a data set obtained for the purpose of the dissertation work, presented in a comprehensive and representative manner, offering sufficiently high-quality and time-acurate information (see Fig. 3.1).

Slot coordination data

Feature	Datatype
CAODB ID	double
Flight Name	string / cat
Call Sign	string / cat
Arrival / Departure	string / cat
Airport	string / cat
Aircraft	string / cat
Terminal	string / cat
SIBT	datetime

Source:
Prague International Airport

Time-stamps and movement data

Feature	Datatype
Call Sign	string / cat
ALDT	datetime
AIBT	datetime
IFPLID	double

Source:
Air Navigation Services of the Czech Republic

Trajectory data

Feature	Datatype
FLT_UID	double
FLT_SOBT	datetime
FLT_ATOT	datetime
FLT_AOBT	datetime
Aircraft_ID	string / cat
Aircraft	string / cat
LOBT	datetime
ADEP	string / cat
Model	string / cat
Flight_Level	double
Time_Over	datetime
Point_Dist	double
LAT	double
LON	double
KM/H	double
ADES	string / cat
SAM_ID	double

Source:
Eurocontrol

Velocity data

Feature	Datatype
Time	datetime
lat	double
lon	double
ws	double
wd	double
alt	double

Source:
Czech Hydrometeorological Institute

Figure 3.1: Structure of data from respective sources.

Trajectory data provided by Eurocontrol containt multiple fields for each single flight record. The sequence of data elements starts with an unique database identifier assigned to a given flight (FLT_UID), continues with actual take-off time (FLT_ATOT), scheduled off block time (FLT_SOBT) and followed by actual off block time (FLT_AOBT). Additional elements include aircraft identifier (Aircraft_IF), aircraft type and flight plan related information such as initial estimated off block time (LOBT). The sequence continues with information on airport of departure (ADEP), airport of destination (ADES), flight plan identifier recorderd in Eurocontrol database (SAM_ID), and indication of the filed or current tactical flight model (FTMF/CTFM). Completeness of the record is conlcuded with flight level information (Flight_level), time over certain point of airspace (Time_Over), distance of the flown portion of trajectory (Point_Dist), and data capturing geographical position expressed in latitude and longitude (LAT/LON) as well as ground speed of the flight (KM/H).

A dataset combining Prague International Airport and Air Navigation Services of the Czech Republic presents additional sequence of fields and data types. This includes airport flight database identifier stored as (CAODB ID), air navigation services identifier as (IFPLID), flight name, call sign, and notion of the movement type stored as arrival The additional information contains airport of departure, aircraft type, and terminal building assigned. Last but not least, scheduled in block time stored as (SIBT), timestamp on the actual runway time as (ALDT), and actual in block time (AIBT) are also present.

Meteorological data constituting essential element of this thesis includes forecasting information for respective region represented by longitude, latitude, and correspoding prediction of velocity elements such as wind speed (ws), and wind direction (wd). These parameteres are further assigned to a corresponding altitude (alt). Further breakdown of the analysis linked to the meteorological considerations is available in chapter 3.1.2.

Datasets, i.e. slot coordination data, time-stamps and movement data, trajectory data (including FTFM and CTFM profiles) and meteorological data, were merged to form a consolidated dataset. Before the actual merging of the data, it was important to clean and prepare the data to ensure that it is consistent and free from errors. This involved removing duplicate records, standardizing data formats, and correcting any missing or incorrect data.

First step, at this point of data merge, was on merging slot coordination data with FTFM profiles (part of trajectory dataset) and time-stamps and movement data, primarily using the departure time as a key criterion. These three datasets unfortunately did not share the same unique flight identifier. Trajectory dataset included a time-dependent flight description, with each flight having multiple waypoints associated with latitude, longitude coordinates, and flight-over times, among other details (as described in Fig. 3.1). By associating slot coordination as well as time-stamps and movement data with FTFM data, variables with a constant nature for a given flight were created. For instance, the SIBT or AIBT remains constant across all waypoints for a single flight.

In this scenario, only the merging of FTFM profiles with other datasets is discussed. This is because CTFM profile may have a different length compared to FTFM for speciffic flight. Consequently, adjustments were made to the CTFM data, and these modifications are detailed in the following chapter (refer to section 3.1.1). For illustration, the merged dataset (in this initial meging phase) would look like Tab. 3.1:

Table 3.1: Exaple of dataset after initial data merging.

FLT_UID	**FLT_ATOT**	...	**ADEP**	**TIME_OVER**	**...**	**LAT**	**SEQ_ID***
5	1.6.2019 1:49	...	KJFK	1.6.2019 1:47	...	40.64	1
5	1.6.2019 1:49	...	KJFK	1.6.2019 1:47	...	40.65	2
5	1.6.2019 1:49	...	KJFK	1.6.2019 1:47	...	40.66	3
5	1.6.2019 1:49	...	KJFK	1.6.2019 1:47	...	40.67	4
⋮	⋮	⋱	⋮	⋮	⋱	⋮	⋮
5	1.6.2019 1:49	...	KJFK	1.6.2019 8:37	...	51.29	61
⋮	⋮	⋱	⋮	⋮	⋱	⋮	⋮
3	1.6.2019 4:22	...	RKSI	1.6.2019 4:32	...	37.31	15
⋮	⋮	⋱	⋮	⋮	⋱	⋮	⋮
3	1.6.2019 4:22	...	RKSI	1.6.2019 11:26	...	59.27	35
⋮	⋮	⋱	⋮	⋮	⋱	⋮	⋮
3	1.6.2019 4:22	...	RKSI	1.6.2019 14:09	...	50.95	75
⋮	⋮	⋱	⋮	⋮	⋱	⋮	⋮

* SEQ_ID is a newly created variable characterizing the sequence order of the flight-over point.

3.1.1 FTFM and CTFM profile data

As mentioned in the preceding chapters, FTFM and CTFM data represent flight models that can be described as the planned flight trajectory (in the case of FTFM) and the updated trajectory (CTFM), which essentially becomes the actual flown trajectory upon completion of the flight.

FTFM and CTFM profiles inherently differ, as shown in Fig. 3.2, not only in the absolute values of latitude, longitude, and altitude but also in their length. This means that a CTFM profile can, and typically does, contain more points along the trajectory over which the flight was conducted compared to an FTFM profile.

For further processing and analytical purposes, it was necessary to align the time points from FTFM and CTFM profiles by generating linearly spaced time vectors of same number of samples. The principle of the aforementioned process is schematically depicted in Fig. 3.3. This step allowed understanding delay along the flown CTFM trajectory in comparison to planned FTFM trajectory from the moment of flight departure until its arrival. FTFM and CTFM times were compared at same phase of fligth, i.e. if FTFM time stamp was at 20 % of planned flight time then corresponding CTFM time stamp reflected 20 % of actually flown time.

Firstly, time stamps in both FTFM and CTFM dataset were converted from datetime format to Unix time. This conversion ensured that timestamp was presented in seconds as Unix time represents seconds elapsed since January 1, 1970 (Matthew & Stones, 2007). Therefore, converted vectors $TimeC$ and $TimeF$ representing time stamps in Unix time for CTFM and FTMF, respectively, were created. The total FTFM time was then crucial for $TimeC$ vector resampling. Total FTFM time $TimeF_{total}$ was calculated as:

$$TimeF_{total} = TimeF(N_F) - TimeF(1) + 1 \tag{3.1}$$

where $TimeF$ is the FTFM time vector (in seconds) and N_F is a total number of $TimeF$ samples. After that, the CTFM time vector was resampled in the way that:

$$TimeC_r(1) = TimeC(1), \text{and} \tag{3.2}$$

$$TimeC_r(TimeF_{total}) = TimeC(N_C) \tag{3.3}$$

where $TimeC_r$ has N_F number of samples and represent lineary resampled vector $TimeC$. The N_C is then number of samples in $TimeC$ vector.

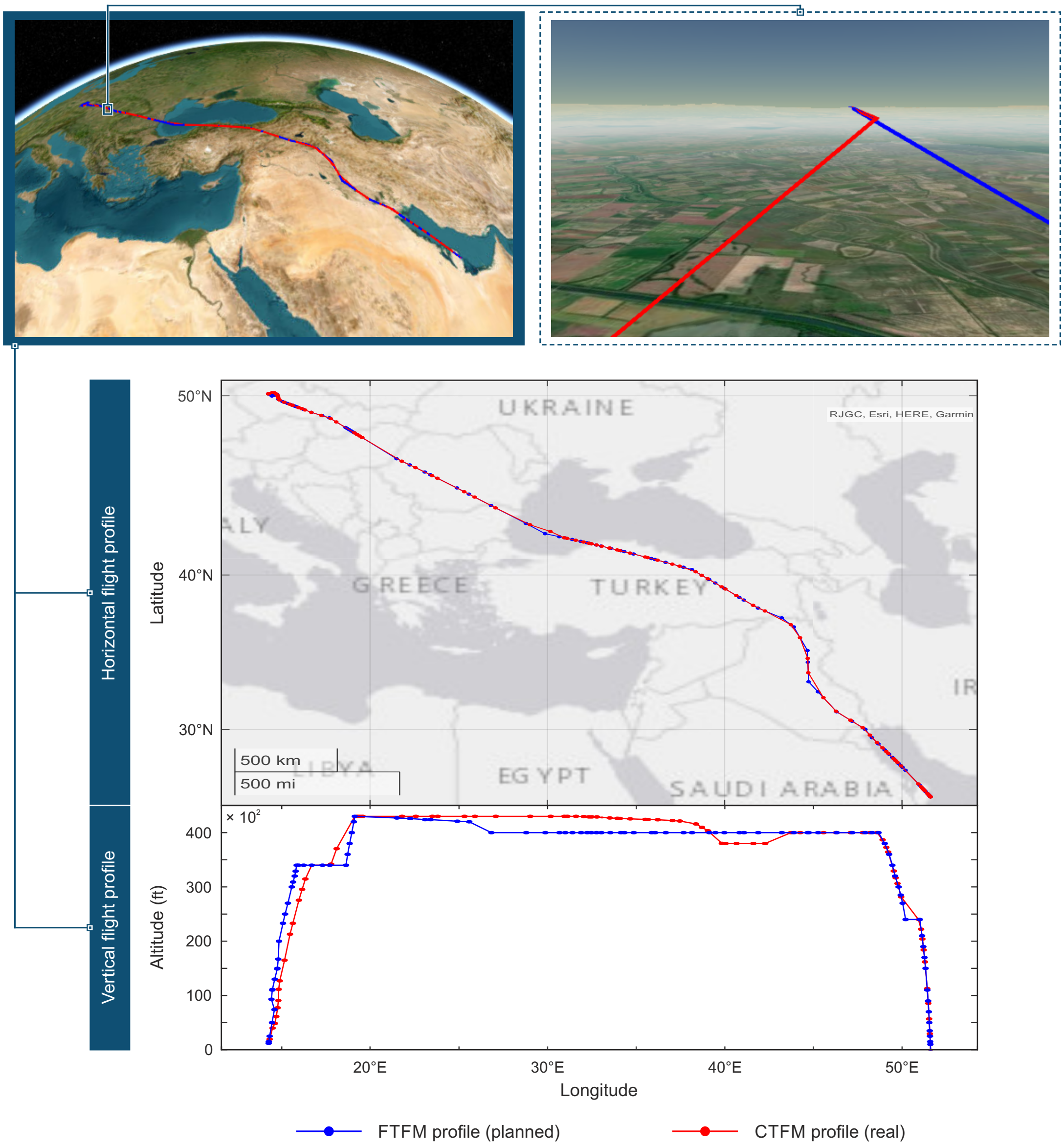


Figure 3.2: Example of FTFM and CTFM profiles for QTR87DF flight from OTHH to LKPR (Terrain source: GMTED2010 7.5 arc-seconds resolution (approx. 250 meters). Terrain data available from the U.S. Geological Survey - Source: Esri, Maxar, Earthstar Geographics and the GIS User Community).

Also $TimeF$ vector was resampled in the same way, therefore

$$TimeF_r(1) = TimeF(1), \text{and} \tag{3.4}$$

$$TimeF_r(TimeF_{total}) = TimeC(N_F) \tag{3.5}$$

represents vector in which each data point represents one second of FTFM flight time (planned flight time). The indexes of vector $TimeF_r$ (idx) of data points that are common to both vectors $TimeF$ and $TimeF_r$ was then identified as

$$\{idx | 1 \leq idx \leq TimeF_{total}, idx \in TimeF, i \in TimeF_r\}. \tag{3.6}$$

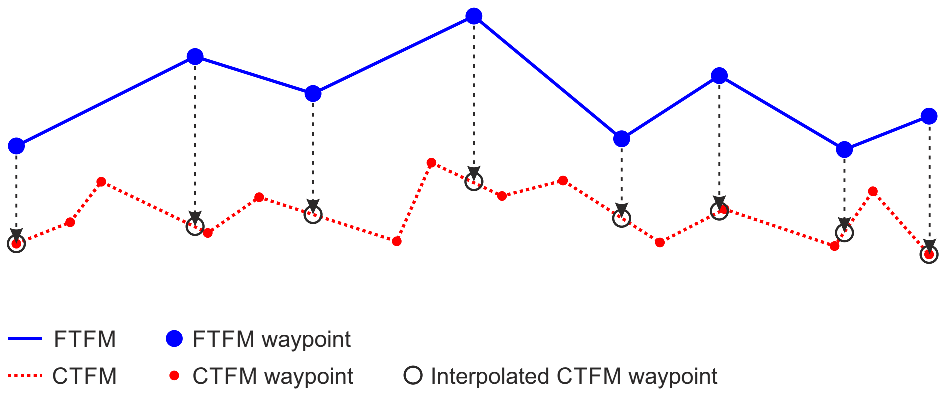


Figure 3.3: Resampling of CTFM data according to FTFM.

The aforementioned principle was also applied for linear interpolation of latitude, longitude, and altitude at individual time points of modified CTFM data. Nevertheless, these variables are not necessary in the context of this work and were derived solely to verify the correctness of this procedure, further data exploration and potential future processing in upcoming research tasks.

The thus resampled data could then be assigned to the overall dataset (as illustrated in the example in Tab. 3.1). This also enabled the calculation of the vector of time differences (FC_{delay}) between CTFM (flown) and corresponding FTFM (planned). This way FC_{delay} vector was calculated as (note: the time differences were converted to minutes):

$$FC_{delay} = \frac{TimeC_r(idx) - TimeF_r(idx)}{60}. \tag{3.7}$$

Described process ensured, that FC_{delay} contains time delay for each time stamp in original FTFM dataset.

Further step involved calculating the delay while taking into consideration the planned and current in-block times. Calculation of delay as $\delta_{IBT} = AIBT - SIBT$ would however result in δ_{IBT} being an dependent variable in sequential data. In other words, sequence dependence within one flight (dependent on the number of waypoints described by independent variables) would address a single overall delay. In addition to the above, a landing delay was defined from the available data, i.e., $\delta_{LT} = ART - SRT$, where SRT is FTFM datapoint P_w and ART is CTFM P_w for given flight. Here, P_w represents the last point on the trajectory.

The final dataset, which includes added meteorological data, along with a list of dependent and independent variables, is detailed in Chapter 3.3.

3.1.2 Meteorological data

Meteorological data refers to a dataset provided by the Czech Hydrometeorological Institute. It contains information about wind direction and speed. These data cover the period from June 1, 2019, to August 31, 2019, and encompass a geographical region extending from 120°W to 120°E and from 80°N to 20°S. The grid divides this region into increments of 1.25 degrees, following the specifications outlined in ICAO Annex 3.

Within this dataset, there are 17 different flight levels, ranging from 050 to 530. For each point defined within this coordinate system, predictions for wind direction and speed are available in 3-hour intervals, with a time step of 3 hours. The predictive interval varies from 6 to 36 hours, and the data is updated four times a day at the following times: 00:00, 06:00, 12:00, and 18:00. The structure of this data, along with its visualization, is depicted in Fig. 3.4.

These data were used to define the meteorological situation at specific points within the FTFM and CTFM profiles. The initial step involved selecting meteorological information, specifically the wind speed and direction, for each point along the flight path. From the previous chapter, it is evident how these profiles are defined. However, for the sake of clarifying the data preprocessing procedure, let's define a flight profile as follows:

$$FP_i = \begin{bmatrix} lat_1 & lon_1 & alt_1 \\ lat_2 & lon_2 & alt_2 \\ \vdots & \vdots & \vdots \\ lat_{w-1} & lon_{w-1} & alt_{w-1} \\ lat_w & lon_w & alt_w \end{bmatrix}, \tag{3.8}$$

where FP_i represents a specific flight profile selected from the set of all profiles based on a unique flight identifier i, where $i = 1 \dots 1730$. This simplified notation contains information about the geographical longitude, latitude, and flight altitude at a specific point along the route. The number of route points varies, and it is therefore defined by the total count of route points, denoted as w.

In simpler terms, the process of associating meteorological data involved attempting to locate the nearest available data point within the meteorological dataset for a specific point along the flight path. Therefore, when given a flight trajectory point P_p (where $P_p \in FP_i$) and a set of meteorological points Q, the nearest point $Q_{nearest}$ to point P_p can be determine as:

$$Q_{nearest} = \arg\min_{q \in Q} ||P_p - q||, \tag{3.9}$$

where $||P_p - q||$ represents the Euclidean distance between point P_p $(p = 1 \ldots w)$ and each point q in the set Q, and $\arg\min_{q \in Q}$ finds the point in Q that minimizes this distance.

Meteorological data were effectively divided into batches based on geographical latitude, longitude, and time, considering their size, as shown in Figure 3.4. This process led to the selection of data subset(s) that covered the geographical latitude and longitude of the flight trajectory at a specific time. For this purpose, i.e., batch processing, a software solution was created in the Matlab environment. The mentioned process, in reality, preceded the actual selection of meteorological data concerning the flight trajectory.

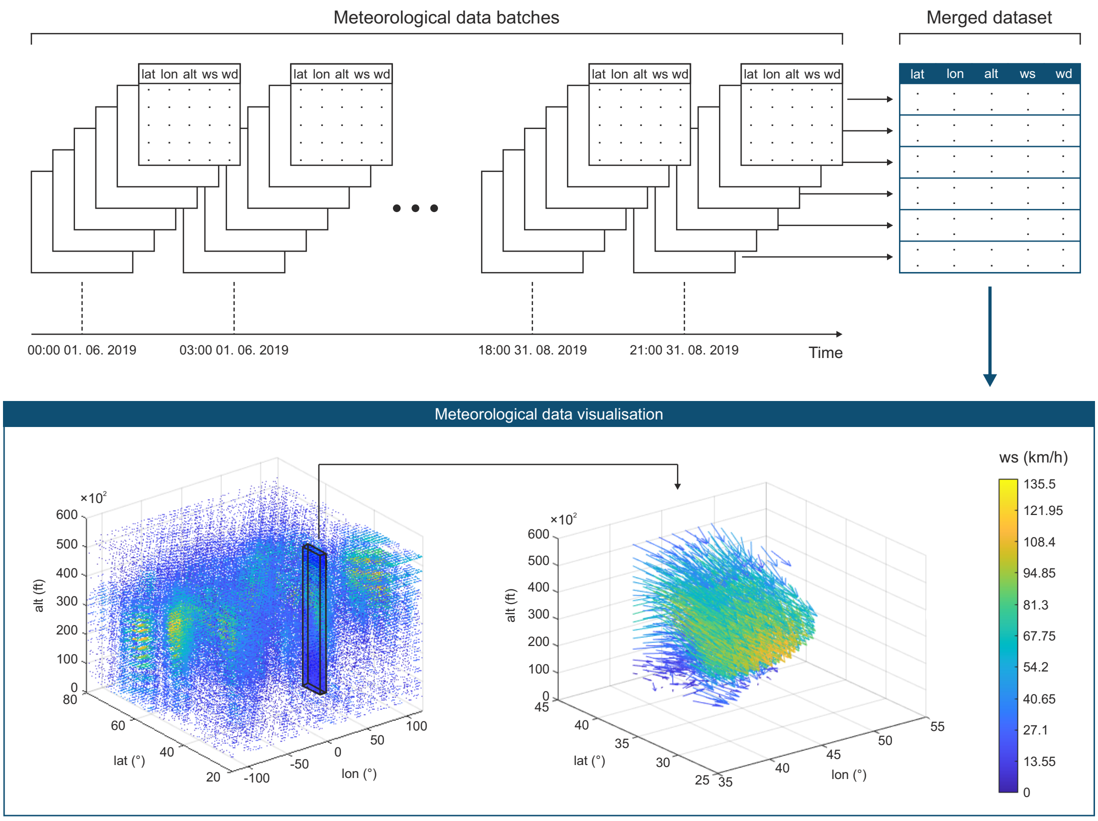


Figure 3.4: Structure of meteorological data and an example of visualisation of this data as a plot of wind direction (wd) and wind speed (ws) relative to geographic latitude (lat), longitude (lon), and altitude (alt).

Within dissertation data processing methodology, a two-step process was employed to ascertain the nearest meteorological data point concerning a specific target point. This approach accommodates both horizontal position (latitude and longitude) and altitude (in feet). This process was iterative and chosen because calculating the Euclidean distance between two points in 3D space would be complicated due to the different units of measurement for altitude (in feet) and coordinates (in degrees). The process was therefore as follows.

In the first iteration nearest point determination based on latitude and longitude took place. The minimum Euclidean distance for latitude and longitude between the flight trajectory point (lat_{P_p} and lon_{P_p}) and all points in the meteorological dataset (lat_Q and lon_Q) was calculated as:

$$d_{horizontal} = \arg\min_{q \in Q} \sqrt{(lat_{P_p} - lat_Q)^2 + (lon_{P_p} - lon_Q)^2}. \tag{3.10}$$

The indexes k of the nearest data points in the dataset, based on the calculated $d_{horizontal}$, were identified. These indexes represent the nearest data points in terms of their horizontal positions from original dataset, and they also define the respective altitudes. Consequently, based on the k, it is possible to select the vector of altitudes corresponding to the nearest data points with horizontal positions defined by latitude and longitude. From the resulting vector A, it is then possible to determine the nearest point in terms of altitude by applying a similar approach as defined in equations 3.9 or 3.10. If we denote the nearest point in terms of altitude as l, then the sought-after information about wind speed and direction in meteorological data will be at position Q_{k_l}.

The above describes an approach for selecting meteorological information for a specific flight point. Subsequently, the headwind or tailwind, i.e., the component of the wind in the direction of or opposite to the each point (P_p) of the flight route, was calculated. Tailwind/headwind was considered a crucial parameter for further modeling. Typically, aircraft operators make use of weather forecasts and wind data to plan their flights and take advantage of favorable wind in terms of its speed and direction. However, the variation of velocity parameters along the flight path may tactically impact the flying time itself. Therefore, the parametrization of velocity behaviour in relation to wind direction in particular, would help understand and reveal potential nuances indicating positive, negative or alternatively none affect on the flight duration.

In this context, the first step was to determine the flight direction for each point along the designated flight path within the FTFM and CTFM profiles. Using the same notation, the heading (hdg) from point P_p to point P_{p+1} can be calculated as follows:

$$hdg = atan2(X,Y), \tag{3.11}$$

where

$$X = cos(lat_{P_{p+1}}) \cdot sin(lon_{P_{p+1}} - lon_{P_p}) \tag{3.12}$$

and

$$Y = cos(lat_{P_p}) \cdot sin(lat_{P_{p+1}}) - sin(lat_{P_p}) \cdot cos(lat_{P_{p+1}}) \cdot cos(lon_{P_{p+1}} - lon_{P_p}). \quad (3.13)$$

To provide additional context, it's worth mentioning that the variable $p = 1 \ldots w - 1$, where w denotes the total number of points along the flight path. Furthermore, it's crucial to emphasize that the determination of flight direction is relative to the initial point, which means that the flight direction is associated with P_p. In other words, the final point does not have a specified hdg. In the case of the ultimate point on the flight trajectory, this corresponds to the instance when the aircraft is on the runway. In such a scenario, it has been posited that the flight direction remains consistent with that of the preceding point, i.e., $hdg_{P_{w-1}} = hdg_{P_w}$. It's worth noting that the 'atan2' function inherently produces values within the range of (-180°, 180°), with 0° denoting North, 90° denoting East, 180° denoting South, and -90° denoting West. The above approach led to the conversion of the hdg into a range from 0° to 360°, where 0° corresponds to the north, and the other cardinal directions are represented in a clockwise manner within this range (e.g., 90° represents east, and so on). This conversion was achieved using modulo as follows:

$$hdg_{360} \equiv (hdg + 360) \mod 360. \quad (3.14)$$

From the above, it is then possible to calculate the component of the wind speed in relation to the direction of flight as:

$$w_{component} = ws \cdot cos(hdg_{360} - wd), \quad (3.15)$$

where negative values of $w_{component}$ indicate headwind and positive ones indicate tailwind.

3.2 Data descriptive statistics

This sub-chapter investigates trends and dependencies within the data set available. This analysis can help identify patterns and relationships between different variables in the data, such as the relationship between flight delays or aircraft performance and wind related conditions along the flight trajectory. Furthermore, the analysis also helps detect a neccesity for a development of synthetic variables. The latter, along with other relevant features, may then be used as inputs for machine learning models aimed at predicting flight delays.

Therefore, the data merging process and this statistical analysis enabling determination of new variables are essential steps in preparing the data for predictive modeling.

The analysis starts with looking at overall number of flights available in the data sample over the period of summer 2019. This reference period captures high summer traffic in standard operations. Certainly, it is possible to perform a relatively robust statistical analysis on the available data; however, this is not the goal of this chapter. The description below aims to outline the main and interesting data characteristics that may influence the prediction.

Furthermore, a section documenting schedules duration of flights that are compared with those flown is also included. There, a reader may also find a view on range of deviations related to difference between scheduled and actual in-block times on respective connections.

The Figure 3.5-A encapsulates the traffic volume recorded during the summer months of June, July, and August 2019 (in total, the dataset included 1 730 flights).

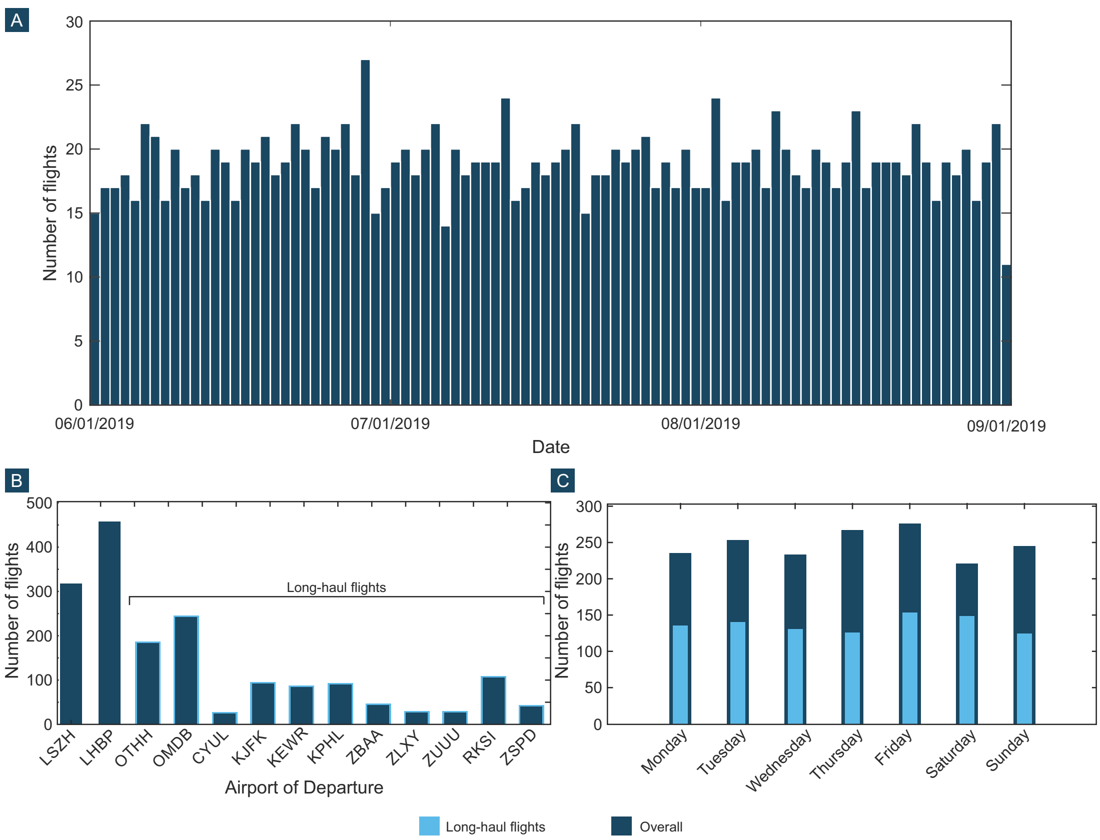


Figure 3.5: Total number of flights to LKPR categorized by date (A), airport of departure (B), and day of the week (C).

In Figure 3.5-B, a deeper exploration is undertaken into the inbound traffic to LKPR by documenting the counts of flights arriving at LKPR from their respective origin airports. This perspective aims to improve the comprehension of the global network converging at LKPR, specifically focusing on all airports outside the ECAC region, but not excluding intra-european origins such as LSZH and LHBP. In total, visibility is provided regarding the sources of international and intercontinental travel to LKPR. The figure offers an illustration of the connectivity with LKPR, revealing that it is linked to a total of 13 origin airports. Although the initial intention was to exclusively include long-haul flights in the analysis, the airports LSZH and LHBP were also added to create a more diverse dataset. In addition to expanding the studied data sample, these airports also contribute to defining the characteristics and behavior of short-haul flights. In any case, with short-haul flights, there is an inherent uncertainty that these "short" flights will be influenced by meteorological conditions along the route, particularly in terms of speed and wind direction, which is assumed more prominently in long-haul flights.

Further analysis of the distribution of traffic counts over the summer months allows for a breakdown to be conducted, examining the distributions on a weekday basis, as depicted in Figure 3.5-C. An interesting pattern is revealed through this detailed examination, with Fridays being identified as the busiest day, characterized by the highest number of overall and long-haul flights (153) arriving at LKPR. In contrast, Sunday and Thursday emerges as the least busy, marked by the lowest count of inbound long-haul flights - totaling 124 and 125 respectively. This weekday-specific breakdown yields valuable insights into the weekly variations in long-haul traffic to LKPR, shedding light on the flow of long-haul arrivals throughout the summer season.

Within the dataset, it is also possible to observe a variable distribution of flight times from various destinations. This is further illustrated in Figure 3.6, where these distributions are represented in the form of boxplots for FTFM and CTFM data. From the differences between median values, it is also observable that the anticipated flight times differ from the actual ones, not necessarily in favor of FTFM concerning shorter flight times.

These times are naturally dependent on the distance from the departure airport but also on the type of aircraft used for the flight. Therefore, Figure 3.6 also illustrates the representation of aircraft types for individual flight categories and their median flight times. For example, from Figure 3.6, a significant increase in flight time can be observed when using an AT72-type aircraft in the case of departure from LHBP. The variability of the flight time itself, the causes of this variability, and the difference between the planned and actual flight, where delays or significantly earlier arrivals may occur, are the subject of this dissertation. Nevertheless, in this section, it is necessary to demonstrate dependencies and data characteristics that may be significant in feeding a machine learning model.

Figure 3.6: Distributions of flight times from individual destinations for FTFM and CTFM data, along with the representation of median flight times from individual destinations depending on the type of aircraft.

From the information provided, it is clear that ADEP and Aircraft type will play crucial roles as predictors. Nevertheless, in this scenario, addressing how this information will be integrated into the training data becomes imperative. Given the sequential nature of the data (each specific flight is characterized by a sequence of variables from its initiation to its termination), it is essential to ensure that features which remain constant across time steps are not present within a single sequence. Consequently, determining the approach to incorporate this information into the input data becomes a key consideration, and this is elaborated upon in Chapter 3.3.

Another parameter that may have an intuitive impact on the prediction is the difference between the planned and actual aircraft pushback time from the stand, denoted as δ_{OBT} and calculated as $\delta_{OBT} = AOBT - SOBT$, where $AOBT$ is the actual off-block time (the actual time the aircraft is pushed back from the stand), and $SOBT$ represents the scheduled off-block time. This data feature is important in that if we assume a significantly meaningful difference between $AOBT$ and $SOBT$, there is an expectation that this difference will also be reflected in the actual landing time and the actual in-block time. The histogram of δ_{OBT}, including all observations in the dataset, along with the histograms of δ_{LT} and δ_{IBT} (see Chapter 3.1.1), is illustrated in Figure 3.7. In this figure, the mutual dependencies of these variables are also displayed.

Upon scrutinizing the data related to the total number of arrivals at LKPR, a striking pattern emerges when examining the δ_{IBT}. Figure 3.7 reveals a considerable range of variations, with some instances showing delays exceeding 2 hours. Further investigation has established that these substantial delays can often be attributed to flights that encountered unforeseen circumstances necessitating diversions from their original routes. After enduring hours of stopovers, these flights eventually resumed their journeys to their final destinations, resulting in the notable disparities between $AIBT$ and $SIBT$. Simultaneously, the data distribution reveals a notable trend of flights arriving ahead of their $SIBT$. The majority of instances show values concentrated within the timeframe of approximately 15 minutes or earlier before the scheduled arrival time.

The above can be observed in the case of both δ_{OBT} and δ_{LT}, although in these cases, the interpretation of these differences would be quite different. While in the case of δ_{LT}, it is conceivable in extreme cases that a significant delay could be caused, for example, by a diversion, in the case of δ_{OBT}, there are likely a multitude of potential causes, ranging from handling issues to exceptional situations at the airport. Nevertheless, Figure 3.7 reveals a linear dependency between δ_{LT} and δ_{OBT}, confirming the previous consideration regarding the relationship between these delays, i.e., if the aircraft is pushed back from the stand earlier/later, it will arrive earlier/later. This dependency can even be considered causal. In this specific case, if we look at the average residual error, it would be at the level of -2.4629 ± 7.6829 minutes.

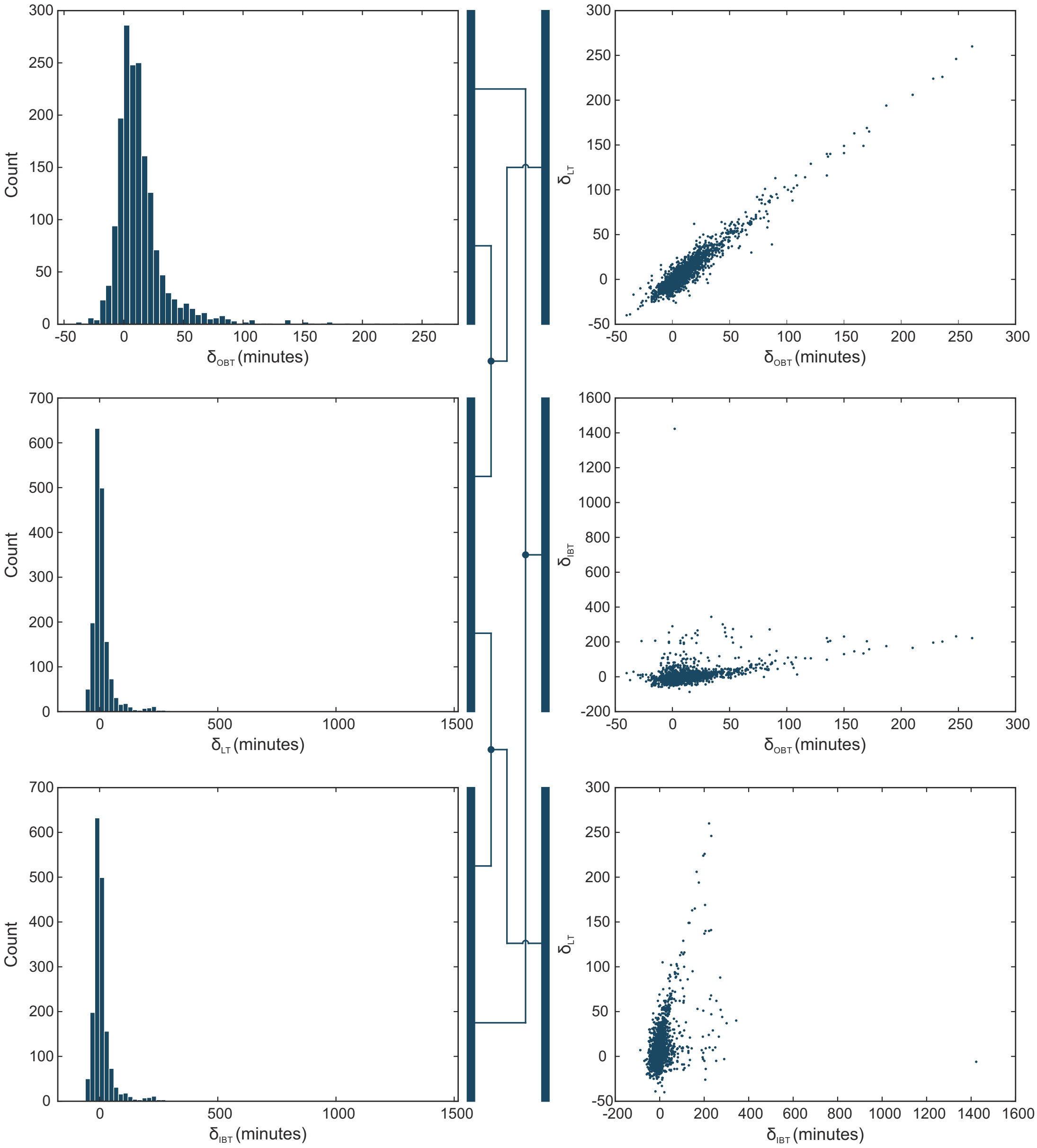


Figure 3.7: Distribution of time differences between actual and planned off-block times (δ_{OBT}), in-block times (δ_{IBT}), and landing times (δ_{LT}) along with the presentation of dependencies between these variables. Negative values indicating early arrivals.

However, this does not hold for the other dependencies related to δ_{IBT}, where a significant linear dependence is not observable. The presumed reason is that the time between landing and parking the aircraft at the stand can be influenced by many factors associated with the current airport situation. In contrast, linear independence between δ_{IBT} and δ_{OBT} is absolutely incomprehensible.

However, from this presentation, on one hand, a very nice linear and intuitively causal dependency is evident; on the other hand, there is a need for additional information for a better description of δ_{IBT}.

Note, that δ_{OBT} is a known quantity during departure and, therefore, will be used in the predictive model as an input variable.

The next, in this context, related information is the flight behavior depending on the route and its length in the context of delay. Of course, this is likely also related to the departure destination. Looking at the distribution of δ_{IBT} and δ_{LT} with respect to departure airports, see Figure 3.8, it can be observed that the departure airport influences both parameters δ. In other words, when testing the hypothesis of statistically significant differences in δ_{IBT} and δ_{LT} among groups characterized by the departure airport, statistically significant differences could be observed between some pairs of airports at the significance level $\alpha = 5\%$. This analysis was conducted through one-way ANOVA Hogg & Ledolter (1987). However, the interpretation of these results, including post-hoc analysis Hochberg & Tamhane (1987), would excessively saturate the information provided with statistical results in the context of this work. Either way, for the purposes of the work and the subsequent machine learning-related aspects, this information is sufficient and once again demonstrates the importance of the departure airport for specifying the target variables δ_{IBT} and δ_{LT}.

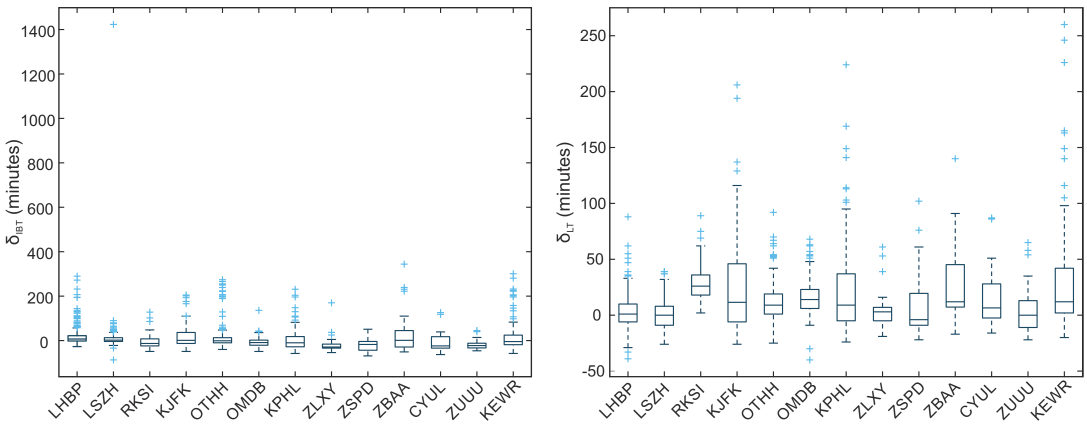


Figure 3.8: Distributions of time differences between actual and planned in-block times (δ_{IBT}) and landing times (δ_{LT}) for specific departure airports.

Figure 3.8 provides information about the distributions of δ_{IBT} and δ_{LT} for individual airports, but it does not offer insights into the evolution of delay on routes with a distinction between individual airports. To understand the flight behavior on specific flight routes, which should be characteristic for individual departure airports, it was necessary to work with a different parameter, namely FC_{delay} (see Eq. 3.7 in Chapter 3.1.1). The variable FC_{delay} signifies the current difference between the planned and actual flight at individual route points. It is clear that the variability of this variable will be substantial across all flights, significantly limiting its graphical representation and making characteristic patterns in the data difficult to discern. Therefore, Z-scores for FC_{delay} were calculated for each route, to address this issue, as:

$$Z_{delay_{FP}} = \frac{FC_{delay_{FP_i}} - \overline{FC_{delay_{FP}}}}{\sigma(FC_{delay_{FP}})}, \tag{3.16}$$

where FP represents a specific flight profile (ranging from 1 to 1730), i denotes a specific route point within FP, $\overline{\cdot}$ denotes the average value, and $\sigma(\cdot)$ represents the standard deviation of all FC_{delay} for a given FP. In such transformed data, the new values represent information about how many standard deviations a given value is away from the mean. In these data, 0 reflects the average value from the original data, 1 represents one standard deviation, and so forth.

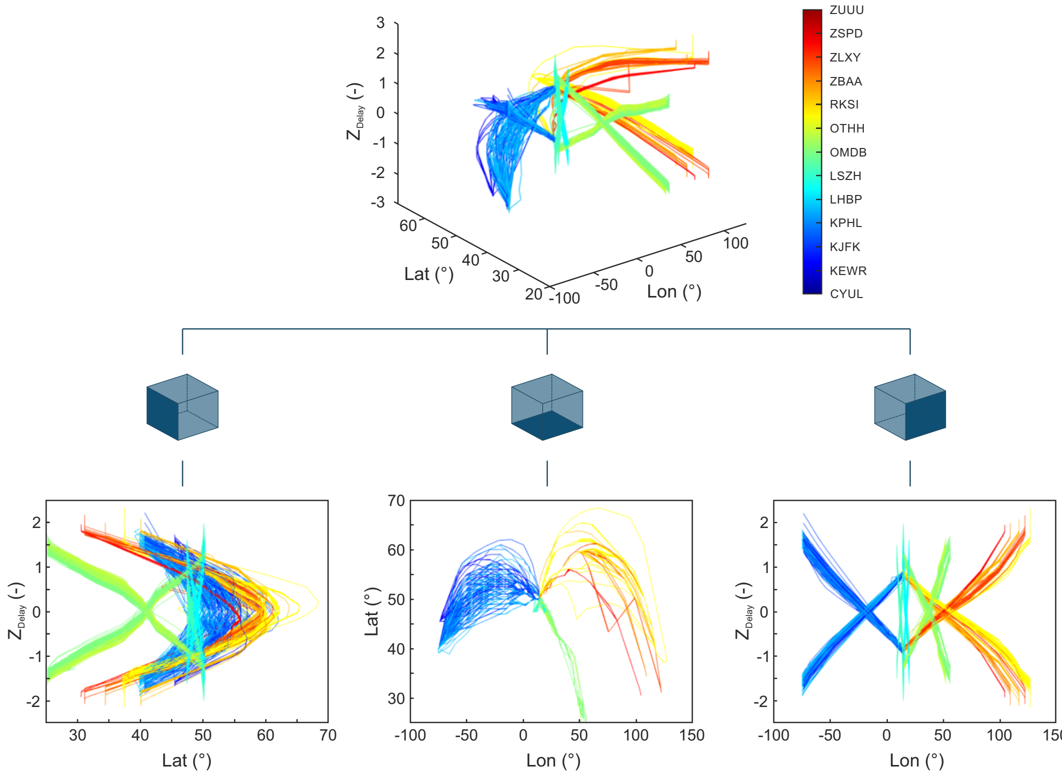


Figure 3.9: Dependency of the standardized time difference between actual and planned flight (Z_{delay}) on geographical coordinates and departure airport.

By visualizing Z_{delay} against geographical coordinates, the representation presented in Figure 3.9 can be obtained. Of course, due to the nature of Z-scores, information about the absolute time differences between FTFM and CTFM profiles is lost. However, this presentation demonstrates characteristic flight behavior depending on FC_{delay}. Even from this image, it is evident that flights from different destinations either continuously accumulate delays or consistently reduce delays.

Taking a closer look at these data by plotting Z_{delay} against distance and SEQ_ID (sequence point number), it is possible to observe that the time differences between FTFM and CTFM profiles behave constantly in some segments, i.e., they do not change (see Figure 3.10). In certain flight segments, the accumulation or reduction of delay is accelerated, while in some specific segments, the delay is decelerated. The mentioned findings indicate that the relative behavior of delays from specific destinations is, among other factors, dependent on the particular flight route. This was further confirmed by the principal component analysis, where variables such as SEQ_ID, Z_{delay}, and $Distance$ demonstrated the ability to cluster flights with respect to the departure airport.

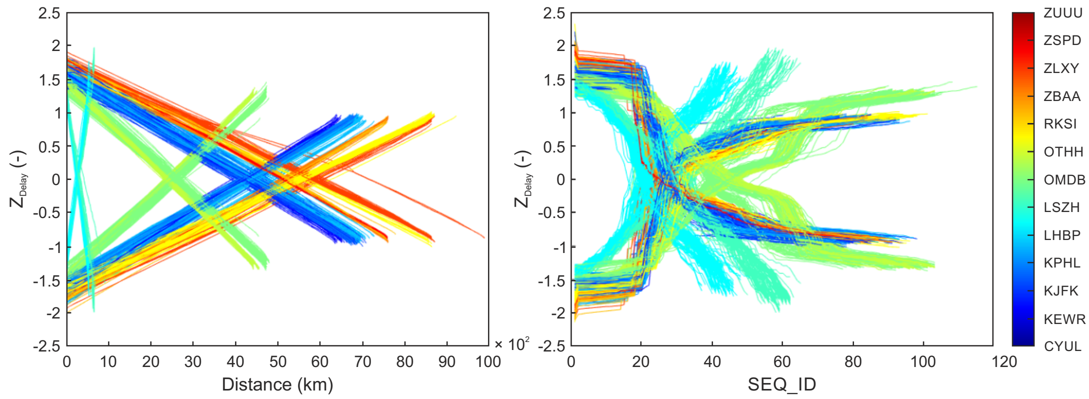


Figure 3.10: Dependency of the standardized time difference between the actual and planned flight (Z_{delay}) on the flown distance and sequence point number (SEQ_ID) along with the departure airport classification.

To link on the above findings, the effort was made to investigate whether these time increments and decrements on given segment may be brought about by wind component. Therefore, we start looking at correlation between flight ground speed and and wind speed along the flown routes. Figure 3.11-A provides a insight into the relationship between flight speed and the wind component, specifically headwind and tailwind experienced en-route. Upon the data scrutiny, no significant trend becomes apparent. As such flights that encounter headwinds along their route tend to not exhibit lower flight speeds but rather maintain the performance. Similarly, if the tail wind is experienced only a slight increase of the flight speed is observed. Although, the fact that wind conditions can have positive or negative impact on the efficiency and performance of the flights is generally recognized, the current data set doesn't not provide sufficient evidence to advocate this narrative. This may be due to the nature of these flights, actual velocity conditions over the given period, or limited amount of traffic sample used within this thesis. Seemingly, the Figure 3.11 further indicates some erroneous data, which may be result of false measurements feeding CTFM reports used in the analysis.

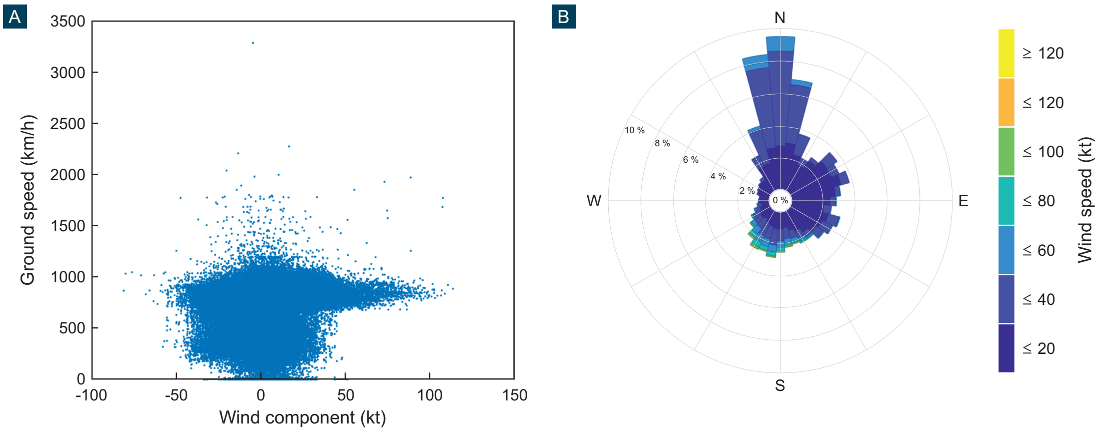


Figure 3.11: Overall view on comparision of wind component versus flight speed (A) and Visualization of prevailing wind direction and speed for all flights (B)

Despite of no real evidence that the wind would impact speed of the flight en-route, some additional insights into the dataset are provided. In Figure **??**-B, a depiction of prevailing winds along all the routes within the observed dataset emerges. The data underscores that northerly winds predominantly prevail in the regions covered by these flight paths. It's also notable that althought some portion of flights experienced also winds exceeding 100 knots during their journeys, the ratio of flights experiencing rather nominal conditions with lighter winds prevail. This insight may indicate reduced impact that high-velocity winds may have on flights within the studied sample.

To conclude on this part where the focus is placed on detection of the relationship between flight speed and wind component, the correlation analysis was conducted Gogtay & Thatte (2017). The Figure 3.12-B shows, that while there is a predominant negative correlation for the majority of flights, indicating that wind conditions typically affect flight speed in the opposite direction, the presence of high positive correlations suggests that in certain cases, the wind component may have a significant impact on flight speed in the same direction. A negative correlation between flight speed and headwind speed can indeed indicate that aircraft are adjusting their speed to counter the effects of headwinds. In other words, when faced with stronger headwinds, aircraft may increase their flight speed to maintain their scheduled arrival times or minimize the impact of adverse wind conditions. While a negative correlation suggests a compensatory effect, it doesn't necessarily imply that wind has no effect on flight duration. Instead, it indicates that the aircraft are responding to the wind conditions, and this response results in a negative correlation between the two variables. Overall, the distribution of the correlation coefficients within the scale suggests no real conclusion on wind effect can be made.

In spite of the reasoning, the Figure 3.12-A indicates slight but progressive increment in positive correlations between flight speed and wind speed in the period of August. This may suggest a potential seasonal trend, which may be influenced by changing weather patterns, wind conditions, or operational practices during different times of the year.

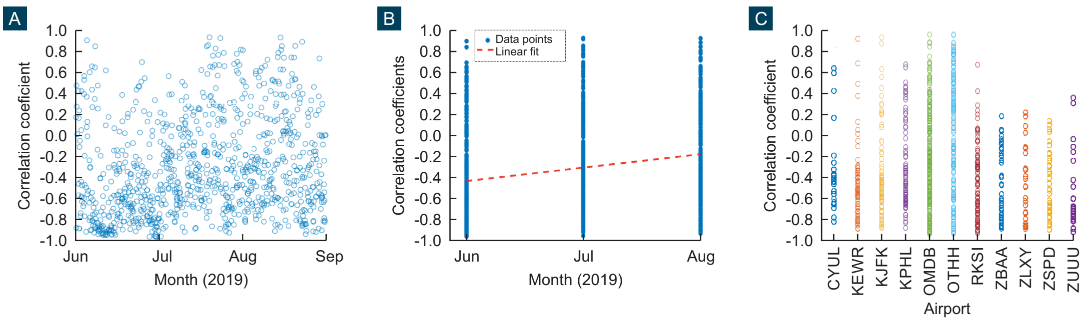


Figure 3.12: Correlation coeficients depicting relationship between wind component and flight speed per date (A), Correlation coeficients and a linear regression model depicting a trendline (B) and Correlation coeficient depicting relationship between wind component and flight speed per airport (C).

This tendency of higher distribution of positive coefficients can also be seen on the Figure 3.12-B. However, and as said, the impact seems to be negligible.

The breakdown observed in Figure 3.12-C, where some airports like Philadelphia (KPHL), Dubai (OMDB), and Hamad (OTHH) show both strong positive and strong negative correlations between flight speed and wind speed, is an interesting finding. It indeed suggests that there is variation in how different flight operations are adapted to the wind component. The presence of both positive and negative correlations for the same group of airports indicates that airlines operating from these airports might have varying strategies or responses to changing wind conditions, which may be translated into manipulation of cost index applied on respective segments.

Stemming from the desriptive analysis, and based on trends and relationships observed, it transpires that the available data manifest certain dependancies. At the same time certain correlations performed on meteorological and flight related data suggest that no or limited relationship may exist. As such, and considering the fact mentioned above, the dataset may not neccessiraly be meeting requirements for a predictive modelling approach that would enable computation of a predicted arrival delay. Thus, it is imperative to analyze the predictive technique used and options to include additional data as explained in Chapter 3.3 below.

3.3 Aditional variables for model

A sequential pattern is followed by our data, where each flight is represented by a sequence of variables (see Table 3.1 for an overview of the data structure.), as mentioned earlier. In total, considering individual flights and the waypoints within them, 122 158 observations are included in this dataset, 1730 filtered arrivals were recorded at LKPR and the dataset is composed of 45 features and two dependent variables, namely δ_{LT} and δ_{IBT}.

Not all features, however, can be used for modeling due to containing information about real flights, i.e., CTFM profiles and details associated with actual flight execution. For prediction, only information available at the departure phase must be selected.

Another thing that needed to be taken into account is dealing with constant features. The mentioned was more or less discussed in Chapter 3.2 (page 47). It is therefore a matter that, by the nature of the data, the prediction approach must be based on a sequence-to-one approach Ye et al. (2022). For this purpose, one can utilize, for example, the Long Short-Term Memory architecture of a recurrent neural network Le et al. (2019) (this is further described in the following chapters). This approach is based on learning through multivariate time series, where these time series essentially represent a sequence describing one flight, with the aim of addressing one corresponding dependent variable (in this case, the predicted delay) for each sequence. Including input features that remain constant throughout a sequence (such as ADEP) may not always be beneficial for several reasons. Firstly, such features provide limited information gain as they don't vary across time steps, and sequential models typically rely on changing patterns for effective predictions. Additionally, the introduction of constant features can lead to redundancy, repeating the same information across all time steps without substantial added value for the model.

Moreover, there's a risk of potential overfitting when constant features are included. The model might memorize specific patterns associated with these constants, compromising its ability to learn more generalized patterns in the data. Furthermore, there's an increase in computational overhead during training when constant features are present, potentially without a proportional improvement in model performance.

Lastly, the presence of constant features may pose a challenge for sequential models like Long Short-Term Memory (LSTM) Le et al. (2019), which excel in capturing temporal dependencies and dynamic patterns. Constant features might divert the model's attention away from the evolving aspects of the data, making it difficult to effectively learn the underlying dynamics.

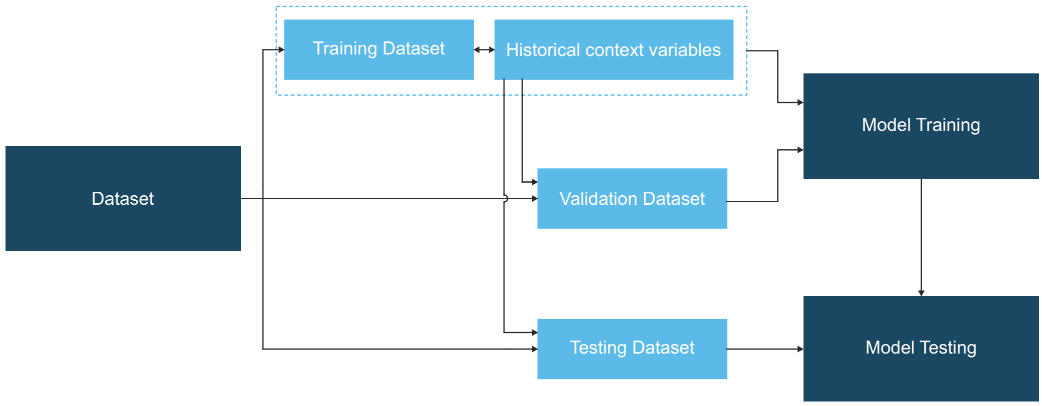


Figure 3.13: Schematic representation of utilizing historical data in model training.

The assumption is that a better model can be obtained by utilizing historical CTFM data. This involves comparing the planned flight time against the average of actually flown flights, considering constant features such as Aircraft type, ADEP, etc. For proper training, these variables need to be calculated from the training data to avoid introducing bias into the prediction during testing (note: the dataset will be standardly divided into training, validation, and testing sets). A schematic representation of this process is shown in Figure 3.13.

The fundamental idea was to calculate the average (historical) flight times from CTFM profiles for each aircraft type, within a specific departing airport. This information was computed from the training dataset, following its definition, and thus:

$$hFT_{ad,ac} = \frac{\sum \mathbb{A}_{FT_{ad,ac}}}{n} \tag{3.17}$$

where hFT is the average historical flight time from the subdataset defined based on the unique departing airport ad and unique aircraft type ac from the training dataset $\mathbb{A}$. $\mathbb{A}_{FT_{ad,ac}}$ je pritom definované ako:

$$\mathbb{A}_{FT_{ad,ac}} = FT_{f_n} - FT_{f_1} \subset \mathbb{A}_{ad_f,ac_f} \subset \mathbb{A} \tag{3.18}$$

where FT_f represents the time vector for individual waypoints of a specific flight f. The flight duration is then defined as the difference in times between the last n and the first waypoint of that specific flight. FT_f is a subset of the dataset $\mathbb{A}_{ad_f,ac_f}$, which is a subset of the overall training dataset $\mathbb{A}$. $\mathbb{A}_{ad_f,ac_f}$ is defined by the departing airport ad and aircraft type ac. The variables ad and ac are categorical in the original dataset and were used for its categorization.

The average values of historical parameters were calculated not only for the overall flight duration but also for the delay increment hFC_{delay} (see equation 3.7 in Chapter 3.1.1), historical $h\delta_{OBT}$, $h\delta_{IBT}$, $h\delta_{LT}$, flight speed hFS and wind component (headwind/tailwind) $hw_{component}$ (see equation 3.15).

As mentioned earlier and will become clearer from the following description, the creation of the predictive model will utilize the LSTM neural network architecture approach. The data for the final dataset were modified to meet the requirements of this approach.

This architecture operates with multivariate time series, so it was necessary to create time series for each flight, describing the flight and containing the necessary information for training. Let F represent the data matrix for a specific flight included in the dataset. The defined predictors will then look as follows (note: the notation of predictors is kept in the form used in modeling in the Matlab environment):

$$dSIBT\_TO = F_{SIBT} - F_{TO}, \tag{3.19}$$

$$dSOBT\_TO = F_{SOBT} - F_{TO}, \tag{3.20}$$

$$dATOT\_TO = F_{ATOT} - F_{TO}, \tag{3.21}$$

$$dAOBT\_TO = F_{AOBT} - F_{TO}, \tag{3.22}$$

where F_{TO} represents the vector of times of passage for individual waypoints on the planned route (from the FTFM profile) for a specific flight. F_{SIBT}, F_{SOBT}, F_{ATOT} and F_{AOBT} represent the scheduled in-block time and off-block time, the actual take-off time, and the real off-block time, respectively. Before calculating these differences, the times were converted to posix time. F_{SIBT}, F_{SOBT}, F_{ATOT} and F_{AOBT} represent a constant feature (represented by one number). This means that the aforementioned variables define time series of the scheduled flight of the aircraft over individual planned waypoints, taking into account already known time deviations.

Furthermore, information about the flight, such as latitude (LAT), longitude (LON), altitude (ALT), ground speed (FS), and the distance of the waypoint from the departing airport ($POINT\_DIST$), was utilized as predictors.

Naturally, information about the weather along the route at the time of departure, including wind speed ($windspd$), wind direction ($winddir$), and wind component ($WindComponent$), was also used. All these features also represent time series as they are defined for each waypoint in F.

The variables mentioned above could be created before dividing them into training, validation, and test datasets. After splitting the data into these types of datasets, the principle of the approach described above regarding historical data was then applied. Let Tr, Va, and Te represent the training, validation, and test datasets, respectively, and let Tr_F, Va_F, and Te_F represent a specific flight in these datasets. Then it is possible to apply the approach described above in equations 3.17 and 3.18. In this case, $\mathbb{A} = Tr$. First of all, "historical" values from Tr were calculated, which were then used to define new variables also in datasets Va and Te. For the example of one flight, the calculations would look as follows (note: the same notation is used as was used in the real solution in the Matlab environment):

$$histDelay\_TO = Va_{F_{TO}} - hFC_{delay}, \;\; histDelay\_TO \in Va, \tag{3.23}$$

$$histD\_IBT\_TO = Va_{F_{TO}} - h\delta_{IBT}, \;\; histD\_IBT\_TO \in Va, \tag{3.24}$$

$$histD\_OBT\_TO = Va_{F_{TO}} - h\delta_{OBT}, \;\; histD\_OBT\_TO \in Va, \tag{3.25}$$

$$histD\_LT\_TO = Va_{F_{TO}} - h\delta_{LT}, \;\; histD\_LT\_TO \in Va, \tag{3.26}$$

$$histFS\_FS = Va_{F_{FS}} - hFS, \;\; histFS\_FS \in Va, \tag{3.27}$$

$$histWC\_WC = Va_{F_{wc}} - hw_{component}, \;\; histWC\_WC \in Va, \tag{3.28}$$

where the above-calculated parameters represent time series of the difference between the planned development of a specific parameter and the average historical development. In all cases, it is the difference of a specific parameter for a flight selected based on the departing airport and aircraft type, with the subtracted value representing the average of that parameter from all flights classified according to this specific departing airport and aircraft type in Tr. This is represented for the Va subdataset, but implemented for the Tr and Te datasets as well.

The overall dataset used for prediction utilizes the predictors described above, which have the format of time series forming a data sequence for a specific flight. Each of these sequences is assigned exactly one dependent variable, either δ_{LT} or δ_{IBT}, depending on the model being built (either for predicting δ_{LT} or δ_{IBT}). The dataset contains 1730 flights, with each flight being described by 18 time series. Emphasis was placed on the possibility of using such an approach in real operations, meaning each flight is described by known information at the time of departure.

3.4 Predictive modelling

For the purposes of the work, the tendency is to primarily predict δ_{IBT}. However, as evident from the previous analysis, such a task can be quite challenging, mainly due to the lack of available information about the flight's behavioral aspects in the time interval between landing and arrival at the stand. Therefore, prediction will also be performed for δ_{LT}.

As outlined in previous chapters, in this case, two approaches can be utilized. While in the case of δ_{IBT} and δ_{LT}, it involves a sequence-to-one approach Ye et al. (2022), prediction can also be performed using a sequence-to-sequence approach Saxena et al. (2008). In the latter case, the target variable would represent the delay increment at each point of the flight.

In other words, the primary difference between the sequence-to-one and sequence-to-sequence approaches lies in the structure of their input and output. In the sequence-to-one paradigm, the model takes a sequence of inputs but produces a single output, often referred to as a many-to-one architecture. This is commonly employed in tasks like sentiment analysis (Chan et al., 2022), where the goal is to classify the sentiment of an entire sequence, such as a sentence or document.

On the other hand, the sequence-to-sequence approach deals with both input and output as sequences, forming a many-to-many architecture. This model configuration is suitable for tasks where the input and output have variable lengths, as seen in machine translation, text summarization, and speech-to-text conversion. In machine translation, for example, the model processes a sequence of words in one language as input and generates a sequence of words in another language as output.

In summary, the key distinction between these approaches is in the nature of the output. Sequence-to-one models produce a single output based on a sequence of inputs, while sequence-to-sequence models are designed for tasks involving the transformation of one sequence into another, where both input and output are sequences of variable lengths (Rao et al., 2022).

In addressing this work, both approaches were employed, but for practical reasons within the context of the work itself, only the sequence-to-one approach will be utilized and further described. This is because the main focus of the work is not based on predicting delays at each point of the route. Although solving such a problem would be interesting, its practical application would likely require updating the predictive model during the flight with real-time flight information. For this reason, it seems more suitable and practical to predict a single final delay based on the information available at "take-off" including initial CTFM timestamp but also available FTFM profile. Additionally, historical CTFM flown profiles on given route may also bring additional insights and improve modelling.

3.4.1 Sequence to sequence approach

Based on the earlier explanation, the sequence-to-sequence approach may not be the most appropriate option for modeling. This approach involves predicting delays across the entire trajectory as a sequence, which may not effectively solve the problem. Put differently, this technique would convert input parameters, including the sequence of FTFM profiles, into an output sequence of CTFM.

In the course of addressing the work, however, a deeper exploration of this approach was undertaken. The philosophy was to predict the delay increment at a point on the flight route. In this case, a neural network was not utilized, but rather a linear regression approach. The dataset was treated as a whole, with each row describing a single waypoint. All types of variables, both categorical and continuous, were considered, but the flight progression itself was not taken into account. The goal, among other things, was to elucidate the importance of predictors capable of capturing the delay increment in regression.

These results were published and presented at the international conference mentioned below.

Appendix D

Špák, M., Socha, V., Hanakova, L., Matowicki, M., Predictability of In-block Time Deviations: An Analysis of Operational Data, Tactical Flight Models and Meteorological Information, Proc. of XVIII. International Scientific Conference New Trends in Aviation Development. The High Tatras – Starý Smokovec: IEEE, 2023. p. 222-227. ISBN 979-8-3503-7040-9.

The thesis findings highlight the significance of SOBT, ATOT (Actual Takeoff Time), and SIBT values in predictive modeling for anticipating in-block time deviations. These parameters offer insights into potential delays, which can be attributed to factors like operational patterns of airlines, air navigation services, airport layout complexity, or weather conditions along the route within a specified time period.

Additionally, the choice of the ADEP as a predictive element emphasizes the role of distance in improving prediction accuracy. Considering gate or terminal information, including variable taxiing times, reduces prediction error and enhances accuracy.

The thesis also introduced a synthetic variable called the wind component, reflecting velocity in relation to flight heading, to enrich the modeling phase. While no significant improvement was observed, there's potential for further research on the impact of meteorological conditions along the route on flight duration and in-block time deviations.

3.4.2 Sequence to one approach

As mentioned several times above, for the purposes of the work in the context of predictions, a recurrent neural network with LTSM architecture was used. Specifically, the Bidirectional LTSM (BiLTSM) was employed. The rationale, basic logic behind choosing this approach, characteristics of LTSM and BiLTSM, and the configuration of the neural network for fine-tuning predictions are described below.

The LSTM neural network architecture has proven effective in addressing the vanishing gradient problem associated with traditional recurrent neural networks (RNNs) (Da Silva & De Moura Meneses, 2023). Particularly suitable for sequential data analysis.

The vanishing gradient problem is a challenge that arises during the training of deep neural networks, particularly in the context of recurrent neural networks (RNNs). It is related to the difficulty of updating the weights of the earlier layers in a deep network during the backpropagation process.

During backpropagation as explained by Gerlinghoff et al. (2023), gradients are calculated and used to update the weights of the network to minimize the error. However, as the gradient is propagated backward through the layers, it can diminish exponentially as it moves toward the input layers. This means that the gradients for the weights in the early layers become very small, approaching zero. As a result, the weights of these early layers are updated very little, if at all, during training.

When the gradients become too small, the network has difficulty learning meaningful representations from the input data. This phenomenon is particularly problematic in deep networks and sequences where information needs to be captured over long distances or time steps. In the context of sequential data processing, such as language modeling or time series prediction, the vanishing gradient problem can make it challenging for the model to capture dependencies that span across many steps.

LSTMs were introduced by Sepp Hochreiter and Jürgen Schmidhuber in 1997. The key idea behind LSTMs is the introduction of a memory cell and gating mechanisms that control the flow of information, allowing for better learning and retention of long-term dependencies.

Sherstinsky (2020) explains that LSTM is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem in traditional RNNs, which makes it difficult for them to capture long-term dependencies in sequential data. LSTMs are well-suited for tasks involving sequences, such as time series prediction, natural language processing, and speech recognition (Senthilkumar et al., 2022).

In a sequence-to-one approach using LSTM, the goal is typically to process an input sequence and produce a single output at the end of the sequence (Sherstinsky, 2020). This is common in tasks like sentiment analysis or image classification where the entire sequence contributes to the final decision.

The basic components of LTSM include Cell State (C_t), Hidden State (h_t), Gates and Cell Input (g_t). The cell state serves as the repository for the network's long-term memory, facilitating the storage and transmission of information across various time steps. Updates to the cell state occur dynamically as the network processes the input sequence.

Representing short-term memory or information pertinent to the ongoing task, the hidden state is updated at each time step based on the input, prior hidden state, and cell state. The final hidden state, or a derived transformation thereof, is instrumental in making predictions or classifications for the entire sequence.

The Input Gate (i_t), Forget Gate (f_t), and Output Gate (o_t) regulate the flow of information within the LSTM. Schematic interpretation of vanilla LSTM is depicted in Figure 3.14.

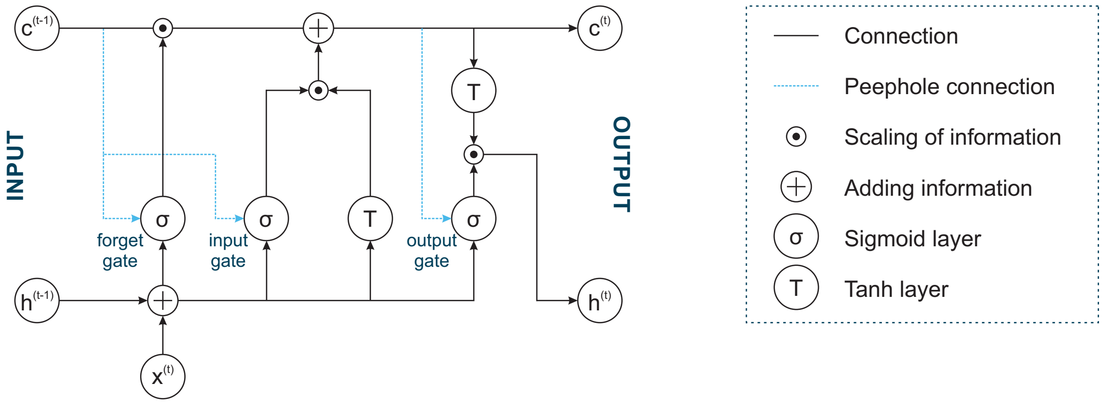


Figure 3.14: Vanilla LSTM architecture

Cell Input represents new information that may be incorporated into the cell state and is computed using the hyperbolic tangent (tanh) activation function. The cell state update (eq. 3.29) and hidden state update (eq. 3.30) is performed using the following approach:

$$C_t = f_t \cdot C_{t+1} + i_t \cdot g_t \tag{3.29}$$

$$h_t = o_t \cdot tanh(C_t) \tag{3.30}$$

At the end of the sequence, there might be a fully connected layer or another type of output layer that takes the final hidden state as input to produce the desired output. In the case of models created in this work, two fully connected layers were used, one with 100 hidden units followed by a layer with 25 hidden units. This configuration was done empirically through repeated testing during the course of the dissertation work. The training process involves adjusting the weights (W and U) and biases (b) through backpropagation and gradient descent.

In the sequence-to-one approach, the LSTM sequentially processes input (X) time steps, updating the cell state (C_t) and hidden state (h_t) at each step. The ultimate prediction for the entire sequence is derived from the final hidden state (h_T).

Bidirectional Long Short-Term Memory (BiLSTM) is an extension of the traditional LSTM architecture (Jamshidzadeh et al., 2023). In a standard LSTM, information is processed from the beginning to the end of a sequence, capturing dependencies from the past to the present. In contrast, BiLSTM processes the sequence in both forward and backward directions simultaneously.

The bidirectional nature of BiLSTM is achieved by maintaining two hidden states for each time step—one for the forward pass and one for the backward pass. The final representation at each time step is a concatenation of these two hidden states, providing a comprehensive understanding of the context from both past and future perspectives.

This bidirectional processing is particularly advantageous in tasks where context from both directions is crucial. For instance, in natural language processing (NLP) tasks like named entity recognition or part-of-speech tagging, the meaning of a word often relies on information from both preceding and following words (Prabakaran et al., 2023).

During training, the backward pass in BiLSTM is typically performed in reverse time order. The model is trained to predict the current time step based on information from both past and future time steps, facilitating the capture of dependencies in both directions.

BiLSTM networks are applied across various domains where capturing context from both past and future information is essential. As worked out by Chai & Li (2019), in NLP, BiLSTM is employed for tasks like sentiment analysis, named entity recognition, and part-of-speech tagging. In speech recognition, BiLSTM models dependencies in acoustic features, improving accuracy. Time series prediction benefits from BiLSTM in capturing temporal dependencies for accurate forecasting.

Handwriting recognition, particularly in the context of sequences of handwritten strokes, utilizes BiLSTM for its ability to learn patterns and dependencies. In biomedical data analysis, BiLSTM is applied to tasks such as DNA sequence analysis and protein structure prediction.

Gesture recognition tasks leverage BiLSTM to understand sequences of movements, enhancing the recognition of complex gestures. In event detection within videos, BiLSTM aids in capturing temporal context, contributing to improved analysis and understanding of dynamic events.

These applications demonstrate the versatility of BiLSTM in capturing contextual information from both ends of a sequence, making it a valuable architecture for tasks requiring a nuanced understanding of context.

In the context of this work, the chosen architecture at the end was specifically the BiLSTM architecture. This decision was made with the goal of working with a perspective on the entire flight behavior described by input parameters, rather than through its gradual development.

In the development of a Bidirectional BiLSTM network, the configuration of hyperparameters plays a pivotal role in optimizing the model's performance. The selection of hyperparameters involves careful consideration of various aspects to strike a balance between model complexity and generalization. The theoretical framework provided below describes key hyperparameters and their impact on the BiLSTM architecture.

The number of LSTM units is a fundamental hyperparameter, dictating the network's capacity to capture intricate patterns. The decision on the number of layers influences the depth of the architecture, with deeper structures capable of grasping more complex dependencies, albeit with an associated risk of overfitting.

The choice between a bidirectional or unidirectional structure is a critical decision. Bidirectional LSTMs, by processing input sequences in both forward and backward directions, excel in capturing information from past and future time steps, providing a comprehensive understanding of sequential data. The dropout rate, as a regularization technique, is pivotal in balancing the trade-off between preventing overfitting and maintaining model performance.

The learning rate and batch size are key optimization parameters. The learning rate determines the step size during optimization, influencing convergence, while the batch size impacts regularization and memory requirements. The number of training epochs, representing the complete traversal of the training dataset, poses implications for both training duration and the risk of overfitting.

The optimizer, a crucial component in the training process, merits careful consideration. Choices such as Adam or RMSprop are contingent on the dataset's characteristics and the specifics of the problem at hand. Additionally, the activation functions within the LSTM cells, such as tanh or ReLU, contribute significantly to the model's capacity to capture temporal patterns.

Tailoring the input sequence length to the inherent characteristics of the sequential data is imperative. This adaptation ensures the network is appropriately attuned to the nature of the input information. Finally, the determination of the output dimensionality, often aligning with the number of classes in classification tasks, completes the hyperparameter configuration.

The configuration of hyperparameters for the neural network was carried out experimentally and empirically, considering continuous monitoring of the behavior of the training process and the predictive capability of the resulting model. The network configuration is shown below in Matlab notation. The network presented below works with an input dataset characterized by the variable $XTrain$ and a target variable contained in the variable $TTrain$ and validation datasets $XValidation$ a $TValidation$. At the output of the training, the function provides the final model (net) selected based on the smallest validation loss in the training, along with information and data from the training process (tr), such as the progression of RMSE, loss, etc.

```
numResponses=1;
numHiddenUnits=50;

layers = [ ...
          sequenceInputLayer(size(X{1},1),...
         'Normalization','zscore')
          bilstmLayer(numHiddenUnits,...
         'OutputMode','last',...
         'StateActivationFunction', 'tanh',...
         'GateActivationFunction', 'sigmoid')
         fullyConnectedLayer(100)
         dropoutLayer(0.8)
         fullyConnectedLayer(25)
         dropoutLayer(0.8)
         fullyConnectedLayer(numResponses)
         regressionLayer
        ];

 options = trainingOptions('adam', ...
        'MaxEpochs', 300, ...
        'ValidationData', {XValidation TValidation}, ...
        'InitialLearnRate', 1e-3, ...
        'SequenceLength', 'longest', ...
        'Shuffle', 'never', ...
        'OutputNetwork', 'best-validation-loss', ...
        'GradientThreshold', 1
        );

[net, tr] = trainNetwork(XTrain, TTrain, layers,
   options)
```

To determine the learning stability, an approach was employed in which the entire training process, along with predictor computation (see Chapter 3.3), was repeated 60 times. Within this process, a training, validation, and test dataset were consistently selected in a 70/15/15 ratio by random selection, and basic performance characteristics of each resulting model were calculated when predicting using the test dataset. Specifically, this involved the computation of Root Mean Square Error ($RMSE$), Mean Absolute Error (MAE), the mean value of residuals ($rMean$), the standard deviation of residuals ($rSTD$), and R^2.

The above was done to avoid selecting data for which the model would work/not work based on a single random selection. Average performance characteristics were then calculated from these 60 iterations. These resulting characteristics indicate the reliability of the final model and the stability of the model-building approach set (see Chapter 4). It is also necessary to note that the same approach was used for predicting δ_{IBT} and δ_{LT}. However, the training results for δ_{IBT} and δ_{LT} will be presented and interpreted separately.

4. Results

During the training of the predictive model for estimating δ_{IBT} and δ_{LT}, performance characteristics were iteratively collected for 60 models. The average values of prediction performance indicators for models trained for δ_{IBT} were at the level of $RMSE = 23.7736 \pm 20.4742$, $MAE = 11.2068 \pm 1.9567$, $rMean = -1.2613 \pm 3.0895$, $rSTD = 23.5353 \pm 20.5743$, and $R^2 = 0.8050 \pm 0.2058$.

The resulting models exhibit a relatively high degree of variability, which is caused by the random selection of the training dataset. This implies that to optimize the training process for more stable results in the context of resulting models, additional information should be added to the dataset. Nevertheless, any model trained on the data defined in Chapter 3.3 should achieve predictive performance within the intervals illustrated above.

On the other hand, for models predicting δ_{LT}, the following metrics were obtained: $RMSE = 9.2779 \pm 1.4897$, $MAE = 6.1911 \pm 0.5155$, $rMean = 0.3517 \pm 1.4294$, $rSTD = 9.1748 \pm 1.5244$, and $R^2 = 0.8999 \pm 0.0243$. Compared to the previous models, the performance in predicting δ_{LT} is much more stable, indicating the independence of the random selection of the training dataset on the resulting model.

In the context mentioned above, it is also possible to observe that the final models for predicting δ_{IBT} and δ_{LT} (trained independently of the previous 60 cases) exhibit similar performance to the respective average performance.

Figure 4.1 offers a graphical depiction of the final model training process throughout individual iterations, taking into account RMSE and loss for both training and validation data. Here, the loss signifies the model's performance, quantifying the disparity between predicted and actual values within a specified input dataset. RMSE on the other hand gauges the average magnitude of errors in predicted versus actual values, calculated as the square root of the mean squared differences.

Before training, the training option was configured to create a model based on the minimum validation loss. In the case of predicting δ_{IBT} (Figure 4.1-A), this model was selected with an RMSE of 31.5 minutes and a loss of 213. This result suggests relatively high error, which may be caused due variations within the training dataset. In this case, it's apparent that the learning process was unable to advance beyond this level.

On the contrary when re-directing our focus on prediction of δ_{LT} and as seen on Figure 4.1-B, the uncertainty seems to significantly decrease, suggesting more stable inputs. The final model for predicting δ_{LT} was selected based on the best validation loss, with an RMSE of 20.64 minutes and a loss of 7.9. Generally speaking, the training and validation loss both decrease and become relatively stabilized at certain point, which may indicate a good fit of the model.

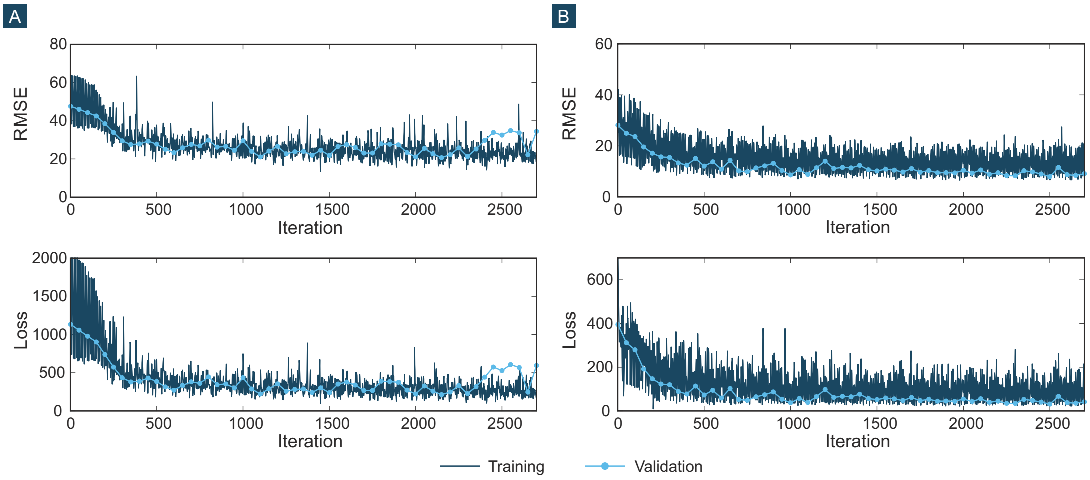


Figure 4.1: Visualization of the training process for the model predicting δ_{IBT} (A) and δ_{LT} (B).

Once the training has successfuly been performed, the testing of the final models took place with results captured in Figure 4.2. When looking at Figure 4.2-A, the outputs of the testing suggest that predictability of in-block deviations may operate with certain volatility. This may be caused due to unavailibility or limited visibility the model has in relation to key factors influencing the time aircraft spends between landing and parking at the stand. In spite of these statements, the RMSE of 18.37 minutes and residual Standard Deviation (rSTD) of 18.40 minutes may still be operationally seen as beneficial when pre-tactical and close to real-time efforts to optimize airport resources, are made.

Furthermore, when looking at Figure 4.2-B, the model performs evidently well when predicting δ_{LT} with RMSE at 8.71 minutes. With such a result, the operational planning becomes resilient as presence of inbound traffic on the runway can be predicted with a rSTD of 8.70 minutes. Thus, the stand and gate allocators together with remaining ground resource planners in the ranks of ground handlers may efficiently anticipate demand and adjust respective capacity.

In summary, when running the exercise, where the performance of the model predicting δ_{IBT} and δ_{LT} is compared, the predictions of the latter are of better quality, thus closer to the ground truth.

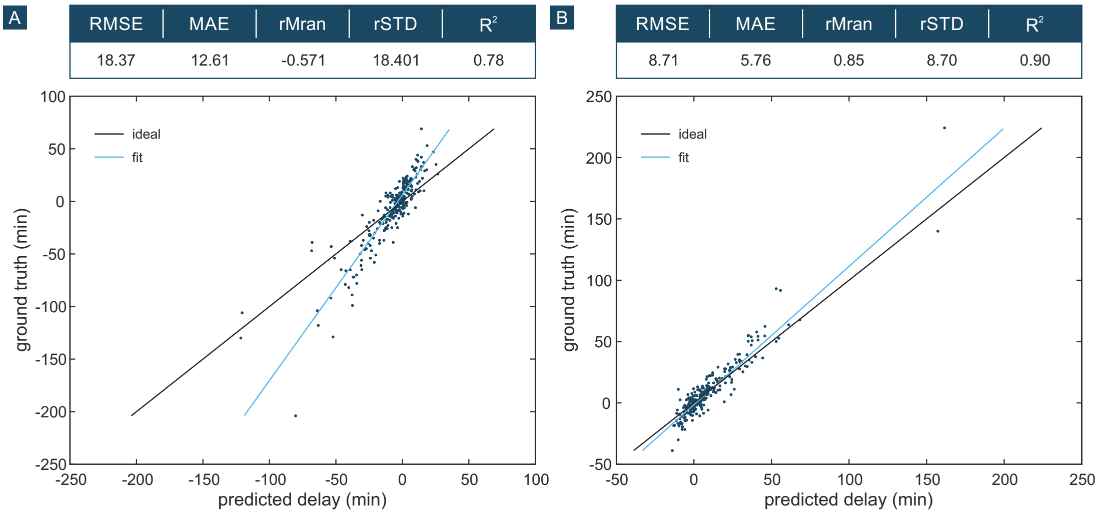


Figure 4.2: The dependency of predicted values on ground truth for predictions of δ_{IBT} (A) and δ_{LT} (B) along with performance characteristics of the models.

5. Discussion

The thesis revealed that the development of an operational playbook in form of a predictive model based on operational, meteorological and trajectory related data may generate satisfactory results. Firstly, the endeveaours were made to use linear regression and predict delays delay increments at each single point on the flight route. Different types of data including categorical and continuous variables were used allowing to focus on predictability of FC_{delay}. The analysis using forward selection process proved that values such as SOBT, ATOT, LOBT play an important role in predicting the said delay increments. Furthermore, progressive adding of predictive elements including terminal information, or wind component improved the model performance and its accuracy as such.

However, the regression approach did not allow studying behavioural changes of input variables along the route as the dataset was treated as a whole, with each flight record stored in a single row. Furthermore, the prediction of FC_{delay}, which constitutes a vector of time difference between planned and flown trajectories, was a diversion from an initially declared target, which focuced on prediction of in-block time delays. This led to additional research, which entailed development of an operational playbook in form of a predictive model applying BiLSTM network approach. The sequence to one technique allowed to study the behavioral evolution of input variables along the flight. Stemming from the experience gained with the regression approach, the BiLSTM was selected due to its ability to concentrate on generation of δ_{IBT} predictations, where this variable constitute a difference between SIBT and AIBT values.

However, as explained in Chapter 4 the elevated error when predicting in-block time delay may imply variations in the training dataset. As such, it indicates a learning process that couldn't progress further due to nature and volume of data elements present. The fact that the input data may have suffered level of inconsistency when undergoing training process, this could cause in-block time predictions being of rather undesired quality with RMSE of 18.37 minutes. This also imply that the model cannot be well generalized from the training dataset.

In light of the suboptimal accuracy observed in predicting in-block times, a subsequent initiative was undertaken to validate the potential for predictive capabilities in relation to forecasting landing times. In contrast, when shifting our attention to predicting δ_{LT}, there is a notable decrease in uncertainty, indicating more consistent inputs and robust training, which is also proved by RMSE value of 8.71 minutes. Moreover, as depicted in Figure 4.2-B, the model demonstrates notable and greater success in predicting δ_{LT} as oppose to δ_{IBT}. An ultimate

RMSE achieving a value of 8.71 minutes suggests that testing was successful. As such, this discovery proves to be intriguing and enlightening. It becomes apparent that the lower quality of predictions for in-block time is compensated by the enhanced accuracy achieved in predicting landing times. This suggests that the decline in predictive performance may be linked to the specific phase of operations between these two variables. In essence, the observed discrepancy underscores the importance of understanding of the operational phases and their impact on predictive outcomes.

The phase in question appears to be when the aircraft taxis into its parking position. Delving into this taxing-to-parking maneuver may be crucial to understanding the factors influencing predictive performance of existing model. Therefore, considerations could be given to incorporating several key factors that typically influence taxi in time. The inclusion of these elements as additional input data may have a potential to enhance the model's accuracy and robustness. Simultaneously, it is also imperative to acknowledge a secondary factor that might significantly influence taxi in time duration and that is stand availability. The real-time availability of parking stands directly affects the course an arriving aircraft takes to reach its designated position, introducing another layer of complexity to the taxiing process. Integrating data on stand availability into the existing model's input parameters may further enhance its capability to generate better output.

Understanding how planes move to parking positions after landing and whether those positions are vacant at the very moment, can be tricky to predict because there are many factors involved. To make things systemetic and conceptual, subsequent research may be built around two additional models: one for predicting taxi in time, and another for predicting parking position availability. As indicated, these two new models may then have the potential to complement the current predictions for δ_{IBT}, and/or δ_{LT} alternatively. As such, this approach allows for a more comprehensive understanding of the entire arrival process called gate-to-gate. The model for taxi in predictions could incorporate the following considerations:

- Typical runway occupancy time:
  Understanding the customary duration an aircraft occupies the runway during its arrival may present a parameter enriching the model. Assiging a typical duration of runway occupancy per aircraft type may shed more light.
- Taxiway layout:
  The design and structure of taxiways are critical as different taxiway layouts

may introduce variations in the route and speed of an aircraft as it navigates toward its parking position.

- Preceding traffic:
  The presence and movement of preceding aircraft on the runway, taxiway, and apron can influence the speed and path of an arriving aircraft.

- Runway in use:
  The specific runway designated for arrivals can impact taxi in times, especially if there are variations in distance and routes from the runway to parking areas. Identifying the runway in use adds a valuable layer of information for more precise predictions.

- Parking position:
  The choice of parking position for an arriving aircraft contributes significantly to the taxi in duration. Different parking positions may involve distinct taxiing paths and distances, influencing the overall time required to reach the designated location.

- Adverse Conditions:
  Integrating data related to weather, especially during instances of reduced visibility, into the model enhances its capability to accurately predict taxi in times by considering the specific challenges posed by adverse weather scenarios.

Considering the large number of factors involved, it may make sense to build a separate model for taxi in predictions. Nevertheless, it's complexity will require mass amount of data from AODB. In this context, further research in this region may build on already existing results of the Taxi-in Taxi-out Predictions (TITOP) project led by Eurocontrol and proposed by Swiss Airlines. The project started in January 2023 and emerged based on the need of users to have a more accurate prediction of taxi time that is necessary between landing and stand for an Arrival flight and between stand and take-off for a Departure flight. Many of the above elements were taken into consideration when computing targets. Additionally, this project also links on another efforts made by Eurocontrol and partners aiming at improving the prediction of the runway in use depending on weather information. Runway in use predictions are also mentioned above as those effecting the taxi in time, and as such it may be beneficial to utilize existing research outcomes too.

On the other side, and when diverting our attention to the development of a second predictive model anticipating the availability of a parking position, a multitude of complex elements need to be regarded too. Key factors include:

- Parking stand occupancy status:
  Determining the current status of parking stands, whether occupied or vacant, is crucial for predicting availability. On top of that what aircraft type is allocated may also enhance the information provided.

- Expected stand occupancy:
  Assessing the anticipated time duration for which aircraft are expected to remain parked provides insights into turnover rates. This may also include availability of ground handling resources.

- Allocation policies:
  Understanding the policies governing the allocation of parking positions, including prioritization rules, influences predictive accuracy as hard and soft business rules may play a crucial role too.

- External factors (e.g. weather):
  Accounting for external factors such as adverse weather conditions or unforeseen disruptions is crucial as they can interrupt usage of particular parking position of entire apron.

As in previous example, majority of input data required will be fed from AODB. From the above mentioned contributing factors, current allocation process and predicted stand occupancy may require highest attention. Here, as in case of taxi in predictions, existing Eurocontrol project called Optimization of Turnaround Times (OpTT) may serve as a good foundation. The OpTT project uses machine learning principles to predict turnaround durations and stand occupancies as such. The aim is to assess the risk of potential ground delay and improve off-block predictability. The methodology was validated at Prague, Geneva and Arlanda airports. The development of new models or enhancement of those existing within the research community with aim to boost predictions of δ_{IBT} is open for discussion, but the clear imperative is to evaluate techniques for effectively merging multiple models ultimately.

While the draft proposal for the next research phase is underway and proposed in this Chapter 5, it is crucial to recognize the current thesis's noteworthy success in achieving quality landing time predictions. Furthermore, the majority of flights, for which the predictions are made pertain long-long to haul traffic. The attained performance stands as a significant accomplishment, opening the door for potential discussions with the industry and airport community regarding the practical application of the predictive model developed. Focus on en-route phase of the flight inluding parameter of velocity as well as flight speed proved to be appropriate and contributed to undertanding of delay increments or decrements produced along the trajectory.

6. Conclusion

The presented thesis reveals details on current status of technical and technological means enabling monitoring of flight progress. The latter becomes a vital element for airport stakeholders when preparing for ground operations. There associated processes with airport resource allocation become a critical task. Timely accomodation of aircraft on stand with appropriate turnaround service is a pre-requisite for on time performance and elimination of reactionarry delay on day of operations. However, to allow for a process of efficient resource allocation, a good visibility on the time as to when to expect acceptance of aircraft on blocks, is a necessity. Typically, when having nominal operations without any extraordinary situation, flights operate on time. However, an early or delayed arrival of an aircraft is a common practice as we operate in high intensity airport and airspace environments. External factors such as technical issues, weather elements, industrial actions or other special events, regularry impact flight operations on ground and en-route. Some of the events mentioned are hapenning ad-hoc and are difficult to predict and such the affect on flight and airport operations is immense. A second category of happenings is formed by events such as weather or operational nuancies such as repeatitve rotational delay, etc.. In fact, there is a number of datalinked to meteorological, flight, or airport operations, which may be used to anticipate departure and arrival delays. On top of that, certain information can be synthetized and used to generate additional variables supporting predictable capability.

Nevertheless, a vast majority of airports in the European network, perhaps excluding a small number of major hubs, struggle to build a capacity enabling anticipation of deviations between scheduled and actual flight performances. They rely on the information available through local Human machine interface (HMI), which is based on manual inputs or automated as aircraft or radar reports kick in. However, the problem is deepened with intercontinental traffic, where a greater portion of the flight trajectory is executed beyond the ECAC region. This provides a very limited flight progress monitoring at airports side, which may result in identification of flight delays at very late stage. This leaves reduced time window for resource allocation changes.

Therefore, this work aims to build a predictable capability supporting airport operations in becoming resilient through anticipation of early and late arrivals. A primary aim is to build a functionality focusing on in-block deviations. However, the results do not manifest a reliable performance in relation to prediction accuracy. This happened despite the efforts made in using appropriate machine learning technique considering the nature of data. It transpires that the available input data, its volume and the nature, seem not to suffice as model struggles to understand interdepandancies having affect on the target variable δ_{IBT}. Although, the main objective to predict the in-block deviations was accomodated, the quality of presented outputs constrains its potential use.

When dealing with the hypothesis of this thesis, the statistical significance of reducing tactical stand allocation changes through a data-driven and machine-learning playbook could not been confirmed. Similarly, it's crucial to acknowledge that the equivalency in the severity of impact on stand availability for delayed and early arrivals by more than 15 minutes could not be validated either. Inability to cope with the verification of these 2 hypothesis is due to the constraints in the model's performance. The latter appears to be a great limitation of the presented work as it impacted accomodation of previously stated goals. Another identified constraint lies in the restricted capacity to conduct feature importance analysis within the neural network framework. Unlike regression models, where the importance of parameters is interpreted through p-value and t-statistics providing the evidence against the null hypothesis, the neural network does not provide intuitive view on feature weights. Having no direct evidence of parameter importance does not allow acceptance or rejection of the last hypothesis, where Airline ID, flight ID, origin airport ID, time of day, wind speed, wind direction are considered as predictors with the highest importance when predicting in-block time deviations. However, the training dataset included wind component information, suggesting that the assumption in the hypothesis considered the impact of velocity on flight duration and eventual arrival time. Having said that, it is presumed that the wind direction and speed may play a role in refining the final prediction.

Ultimately, the limitations of this thesis turned out to be an opportunity to re-direct towards a secondary deliverable of this thesis and that is predictability of landing times. The ability to predict those not only overcame an initial setback linked to poor in-block time predictions but also emerged as a crucial and successful aspect in the methodology. This success is particularly significant when considering the prediction of early or late arrivals, even after adapting the approach to detecting presence of aircraft on runway rather than on stand. The low erorr in predictability of δ_{LT}, reflected in the RMSE of 8.71 minutes, presents a valuable opportunity for airport stakeholders. This accuracy allows them to foresee the arrival of flights, regardless of their profile, whether long or short haul, within the specified on-time performance window of -15 to +15 minutes. Despite the fact, that this deliverable is a clear success, it did not help in hypothesis verification process. Either way, the achieved precision enhances operational planning and responsiveness. As such, this byproduct of the thesis poses additional questions as to how further extend the research in this region so that improvements of in-block time predictions are addressed. As Chapter 5 suggests and to enhance predictions for arrival processes, two new models could be developed: one for taxi-in time and another for parking position availability. These models could complement existing model capabilities, offering better anticipation of (δ_{IBT}). At the same time, the assumption is made that the final model could be applicable to any airport environment, given that input data are available. Furthemore, having such model at disposal may provide airport operators with a flexibility in regards with the utter reliance on CPR, APR and other position related messages fed by ETFMS into airport systems. At the same time, the model outputs cater for applications of TTAs or similar pre-emptive measures to protect airport capacity at times of DC imbalances. As result of it, improved knowledge of arrival time deviations, may improve overall predictability of departuring traffic trough stable TOBT and TTOT declarations. Undeniably, the truth is that results of presented work aiming to optimize operations localy may benefit wider ATM network. This thesis successfully achieves its primary objective of predicting in-block and ultimately also landing time deviations. Despite the inability to conclusively verify certain hypotheses, the demonstrated precision of landing time predictions, provides a reliable basis for evaluating operational impact in resource allocation, should additional information become available. The thesis's outcomes may thus be considered academically significant and operationally consequential.

References

Abdel-Aty, M., Lee, C., Bai, Y., Li, X., & Michalak, M. (2007). Detecting periodic patterns of arrival delay. *Journal of Air Transport Management*, *13*(6), 355–361.
URL https://doi.org/10.1016/j.jairtraman.2007.06.002

Abdelghany, K. F., Shah, S. S., Raina, S., & Abdelghany, A. F. (2004). A model for projecting flight delays during irregular operation conditions. *Journal of Air Transport Management*, *10*(6), 385–394.
URL https://doi.org/10.1016/j.jairtraman.2004.06.008

AhmadBeygi, S., Cohn, A., Guan, Y., & Belobaba, P. (2008). Analysis of the potential for delay propagation in passenger airline networks. *Journal of Air Transport Management*, *14*(5), 221–236.
URL https://doi.org/10.1016/j.jairtraman.2008.04.010

Al-Tabbakh, S. M., Mohamed, H. M., & El, Z. H. (2018). Machine learning techniques for analysis of egyptian flight delay. *International Journal of Data Mining & Knowledge Management Process*, *8*(3), 01–14.
URL https://doi.org/10.5121/ijdkp.2018.8301

Arikan, M. (2011). *The impact of airline flight schedules on flight delays: An analysis of block-time, delay propagation, and schedule optimization using stochastic models*. Ph.D. thesis, Purdue University.

Bacena, A. L. B., Bihasa, A. M. B., Cadayong, L. J. A., Romulo, P. M. A., & Guzman, A. B. D. (2019). An intergenerational investigation of air passengers' emotions during tarmac delay. *Anatolia*, *31*(1), 19–30.
URL https://doi.org/10.1080/13032917.2019.1684960

Barnhart, C. (????). Airline scheduling: Accomplishments, opportunities and challenges. In *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, (pp. 1–1). Springer Berlin Heidelberg.
URL https://doi.org/10.1007/978-3-540-68155-7_1

Chai, J., & Li, A. (2019). Deep learning in natural language processing: A state-of-the-art survey. In *2019 International Conference on Machine Learning and Cybernetics (ICMLC)*. IEEE.
URL http://dx.doi.org/10.1109/ICMLC48188.2019.8949185

Chan, J. Y.-L., Bea, K. T., Leow, S. M. H., Phoong, S. W., & Cheng, W. K. (2022). State of the art: a review of sentiment analysis based on sequential transfer learning. *Artificial Intelligence Review*, *56*(1), 749–780.
URL https://doi.org/10.1007/s10462-022-10183-8

Da Silva, D. G., & De Moura Meneses, A. A. (2023). Comparing Long Short-Term Memory (LSTM) and bidirectional LSTM deep neural networks for power consumption prediction. *Energy Reports*, *10*, 3315–3334.
URL https://doi.org/10.1016/j.egyr.2023.09.175

Deepudev, S., Palanisamy, P., Gopi, V. P., & Nelli, M. K. (2020). A machine learning based approach for prediction of actual landing time of scheduled flights. In *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*, (pp. 755–766). Springer Singapore.
URL https://doi.org/10.1007/978-981-15-7234-0_72

Esmaeilzadeh, E., & Mokhtarimousavi, S. (2020). Machine learning approach for flight departure delay prediction and analysis. *Transportation Research Record: Journal of the Transportation Research Board*, *2674*(8), 145–159.
URL https://doi.org/10.1177/0361198120930014

Eurocontrol (2005). *The Manual: Airport CDM Implementation*. Brussels: Eurocontrol.
URL https://www.viennaairport.com/jart/prj3/va/uploads/data-uploads/CDM/cdm_implementation_manual[1].pdf

Eurocontrol (2023). *ATFCM Operating procedures for flow management position*. Brussels: Eurocontrol, 27.0 ed.

Evans, A., Schafer, A., & Dray, L. (2008). Modelling airline network routing and scheduling under airport capacity constraints. In *The 26th Congress of ICAS and 8th AIAA ATIO*. American Institute of Aeronautics and Astronautics.
URL https://doi.org/10.2514/6.2008-8855

Fricke, H., & Schultz, M. (2009). Delay impacts onto turnaround performance optimal time buffering for minimizing delay propagation.
URL https://api.semanticscholar.org/CorpusID:28942704

Gatsinzi, D., Saez Nieto, F. J., & Madani, I. (2018). Ecac use case of optimised pre-tactical time of arrival adjustments to reduce probability of separation infringements. *IFAC-PapersOnLine*, *51*(9), 186–192. 15th IFAC Symposium

on Control in Transportation Systems CTS 2018.
URL https://www.sciencedirect.com/science/article/pii/S2405896318307547

Gerlinghoff, D., Luo, T., Goh, R. S. M., & Wong, W.-F. (2023). Desire backpropagation: A lightweight training algorithm for multi-layer spiking neural networks based on spike-timing-dependent plasticity. *Neurocomputing*, *560*, 126773.
URL https://doi.org/10.1016/j.neucom.2023.126773

Glover, C. N., & Ball, M. O. (2013). Stochastic optimization models for ground delay program planning with equity–efficiency tradeoffs. *Transportation Research Part C: Emerging Technologies*, *33*, 196–202.
URL https://doi.org/10.1016/j.trc.2011.11.013

Gogtay, N. J., & Thatte, U. M. (2017). Principles of correlation analysis. *Journal of the Association of Physicians of India*, *65*(3), 78–81.

Günther, Y., Werther, B., Temme, M., Böhme, G. D., Niederstrasser, H., Inard, A., Marsden, A., Bonnier, M., & Lane, R. (2006). Total airport management (operational concept and logical architectur).
URL https://www.researchgate.net/publication/224997997_Total_Airport_Management_Operational_Concept_and_Logical_Architectur

Hansen, M. (2002). Micro-level analysis of airport delay externalities using deterministic queuing models: a case study. *Journal of Air Transport Management*, *8*(2), 73–87.
URL https://doi.org/10.1016/s0969-6997(01)00045-x

Hochberg, Y., & Tamhane, A. C. (1987). *Multiple Comparison Procedures*. Wiley Series in Probability & Mathematical Statistics: Applied Probability & Statistics. Nashville, TN: John Wiley & Sons.

Hogg, R. V., & Ledolter, J. (1987). *Engineering Statistics*. London, England: Macmillan.

Holoda, S., Kandera, B., Jancik, M., & Zacik, N. (2019). Digital transformation of ATM - improving EUROCONTROL network manager b2b. In *Proc. of 2019 New Trends in Aviation Development (NTAD)*, (pp. 68 – 72). IEEE.
URL https://doi.org/10.1109/ntad.2019.8875558

IATA (2018). *IATA Airport Handling Manual*. Montreal: International Airline Transport Association, 38 ed.

Jamshidzadeh, Z., Ehteram, M., & Shabanian, H. (2023). A new hybrid model for predicting water quality parameters. *Ain Shams Engineering Journal*, (p. 102510).
URL https://doi.org/10.1016/j.asej.2023.102510

Kim, Y. J., Choi, S., Briceno, S., & Mavris, D. (2016). A deep learning approach to flight delay prediction. In *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. IEEE.
URL https://doi.org/10.1109/dasc.2016.7778092

Klein, A. (2010). Airport delay prediction using weather-impacted traffic index (WITI) model. In *29th Digital Avionics Systems Conference*. IEEE.
URL https://doi.org/10.1109/dasc.2010.5655493

Koolen, H., & Coliban, I. (2020). *Flight Progress Messages Document*. Brussels: Eurocontrol.
URL https://www.eurocontrol.int/sites/default/files/2020-06/flight-progress-msg-update-230620.pdf

Le, Ho, Lee, & Jung (2019). Application of long short-term memory (lstm) neural network for flood forecasting. *Water*, *11*(7), 1387.
URL http://dx.doi.org/10.3390/w11071387

Leeman, E., Popova, T., Phythian, D., & Birenheide, M. (2018). *Airport Network Integration: Concept for establishment of an Airport Operations Plan (AOP)*. Brussels: Eurocontrol, 1.1 ed.
URL https://www.eurocontrol.int/sites/default/files/publication/files/airport-network-integration-apoc-aop-v1.pdf

Liu, Y.-J., & Ma, S. (2008). Flight delay and delay propagation analysis based on bayesian network. In *2008 International Symposium on Knowledge Acquisition and Modeling*. IEEE.
URL https://doi.org/10.1109/kam.2008.70

Matthew, N., & Stones, R. (2007). *Beginning Linux Programming*. Wrox Press, 4 ed.

Mueller, E., & Chatterji, G. (2002). Analysis of aircraft arrival and departure delay characteristics. In *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*. American Institute of Aeronautics and Astronautics.
URL https://doi.org/10.2514/6.2002-5866

Pick, A., & Rawlik, T. (2011). Optimization of airport processes: Support system for human decision making in total airport management. In *Proc. of ASCC 2011 – 8th Asian Control Conference*, (pp. 1054–1059).
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-80052006448&partnerID=40&md5=108f9ac78842a86d1c4a953e5bcef598

Pilon, N., Guichard, L., Bazso, Z., Murgese, G., & Carré, M. (2021). User-driven prioritisation process (udpp) from advanced experimental to pre-operational validation environment. *Journal of Air Transport Management*, *97*, 102124.
URL https://www.sciencedirect.com/science/article/pii/S096969972100106X

Prabakaran, N., Kannadasan, R., Krishnamoorthy, A., & Kakani, V. (2023). A Bidirectional LSTM approach for written script auto evaluation using keywords-based pattern matching. *Natural Language Processing Journal*, *5*, 100033.
URL https://doi.org/10.1016/j.nlp.2023.100033

Rao, A. C., Rao, A. C., & Kulkarni, C. (2022). A survey on sentiment analysis methods, applications, and challenges. *Artificial Intelligence Review*, *55*(7), 5731–5780.
URL https://doi.org/10.1007/s10462-022-10144-1

Reyna, S. S., Mota, M. M., Delahaye, D., & Ortiz, J. (2020). Improvement of APOC operations by using simulation and experimental economics: Conceptual approach. In *Proceedings of the 32nd European Modeling & Simulation Symposium (EMSS 2020)*. CAL-TEK srl.
URL https://doi.org/10.46354/i3m.2020.emss.029

Reynolds-Feighan, A., & Button, K. J. (1999). An Assessment of the Capacity and Congestion levels at European Airports. ERSA conference papers ersa99pa241, European Regional Science Association.
URL https://ideas.repec.org/p/wiw/wiwrsa/ersa99pa241.html

Saxena, A., Goebel, K., Simon, D., & Eklund, N. (2008). Damage propagation modeling for aircraft engine run-to-failure simulation. In *2008 International Conference on Prognostics and Health Management*. IEEE.
URL http://dx.doi.org/10.1109/PHM.2008.4711414

Senthilkumar, N., Karpakam, S., Devi, M. G., Balakumaresan, R., & Dhilipkumar, P. (2022). Speech emotion recognition based on Bi-directional LSTM architecture and deep belief networks. *Materials Today: Proceedings*,

57, 2180–2184.
URL https://doi.org/10.1016/j.matpr.2021.12.246

Sherstinsky, A. (2020). Fundamentals of Recurrent Neural Network (RNN) and Long Short-Term Memory (LSTM) network. *Physica D: Nonlinear Phenomena*, *404*, 132306.
URL https://doi.org/10.1016/j.physd.2019.132306

Spies, G., Piekert, F., Marsden, A., Suikat, R., Meier, C., & Eriksen, P. (2008). Operational concept for an airport operations center to enable total airport management. In *Proc. of 26th Congress of International Council of the Aeronautical Sciences 2008, ICAS 2008*, vol. 2, (pp. 3242–3251).
URL https://www.scopus.com/inward/record.uri?eid=2-s2.0-84878922713&partnerID=40&md5=691a104d0613d28f20434c253675047d

Thiagarajan, B., Srinivasan, L., Sharma, A. V., Sreekanthan, D., & Vijayaraghavan, V. (2017). A machine learning approach for prediction of on-time performance of flights. In *2017 IEEE/AIAA 36th Digital Avionics Systems Conference (DASC)*. IEEE.
URL https://doi.org/10.1109/dasc.2017.8102138

Tu, Y., Ball, M. O., & Jank, W. (2005). Estimating flight departure delay distributions - a statistical approach with long-term trend and short-term pattern. *SSRN Electronic Journal*.
URL https://doi.org/10.2139/ssrn.923628

Venkatesh, V., Arya, A., Agarwal, P., Lakshmi, S., & Balana, S. (2017). Iterative machine and deep learning approach for aviation delay prediction. In *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*. IEEE.
URL https://doi.org/10.1109/upcon.2017.8251111

Walker, C. (2022). *CODA Digest 2021: All-causes delay and cancellations to air transport in Europe*. Brussels: Eurocontrol.
URL https://www.eurocontrol.int/publication/all-causes-delay-and-cancellations-air-transport-europe-2021

Wan, Y., & Roy, S. (2008). A scalable methodology for evaluating and designing coordinated air-traffic flow management strategies under uncertainty. *IEEE Transactions on Intelligent Transportation Systems*, *9*(4), 644–656.
URL https://doi.org/10.1109/tits.2008.2006813

Ye, R., Feng, S., Li, X., Ye, Y., Zhang, B., & Luo, C. (2022). Splnet: A sequence-to-one learning network with time-variant structure for regional wind speed prediction. *Information Sciences*, *609*, 79–99.

Zanin, M., Zhu, Y., Yan, R., Dong, P., Sun, X., & Wandelt, S. (2020). Characterization and prediction of air transport delays in china. *Applied Sciences*, *10*(18), 6165.
URL https://doi.org/10.3390/app10186165

Appendices

A Appendix 1

Špák, M., Olexa, P.

Enhancement of the diversion airport selection methodology.

Transportation Research Procedia. Linz: Elsevier BV, 2020. p. 232-242. ISSN 2352-1465.

Available online at www.sciencedirect.com

ScienceDirect

Transportation Research Procedia 00 (2019) 000–000

Transportation Research Procedia

www.elsevier.com/locate/procedia

9th International Conference on Air Transport – INAIR 2020, CHALLENGES OF AVIATION DEVELOPMENT

Enhancement of the diversion airport selection methodology

Ing. Miroslav Spak, Ing. Peter Olexa

Ing. Miroslav Spak, Konvitska 20, 110 00, Stare Mesto, Prague ,Czech Republic
Ing. Peter Olexa, Prague, Konvitska 20, 110 00, Stare Mesto, Prague, Czech Republic

Abstract

The paper deals with the airlines diversion management as this entails a complex decision-making process that comes into action during the adverse events, preventing the execution of the regular flight operations. Diverting from the intended route can be caused by several reasons both when the safety of the flight or the wellbeing of the passengers is concerned. The most frequent reasons to divert a flight are the repugnant weather, medical emergency or the technical fault. A flight diversion can severely disrupt the passenger's comfort, jeopardize the fleet or the crew planning and is generally accompanied with the significant expenses. The intention of every airline in a case of a diversion is to assure the fastest turnaround time and/or provide the passengers with the highest level of services to mitigate the negative effects.

A wrong selection of an alternate airport, that is not having enough capacity, personnel or material resources to handle a flight diversion in a timely manner can negatively influence the airline in several ways. Both strategic and tactical diversion management process is a collaborative decision-making procedure, heavily dependent on the base of the information available, where the main actors are not only flight crew and the airline OCC but also the ATC, airport operators and the local service providers.

Airline strategic diversion planning shall comprehend a well-defined data set and its data quality procedures shall ensure the information accuracy in the pre-tactical phase. The consequent real time data processing can enhance the decision-making process in a tactical phase of the flight diversion. The data and information required comprise of available airport capacities and other relevant ground resources needed to handle a diverted flight. As stated, such data and information are mostly tactically critical, therefore continuous monitoring and updating shall be ensured. The paper suggests that the overall data management is managed via a local Airport Operations Center (APOC), which will assure provision of actual airport diversion capabilities (incl. weather situation) needed for accommodation of an affected flight.

Keywords: Diversion; Airport Operations Centre (APOC); Index system

* Corresponding author: Ing. Peter Olexa.
Tel.:
E-mail address: olexapet@fd.cvut.cz

Peer-review under responsibility of the scientific committee of the 9th International Conference on Air Transport – INAIR 2020, CHALLENGES OF AVIATION DEVELOPMENT

1. Introduction

The problematics of the aircraft diversions is becoming gradually more significant in line with the growing volume of the air traffic hitting the limits of the current infrastructure. From the publicly available data (BTS, 2020) a portion of the diverted flights in the US is approximately 0.33 %. When this percentage is extrapolated over the annual number of worldwide flights, which is app. 40 million yearly (ICAO, https://www.icao.int/annual-report-2018/Pages/the-world-of-air-transport-in-2018.aspx, 2019), roughly estimated number of 132 000 of flights are affected annually (estimation of the paper authors). Every diversion is considered as an adverse event connected with the negative impacts mainly on the aircraft operator, secondary on the airport stakeholders and as a knock down effect on the other industry members linked with the air traffic. A delay caused by a flight diverting from its intended route can be directly expensed and expressed monetarily in the additional costs decreasing the rentability of flight operation. These expenses and the following causal consequence of impacts is gradually deepened as the delay is prolonged. The events triggering an operative flight diversion (Ryerson et.al., 2015) are caused by force major and although significant risk reducing actions can be achieved by careful foreseeing and planning, the main mitigation efforts of the industry must be aimed on the most effective operational handling of diverted flights.

In the airline business, the so called IROPS - Irregular Operations, especially the diversion handling and consequent recovery from the point of view of the aircraft operator must be executed tactically based on the current condition of the diversion airports. (Myeonghyeon Kim, YuriChoi, Ki HanSong, 2017) The operators are preselecting the diversion locations for all phases of flights, after the departure, enroute and as the destination alternates based on the information and data available with the aim of the least disruption effect in a diversion occasion. Thus, a so-called “diversion risk index” (see paragraph: *4.1.1 Airport diversion risk index)* for each location can be extrapolated. Currently this index is based only on static data collected in the pre-tactical phase of the flight, often even in strategic one. However, it is typical the aircraft will already be airborne when the decision will be taken to divert, therefore it is inevitable to aim our focus into the tactical time horizon. Here the data variables, such as actual airport diversion capacity, ground handler availability of personnel and equipment, fueler capacities and others are floating in time and have a dramatic influence on the factual TARMAC delay, especially in events, when several flights are diverted to same location. Therefore, a benefit has been identified in sharing these data variables in a real time.

An important role here lies in the framework of a local coordination arrangement, developed within European SESAR program, called APOC (Airport Operations Centre) and its data exchange with the airspace users using direct or indirect communication means. Within the APOC given operational stakeholders (actors) collaborate for the effective/efficient establishment and execution of an agreed Airport Operations Plan (AOP), in a structured manner with agreed processes, either through physical or virtual interaction or their combination. In the European theatre, the APOC is the prime interface between the Airport and the Network Manager Operations Center (NMOC) established in the States within, and adjacent to the ECAC area. (EUROCONTROL, 2018)

It is vital that tactical changes to an actual airport diversion capacity are provided in order to safeguard dynamic adjustment of the diversion risk index values. The research paper aims to address a necessity to produce a complex system able to gather the airport data through the APOC, and subsequently evaluate them using a standard taxonomy. As a result of this, airlines may produce real-time decisions and select the most suitable alternate airport based on its calculated diversion risk index value, which is a proposed variable elaborated in following paragraphs. The concept then leads to a significant reduction of the flight diversion costs and mitigation of other negative effects of the diversion on the network management.

2. Methodology

To justify a need for change in the selection process of alternate airports a qualitative investigation of current techniques is to be performed. The focus will be aimed on the existing means of securing a real-time data exchange. The investigation of this area of the analysis shall unveil current limitations hampering effective and efficient diversion airport selection methodology. The limitations shall then be subject to examination to identify both immediate and future improvements leading to operationally and financially feasible flight diversion management.

3. Current diversion airport selection methodology

There is no given methodology for the diversion airport selection that would be determining and mandating to use the alternate airport as per given set of hard rules. Therefore, for every flight a flight planner can select an alternate airport as per the current requirements of dedicated flight or preferences of the airline operator. Especially for small airlines, only basic characteristics of the airport are examined such as the opening hours, current NOTAMs and physical characteristics. As a best practice, various factors playing a role in the diversion location selection are evaluated and can be divided into two groups:

Basic factors (determining if an aircraft can land on an airport):

- Airport operational, physical and technical characteristics to accept the aircraft type at given time (RWY length, equipment, PCN, opening hours, curfews, NOTAMs, etc.)
- Weather conditions (Marco-Michael Temme, Charlotte Tienes, 2018)
- Distance from the intended route and destination (fuel planning)
- Other (pilot licenses, etc.)

Commercial factors (influencing the diversion recovery):

- Relative proximity of the diversion location (both along the planned flight route and from ADEP/ADES) - as a passenger wellbeing factor (Ryerson, 2018)
- Level of airport infrastructure (airport size)
- Airline own infrastructure (own or contracted airline staff present on the location)
- Existence of agreements with the handlers, fueler and other stakeholders
- Experiences with the given location
- Other commercially important and detailed factors (hotels availability, transportation possibilities, etc.)

Based on the evaluation of the above, the flight planner decides about the alternate locations, files them in the flight plan and provides to the crew at the time of the pre-flight briefing. At least two scenarios consequent to the diversion landing shall be assessed:

- **Technical stop** - the aircraft is refueled and dispatched as soon as possible, the passengers are often remaining on board
- **Full termination of the flight** – when an aircraft is forced to stay on the diversion airport for an extended period time

3.1 Diversion event flow of actions

The most common cause for diversion can be divided into:

- Safety and security reasons:
  - Not assessed, as such diversion has an utmost priority and economic factors are negligible
- Flight internal factors:
  - Medical - airports that enable the medical assistance are selected
  - Technical - minor technical fault
- External factors of a flight:
  - Capacity constraints (see paragraph *3.1.1 Trigger events*) at the destination airport. Under such circumstances the air crew will typically get information from the ATC that the approach cannot be completed at the destination airport and therefore the arriving aircraft will be directed into a holding pattern. However, it is very likely that the air crew will contact their OCC to consult options.

It may happen that holding will take a significant amount of time and that can negatively influence the level of remaining fuel on board. Furthermore, late arrival at the destination airport will most likely result in late departure, which may potentially fall into the night curfew period, and that can considerably disrupt the flight continuation. Having said that, after a thorough situation assessment, the air crew under consent of their OCC[1] (if available) may decide to divert to an alternate airport for landing. A diversion destination is then usually one of predefined locations that have been filled in the flight plan.

3.1.1 Trigger events

Capacity constraints:

- ○

[1] Operations Control Centre

- Weather issues
- Ground handling issues
- Industrial action
- System failure (airport, ANSP)
- Night curfew issues
- Others

3.2 Data collection

In order to facilitate air crews to take better business decisions in case of diversion circumstances, there is a provision in place to inform airlines.

3.2.1 Pre-flight horizon

In the European network, there is a EUROCONTROL platform called Airport Corner, where airports share information typically related to capacity topics. One element of the Airport Corner is dedicated to diversions and provides information about handling capabilities for various aircraft types, night curfews, etc. The Airport Corner is also made available to the airlines. As an alternate, the private databases of the airline operators with their subjectively defined set of basic and commercial factors (see paragraph 3.) are used and flight planners bear the decision-making responsibility upon the most suitable diversion location for a dedicated flight.

3.2.2 Real-time horizon

Currently in the European framework, there is a trial running, when NMOC contacts airports in the network with the request to update their diversion-related information provided in the Airport Corner. This occurs, when it is anticipated that multiple diversion can take place. Although the trials are successful, airports are often found unable to share their available diversion capabilities needed to ensure commercial handling of the affected traffic with a minimum TARMAC time. In general, there is no consolidated approach in place for collection of real-time data reflecting the airport's capability to handle any diverted flight, and so the airlines are typically left to make their final judgement about an alternate airport based on the data available prior to the flight plan activation.

There are publicly available airport delay indexes of the flight trackers such as the one of FlighRadar24 platform (FlightRadar24, 2020) whereas the index is extrapolated from the difference of the scheduled and actual times of the operations based on the data gathered by this platform. Similarly, the EUROCONTROL NMOC is providing the real time information of the airports generating the highest delays.

3.3 Problem definition

The boundaries of the current diversion airport selection methodology lay in the fact that operational situation at the alternate airport and its vicinity may change during the flight and these real-time updates will not be instantly reflected in the list of predefined alternate airports contained in the flight plan. Instances may vary; however, one common is when multiple diversions are directed to the same alternate airport of which the ground capacity is quickly consumed. This transpires that TARMAC delays will be excessively prolonged as ground resources are not available to be allocated for the affected flight. To enable reaction to these dynamic changes and select the most suitable alternate airport, more real-time diversion-related data must be gathered and distributed to the airspace users at any moment of the flight, if required so.

4. Proposed diversion airport selection methodology

The biggest challenge remains in having the real-time airport diversion capacity data available. To achieve that, it is proposed that the data collection is performed through a local APOC establishment, where centralization of airport services takes place and where operational data and information is concentrated. (1) As a second step to overall success, it is necessary to provide this diversion-related data to the airspace users who may then effectively and real-time select a diversion airport in case such a need occurs. The data distribution and its overall flow may be done either directly to a given airspace user or indirectly via the NMOC.

4.1 Flow of actions

To achieve information sharing related to airport diversion capabilities, air to ground communication channels must be used. The process starts at the airport, where appropriate data is collected and airport diversion risk index (see below) is generated. Subsequently, the latter is distributed either directly or indirectly to a given air crew. The means through which this is done, very much depends on the communication infrastructure available. This paper tries to look at both currently available and foreseen information management. Either way, there is no difference anticipated in the nature of data needed to be collected to serve in determination of the airport diversion risk index, now and in the future.

APOC as a platform uniting all the airport's key players involved in managing passenger, aircraft and baggage processes, will remain the very foundation of clear and uniform view on all ground processes and available capacities. (International Airport Review, 2018) Future diversion airport selection process is to be based on real-time airport's diversion risk index management and updates provision. All alternate airports included in the flight plan will be selected based on the preflight airport diversion risk index as a product of the APOC arrangement, and on the aforementioned basis and commercial factors that airlines consider too. However, a decisive element here remains the value of the airport diversion risk index, and should the latter be changed after the aircraft is reported airborne, updates must be provided, so that a right decision when choosing an alternate airport can be made. Furthermore, for the purpose of this paper, no changes to future diversion trigger events are anticipated.

4.1.1 Airport diversion risk index

As stated above, the airport diversion risk index generated by APOC, is a new proposed variable that shall be shared to the actors based on the real time or frequently updated data. There are many scenarios that determine how long an airplane will be delayed in an event of a diversion and many elements that do play a role in the delay estimation. The most important key triggers factors contributing to the delays that need to be looked at, are mostly associated with ground resources availability.

As their evaluation per-partes in a diversion event might be complicated, it is beneficial to assign every data item with a specific severity index based on the possible impact it might have on a delay and multiply them with the current availability index. The sum of all diversion risk elements shall be resulting in the airport diversion risk index.

<u>Airport diversion risk index calculation formula:</u>

Item 1 (severity index) x Item 1 (Current availability index) +
Item 2 (severity index) x Item 2 (Current availability index) +
= Airport diversion risk index

Proposal of minimum list of ground resources required for a diversion risk index calculation:

Table 1: Diversion risk elements

Diversion risk elements		**Details**	**Delay Impact Severity Index**	**Availability index**
Item 1	ATC Capacity	RWY capacity, adjacent airspace capacity	X	Y
Item 2	Available pier / remote stands	To be assessed per the aircraft dimensions and in regard to the air bridges availability	X	Y
Item 3	Terminal capacity	Shall be discussed if the deplaning of passengers is likely	X	Y
Item 4	Ground handler GSE and personnel availability	Complex variable that shall be discussed in detail per partial services	X	Y
Item 5	Fueling service and equipment availability	Availability of tankers and/or hydro trucks and staff that can be allocated in a given timeframe	X	Y
Item 6	Deicing availability	Can be considered based on the deicing waiting times	X	Y
Item 7	Other factors	Caterer, Customs and Immigration, Airport ambulance, etc.	X	Y

Proposal of current availability index (incl. TARMAC delay code) and delay impact severity index:

Table 2: Current Availability Index and associated delay risk

Current Availability Index		**TARMAC delay**
Available	1	None
Limited	2	Moderate
Unavailable	3	Long

Table 3: Delay impact and Severity index

Delay Impact	**Severity Index**
Severe impact	3
Intermediate	2
Small influence	1

Use case (example):

Table 4: Airport Diversion Risk Index

Diversion Risk Element	Delay Impact Severity Index	Current Availability Index per airport		
		London Heathrow (LHR)	London Gatwick (LGW)	London Luton (LTN)
Item 1	2	1	1	1
Item 2	3	2	3	3
Item 3	3	1	3	2
Item 4	1	2	3	2
Item 5	3	3	2	1
Item 6	3	1	1	2
Item 7	1	2	3	1
Item 8	1	1	1	2
Airport Diversion Risk Index		**28**	**36**	**31**

A resulting number shall be called the airport diversion risk index and scaled based on the overall result of all the airports in the network or based on the predefined measure scale. The airport with the lowest diversion index has a potential to recover the diverted aircraft with the shortest delay (LHR), whilst the locations where the risk is high are to be considered as those with the highest potential to generate the delay (LGW).

4.2. Data collection

As seen above, to support and to enhance the diversion airport selection process using the airport's diversion risk index, the attention must be placed on the airport's ability to collect right data in an effective manner. The data elements (Table XY) to be collected which ultimately determine the diversion risk index are continuously recorded and updated, if any changes occur. Looking at the nature of these elements, it is obvious that collaboration in the data collection entails inputs from a broad variety of airport stakeholders.

Table 5: Data providers list

Airport stakeholder	Items from the Diversion risk index assessment
ANSP	Item 1
Airport Operator	Item 2 – 4
Ground Handler	Item 5
Fueler	Item 6
De-icing Company	Item 7
Other actors	Item 8

4.2.1. Real-time horizon

Despite the EUROCONTROL efforts to enhance the scope of the Airport Corner in relation to data collection associated with airports' diversion capabilities, there is a rising need to assure more flexible approach catering for reflection of dynamic changes linked to the airport diversion capabilities. These capabilities have to be determined, monitored and managed through the APOC arrangements. (EUROCONTROL, 2018)

The airport diversion capabilities shall ideally be determined a day before the day of operations and form part of the Airport Operations Plan (AOP). As the operational hours draw closer and flight plans are getting activated, it is then a role of the APOC team to provide updates to previously determined airport diversion risk indexes, if required so. The airport diversion capabilities always have to be as accurate as feasibly possible for the benefit of the entire network. The APOC HMI (Human-Machine interface) shall enable the airport stakeholders to provide real-time information inputs on the resources available known as the diversion risk elements (Table 4).

4.2.2. Service Level Agreement (SLA)

Levels of service shall be jointly agreed between all airports' stakeholders. It is expected that the SLA shall be established to ensure an approach of openness, transparency and collaboration. (IATA, 2019) The stakeholders shall in the first instance jointly agree to the framework of the diversion risk index assessment elements which are to be measured, provided, monitored and managed through the collaborative approach within the APOC. Secondly, this document shall provide policy guidance on how these elements representing a broad variety of airport ground resources, are to be made available in case a diversion takes place.

The agreement shall always grant transparent provision of resource availability without discrimination. Meaning, there shall not be a space for prejudice in business and all unutilized resources indicated in the Table 4 shall be made available to handle a flight of any airline without other business bonds considerations.

4.3. Data distribution

Once the data has been collected and the airport diversion risk index generated, it is necessary to publish it and make available for further use. This is the moment when ground-ground and air-ground communication channels shall be used to complete the flow of actions. As stated previously, the paper depicts both immediate and future prospective applications for the information exchange. For both instances, now and then, it is the System Wide Information Management (SWIM) (ICAO, ICAO-SWIM, 2020) that will play a major role.

4.3.1. Immediate application in EU network - APOC to Network Manager (NM)

As SWIM enables seamless information access and interchange between all providers and users of ATM information and services, it is the right candidate to cater for real-time airport diversion risk index dissemination. Currently, it is only the yellow and blue concept infrastructure profiles (SESAR, SESAR Joint Undertaking, 2020), which may be of use for real-time information exchange in the context of dynamic evolution of the airport diversion capabilities.

In principle, the SWIM yellow and blue profiles are limited to ground-ground communication and therefore a direct link between the ground and the air crew would not be possible using the latter. Immediate distribution of the airport diversion risk index would today only be feasible using Very High Frequency (VHF), or alternatively a two-way data link system called Controller Pilot Data Link Communications (CPDLC) by which controllers can transmit non urgent strategic messages to an aircraft. However, there is a full range of downsides for using VHF or CPDLC to communicate vital and real-time operational messages linked with the airport diversion capabilities.

Therefore, it is proposed that the operational information containing the airport diversion risk index value is sent out by the APOC team using EUROCONTROL Network Manager Business t0o Business (B2B) Services that are here and available to provide real-time situational awareness and to support Collaborative Decision Making (CDM) processes. NM B2B Services form part of the SWIM concept and their main advantage lies in ability to provide conditioned subscription to a topic or selected data related to flight, airspace or flow services. (EUROCONTROL, Network Manager B2B, 2020)

It is therefore proposed that the airport diversion risk index becomes an ATM information element belonging to a group of flight related services, and is then made available to subscribed airlines' Operations Control Centre (OCC), or to anyone else who may consider this information element vital for the operations. OCC retrieves the pertinent

ATM information element and communicates the latter to the respective air crew.

The idea of providing real-time changes to data contained in the flight plan further also supports the concept of Trajectory Based Operations (TBO), (Air Traffic Management, 2020) which aims to enable the Air Traffic Management (ATM) systems to know and, where appropriate, modify the flight's planned and actual trajectory, before or during flight, based on accurate information that has been shared by all stakeholders. In this context, last minute changes to the selection process of an appropriate alternate airport as per the actual airport diversion risk index may be instantly reflected in projection of a new flight trajectory, thus improving the overall ATM. https://www.sesarju.eu/sesar-solutions/trajectory-based-operations

4.3.2. Future application on global level - ATC to Cockpit

In context with SWIM and its use for direct air-ground communication exchange, the purple infrastructure profile constitutes a right candidate for application. As stated, SWIM allows the distribution of information through the aeronautical telecommunications network/Internet protocol suite (ATN/IPS) in place of legacy point-to-point contracted services.

And here comes the SWIM purple profile, which aims to replace legacy services – such as those in place to support CPDLC or Automatic Dependent Surveillance - Contract (ADS-C). (SESAR, SWIM PURPLE PROFILE, 2020)

Research and specification documents are being currently drafted under the European SESAR research program aiming to define the functional and non-functional requirements of the air-ground service infrastructure. Having said that, once this work has been concluded, it is proposed that the airport diversion risk index is communicated directly between the Air Traffic Controller (ATC) and the air crew through the datalink based on the SWIM purple profile functional requirements. The ATC is expected to retrieve the airport diversion risk index value through the APOC interface and communicate it to the air crew, if required so.

An alternate solution to ensure that the airport diversion risk index is directly communicated to the air crew consists in deployment of the concept called Data link - Operational Terminal Information Service (D-OTIS). The D-OTIS service is to provide automated assistance in requesting and delivering compiled meteorological and operational flight information derived from ATIS[2], METAR[3] and NOTAMs / SNOWTAMs[4], specifically relevant to the departure, approach and landing flight phases. (WG-78, RTCA SC-214 / EUROCAE, 2013) In principle, the D-OTIS service encompasses three different sub-services, namely ATIS, METAR and OFIS[5]. ATIS and METAR contain standardized meteorological information and data which are made available to all Air Traffic Services Units (ATSU) and further to air crews by requesting and receiving voice transmissions from ATC. Some airline operators might also provide access to METAR reports through data link service.

Either way, the paper investigates possibilities of transmitting real-time operational information to aircrews on request. For this purpose, the OFIS element of D-OTIS is being developed and is intended to issue information promptly whenever direct operational significance to the approach and landing phases has been detected.

To fulfill the objective of the submitted paper, the OFIS shall not be only limited to data such as approach minima, reduced runway length, runway contamination, or NAVAIDs operating status (information derived from NOTAM/SNOWTAM). The extended scope of OFIS shall also contain information on the airport diversion capabilities. The former shall be broadcasted to air crews through the ATC using a data link.

In case of data link failure, the ATM information linked to the airport diversion capabilities shall be an element of conventional ATIS and provided to aircrews through ATC using customary VHF broadcasting method - voice transmission. ATC is to feed D-OTIS, alternatively conventional ATIS with the information on the airport diversion capabilities retrieved from the APOC platform. (EUROCONTROL, 2018)

5. Results and discussion

[2] Automatic Terminal Information Service (ATIS)
[3] Meteorological Terminal Air Report (METAR)
[4] Notice to Airmen (NOTAM)
[5] Operational Flight Information Service (OFIS)

The proposal to collect, process and share airport data relevant to the diversion has direct and immediate benefit to the air traffic stakeholders even when shared in the currently available environment. The willingness to enforce the real time diversion capability sharing has been already recognized in the European ATM network, however a clear framework has not yet been proposed. This article sets a proposal of the basic airport data elements to be monitored, collected and distributed throughout the network and suggests the assessment methodology algorithm resulting in the airport diversion risk index.

6. Conclusion

Currently there is no real time data available, that would indicate to the aircraft operators if a dedicated airport - selected as the one to be diverted to, will be able to comfortably handle the diverted flight without severe delays caused by the airport capacity or airport services unavailability. There is also no, either nationally, regionally or globally driven methodology relying on real time data used to determine diversion strategies for the airlines. The diversion destination selection is relatively flexible based on the operator needs, provided the ATC capacity and safety margins are met.

To estimate the number of impacted flights that could potentially benefit the enhanced diversion handling the publicly available FAA (BTS, 2020) and sourced EUROCONTROL data have been used. As an example, in July 2019 (selected as a statistically busiest month), the number of flights routed through the European network was 1 093 734. Out of this 0.507 %, have been diverted from their intended route, which is 5548 flights in total. Looking overseas, in the same month in the US, 2517 flights have been diverted out of 717 684, representing 0,3 % of flights. The analysis of the single airline operator (Delta Air Lines) revealed the 2119 diverted flights in 2019 to or from the US airport which is 177 diverts per month in average. Most diversion in July (358) and June (338) and least in November (69) and March (77). Daily average number of the diverts is- 6 (6,42), ranging from 0 diverts a day to maximum 58, concerning only a single aircraft operator. An estimated cost of a diversion in 2014 was ranging from 5 000 USD to 100 000 USD (Johnson, 2014). When multiplied over the global number of diverted flights, even a small contribution to the reduction of aircraft delays can represent significant cost reductions.

Shall the direct delay impact of a diversion be assessed; the airline operator could benefit from smaller delay times on ground and their direct and mainly indirect diversion cost could be significantly reduced. As the diversion events and their effect have the domino effect, the list of direct benefiters from the operatively driven airport diversion selection can be widely enlarged. The clients of the aircraft operators can experience less disturbance shall their diverted flight be dispatched or handled expeditiously. The distribution of movements throughout the ATC network could be used more evenly and even the more distant airports that are not frequently diverted to but could potentially offer a quick turnaround due to a lower diversion risk index could increase their annual income.

References

Air Traffic Management. (2020). *Air Traffic Management*. Retrieved from Air Traffic Management: https://airtrafficmanagement.keypublishing.com/2020/01/30/trajectory-based-operations-are-one-step-closer/

BTS. (2020). *Burreau of Transportation Statistics*. Retrieved from BTS: https://www.bts.gov/

EUROCONTROL. (2018). Retrieved from https://www.eurocontrol.int/sites/default/files/publication/files/airport-network-integration-apoc-aop-v1.pdf

EUROCONTROL. (2020). *Network Manager B2B*. Retrieved from https://www.eurocontrol.int/service/network-manager-business-business-b2b-web-services

FlightRadar24. (2020). Retrieved from https://www.flightradar24.com/data/airports

IATA. (2019). *IATA*. Retrieved from IATA: https://www.iata.org/contentassets/4eae6e82b7b948b58370eb6413bd8d88/airport-service-level-agreement.pdf

ICAO. (2019). *https://www.icao.int/annual-report-2018/Pages/the-world-of-air-transport-in-2018.aspx*. Retrieved from https://www.icao.int/annual-report-2018/Pages/the-world-of-air-transport-in-2018.aspx

ICAO. (2020). *ICAO-SWIM*. Retrieved from https://www.icao.int/APAC/Pages/swim.aspx

International Airport Review. (2018, June 19). Retrieved from https://www.internationalairportreview.com/article/79140/airport-management-apoc/

Johnson, B. (2014). Diversion Management. *Airline Cost Conference.* Geneva: IATA.

Marco-Michael Temme, Charlotte Tienes. (2018). Factors for Pilot's Decision Making Process to Avoid. *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*.

Myeonghyeon Kim, YuriChoi, Ki HanSong. (2017). Identification model development for proactive response on irregular operations (IROPs). *Journal of Air Transport Management*.

Ryerson et.al. (2015). Landing on empty, estimating the benefits from reducing fuel uplift in US Civil aviation. *Environ. Res. Lett.,, 10*(9).

Ryerson, M. S. (2018, March). Diversion Ahead: Modeling the Factors Driving Diversion Airport Choice. *Journal of Infrastructure Systems, 24*(1). Retrieved from https://ascelibrary.org/doi/abs/10.1061/%28ASCE%29IS.1943-555X.0000407

SESAR. (2020). *SESAR Joint Undertaking*. Retrieved from https://www.sesarju.eu/sites/default/files/SESAR-Factsheet-2015-SWIM-Profiles.pdf

SESAR. (2020). *SWIM PURPLE PROFILE*. Retrieved from https://www.eurocontrol.int/service/network-manager-business-business-b2b-web-services

WG-78, RTCA SC-214 / EUROCAE. (2013). Advanced Data Communication. *Advanced Data Communication - Baseline 2.* Melbourne: ICAO. Retrieved from https://www.icao.int/APAC/Meetings/2013_FIT_Asia2_RASMAG18/WG78-SC214%20Baseline%202%20Tutorial%20-%20Final%20(25March2013).pdf

B Appendix 2

Olexa, P., Špák, M., Stojić, S., Lán, S., Hamza, M.

Static Validation of the Enhanced Diversion Airport Selection Methodology.

2022 New Trends in Civil Aviation (NTCA). Praha: České vysoké učení technické v Praze, 2022. p. 141-146. ISBN 978-80-01-06985-1. ISSN 2694-7854.

Static Validation of the Enhanced Diversion Airport Selection Methodology

Peter Olexa
Department of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
olexapet@fd.cvut.cz

Miroslav Špák
Department of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
spakmir1@fd.cvut.cz

Slobodan Stojić
Department of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
stojislo@fd.cvut.cz

Sébastien Lán
Department of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
lansebas@fd.cvut.cz

Marek Hamza
Department of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
hamzamar@cvut.cz

***Abstract*—This paper describes the elaboration of Enhanced diversion airport selection methodology, which can be used for the optimization of flight planning by airline operators and the better diversion decisions in the real time operations. The paper is based on the analysis of the problematics of flight diversion from different perspectives and determining the factors influencing the diversion recovery. Based on these factors an methodology is proposed for decision making in the process of alternate airports selection during the planning process and real time operations. The selection methodology incorporates operational and commercial factors determining the suitability of an airport as an alternate destination. Emphasis is put on the economic suitability of the airport in terms of minimizing costs in case of flight diversion. In the validation process the methodology was applied on selected airports using real data and compared with actual operations, and as such presented to the experts. The result of the work is an optimized system for selecting an alternate airport in case of diversion from the destination, with the possibility of wider application and further elaboration when using the real-time shared data.**

***Index Terms*—airport data, diversion risk index, diversion risk factor, flight diversion**

I. Introduction

Aircraft diverts and the management of diverted flights remains an underestimated problem mainly due to the lower frequency of such occurrences in comparison to the total number of all flights. In accordance with the US Bureau of Transportation Statics, Diversions of US air carriers 1990-2017 averages on 0.0022 % of flights (out of over 5.67 million operations, 12 530 of such flights were diverted.) [1]. The associated costs are however significant and seeking the means of minimizing the adverse effect offers room for improvement in the efficiency of commercial aircraft operations. In the cases, when an aircraft is forced to divert from its intended route and the flight destination cannot be reached in accordance with the planned schedule, the adverse effects are not only disrupting the airline itself but also aircraft traffic and the associated services at the alternate airport [2]. The selection of an appropriate airport to divert to, plays a crucial role in the process of diversion recovery and the adverse effect might be significantly reduced. There is a proposal of the Enhanced diversion airport selection methodology based on indexing the alternate airports in accordance to their Diversion Risk Factor, driven by airports scoring based on the actual real-time availability of the given factors (Ground service equipment availability, Fueler availability, Apron Stands availability, etc.) and their relevance to the aircraft operator [2]. As this airport selection methodology proposal is expecting the usage of the real time data, its validation is limited due to the current non-existence of the network and the data flow that would enable its validation.

It is important to differentiate emergency landing from flight diversion. While the classification of every specific diversion occasion is up to the operating airline, some general characteristics of each event can be determined. Emergency landing is considered to be a manoeuvre following a severe safety-threatening event preventing the flight from continuing (i.e. double engine failure, structural damage to the aircraft, pilot incapacitation etc.) [3]. From the Air Traffic Control (ATC) point of view, an emergency landing is when the flight crew declare a state of emergency (either via voice “Mayday” call, Secondary Surveillance Radar (SSR) transponder code 7700 or Controller-Pilot Data Link Communications (CPDLC) message DM 56) and this is not cancelled by the time the aircraft touches down. A flight diversion, on the other hand, is performed as a precaution of possible further threats but with no presence of a direct danger. such case, most likely to be misinterpreted as an emergency landing, should be a single engine failure, one system failure or similar abnormality. A flight diversion also ends on an airfield while an emergency landing might be terminated by a forced landing in the terrain or on water, respectively [3].

Generally, there are two main reasons leading the crew decision to alter from their intended flight destination. The first of them occurs when there are factors degrading the airports characteristics, making the flight destination unsuitable for landing. This can be caused either by the airport closure, runway contamination, weather conditions or due to the decision made by the operating company/state of registration of the operator. All these reasons can be named ground located causes. In the second case, the aeroplane becomes unable to reach its destination due to in-flight cause (e.g., change of aircraft technical state) [4], [5].

Alternate airports are always planned for every flight so that pilots know where they can divert to in every stage of flight even before getting on board. In airlines and major aircraft operator companies flight plans are created by Dispatch departments or Operations Control Centres (OCC). Weather situation and Notices to Airmen (NOTAMs) at destination and at the alternate airports as well as significant weather en-route are considered. In either case, when there is a need to divert, there is always a suitable airport included in the flight plan providing all the required properties guaranteeing safe diversion despite the reason of diverting. Even though adhering to the diversion plan is not obligatory and flight crew have the right to divert to an airport they consider most suitable, choosing a non-planned alternate is very unusual.

This paper is summarizing the results of further elaboration of the Enhanced diversion airport selection methodology, as part of early research done on this subject by Olexa and Spak [6]. Pre-validation was based on the static data usage and detailed and focused research performed in the aircraft operators and airlines environment.

II. Methods

A. Expert Assessment

The Enhanced Diversion Airport Selection Methodology has been assessed with the expert's evaluation. The authors of the paper have presented the methodology and problem statement to major airlines (Delta Air Lines, United Airlines, Smartwings, Vueling) – their operational and dispatch experts and pilots, airport authorities and the ATC representatives – all former or active air traffic controllers (on the platform of EUROCONTROL focus group meeting) to determine the common diversion problematics denominators.

- The methodology itself has been presented, discussed, and evaluated.
- The diversion factors have been presented, discussed, and evaluated.
- The limitations of the methodology have been discussed.

The full methodology assessment and testing in a real time is momentarily limited by the nonexistence of the data sources that would be usable for the direct computation of the risk indexed and associated measurements of the delays. Therefore, the methodology was first evaluated by the experts and the testing platform is proposed in the following chapters of this article.

Similarly for demonstration purposes, the methodology has been applied with the usage of the static data on the series of airport both for the en-route diversion scenario (occurrence of the flight diversion in the en-route phase) and the scenario, when the aircraft is forced to divert when close to its original destination. State of Florida (USA) was chosen for this purpose as a polycentric region with several international airports located relatively close to each other. Unlike in other regions with fewer airports suitable for diversions, the density of possible alternate airports in this state reduces the advantage of proximity to the destination, which is usually one of the most considered variables. Therefore, other criteria must be considered to select the most suitable alternate airport. Several diverting flights of Delta Air Lines were chosen and analyzed using the suggested method and then compared to the actual course of the diversion. Delta was chosen because it is one of the biggest operators in the North American region with statistically relatively high chance of flight diversion. It also operates several flights to various airports in Florida and there is an active relation between the airline and the US Bureau of Transportation Statistics providing access traffic information.

B. Problem Scope Definition

The main problem of selecting the alternate airport lies in:

- Considering the most suitable location for the flight to divert to.
- Monitoring the dynamic changes of the airport conditions and the fluctuation of the availability of the airport resources [6].

In the model situation, we will be referring to within the scope of this article, there is an external cause of diversion, adverse weather situation on the destination airport, posing a hazard for the incoming air traffic and forcing the aircraft to divert for a finite time.

On the Fig 1, A model situation is depicted, with the inbound traffic for the original destination airport (Dap). Where there are the adverse weather present. The patter of incoming traffic is dispersed in the diversion locations – alternates (A1, A2 and A3). Such situation can be named as a massive diversion and depicts one of the most complex air traffic situations.

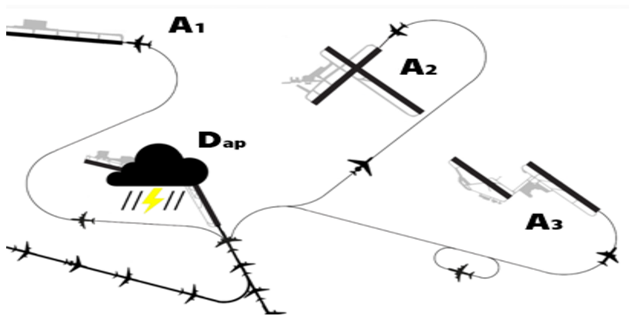


Fig. 1. Destination airport (Dap) is not available due to adverse weather conditions, with traffic being diverted to alternates A1, A2 and A3.

C. Problem Definition

Problem can be better defined as shown as described in the model situation (see Fig. 1). The external diversion factors such as adverse weather conditions are influencing an airport with the high volume of the inbound traffic are creating situations, where a higher demand is put on the nearly located airport that are often filed as the alternate airports. in the strive for the resource's optimisation, the airports are not keeping sufficient buffer of the resources (manpower, stand and equipment capacity) to accommodate higher volumes of the non-plannable traffic (diverted flights) and hence the free capacity can be quickly occupied. Any incoming flight above the available airport capacity (requiring the resources that are occupied at the moment) will suffer from the delay with obtaining the airports services (parking spot, ground handling, fuel tanking and eventually also the passenger related services) [5]. An optimisation can be sought of diverting the aircraft in-flight in a need to divert to a location, that is having an available capacity to accommodate them. The study is focusing on commercial passenger air traffic, as it holds the biggest share in the total number of flights [7].

D. Listing the Diversion Factors and Assesment Methodology

The paper authors have cross examined the factors influencing a so called "diverted flight resolution". In a diversion situation, there are generally two main scenarios to be assessed – a "fuel and go" when an aircraft can proceed either further to the destination airport (providing the reason for the flight continuation has been resolved) or elsewhere (based on the requirements of the aircraft operator) or the "full termination of the flight" – which is requiring the passengers to be disembarked and the aircraft needing a prolonged period of parking on ground. In the defined problem scope, the most frequent resolution scenario would be the "fuel and go" [8]–[10].

There were however common factors nominated for both scenarios as an inbound aircraft is utilizing the same airport resources. Out of which the initial set of the static data used for the methodology assessment has been pre-selected and evaluated as shown on the Table I.

The Diversion Risk Formula (1) consists of two elements. The first one is Severity s and that evaluates how important of chosen factor (or vendor such as ground handling, fueller etc.) is important for the operator. The second element is Availability a of chosen factor or vendor. The basic distinctiveness are three levels of state for each element. For Severity the states of importance are low, medium and high. For Availability the states are available, limited and unavailable. The factors are determined by experts and those select the level of Severity. This formula allows to anyone to utilize each element of severity to match any company priorities [6].

$$Diversion\,Risk\,Index = \sum_{i=1}^{e} s_i \cdot a_i \tag{1}$$

TABLE I
EXAMPLE OF THE INITIAL SET OF THE STATIC DATA USED FOR THE METHODOLOGY ASSESSMENT.

Order	Name	Severity Index	Current Availability Index
1	Airport characteristics	3	1
2	Weather	3	2
3	Fuel on board, ETOPS	3	1
4	Training, licences and other	2	1
5	Company polices	1	1
6	Transport to the destination	1	1
7	Airport infrastructure	2	1
8	Flight operator's infrastructure	1	1
9	Services agreements	1	1
10	Local familiarity and difficulty	1	2
11	Services for passengers	2	1

III. RESULTS

The main result of the consequent research of the Enhanced Diversion Airport Selection Methodology [6] and the aircraft diversion problematics stem in comperhensive understanding of the operational and commercial requirements and limitations. As such and following the expert assessment, it has been confirmed, that more significant benefits can be obtained, when applying this methodology during the final phase of flight. The main reason is that that en-route types of diversion, and diversions shortly after the departure are more often caused by the "internal factors" (technical malfunction, medical problem, etc.).

The diversions factors have been divided Into the 3 categories, the primary – determining if the flight can land on the airport, secondary – commercially significant – determining, if whether there is a risk of delaying the aircraft due to the non-availability of the airport resources and tertiary – not having a significant influence on the aircraft landing and timely turnaround of the flight but rather respecting the airline contractual relations with the airport or handling companies and other general circumstances.

A. Go & No-Go Factors

The primary factors can be named in a simplified language as "Go & No-Go factors. There are static information and airport characteristics that are determining if the aircraft can land on the airport, comprising the runway and taxiway characteristics to withhold the aircraft and its weight, the sufficient navigational equipment. These are to be assessed prior and maintained in a static database, however in the operational environment it is important to examine the dynamic information (NOTAMs) to confirm that the airport characteristics have not been compromised. An important part is also the so-called minimum ground service equipment (GSE), that is relative to the aircraft type. The primary dynamic factors have been limited to:

- Airport technical characteristics – information obtained from the NOTAM. Determining any downgrade to the

state of the runway, available taxiways, and navigational equipment.
- Current weather at the airport.
- Proximity of the airports (depending on the current live position of the aircraft).
- Aircraft stand (there must be an available place to park the aircraft at the airport).
- Minimum GSE – the minimum ground service equipment list has been determined as follows:
  - First it is assessed if a self-manoeuvring aircraft stand can be allocated. If not there has to be either a tractor and a towbar or a towbar-less tug able to operate a given aircraft type.
  - At least one stair suitable for the aircraft type (or a jetway as an alternative).

The minimum GSE list can be altered in case there are additional downgrades to the aircraft system. An inoperative APU can result in the necessity of having a suitable GPU and the air-starter, etc. A non-availability of any of the above-mentioned factor results is the un-usability of the airport and the airport is to be considered only in a case of emergency as the likelihood of the recovery of an aircraft from the airport (continuation of the flight) is unguaranteeable [11].

B. Operationanly Significant Factors

Based on the current availability of the variables as defined in the set of operationally significant factors is to determine a possible risk of additional delay in the provision of the basic handling and fuelling services and furthermore the passenger services. Following the expert assessment, it has been confirmed, that the operationally significant factors are:

- ATC capacity – this determines if the runway through put is allowing a direct landing of the aircraft without the expectation of prolonged holding to get into the landing sequence.
- Fuel availability – determined as the sufficient capacity of the fuel stock at the airport and the ability of a fuel truck or the hydro trucks to refuel the aircraft. It is described by the number of the currently available refueling machinery.
- Advanced GSE availability – the list of ground service equipment suitable for the aircraft type available at the airport and its count. (greater number of available GSE can be directly linked with their allocation to the diverted aircraft).
- De-icing capability and availability.

C. Other Factors

After evaluating the primary factors, the Go, No-Go and secondary factors there are tertiary factors or other factors. Based on their nature it is also possible to determine the suitability of the airport for landing, however, are of a lesser significance than the aforementioned factors. Tertiary factors mostly come from business side of aircraft operator. One of them is flight operator's infrastructure. If an airline is operating other flight to this location, selected as the alternate, it can be considered as a more suitable due to higher level of familiarity

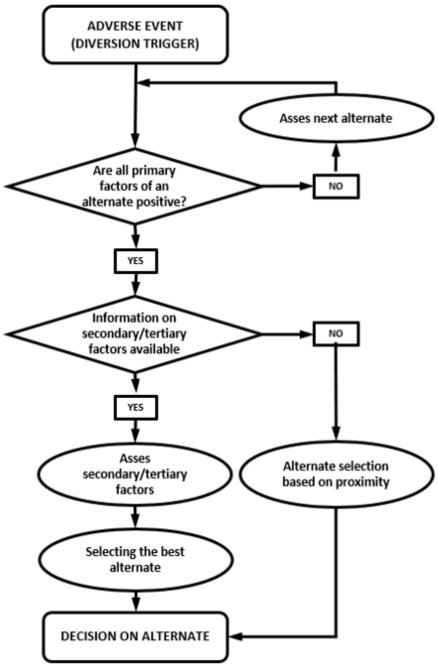


Fig. 2. Decision making algorithm for a selection of the most appropriate alternate location following the adverse event that triggered the need to reroute an aircraft.

and possible presence of the contractual links with the service provider. It can be named Location familiarity. Different to company's business factors are passengers' requirements. This creates a demand on carrier to maintain the level of service even if there is a disruption of flight – level of airport infrastructure – hence the greater ability for the provision of the passenger supplies in case there will be prolonged waiting times, the general traffic situation and ability of the alternative means of transport, the hotels availability in the nearest area. The significancy of these vary with the nature of the diversion, however for the use case as described in the model situation, when for most of the flights a "Fuel and Go" diversion scenario is desired.

The detailed information, once collected and provided to the decision operational authority of the aircraft (crew or the OCC) resulting decision-making algorithm is represented on the Fig. 2, where the flow of actions shall assure the selection of the most appropriate diversion location in the given situation.

D. Database Prototype

Based on the aforementioned factors and needs, a database prototype is being developed, which will be used for the collection of airport data. This is in the sense of gaining awareness of the current occupancy of all necessary entities of the airport. The data model therefore includes data item for evaluating the occupancy of the airport area in terms of stand occupancy. Furthermore, data item for determining the workload of ground handling staff and, independently, the availability of ground equipment (GSE). A low occupancy mark does not necessarily indicate the availability of the necessary GSE for a specific aircraft type. Another item in the data model is Fueler. In most cases, Fueler and Handling are two different companies, so it is necessary to know this entity as well.

After uploading the data to the data model, the Diversion Risk Index is calculated for each monitored airport. Those DRI are then represented on the User Dashboard.

E. Data Model

A specific data model has been developed for the purposes of the database prototype. The data model has been developed in the software PhpMyAdmin and it's based as a relational SQL database, bearing the factors as listed in the previous sections. The data model enables a dynamic update of the information with the inheritance logic, interlinking the values with the e.g. contact details of the handler, these could become useful for possible verification of the situation and information and quality check.

IV. Discussion

Usage of the static data and expert assessment methodology is limited due to several constraints [12]. The scope of experts could have been broader and covering more aircraft operators, which would possibly result in further refinements of the discussed data items. Similarly, due to the non-existing data platform momentarily allowing the additional determination of benefits or limitations the expert assessment was conducted in non-operational or simulated environment, rather than via the static presentations. A further development of the Enhanced Diversion Airport Selection Methodology requires the testing of the data collection feasibility to fulfil our factors as the data items. Another testing requirement is a computational simulation testbed, enabling the dynamic testing of the aircraft distribution in a diversion scenario, based on the dynamic availability and diversion suitability of the alternate locations.

The authors of the thesis have decided to prepare a simulation platform in the AnyLogic simulation software, enabling the dynamic simulation of the air traffic distribution of the massive diversion scenario (see Fig. 1) with the possibility of automated alternate airport selection of the aircraft into the diversion location as per following logic: The aircraft will be choosing the diversion locations based on the most optimal Diversion Risk Index that will include the availability of the factors coming into the Diversion Risk equation (see Fig. 2). With more incoming aircraft the available resources will be spent, and the airports indexes are about to change accordingly, steering the traffic into more suitable location. The data sources for the simulation testbed are planned to be simplified. Aircraft data model shall be bearing the ICAO aircraft type and dimensions (as the width of the aircraft can differ in case the winglets are installed), the aircraft code (with the reference to the ICAO Aerodrome Reference Code); it's position (from where the time to alternate destination shall be computed) and the aircraft pre-set preference for the diversion location, for further testing scenarios. The airport data model shall bear the data items as defined by the methodology as described in this paper, in a simplified manner. The default values of the airport factors will be pre-set with the potential of their dynamic changes within the course of the simulation. The simulation software shall enable the input and modification of the basic scenarios, the air traffic and the airport situation in order to compare them with the real traffic data and a consequent comparison of the simulated behaviour using the Enhanced Diversion Airport Methodology with the real-life situations.

V. Conclusion

It can be concluded, that following the expert's assessment of The Enhanced Diversion Airport Selection Methodology, its comparison with the other research in the area of the aircraft diversions and the diversion impacts assessment, there is a potential of the methodology to dampen the negative impact of the flight irregularities on the aircraft operators, passengers and other concerned entities. A further testing of usage of the database prototype shall confirm the ability of the data collection in the airport operational environment. The simulation testbed shall show the benefits and further limitations of the proposed methodology. The simulation and comparison with the real-life diversion occurrences shall also estimate the dependency of the Diversion Risk Index (DRI) on the time delay and further evaluate the hypotheses about the influence of the DRI on the flight operations economics.

Acknowledgment

This work has been supported by the project SGS21136OHK22T16 – "Enhancement of the diversion airport selection methodology".

References

[1] U.S. Department of Transportation, *National Transportation Statistics*. Washington, DC, USA: U.S. Department of Transportation - Bureau of Transportation Statistics, 2018. [Online]. Available: https://www.bts.gov/sites/bts.dot.gov/files/docs/browse-statistical-products-and-data/national-transportation-statistics/223001/ntsentire2018q3.pdf

[2] Sheffield School of Aeronautics. (2019) Reasons an aircraft dispatcher may divert a flight. [Online]. Available: https://www.sheffield.com/articles/reasons-an-aircraft-dispatcher-may-divert-a-flight

[3] SKYbrary. (2019) Emergency landing. [Online]. Available: https://skybrary.aero/articles/emergency-landing

[4] SKYbrary. (2021) Medical emergencies - guidance for flight crew. [Online]. Available: https://skybrary.aero/articles/medical-emergencies-guidance-flight-crew

[5] N. Kohl, A. Larsen, J. Larsen, A. Ross, and S. Tiourine, "Airline disruption management—perspectives, experiences and outlook," *Journal of Air Transport Management*, vol. 13, no. 3, pp. 149–162, May 2007. [Online]. Available: https://doi.org/10.1016/j.jairtraman.2007.01.001

[6] M. Špák and P. Olexa, "Enhancement of the diversion airport selection methodology," *Transportation Research Procedia*, vol. 51, pp. 232–242, 2020. [Online]. Available: https://doi.org/10.1016/j.trpro.2020.11.026

[7] EUROCONTROL. (2022) Eurocontrol market segment update 2022. [Online]. Available: https://www.eurocontrol.int/sites/default/files/2022-05/eurocontrol-market-segment-update-2022-05.pdf

[8] C. Malandri, L. Mantecchini, and V. Reis, "Aircraft turnaround and industrial actions: How ground handlers' strikes affect airport airside operational efficiency," *Journal of Air Transport Management*, vol. 78, pp. 23–32, Jul. 2019. [Online]. Available: https://doi.org/10.1016/j.jairtraman.2019.04.007
[9] C. Malandri, L. Mantecchini, F. Paganelli, and M. N. Postorino, "Impacts of unplanned aircraft diversions on airport ground operations," *Transportation Research Procedia*, vol. 47, pp. 537–544, 2020. [Online]. Available: https://doi.org/10.1016/j.trpro.2020.03.129
[10] A. Massidda and S. P. Mattingly, "Addressing irops diversion-related extended delay events through a coordinated regional airports/airlines diversion network," 2011.
[11] B. Li, L. Wang, Z. Xing, and Q. Luo, "Performance evaluation of multiflight ground handling process," *Aerospace*, vol. 9, no. 5, p. 273, May 2022. [Online]. Available: https://doi.org/10.3390/aerospace9050273
[12] E. A. Burkov, P. L. Lyubkin, and P. I. Paderno, "Intellectual systems - the future of expert assessment," in *2017 XX IEEE International Conference on Soft Computing and Measurements (SCM)*. IEEE, May 2017. [Online]. Available: https://doi.org/10.1109/scm.2017.7970487

C Appendix 3

Dalmau R., De Falco P., Spak M., Rodriguez Varela J. D.

Probabilistic pre-tactical arrival and departure flight delay prediction with quantile regression.

15th USA/Europe Air Traffic Management Research and Development Seminar. 2023.

Fifteenth USA/Europe Air Traffic Management Research and Development Seminar (ATM2023)

Probabilistic Pre-tactical Arrival and Departure Flight Delay Prediction with Quantile Regression

A case study for Geneva international airport using operational data

Ramon Dalmau & Paolino De Falco
EGSD/INO/ENG
EUROCONTROL Innovation Hub (EIH)
Brétigny-Sur-Orge (France)

Miroslav Spak
NMD/TRG/TDD
EUROCONTROL
Brussels (Belgium)

José Daniel Rodriguez Varela
Geneva Airport
Geneva (Switzerland)

***Abstract*—Airports plan their resources well in advance based on anticipated traffic. Currently, the only traffic information accessible in the pre-tactical phase are the flight schedules and historical data. In practice, however, flights do not always depart or arrive on time for a variety of reasons, such as air traffic flow management or reactionary delay. Because neither air traffic flow management regulations nor aircraft rotations are known during the pre-tactical phase, predicting the precise arrival and departure delay of individual flights is challenging given current technologies. As a result, probabilistic flight delay predictions are more plausible. This paper presents a machine learning model trained on historical data that learned the various quantiles of the departure and arrival delay distributions of individual flights. The model makes use of input features available during the pre-tactical phase, such as the airline, aircraft type, or expected number of passengers, to provide predictions of the delay distribution several days before operations. The performance of the model trained on operational data from Geneva airport is compared to a statistical baseline, providing evidence that machine learning is superior. Furthermore, the contribution of the various input features is quantified using the Shapely method, stressing the importance of the expected number of passengers. Finally, some examples are presented to illustrate how such a model could be applied in the pre-tactical phase.**

***Keywords*—Flight delay; machine learning; quantile regression**

I. Introduction

Flight delay is commonly defined as the difference between actual and scheduled times of departure or arrival of a flight from or to an airport, respectively[1]. As a result of the crisis related to the COVID-19 pandemic, aviation activities drastically dropped in 2020 [1]. However, the last few months have recorded a recovery of air traffic, characterised by increasing flight delays. During the 3rd quarter of 2022, the average delay per flight in the European Civil Aviation Conference (ECAC) area was 23 min, the highest value recorded in the last 5 years [2]. Flight delay is one of the key performance indicator of air transportation since it can impact negatively the airline and airport management as well as the level of passengers' satisfaction [3]. Better prediction of flight delays could aid in the implementation of mitigation measures before they occur.

Machine learning algorithms have proven to be effective in predict flight delays during the tactical phase (i.e., during the day of operations) [4], [5], when both aircraft rotations and air traffic flow management (ATFM) measures are fully (or partially) available to assist the models. A critical aspect when developing machine learning models is the type and quality of data that are available at the time horizon of interest. From the Network Manager (NM) and airports point of view, the only data available during the pre-tactical phase (i.e., several days before operations) are the flight schedules, as aircraft rotations and exact ATFM measures are still unknown. In such an uncertain time horizon, it is more reasonable to approach the flight delay prediction problem with probabilistic models capable of providing not only the expected value of the delay (arrival or departure) but also its probability distribution.

From the airline perspective, the duty manager in charge of monitoring and manage the fleet decides whether a flight should be cancelled or revised [6]. In order to perform this task, he/she needs to access information about the costs of different alternatives, which are non-linear with respect to the (uncertain) delay values [7]. Capturing not only the expected values of the flight delay but also its likelihood might improve the decision-making process of the duty managers.

From the airport perspective, airport management implies decision making under uncertainty, which becomes critical especially for long look-ahead times [8]. As an example, strategic airport capacity planning is typically not sufficiently accurate because of the inherent uncertainty of weather forecasts [9]. Although flight schedules provide an indication on when aircraft might depart from or arrive at an airport, this information always carries a certain amount of uncertainty which makes airport planning operations very challenging (e.g., to decide where and when to allocate the ground handling resources, or to efficiently plan the shifts for the staff), especially several days before the day of operations.

This paper presents a probabilistic model that utilises historical flight data to predict arrival and departure flight delays several days in advance. The model is based on multi-

[1]Throughout this paper, departure time refers to off-block time, whereas arrival time refers to in-block time.

quantile regression, which is a method for estimating how the different quantiles of a distribution (in this context, the departure and arrival delay distribution) change as a function of a set of predictors. It should be noted that all predictors used in this paper (also known as features in the machine learning jargon) are available during the pre-tactical phase. The predictions in the test set, which includes observations never seen by the model during training, are compared to a dummy baseline that assumes flights will depart and arrive on time, as well as a baseline based on standard statistics.

This paper is organised as follows: a literature review on flight delay prediction, with a particular focus on probabilistic models, is performed in Section II; Section III provides the description of the generic multi-quantile regression model that was tailored to the departure and arrival flight delay prediction problem at Geneva Airport (GVA); the details of the experiment and the results are presented in Sections IV and V, respectively; Section VI provides a discussion of the results and an overview of the implementation at GVA.

II. Literature review

In recent years, along with the development of sophisticated machine learning models, there has been a lot of interest in probabilistic flight delay prediction. The emergence of probabilistic flight delay prediction models is also likely due to the fact that point predictions are not sufficiently accurate given the uncertainty of the air transportation system, in which many agents interact (passengers, ground handlers, air traffic controllers, flight dispatchers, etc.) in addition to the weather.

For example, [10] used random forest (i.e., an ensemble of decision trees trained with the *bagging* method) and clustering algorithms to predict departure delays at US airports with a look-ahead time up to 24 hours. Promising results revealed that combining clustering and ensembles of decision trees is effective at predicting flight delays several hours in advance.

The effectiveness of alternative machine learning models has also been investigated. For instance, [11] compared the performances of random forest and recurrent neural networks (RNNs) when predicting the flight delay at Chinese airports. The authors also approached the flight delay prediction problem as a classification task, in which the model learns the probability of the delay falling into one of several predefined categories rather than forecasting the precise delay in minutes.

RNNs were also used by [4] to predict the arrival delay propagation along a sequence of flights (i.e., rotation) operated by an aircraft along the day. Specifically, the model was trained to predict the parameters of the arrival delay distribution, which was modelled as a Gaussian function for the sake of simplicity. The proposed model requires rotations data to propagate the (predicted) delay, which, as previously stated, are not available during the pre-tactical phase.

In parallel, [5] also addressed the probabilistic flight delay prediction problem. The authors presented a machine learning model to categorise flight departure times as *early*, *on-time*, or *delayed*. Similar to [4], however, the proposed model requires knowledge on the prior flight operated by the same aircraft, and thus cannot be used during the pre-tactical phase.

Recently, [12] developed two probabilistic models for individual flight delay prediction model using mixture density networks (MDN) and random forest, respectively. In reality, however, the generic random forest model was designed to perform point predictions, not probabilistic. In order to obtain the flight delay distribution from a random forest, the predictions of the individual decision trees of the ensemble were not averaged, but collected, and a kernel density estimation (KDE) was performed. Using this approach, however, the model is still trained to minimise a loss function designed for point predictions, like the mean absolute error (MAE).

Regarding the MDN proposed by [12], it comprises a neural network that predicts the parameters for each Gaussian component in the mixture. The parameters (i.e., weights and biases) of the neural network are trained to minimise the negative log-likelihood. Consequently, the MDN assumes that the delay can be represented by a multi-modal Gaussian distribution. On the other hand, the model proposed in this paper does not make any specific assumptions regarding the shape of the delay distribution. Furthermore, it is important to note that neither the random forest nor the MDN proposed by [12] could used several days before operations since they rely on weather information at the destination or origin airport, which is only available 24 hours before operations.

In a similar vein, [13] explored probabilistic flight delay predictions using Bayesian artificial neural networks (ANNs) to predict aggregate flight delays in the United States, broken down by airport. Their study highlights the difficulty of predicting even aggregate-level flight delays, underscoring the importance of uncertainty quantification. Similar to [12], the model requires weather features (e.g., visibility, temperature) that are not available several days before operations.

Finally, [14] assessed the performance of various machine learning models for probabilistic flight delay prediction, including ANNs, random forest and gradient-boosted decision trees (GBDTs). Like [4], the authors assumed a Gaussian distribution that was fitted to the flight delays. The various machine learning models were then trained on historical data to predict the parameters of the distribution. For the GBDTs instance, two models were trained: one to predict the mean and the other to predict the standard deviation. Different from [4], [12] and [13], the features used by the models are available in the pre-tactical phase.

III. Generic model

Unlike classical regression models, which estimates the conditional mean of the target (i.e., the output) across features (i.e., the inputs) using the least squares approach, quantile regression determines the relationship between the features and a quantile (or quantiles) of the target distribution. It should be noted that, in contrast to previous works that parameterised the presumed delay distribution and then learned its parameters using machine learning [4], [14], quantile regression makes no assumptions about the distribution of the target and is robust to the influence of outliers.

There are various machine learning models that can be extended to quantile regression tasks. Gradient descent-based learning algorithms, such as ANNs, can learn a specific quantile by switching from the classical MAE or mean squared error (MSE) loss to the mean pinball error (MPE):

$$\text{MPE} = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} \text{PE}\left(y_i, \hat{y}_i, \alpha\right), \tag{1}$$

where n_{train} is the number of training observations, y_i and $\hat{y}_i$ are the actual and predicted target for the i^{th} observation, respectively, $\alpha \in [0, 1]$ is the quantile to be learned, and

$$\text{PE}\left(y, \hat{y}, \alpha\right) = \alpha \max(y - \hat{y}, 0) + (1 - \alpha) \max(\hat{y} - y, 0) \tag{2}$$

is the pinball error (PE) for one observation.

In many practical applications, the goal is to determine not just one, but several quantiles. There are two methods for accomplishing this goal. The first approach involves training a separate model for each quantile. Because the models corresponding to the different quantiles are trained independently, the consistency of the predictions cannot be guaranteed [15]. Furthermore, this strategy necessitates the development and maintenance of many models, making it a time-consuming and inefficient option in practise.

The second approach consists of training just one model with a outputs, each one associated to one quantile, to minimise the mean multi-quantile pinball error (MMQPE):

$$\text{MMQPE} = \frac{1}{n_{\text{train}}} \sum_{i=1}^{n_{\text{train}}} \sum_{j=1}^{a} \text{PE}\left(y_i, \hat{y}_i, \alpha_j\right). \tag{3}$$

Many machine learning models can be configured to handle multi-quantile regression tasks. The generic model proposed in this study is based on ensemble methods, which produce a strong learner from a group of weak learners. Boosting is a well-known ensemble method that involves training a series of weak learners (e.g., rudimentary decision trees) sequentially. The training observations for the next learner in traditional adaptive boosting (AdaBoost) [16] are weighted based on how well the previous learners performed, i.e., observations that correspond to wrong predictions are assigned more weight in order to concentrate the model's attention on correcting them. Gradient boosting differs from AdaBoost in that, instead of assigning weights to observations based on performance, a new learner is trained at each iteration to fit the residual errors of the preceding learners. The ensemble is known as GBDTs model when decision trees are used as weak learners.

GBDTs can outperform ANNs in many practical applications, notably on tabular datasets where each row corresponds to one observation and each column represents a feature [17]. Furthermore, GBDTs are easier to interpret than ANNs and have very attractive properties such as the ability to handle missing data and categorical features with high cardinality. The GBDTs model was chosen for the problem addressed in this study because of the numerous benefits it provides.

Sections III-A and III-B list the features that compose the observation vector $\boldsymbol{x}$ and define the target y of the generic model developed during this research, respectively. It should be emphasised that this generic model could be trained using any of the traditional GBDTs algorithms (e.g., lightGBM, CatBoost, XGBoost) on historical data gathered by any airport. Section IV will present the specific GBDTs algorithm used to train the model as well as the dataset. Furthermore, two independent GBDTs models were trained: one to predict the quantiles of the arrival delay distribution and the other to predict the quantiles of the departure delay distribution. The set of features used by these models, however, is similar.

A. Input features

There are various limitations on the set of features that can be incorporated when building a model for usage during the pre-tactical phase. Predictions cannot, of course, be made using information from the future. For example, the majority of ATFM regulations are defined either the day before operations (so-called pre-tactical regulations) or tactically the same day. As a result, this information is unknown several days or weeks in advance. Similarly, the sequence of flights operated by each aircraft (i.e., registration number) must be known in order to anticipate rotational reactionary delays. The registration number that is going to operate a certain flight is only known when the airline submits the flight plan to the Network Manager (NM). Airlines, however, tend to wait for the most accurate weather and network information before submitting the flight plan. As a result, several days or weeks in advance, only the aircraft type that will be used for a flight can be speculated, but not which will be the inbound flight.

Based on the preceding discussion, it is understandable that the set of (more or less certain) features available for making predictions in the pre-tactical phase is rather limited. The model proposed in this paper uses the following 14 features: (1) airline, (2) handling agent who will process the flight, (3) destination (resp. origin) airport for departures (resp. arrivals), (4) aircraft type (e.g., A320), (5) flight service type, (6) type of flight (e.g., scheduled), (7) whether or not is a Schengen flight, (8) hour of the day, (9) day of the week, (10) month of the year, (11) great circle distance (GCD), as well as (12) the number of departures and (13) arrivals scheduled in the same hour. The last feature of the model is the expected number of passengers, which is estimated based on historical load factors[2] according to a model executed by the operations performance & forecasting department of GVA.

It is worth noting that the notion that circular features, such as the hour of the day, day of the week, or month of the year, always require transformation using sine and cosine functions is often misunderstood. While this transformation is commonly used in neural networks to capture periodicity, decision tree-based algorithms, such as random forest and GBDTs, can effectively handle circular features without the need for explicit transformation. The authors conducted experiments using both approaches and found that the categorical approach generally yields superior results.

Furthermore, it may seem that GCD and airport are highly correlated and provide duplicate information, but this is not the case. Different airports may have different operating

[2]The load factor is an aviation industry indicator that represents the proportion of available seating capacity that has been filled with passengers.

methods, leading to different contributions to the predictions. Meanwhile, the GCD feature was included to allow the model to learn the correlation between delay and the length or duration of the flight. Additionally, observations with airports that have few observations and are not representative in the training set could benefit from the more generic GCD feature.

B. Output target

The departure (resp. arrival) delay prediction model outputs multiple quantiles of the predicted departure (resp. arrival) delay distribution. Specifically, the models were trained to minimise the MMQPE and predict the 5^{th}, 25^{th}, 50^{th} (i.e., the median), 75^{th} and 95^{th} quantiles of their respective targets. The quantiles were selected to represent the entire delay distribution, including both regular and extreme events.

It should be noticed that during the training phase, the model generates five values (one for each quantile) for each training observation but only requires one ground truth y (the actual delay) to compute the multi-quantile pinball error.

IV. Experiment

This section describes the experiment carried out in this research to evaluate the performance of probabilistic models in predicting departure and arrival delays during the pre-tactical phase. Section IV-A describes the datasets used for the experiment, while Section IV-B covers the specific GBDTs algorithm as well as the hyper-parameters of the models.

A. Specific dataset

A dataset is a collection of n observations $\boldsymbol{X} := (\boldsymbol{x}, y)^n$ used to train a model and assess its performance. In this work, two datasets were created: one for departures and one for arrivals, with each observation belonging to one flight departing from or arriving to GVA from the 28^{th} of October 2018 to the 11^{th} of December 2022, respectively. It should be noted that the traffic from March, 1^{st} 2020 to July, 1^{st} 2021 was excluded from the dataset because it was strongly affected by the COVID-19 pandemic. The raw data used to generate the features and target of each observation were kindly provided by GVA. A portion of the data, such as the predicted number of passengers per flight, is confidential and therefore cannot be publicly disclosed.

Table I lists and describes the features that compose the observation vector $\boldsymbol{x}$ in the two datasets. The columns of this table show basic statistics computed on the entire datasets, including both train and test sets. For each categorical (i.e., discrete) feature, like the departure airport, Table I shows the number of unique values, the most frequent value (Top) as well as its frequency (Freq.). For each numerical (i.e., continuous) feature, like the great circle distance (GCD), three quartiles are presented: 25^{th} (Q1), median (Q2) and 75^{th} (Q3).

It is worth noting that the model presented in this paper does not incorporate weather features since it is designed to be used several days before operations, when weather forecasts are often inaccurate. However, a variant of this model could be designed for use one day before operations, when Terminal Area Forecasts (TAF) for the next 24 hours are available. This variant could incorporate features such as visibility, cloud ceiling, wind speed and direction, gusts, and significant weather phenomena like thunderstorms, fog, and snow. With these additional features, this variant could better capture the effect of weather on departure and arrival delays.

Furthermore, The model does not incorporate features related to reactionary delay because the sequence of flights operated by each aircraft is only known on the day of operations when airlines submit their flight plans to the NM. Nonetheless, the authors of this paper encourage the air traffic management (ATM) research community to explore the possibility of developing a model that can predict the rotations of a particular aircraft days in advance based on its recent history. The predicted rotations could be incorporated as additional information into the model presented herein, likely improving the quality of the model's predictions.

In many applications, dataset splitting is done randomly by taking 80% of the data for training and using the rest for assessing the performance on *unseen* data (i.e., testing). When dealing with time-related and dynamically changing environments, such as the air transportation system, it is preferable to employ time-based splitting to provide statistically robust model evaluation and better imitate real-life scenarios. Accordingly, the first 80% of the flights (ordered by time) where used for training, and the rest for testing.

B. Specific algorithm

In this paper, the CatBoost implementation of the GBDTs model by Yandex [18] was used. CatBoost has gained more momentum than other GBDTs implementations (e.g., XGBoost and LightGBM) mainly because its native ability to handle high-cardinality categorical features like the departure and destination airports, as well as the use of ordered boosting and symmetric trees, which help to overcome over-fitting.

Many hyper-parameters can be used to optimise the CatBoost model, which allow to control the entire ensemble (e.g., the number of decision trees) as well as individual decision trees (e.g., the maximum depth). In the experiment conducted in this study, only the maximum depth and the number of decision trees were optimised because they were found to have the most significant impact on the loss function. The learning rate was determined automatically using the CatBoost framework's heuristic, which is dependent on the dataset attributes and the number of decision trees.

A widely used procedure to assess the performance of a model given a combination of hyper-parameters is the cross-validation (CV). The most basic k-fold CV, for instance, consists of splitting the train set into k subsets, also known as folds. Then, the following procedure is applied to each of the k folds: a copy of the model is trained using the other $k-1$ folds as train set, while the fold in hand is used as test set to compute a performance score. The average of the k scores is the CV score of the model using the combination of hyper-parameters under consideration. In this paper, the CV procedure was performed by respecting the temporal order of the observations with a `TimeSeriesSplit` [19]. Specifically, this variation returns first (order by time) i folds as train set and the $(i+1)^{\text{th}}$ fold as validation set (with $i \in \{1, k-1\}$), and averages the resulting $k-1$ scores.

TABLE I. Input features and statistics on the entire dataset (train & test)

Type	Name	Departures (170K observations)			Arrivals (170K observations)		
		Unique	Top	Freq (%)	Unique	Top	Freq (%)
Categorical	Airline	159	EZS	23	153	EZS	23
	handling agent	2	SWISSPORT	74	2	SWISSPORT	74
	Airport	266	LHR	6	267	LHR	6
	Arctyp	68	A320	42	68	A320	42
	Flight service type	14	J	97	14	J	97
	Type of flight	7	S	55	5	S	54
	Schengen flight	2	Y	66	2	Y	66
	Dayofweek	7	6	15	7	4	15
	Hour	19	9	8	20	8	7
	Month	12	12	10	12	12	10
		Q1	Q2	Q3	Q1	Q2	Q3
Numerical	Pax total (#)	82	123	155	83	123	156
	Hourly arrivals (#)	6	10	13	8	11	15
	Hourly departures (#)	9	12	15	7	10	14
	Great Circle Distance, GCD (km)	532	754	1309	532	754	1309

There exist several methods to search the hyper-parameter space for the best CV score. The most popular method is the `GridSearchCV` [19], which consists of exhaustively evaluating all the possible candidates (i.e., combinations of hyper-parameter) and returning that minimising the CV score. In this study, a more refined method called `HalvingGridSearchCV` [19] was used. The `HalvingGridSearchCV` consists of evaluating all possible candidates with a small amount of *resources* at the first iteration. In the second iteration, only some of these candidates are selected for the next iteration, which will be allocated more resources, and so on. In this paper, the number of decision trees in the ensemble was used as resource. The reader should keep in mind that when using the number of decision trees (also known as estimators) as resource, this hyper-parameter cannot be included in the search grid. It is optimised intrinsically by the `HalvingGridSearchCV` method, and including it in the search grid (which would be incorrect) will result in an exception being thrown.

For both departure and arrival delay prediction models, the best maximum depth and number of decision trees were found to be 9 and 1K, respectively.

V. Results

This section presents the results of the experiment described in Section IV. Specifically, Section V-A compares the performance metrics of the model with those of two baselines. Then, Section V-B unravels the attribution of the features according to the Shapley values computed with the trained models. Finally, Section V-C shows illustrative examples.

A. Performance

The performance of the proposed models must be compared to some reference values, i.e., the baseline. The simplest baseline is to assume that all flights will depart and arrive on time at their scheduled departure and arrival time, respectively. That is, regardless of the values in the observation vector $\boldsymbol{x}$, the predicted quantiles are zero for all observations in the test set. For the remainder of the paper, this baseline will be referred to as the 'zero delay'. This baseline is practical because it is extremely simple, and it also nearly replicates the current system when delays are completely disregarded.

A more principled baseline consists of predicting the quantiles of the delay distribution based on historical data. In this paper, flights in the train set were grouped by season (winter, spring, summer or autumn), period of the day (morning, afternoon, evening, late or late night) as well as airport (departure or destination depending on the model). For each one of these groups, the various quantiles of the departure and arrival delay distributions were computed, and these quantiles were used as predictions for the observations in the test set belonging to the same group. It is worth mentioning that groups with fewer than 250 observations were declared underrepresented, implying that the predicted quantiles are not statistically significant. Observations in the test set that were missing predictions because their group was underrepresented in the train set were assigned the quantiles of the delay distribution resulting from grouping by season and period of year, omitting the airport. For the remainder of the paper, this baseline will be referred to as the 'statistics'.

The authors selected these two rather naive baselines because, to the best of their knowledge, none of the models proposed in the literature, except for [14] - who also used a GBDTs model, but assumed a Gaussian distribution - are able to perform probabilistic predictions several days before operations due to the need for weather data, information about aircraft rotations, and/or ATFM regulations.

Table II shows the performance metrics on the test set for the two baselines and the machine learning models. Results indicate that the statistics baseline outperforms the zero delay baseline, particularly for the high quantiles. It reduces the 95$^{\text{th}}$ quantile's MPE of the departure and arrival delay distributions by 11.1 min (64%) and 9.6 min (63%), respectively. However, the performance is very similar at the low quantiles. The statistics baseline reduces the MMQPE of the zero delay baseline by 15.9 min (33%) and 15 min (31%) for the departure and arrival delay predictions, respectively.

TABLE II. Performance metrics on the test set (min)

Operation	Model	MPE for the various quantiles					MMQPE
		5^{th}	25^{th}	Median	75^{th}	95^{th}	
Departures	Zero delay	1.9	5.3	9.6	13.9	17.4	48.1
	Statistics	1.3	5.3	8.9	10.3	6.3	32.2
	CatBoost	1.2	4.9	7.9	8.7	5.2	28.0
Arrivals	Zero delay	4.2	6.6	9.7	12.8	15.3	48.6
	Statistics	1.8	6.3	9.7	10.2	5.7	33.6
	CatBoost	1.6	5.8	8.8	9.2	5.0	30.3

The machine learning models proposed herein perform even better. It should be noted, however, that the performance gap between the statistics and zero delay baselines is higher than the performance gap between the CatBoost regressor and the statistics baseline, indicating that relatively simple statistical methods can indeed deliver decent predictions.

Specifically, the CatBoost model reduces the MMQPE of the zero delay baseline by 20.1 min (42%) and 18 min (38%) for the departure and arrival delay prediction tasks, respectively. The relative benefit in comparison to the statistical baseline, however, is not that extraordinary: the MMQPE for the departure delay prediction task improves by 4.2 min (13%), whereas the improvement is about 3.3 min (11%) for the arrival delay prediction task. In any instance, the machine learning approach yields more reasonable estimates for all individual quantiles, particularly for the high quantiles.

The reader should be aware that the interpretation of the MMQPE may be misleading, since it is simply the sum over all the individual quantile's MPE. Adding more quantiles would increase each of the MMQPE values accordingly, (potentially) leading to more dramatic differences across models.

B. Feature attribution

Principles from game theory can be used to interpret the prediction of a model for a given observation vector $\boldsymbol{x}$, assuming that each one of the d features is a player and the model output $\hat{y}$ is the payout. Let us consider the following scenario: all features participate in the game (i.e., contribute to the model output), and the features enter the room where the game is played in a random order. The contribution of a feature could be calculated as the average change in the payout received by the coalition already in the room when the corresponding player (feature) joins them. This contribution measure is commonly known in the literature as the Shapley value. Specifically, the Shapley value $\phi_i(\boldsymbol{x})$ of the feature i for a given observation vector $\boldsymbol{x}$ represents the average marginal contribution of i on the output of the model across all possible combinations of features. It can be proven that the Shapley value is the only contribution measure that simultaneously satisfy local accuracy, consistency, and missingness [17].

In practical applications, however, Shapley values can only be approximated because computing them precisely is an NP-hard problem. `TreeExplainer` is a novel explanation method for tree-based models (including GBDTs) that allows for the tractable computation of Shapley values in polynomial time [17]. The `TreeExplainer` was used in this paper.

Figures 1 and 2 aggregate Shapley values for all the features and observations in the test sets, which were computed by using the `TreeExplainer` with the trained models. Because each model produces five outputs (the various quantiles of the predicted delay distribution), Shapley values can be computed independently for each quantile. Only the 5^{th}, median and 95^{th} quantiles are examined for the sake of clarity.

In Figs. 1 and 2, the vertical axis indicates the name of the features, in order of importance from the top to the bottom in terms of mean absolute Shapley value. Each dot in the horizontal axis shows the Shapley value of the associated feature on the prediction for one observation, and the colour indicates the magnitude of that feature: red indicates high, while blue indicates low. Note that colour has no meaning for categorical features such as the airline or the aircraft type.

According to Fig. 1a, the most influencing feature when predicting the 5^{th} quantile of the departure delay distribution, in terms of mean absolute Shapley value, is the (expected) number of passengers. Results indicate that the higher the number of passengers, the greater the output of the model. It is important to remember that the expected value (i.e., mean) of the target in the train set plus the Shapley values of the individual features equals the model's output. As a result, a positive Shapley value indicates that the corresponding feature is influencing the model's output to be greater than the expected value in the train set. The month of the year and the aircraft type are closely followed by the airline, hour of the day and airport in the list of the most relevant features.

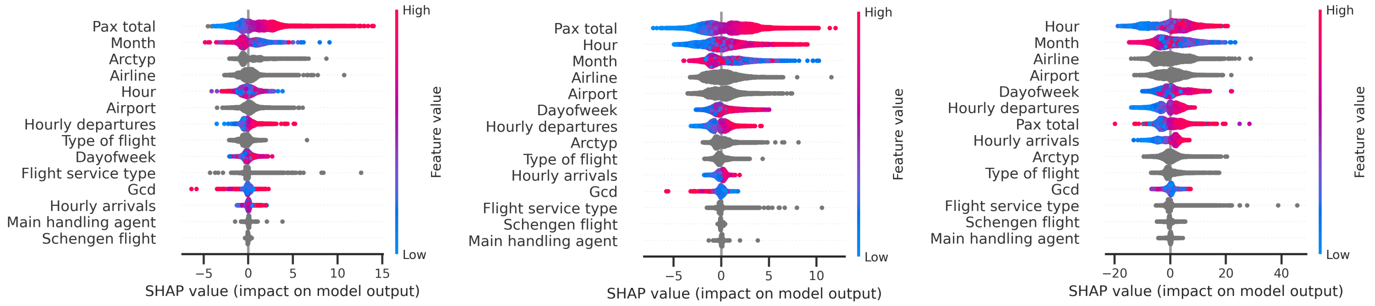
Figure 1b shows how the feature ranking changes when predicting the median. The findings indicate that calendar features are extremely important for higher quantiles.

The preceding statement is further supported by Fig. 1c, which shows that the most important features when predicting extreme departure delays are the hour, the month and the airline. Figure 1 reveals that there is no dominant feature (also known as *golden* feature) in the model, and that multiple features contribute to the output with about the same impact.

Another conclusion that can be derived from Fig. 1 is that, as expected, the higher the number of hourly departure operations (a proxy for airport congestion), the greater the predicted departure delay, albeit with a relatively minor contribution.

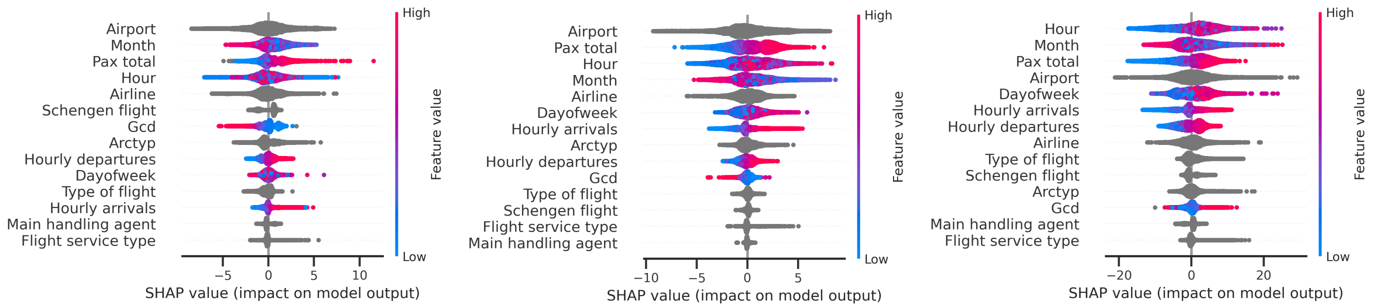
Curiously, Fig. 2 shows that the most important features of the arrival delay prediction model are ranked differently. Specifically, Fig. 1a indicates that the departure airport plays the most important role when predicting the 5^{th} quantile of the arrival delay distribution, while the number of passengers is placed 3^{rd} in terms of mean absolute Shapley value. Similar to Fig. 1, Fig. 2 indicates that calendar features are increasingly crucial at higher quantiles, and that the higher the number of arrival operations, the greater the predicted arrival delay.

Last but not least, Figs. 1 and 2 show that the absolute value of the Shapley value increases with the GCD. This indicates that, for long-haul flights, the delay strongly depends on the distance - according to the data used for training the models. The results presented in this section are only applicable to GVA, and other airports or countries may exhibit different figures. For example, in the United States, delays on connected short-haul flights are often more uncertain as the domino effect can grow from one leg to the next.



(a) 5th quantile (b) Median (c) 95th quantile

Figure 1: Feature attribution distribution for departure delay prediction model. It should be noted that the x-axes of the figures are represented with different scales to enable a visual assessment of the trends



(a) 5th quantile (b) Median (c) 95th quantile

Figure 2: Feature attribution distribution for arrival delay prediction model. It should be noted that the x-axes of the figures are represented with different scales to enable a visual assessment of the trends

The reader should keep in mind that each point in Figs. 1 and 2 corresponds to a single observation, i.e., flight. Accordingly, these figures represent the global interpretation of the departure and arrival delay prediction models, respectively, based on local feature attributions. Figures 3 and 4, on the other hand, illustrate the Shapley value of the expected number of passengers feature as a function of its value for the departure and arrival delay prediction models, respectively. As in the previous figures, each point represents one observation.

Figure 3a shows that, for the 5th quantile of the departure delay distribution, the relationship between the number of passenger and its Shapley value is linear. The results also show that when the number of passengers is below (roughly) 100, the Shapley value corresponding to this feature is negative (i.e., this feature contributes to predicting early departure when compared to the expected value in the training set), whereas it is positive when the number of passengers is above. A similar pattern can be observed for the median in Fig. 3b. For the 95th quantile (see Fig. 3c), however, the relationship is linear only when the number of passengers is below 200. The attribution of this feature is similar for the arrival delay prediction model, according to Fig. 4.

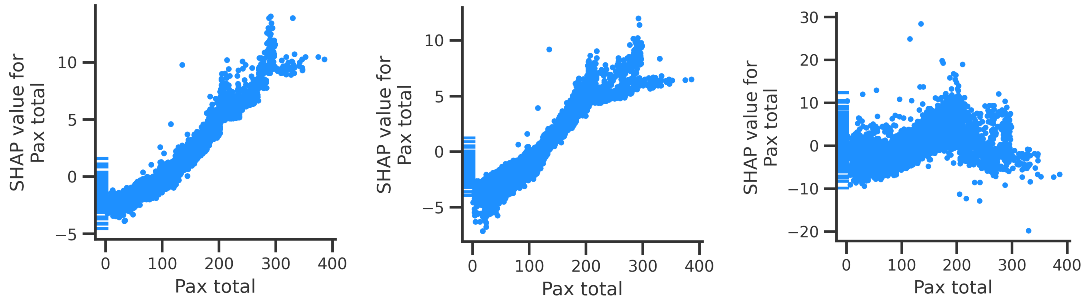
Figures 5 and 6 show the Shapley value of the hourly departures and arrival features as a function of their value for the departure and arrival delay prediction models, respectively. As expected, Figs. 5 and 6 indicate that the higher the number of hourly departures and arrivals, the higher the quantiles of the predicted departure and delay distribution, respectively. For instance, Figs. 5b shows that when the number of hourly departures is lower than around 10, the contribution of this feature is null or negative, whereas higher values tend to increase the median of the predicted delay distribution. Similar conclusions can be derived for the arrival delay prediction as well as for the rest of quantiles.

The dispersion of Shapley values for a specific value of a feature in Figs. 3 to 6 is due to the fact that the Shapley value depends on the value of the other features.

C. Illustrative applications

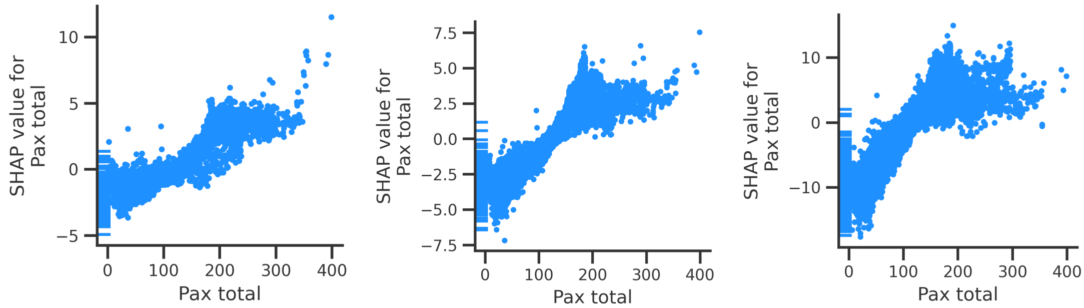
This section presents some illustrative examples of how the departure and arrival delay prediction models covered in previous sections could be used in real operations. In hierarchical order, Sections V-C1 to V-C3 show how to pinpoint flights that are likely to not depart or arrive on time starting for an aggregated prediction over the next months.

1) Detection of problematic days: Let us start with the most basic use case, in which airport operators plan their resources (like number of staff and handling agents, stand and gate allocation, etc.) several days in advance. Figure 7 (resp. 8) shows the mean absolute hourly mismatch between scheduled and potential number of departures (resp. arrivals). For instance, a date marked with the number 3 means that, in average during that day (considering the 24 hours), the absolute difference (positive or negative) between the number of scheduled and potential hourly operations is 3.



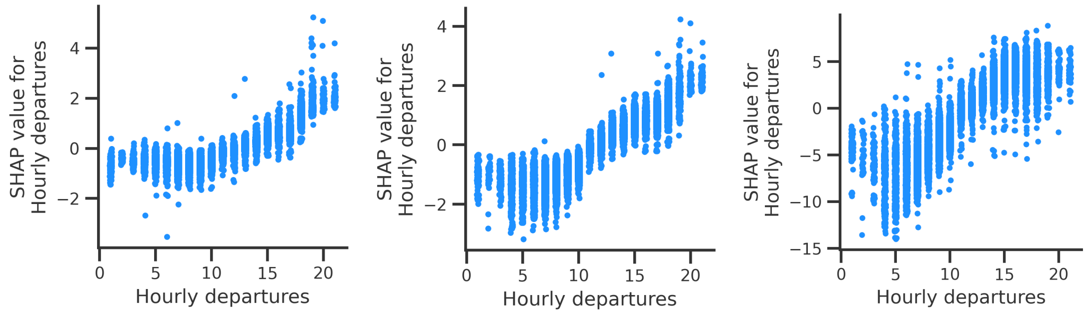
(a) 5th quantile (b) Median (c) 95th quantile

Figure 3: Attribution of the total pax. feature for departure delay prediction model as a function of its value. It should be noted that the y-axes of the figures are represented with different scales to enable a visual assessment of the trends



(a) 5th quantile (b) Median (c) 95th quantile

Figure 4: Attribution of the total pax. feature for arrival delay prediction model as a function of its value. It should be noted that the y-axes of the figures are represented with different scales to enable a visual assessment of the trends



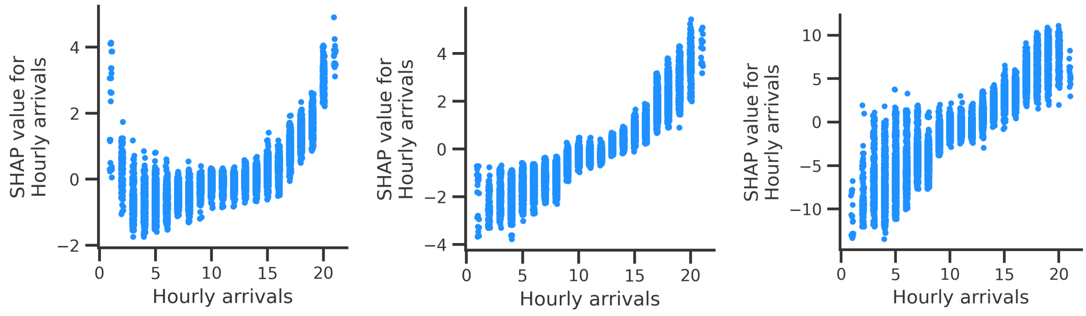
(a) 5th quantile (b) Median (c) 95th quantile

Figure 5: Attribution of the hourly departures feature for departure delay prediction model as a function of its value. It should be noted that the y-axes of the figures are represented with different scales to enable a visual assessment of the trends

Each cell within Figs. 7 and 8 shows the mismatch considering that flights could arrive at any time within a predicted quantile range. For example, a flight with a scheduled departure time of 10:30 and which 5th and 95th quantiles of the predicted departure delay distribution are -45 and 60 min, respectively, may depart at any time between 9:45 and 11:30, and thus should be considered when computing counts for the windows [9:00,10:00), [10:00,11:00), and [11:00,12:00). Accordingly, depending on the quantile range under consideration, a single flight could be counted in several windows.

Figures 7a and 8a show, respectively, the mismatch when flights depart and arrive as late (or early) as the median of the predicted delay distribution. It should be noted that because the median is a single value rather than a range, flights are counted only in one window. In this situation, Figs. 7a and 8a show that the mean absolute hourly mismatch never exceeds two operations (either departures or arrivals), and, as expected, dates across the whole summer season and on weekends are the most uncertain, particularly for the arrivals.

When considering that flights could depart or arrive at any time between the 25th and 75th quantiles of the predicted delay distribution, the discrepancy between the scheduled number of hourly operations and the potential number of operations begins to increase (see Figs. 7b and 8b). Obviously, the most

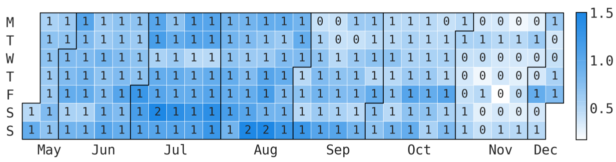


(a) 5th quantile (b) Median (c) 95th quantile

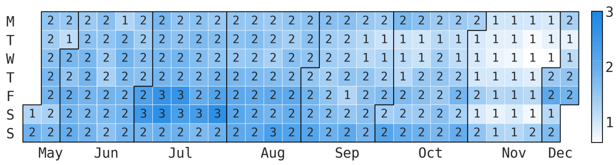
Figure 6: Attribution of the hourly arrivals feature for arrival delay prediction model as a function of its value. It should be noted that the y-axes of the figures are represented with different scales to enable a visual assessment of the trends

extreme difference is observed when considering that flights could depart or arrive at any time between the 5th and 95th quantiles of the predicted delay distribution. In that situation, which is shown in Figs. 7c and 8c, the mean absolute hourly mismatch could be as high as 8 departures and 9 arrivals, respectively.

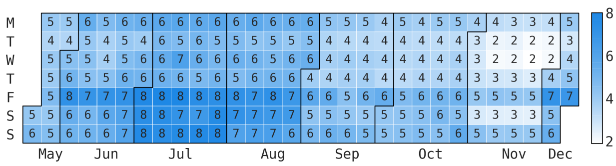
Airport operators could use this simple calendar view to identify dates with potential pitfalls caused by a difference between the number of scheduled and potential operations per hour. It is worth noting that because all of the model's features are accessible during the pre-tactical phase, airports might do this assessment months in advance. Once the most critical days have been found, operators could zoom in and identify the hours with the most (predicted) disparities.



(a) Median



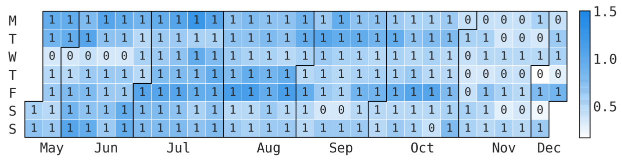
(b) Any time between the 25th and 75th quantiles



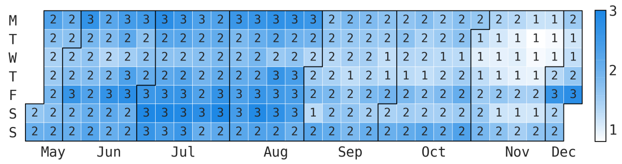
(c) Any time between the 5th and 95th quantiles

Figure 7: Mean absolute hourly mismatch between scheduled and potential number of departures in the test set

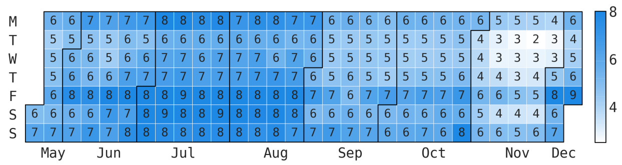
2) Detection of problematic hours: Based on Figs. 7c and 8c, the 30th of July has been selected to illustrate how problematic hours could be identified. This day showed the highest mean absolute hourly mismatch between scheduled and potential number of operations. Figures 9a and 9b show the detailed hourly departures (resp. arrivals), considering that flights depart (resp. arrive) at the scheduled departure (resp. arrival) time, that realise the median delay, and that depart (resp. arrive) at any time between the 25th and 75th quantiles as well as between the 5th and 95th quantiles.



(a) Median



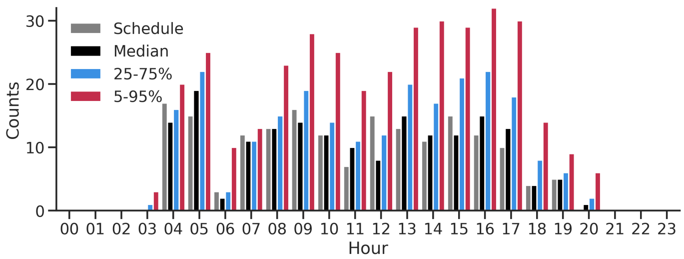
(b) Any time between the 25th and 75th quantiles



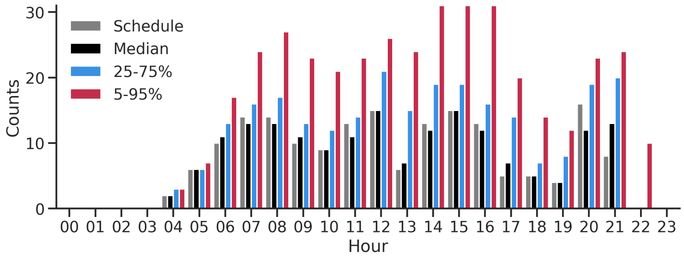
(c) Any time between the 5th and 95th quantiles

Figure 8: Mean absolute hourly mismatch between scheduled and potential number of arrivals in the test set

In Figure 9, the extension of the bar showing the number of probable events included by the 5th and 95th quantile values may be of particular interest (i.e., red bar) when conservative decisions might be taken by the user. Even more interesting could be to detect the most critical periods of a day by looking at the difference between the planned and predicted amount of operations (i.e., difference between the extension of the grey and red bars). However, the user might decide to plan an action without considering any uncertainty. In this case, the count provided by the median predictions (i.e., the black bar) should be adopted. An intermediate approach could be to consider a more likely range of predictions that are provided



(a) Departures



(b) Arrivals

Figure 9: Potential number of departure and arrival operations at GVA during the 30th of July 2022

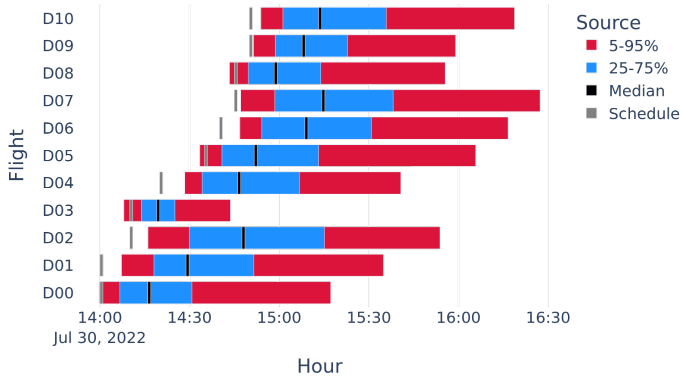
by the 25-75th quantile values (i.e., blue bar).

3) Detection of problematic flights: Once a critical period of the day has been identified, it might be desired to detect which flights require more attention since their arrivals or departures predictions incorporate more uncertainty and present higher mismatched with the scheduled in-block and off-block times. As an example in Figure 10, on a day characterised by high delays such as the 30th of July 2022, flights D10, D09, D07, D06, D04, D02 are very likely to depart later than scheduled since their planned time does not even fall within the red bar (5-95th quantile) which covers 90% of possible occurrences (Figure 10a). On the contrary, it is possible to observe that the mismatch between planned and predicted arrival times is lower since the grey lines fall within or are in close proximity to the blue bars (Figure 10b) for most of the represented flights. An intuitive indication of criticality is the distance between the schedule time (grey lines) and both, the ends of the red bar (5-95th quantile) and the median value (black lines). An undesired scenario is represented indeed by flights scheduled much earlier than these two values. With these ad hoc predictions the assignment of airport resources for each single flight could be more efficient.

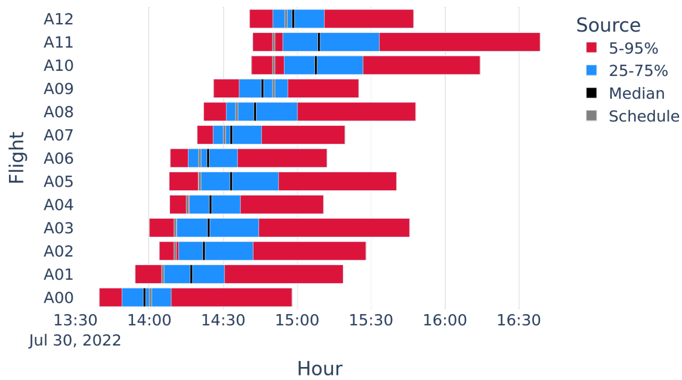
VI. Discussion & Conclusions

Once a machine learning model is trained, specific metrics, such as the average error between predicted and actual realisations of the target variable, can be computed accounting for both aleatory and epistemic uncertainty [20]. However, while these averaged statistics of the error can be used to assess the overall quality of the model, they do not provide a quantification of the uncertainty of a single prediction.

There are various models and methodologies for predicting flight delays in the literature. However, some of them provide punctual predictions, leaving to the user the assessment of



(a) Departures



(b) Arrivals

Figure 10: Predicted quantiles of the flights scheduled to depart or arrive at GVA during the 30th of July 2022 from 14:00 to 15:00

possible deviations from the predicted values that might derive from the complex and uncertain environment in which flights are operated. Other approaches have been suggested in the literature to estimate the uncertainty of the individual predictions, such as sensitivity analyses [21], bootstrapping methods [22], Bayesian methods [23] and Gaussian processes [12]. Most of these methods provide an estimation of the variance of the error but they are not able to provide a range of probabilistic occurrences of predicted values.

In this paper, the uncertainty of individual predictions can be quantified by the quantile values provided by the model's outcomes. As an example, the difference between two (predicted) quantiles, one relatively high and another relatively low (e.g., 95th and 5th, respectively), represents the extension of the time domain over which departures and arrivals are predicted to take place. This quantification allows to assess the *criticality* (or miss-match between plan and prediction) of specific periods of the year at an aggregated level (see Figs. 8 and 7) and, more in detail, of specific days of the year (see Fig. 9) and of individual flights (see Fig. 10)

An important methodology to quantify the contributions of each single input feature to the predictions is the Shapley analysis, which results are shown in Section V. As a main outcome of this analysis, it has been observed that the number of passenger highly affects the predictions. Specifically, the higher the number of passengers (*Pax total* in Figs. 1 and 2'), the higher the arrival and departure delays, showing that particular attention should be paid to operations involving passengers, such as boarding and de-boarding. For a flight showing high positive Shapley values of the total passenger input feature, further analysis and sets of data might allow to identify the operations involving passengers that are more likely to cause delays.

This study has been developed in response to a proposal of the operations performance & forecasting department of Geneva airport (GVA) within one of the EUROCONTROL Air Transport Innovation Network (EATIN) initiative (https://www.eurocontrol.int/project/eatin). As such, it is of particular interest to understand how this probabilistic approach can satisfy the needs of already complex and demanding airport operations. The model is currently under trial at GVA, and in the following months a survey will be conducted to study the impact that the model is making on the planning of the operations at GVA. As a result of the survey, a suitable human-machine interface might be developed and implemented. Alternatively, the schedule arrival and departure values could be replaced in the systems (planning, demand & capacity balance in the land side as well as in the air side, etc.) by the model values. This approach could be extended to other airports and adjusted to serve the needs of any other ATM stakeholder.

In future work, the performance improvement of incorporating weather features into the model, such as visibility, cloud ceiling, or wind speed, could be assessed. However, this variant would only be usable when weather forecasts for the airport are available, which is typically 24 hours before operations. A similar discussion applies to aircraft rotations and ATFM regulations.

References

[1] A. Abu-Rayash and I. Dincer, "Analysis of mobility trends during the covid-19 coronavirus pandemic: Exploring the impacts on global aviation and travel in selected cities," *Energy research & social science*, vol. 68, p. 101693, 2020.

[2] C. Walker, "All-causes delays to air transport in europe - quarter 3 2022," *CODA DIGEST Q3 2022*, pp. 1–15, 2022.

[3] A. Sternberg, J. Soares, D. Carvalho, and E. Ogasawara, "A review on flight delay prediction," *arXiv preprint arXiv:1703.06118*, 2017.

[4] R. Dalmau, G. Murgese, Y. De Wandeler, R. Correia, and A. Marsden, "Early Detection of Night Curfew Infringements by Delay Propagation with Neural Networks," in *14th USA Europe Air Traffic Management Research and Development Seminar*, (Virtual Event), 2021.

[5] D. Sahadevan, P. Ponnusamy, M. K. Nelli, and V. P. Gopi, "Predictability improvement of scheduled flights departure time variation using supervised machine learning," *International Journal of Aviation, Aeronautics, and Aerospace*, vol. 8, 2021.

[6] H.-W. M. Vos, B. F. Santos, and T. Omondi, "Aircraft schedule recovery problem–a dynamic modeling framework for daily operations," *Transportation Research Procedia*, vol. 10, pp. 931–940, 2015.

[7] A. J. Cook and G. Tanner, "European airline delay cost reference values," 2011.

[8] A. Spencer, P. J. Smith, and C. Billings, "Airport resource management and decision aids for airlines," in *2005 International Symposium on Aviation Psychology*, p. 700, 2005.

[9] R. Kicinger, J.-T. Chen, M. Steiner, and J. Pinto, "Airport capacity prediction with explicit consideration of weather forecast uncertainty," *Journal of Air Transportation*, vol. 24, no. 1, pp. 18–28, 2016.

[10] J. J. Rebollo and H. Balakrishnan, "Characterization and prediction of air traffic delays," *Transportation research part C: Emerging technologies*, vol. 44, pp. 231–241, 2014.

[11] G. Gui, F. Liu, J. Sun, J. Yang, Z. Zhou, and D. Zhao, "Flight delay prediction based on aviation big data and machine learning," *IEEE Transactions on Vehicular Technology*, vol. 69, no. 1, pp. 140–150, 2019.

[12] M. Zoutendijk and M. Mitici, "Probabilistic flight delay predictions using machine learning and applications to the flight-to-gate assignment problem," *Aerospace*, vol. 8, no. 6, 2021.

[13] T. Vandal, M. Livingston, C. Piho, and S. Zimmerman, "Prediction and uncertainty quantification of daily airport flight delays," in *Proceedings of The 4th International Conference on Predictive Applications and APIs*, vol. 82 of *Proceedings of Machine Learning Research*, pp. 45–51, October 2017.

[14] Z. Wang, C. Liao, X. Hang, L. Li, D. Delahaye, and M. Hansen, "Distribution prediction of strategic flight delays via machine learning methods," *Sustainability*, vol. 14, no. 22, 2022.

[15] S. J. Moon, J.-J. Jeon, J. S. H. Lee, and Y. Kim, "Learning multiple quantiles with neural networks," *Journal of Computational and Graphical Statistics*, vol. 0, no. 0, pp. 1–11, 2021.

[16] R. E. Schapire, "Explaining adaboost," in *Empirical inference*, pp. 37–52, Springer, 2013.

[17] S. M. Lundberg, G. Erion, H. Chen, A. DeGrave, J. M. Prutkin, B. Nair, R. Katz, J. Himmelfarb, N. Bansal, and S.-I. Lee, "From local explanations to global understanding with explainable ai for trees," *Nature machine intelligence*, vol. 2, pp. 56–67, 2020.

[18] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, and A. Gulin, "Catboost: unbiased boosting with categorical features," in *Proceedings of the 31st Advances in Neural Information Processing Systems (NIPS)*, (Montréal, Canada), 2018.

[19] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[20] Y. Li, J. Chen, and L. Feng, "Dealing with uncertainty: A survey of theories and practices," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 25, pp. 2463–2482, 11 2013.

[21] Z. Bosnić and I. Kononenko, "Estimation of individual prediction reliability using the local sensitivity analysis," *Applied Intelligence*, vol. 29, pp. 187–203, 2008.

[22] T. Heskes, "Practical confidence and prediction intervals," in *Proceedings of the 9th International Conference on Neural Information Processing Systems*, NIPS'96, (Cambridge, MA, USA), p. 176–182, The MIT Press, 1996.

[23] A. Khosravi, S. Nahavandi, D. Creighton, and A. F. Atiya, "Comprehensive review of neural network-based prediction intervals and new advances," *IEEE Transactions on Neural Networks*, vol. 22, no. 9, pp. 1341–1356, 2011.

D Appendix 4

Spak M., Socha, V., Hanakova, L., Matowicki, M.

Predictability of In-block Time Deviations: An Analysis of Operational Data, Tactical Flight Models and Meteorological InformationEnhancement of the diversion airport selection methodology.

Proc. of XVIII. International Scientific Conference New Trends in Aviation Development. The High Tatras – Starý Smokovec: IEEE, 2023. p. 222-227.
ISBN 979-8-3503-7040-9.

Predictability of In-block Time Deviations: An Analysis of Operational Data, Tactical Flight Models and Meteorological Information

Miroslav Spak
Dept. of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
spakmir1@fd.cvut.cz

Vladimir Socha
Dept. of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
sochavla@fd.cvut.cz

Lenka Hanakova
Dept. of Air Transport
Czech Technical University in Prague
Prague, Czech Republic
hanakle1@fd.cvut.cz

Michal Matowicki
Dept. of Applied Mathematics
Czech Technical University in Prague
Prague, Czech Republic
michal.matowicki@cvut.cz

***Abstract*—Efficient airport operations are vital for ensuring the seamless flow of aircraft and passengers. However, real-world airport operations often deviate from meticulously crafted plans, necessitating dynamic resource allocation. These challenges become particularly pronounced during peak operational periods when buffer capacity is limited. To address these issues, we conducted an analysis utilizing operational data from the airport, along with Filed Tactical Flight Model and Current Tactical Flight Model data, as well as meteorological information. Our study primarily focuses on incoming flights at Prague Airport, where we employed regression analysis, forward parameter selection, and a brute-force approach to identify key parameters. These parameters significantly enhance the predictability of in-block time deviations and delay caused in individual phases of the flight. Our model successfully explains up to 96 % of the variability in the data. In summary, this research aims to optimize airport operations by bridging the gap between planned and actual in-block times, ultimately enhancing airport efficiency. The comprehensive analysis of various data elements offers valuable insights for airport management and decision-makers.**

***Index Terms*—prediction, regression, in-block time, delay, airline operations**

I. Introduction

The efficient operation of airports is a critical component of the global aviation network, ensuring the smooth movement of aircraft and passengers. However, the reality of airport operations often deviates from carefully laid out plans, necessitating on-the-fly adjustments to resource allocation. These adjustments become particularly challenging during peak operational periods, where buffer capacity is often nonexistent. This predicament is exacerbated by the limited ability of airports to predict deviations in pre-tactical and tactical block time (gate slots) for long-haul routes, hampering their ability to optimize their resource planning.

The primary reason for this limitation is that substantial insight into flight progress becomes available only once the flight plan is activated. This leaves airports with a very narrow window of opportunity to react, considering the uncertainties associated with the execution of the flight trajectory and the limited radar coverage along long-haul routes. For instance, a significant portion of the trajectory for inbound flights from far east regions, like those approaching e.g. Prague Airport (LKPR), remains devoid of real-time positional reporting for much of the journey. Any deviations resulting from velocity conditions along this trajectory can pose challenges for ground handlers and airline operators, ultimately leading to a deterioration in the level of service provided to airline operations centers (APOC) and their passengers.

As highlighted by previous research [1], [2], flight delays have far-reaching economic and operational implications for airlines, passengers, network connectivity, and, importantly, airports themselves. Consequently, any measures taken to reduce the impact of these delays hold the promise of making air travel more efficient and effective. As suggested, the path forward involves improving flight predictability, and this can be achieved by reinforcing pre-tactical planning using principles guided by the total airport management concept, with APOC and airline operations planners working in tandem [3], [4].

A substantial body of research has already been dedicated to the study of flight predictability and delay analysis. Scholars have employed various prediction techniques, including statistical methods, probability models [5]–[7], and network-based approaches [8]–[10]. Some have ventured into operational methods [11], [12], while others have harnessed the power of machine learning [13]–[17]. This diverse research landscape underscores the significance of improving delay prediction and control, given that flight delays are rarely arbitrary and can be attributed to a multitude of factors, including weather

979-8-3503-7041-6/23/$31.00 ©2023 IEEE

conditions, airport congestion, aircraft mechanical issues, and airline scheduling [18].

Of particular interest is the role of airline scheduling, especially when dealing with flights arriving ahead of or after their scheduled slots. A noteworthy study by Deepudev et al. [19] delved into this matter by using supervised machine learning and data mining techniques to improve the predictability of actual landing times for scheduled flights. The study identified historical data as a valuable resource for predicting arrival delays and employed a multi-linear regression (MLR) model to forecast variations based on departure-time information. The study also shed light on the root causes of early arrivals, with flight duration being a crucial factor in the MLR equation.

Additionally, research has been conducted to evaluate the impact of delays on the turnaround process [20]. Authors found that the scheduled turnaround process is consistently disrupted when aircraft fail to arrive at their designated gates or apron positions on time. To address this issue, they recommended the integration of time buffers during the gate allocation planning phase, as these buffers enhance system reliability. This study highlighted the absence of a systematic buffer concept and the reliance on empirical experiences to initiate buffer allocation. Therefore, it aimed to develop a model that optimizes time buffer size based on expected average delays, ultimately contributing to more efficient airport operations.

Considering these challenges, this paper presents an initial solution and ongoing discussion focused on elevating operational performance through the application of mathematical modeling. The primary goal of this modeling is to improve the predictability of in-block time deviations. Broadly speaking, such approaches hold promise in enhancing visibility for APOC or comparable operational setups, thereby providing crucial support at both pre-tactical and tactical stages, regardless of whether flights are currently underway.

II. Methods

This article is dedicated to the evaluation of factors influencing the predictability of flight delays. To achieve this objective, the analysis focuses primarily on incoming flights and their in-block times at Prague Airport (LKPR) during the period spanning from June 1, 2019, to August 31, 2019. This particular time frame was chosen due to the availability of a comprehensive and intricate dataset, specifically suited for the intended investigation.

A. Data Acquisition and Description

To achieve our research objectives, we obtained three distinct datasets. The first dataset contained operational data, encompassing details such as flight numbers, departure dates, or flight call signs. However, the most pivotal information within this dataset pertained to the actual in-block time (AIBT), signifying the moment when an aircraft came to a halt at its designated stand. This data was provided by the airport.

Another dataset comprised information related to a Filed Tactical Flight Model (FTFM), which included a point-by-point and airspace volume profile generated within the Enhanced Tactical Flow Management System. This profile is generated when flight plan details and any subsequent updates are received from the APOC. These data essentially provide a description of the anticipated flight path prior to take-off. Consequently, they include details about the planned waypoints, identified by their latitude and longitude coordinates, the intended flight altitude at these waypoints, the projected times of arrival at these points, and related information such as planned in-block time.

The Current Tactical Flight Model (CTFM) was also available as a complement to the FTFM. CTFM also containing information about point and airspace volume profile but for a flight which has been activated.

The comparison of FTFM and CTFM flight profiles helps understand deviations between what is filed as the planned trajectory and what is flown during the flight. As mentioned above, the filed trajectory represents the flight plan that the airline representative submits before the flight, detailing the intended route, waypoints, altitudes, and estimated times. On the other hand, the flown trajectory represents the actual path taken by the aircraft during the flight, which may differ due to various factors such as weather conditions, air traffic control instructions, and pilot decisions.

Eurocontrol supplied both the FTFM and CTFM data for the investigated time period.

The last dataset used was meteorological data provided by the Czech Hydrometeorological Institute for the period June 1, 2019, to August 31, 2019. The data covers a geographical region spanning from 120°W to 120°E and from 80°N to 20°S. The grid is divided into increments of 1.25 degrees, following the specifications outlined in ICAO Annex 3. Within this dataset, there are 17 flight levels: 050, 080, 100, 140, 180, 210, 240, 270, 300, 320, 340, 360, 390, 410, 450, 480, and 530.

For each point defined within this coordinate system, wind direction and speed predictions are provided in 3-hour intervals, with a time step of 3 hours. The predictive interval ranges from 6 to 36 hours, and the data was refreshed four times a day at the following times: 00:00, 06:00, 12:00, and 18:00.

B. Data Preprocessing

Datasets, i.e. operational data, FTFM/CTFM and meteorological data, were merged to form a consolidated dataset. Before the actual merging of the data, it was important to clean and prepare the data to ensure that it is consistent and free from errors. This involved removing duplicate records, standardizing data formats, and correcting any missing or incorrect data.

The primary focus was on merging operational data with FTFM/CTFM data, primarily using the departure time as a key criterion. Operational data and FTFM/CTFM data unfortunately did not share the same unique flight identifier. FTFM/CTFM data included a time-dependent flight description, with each flight having multiple waypoints associated with latitude, longitude coordinates, and flight-over times, among other details. By associating operational data with FTFM/CTFM data, variables with a constant nature for a given flight were created. For instance, the Scheduled In-block Time (SIBT) or AIBT remains constant across all waypoints for a single flight. In addition to the mentioned data integration, meteorological conditions were associated with each waypoint, derived from meteorological data. This included information about wind direction and speed. Subsequently, the wind speed and direction were recalculated to represent the wind speed along the direction of flight, where positive values indicated a tailwind and negative values indicated a headwind.

Taking into account the calculation of delay as $\delta = AIBT - SIBT$, this would result in δ being an dependent variable in sequential data. In other words, sequence dependence within one flight (dependent on the number of waypoints described by independent variables) would address a single overall delay. Although it is possible to employ a recurrent neural network for such a problem, we believe that a comprehensive explanation of such modeling is beyond the scope of a conference paper. Therefore, a different type of analysis was chosen, namely regression.

However, to create a linear regression model, it was necessary to redefine the concept of delay. The purpose of this process was to align the time points from two flight datasets, FTFM and CTFM by generating linearly spaced time vectors of same number of samples. This step allowed understanding delay along the flown CTFM trajectory in comparison to planned FTFM trajectory from the moment of flight departure until its arrival. FTFM and CTFM times were compared at same phase of fligth, i.e. if FTFM time stamp was at 20 % of planned flight time then corresponding CTFM time stamp reflected 20 % of actually flown time.

Firstly, time stamps in both FTFM and CTFM dataset were converted from datetime format to Unix time. This conversion ensured that timestamp was presented in seconds as Unix time represents seconds elapsed since January 1, 1970. Therefore, converted vectors $TimeC$ and $TimeF$ representing time stamps in Unix time for CTFM and FTMF, respectively, were created. The total FTFM time was then crucial for $TimeC$ vector resampling. Total FTFM time $TimeF_{total}$ was calculated as:

$$TimeF_{total} = TimeF(N_F) - TimeF(1) + 1 \tag{1}$$

where $TimeF$ is the FTFM time vector (in seconds) and N_F is a total number of $TimeF$ samples. After that, the CTFM time vector was resampled in the way that:

$$TimeC_r(1) = TimeC(1), \text{and} \tag{2}$$

$$TimeC_r(TimeF_{total}) = TimeC(N_C) \tag{3}$$

where $TimeC_r$ has N_F number of samples and represent resampled vector $TimeC$. The N_C is then number of samples in $TimeC$ vector. Also $TimeF$ vector was resampled in the same way, therefore

$$TimeF_r(1) = TimeF(1), \text{and} \tag{4}$$

$$TimeF_r(TimeF_{total}) = TimeC(N_F) \tag{5}$$

represents vector in which each data point represents one second of FTFM flight time (planned flight time). The indexes of vector $TimeF_r$ (idx) of data points that are common to both vectors $TimeF$ and $TimeF_r$ was then identified as

$$\{idx | 1 \leq idx \leq TimeF_{total}, idx \in TimeF, i \in TimeF_r\}. \tag{6}$$

Vector of time differences (FC_{delay}) between CTFM (flown) and corresponding FTFM (planned) time was then calculated and at the same time, delays were converted to minutes. This way FC_{delay} vector was calculated as

$$FC_{delay} = \frac{TimeC_r(idx) – TimeF_r(idx)}{60}. \tag{7}$$

Described process ensured, that FC_{delay} contains time delay for each time stamp in original FTFM dataset.

The final dataset than consisted of 45 features and one dependent variable, i.e FC_{delay}. In total, the dataset included 1 730 flights, and these flights, considering the waypoints contained within them, provided a total of 122 158 observations.

C. Data Analysis

As previously mentioned, the data analysis conducted in this study is primarily focused on regression analysis, wherein an dependent variable represents the delay at all waypoints. The aim of this analysis was to determine whether it is feasible to predict delays or delay increments relative to the scheduled in-block time (SIBT) based on available parameters.

A fundamental assumption underlying this analysis was that delays would naturally be influenced by differences in timing between the anticipated and actual departure from the gate (referred to as LOBT - flight plan off-block time and AOBT - actual off-block time) and the actual takeoff time (ATOT). Naturally, this could logically affect the aircraft's arrival time at its destination.

In any case, the objective was to ascertain whether this relationship is linear, whether it can be generalized, and whether additional features (known and available before or during the flight) could potentially mitigate prediction errors based on a linear model.

For the purpose of this study's data analysis the stepwise regression approach [21] was employed to evaluate the significance of predictors within the regression model. Stepwise regression hinges on specific criteria that gauge enhancements in model fitness, predominantly leveraging statistical measures. The process entails two core phases: forward selection and backward elimination.

In the forward selection phase of stepwise regression, predictors are introduced into the model one at a time. Conversely, the backward elimination phase systematically removes predictors one by one. The basis for adding/removing predictors revolves around the evaluation of their influence on the model's fit. At each step of this process, a p-value is calculated for the F-statistic, comparing the models with and without a candidate term.

This method allows for a step-by-step refinement of the regression model, progressively selecting or excluding predictors

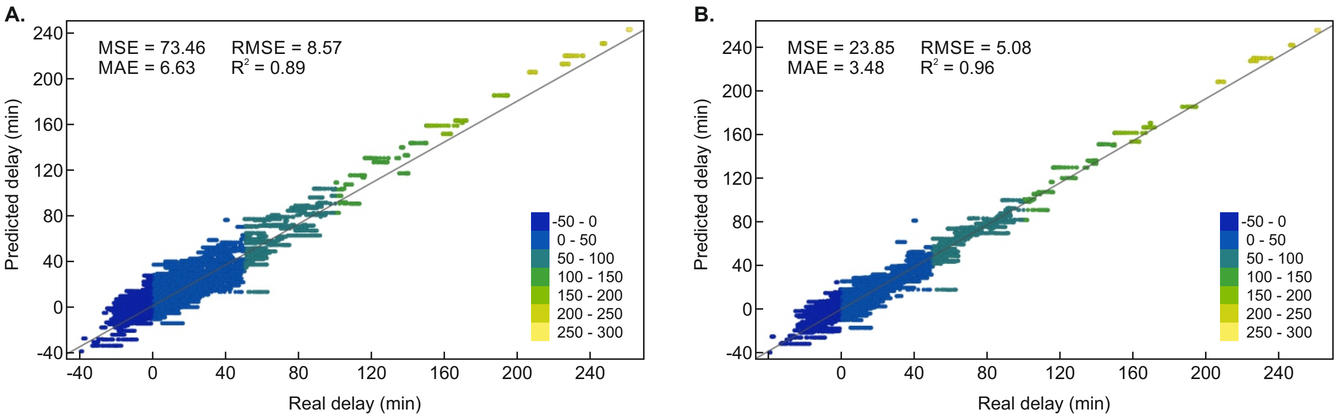


Fig. 1. Relationship between predicted and actual delays for two regression models. Model (A) utilizes the predictors airport of departure, actual take-off time, and flight plan off-block time, while the model (B) incorporates all selected predictors.

to enhance the model's overall accuracy and suitability. The primary objective was to enhance the model's predictive power while maintaining simplicity.

To create the final model based on predictors selected, 70 % of the data was utilized. These data were selected using stratified shuffle split [22]. The remaining 30 % of the data was used for model testing. The process was repeated over 50 iterations, enabling the computation of standard model metrics, which include mean squared error (MSE), root mean squared error (RMSE), mean absolute error (MAE), and R^2 [23]. These metrics are, therefore, the average performance indicators derived from the 50 iterations. This was done to describe the model in a generalized manner, ensuring that the model's performance does not solely depend on a single random selection of test and training data.

III. Results

Forward selection identified parameters, along with their regression coefficients, which are presented in Table I. A more in-depth analysis was then conducted to determine the key parameters responsible for the majority of the variability in the regression model. For this purpose, a brute-force [24] approach was used, involving iterative removal and addition of parameters from the model. This approach revealed that the fundamental model could be defined using LOBT, ATOT, and SIBT. This model is capable of explaining 0.89 % of the variability in the data, with metrics of $MSE = 73.46$ minutes, $RMSE = 8.57$ minutes, and $MAE = 6.63$ minutes. The graphical representation of this model's prediction capability is illustrated in Fig. 1A.

The selected additional parameters modify the model in a way that leads to significantly improved predictability of FC_{delay}. Therefore, the model incorporating all the selected parameters can account for up to 96 % of the variability present in the data, with metrics of $MSE = 25.84$ minutes, $RMSE = 5.08$ minutes, and $MAE = 3.47$ minutes (refer to Fig. 1B).

TABLE I
IDENTIFIED IMPORTANT PARAMETERS ALONG WITH THEIR CORRESPONDING REGRESSION COEFFICIENTS AS THE RESULTS OF FORWARD SELECTION FOR REGRESSION.

Parameter	Coefficient
intercept	72.8089
ADEP=CYUL	-3.8545
ADEP=KEWR	-12.1494
ADEP=KJFK	-17.8268
ADEP=KPHL	-6.50698
ADEP=LHBP	2.97733
ADEP=LSZH	-3.72442
ADEP=OMDB	5.62139
ADEP=OTHH	2.03308
ADEP=RKSI	12.88
ADEP=ZBAA	9.00089
ADEP=ZLXY	7.937
ADEP=ZSPD	4.16353
ADEP=ZUUU	-0.599828
Terminal=T1	-0.650553
Terminal=T2	-1.36129
Terminal=T3	1.40549
Terminal=T4	10.54
Terminal=TC1	-0.960873
Terminal=TC2	-8.70814
FLT_ATOT	0.016525
LOBT	-0.0163966
WindComponent_C	-0.00421778
SIBT	-0.00012843

ADEP - airport of departure; FLT_ATOT - actual take-off time, LOBT - flight plan off-block time; Terminal - the identifier of the aircraft parking stand terminal; SIBT - scheduled in-block time; WindComponent_C - the wind speed recalculated into the flight direction.

IV. Discussion

The study shows that LOBT, ATOT, and SIBT values are important when applying predictive modelling in effort to anticipate in-block time deviations.

Interestingly, these parameters provide some information about the potential for delays, which may be linked to various reasons such as operational patterns of airlines and air navigation services, to airport layout complexity or to a state of weather along the route within given time period.

Selecting airport of departure (ADEP) as a predictive element indicates that distance plays also its role when investing efforts in prediction accuracy. Furthermore, if model works with a gate or terminal information, impact of variable taxing times is also evident and generates an added value as it reduces prediction error.

Syntetic variable called wind component that encapsulates velocity in relation to flight heading was also used to enrich the modelling phase. Although, no signifact improvement was manifested, there is a potential to continue studying impact of the meteorological conditions along the route on flight duration, and in-block time deviations as such.

There are strong assumptions made, that availability of additional information f.e. cost indices or schedule buffering could also shed more light on causes for in-block time deviations. This remains to be a primary objective for continued research in this region.

V. Conclusion

Real-world airport operations often deviate from meticulously crafted plans due to a variety of factors, including weather, congestion, and aircraft issues. These deviations become particularly problematic during peak operational periods when buffer capacity is limited. The issue lies in the limited ability of airports to predict deviations in pre-tactical and tactical block times for long-haul routes. This unpredictability hampers resource planning and operational optimization.

The root cause of this limitation is the delayed availability of flight progress information, particularly for long-haul flights, where real-time positional data is sparse for significant portions of the journey. One intriguing aspect is the impact of delays on the turnaround process. Studies have shown that scheduled turnaround processes are disrupted when aircraft fail to arrive on time at their designated gates. Recommendations include integrating time buffers into gate allocation planning to enhance system reliability. This underscores the need for a systematic approach to buffer allocation based on expected average delays, which can contribute to more efficient airport operations.

In response to these challenges, the paper proposes a predictive model to improve the predictability of in-block time deviations. This model aims to provide crucial pre-tactical and tactical support to operational planning and management. By studying various data elements, the model seeks to anticipate differences between scheduled and actual in-block times.

Our study narrows its focus to incoming flights and their in-block times at Prague Airport during a specific timeframe. The analysis primarily employs regression analysis to determine if it's feasible to predict delays or delay increments relative to the scheduled in-block time based on available parameters linked to meteorological information, trajectory data as well as post operational airport data. In result, the study underscores the significance of LOBT, ATOT, and SIBT as fundamental determinants of flight delay increments. Furthermore, the incorporation of additional parameters has shown the potential to refine and improve predictive accuracy. Although the results are positive, the future research focusing on exploration of advanced modeling techniques such as recurrent neural networks, might be considered.

Acknowledgment

Authors would like to thank Eurocontrol, Czech Hydrometeorological Institute, Prague Airport and Czech Navigation Services for providing the data. This work presents partial results of M. Spak's dissertation entitled "Data driven and machine learning playbook supporting identification of block time deviations and the assessment of their operational impact".

References

[1] C. Barnhart, "Airline scheduling: Accomplishments, opportunities and challenges," in *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, L. Perron and M. Trick, Eds. Springer Berlin Heidelberg, 2008. [Online]. Available: https://doi.org/10.1007/978-3-540-68155-7_1

[2] A. L. B. Bacena, A. M. B. Bihasa, L. J. A. Cadayong, P. M. A. Romulo, and A. B. D. Guzman, "An intergenerational investigation of air passengers' emotions during tarmac delay," *Anatolia*, vol. 31, no. 1, pp. 19–30, Nov. 2019. [Online]. Available: https://doi.org/10.1080/13032917.2019.1684960

[3] G. Spies, F. Piekert, A. Marsden, R. Suikat, C. Meier, and P. Eriksen, "Operational concept for an airport operations center to enable total airport management," in *Proc. of 26th Congress of International Council of the Aeronautical Sciences 2008, ICAS 2008*, vol. 2, 2008, pp. 3242–3251.

[4] O. Delain, "Total airport management," 2019.

[5] Y. Tu, M. O. Ball, and W. Jank, "Estimating flight departure delay distributions - a statistical approach with long-term trend and short-term pattern," *SSRN Electronic Journal*, 2005. [Online]. Available: https://doi.org/10.2139/ssrn.923628

[6] M. Abdel-Aty, C. Lee, Y. Bai, X. Li, and M. Michalak, "Detecting periodic patterns of arrival delay," *Journal of Air Transport Management*, vol. 13, no. 6, pp. 355–361, Nov. 2007. [Online]. Available: https://doi.org/10.1016/j.jairtraman.2007.06.002

[7] A. Evans, A. Schafer, and L. Dray, "Modelling airline network routing and scheduling under airport capacity constraints," in *The 26th Congress of ICAS and 8th AIAA ATIO*. American Institute of Aeronautics and Astronautics, Jun. 2008. [Online]. Available: https://doi.org/10.2514/6.2008-8855

[8] E. Mueller and G. Chatterji, "Analysis of aircraft arrival and departure delay characteristics," in *AIAA's Aircraft Technology, Integration, and Operations (ATIO) 2002 Technical Forum*. American Institute of Aeronautics and Astronautics, Oct. 2002. [Online]. Available: https://doi.org/10.2514/6.2002-5866

[9] S. AhmadBeygi, A. Cohn, Y. Guan, and P. Belobaba, "Analysis of the potential for delay propagation in passenger airline networks," *Journal of Air Transport Management*, vol. 14, no. 5, pp. 221–236, Sep. 2008. [Online]. Available: https://doi.org/10.1016/j.jairtraman.2008.04.010

[10] Y. Wan and S. Roy, "A scalable methodology for evaluating and designing coordinated air-traffic flow management strategies under uncertainty," *IEEE Transactions on Intelligent Transportation Systems*, vol. 9, no. 4, pp. 644–656, Dec. 2008. [Online]. Available: https://doi.org/10.1109/tits.2008.2006813

[11] Y.-J. Liu and S. Ma, "Flight delay and delay propagation analysis based on bayesian network," in *2008 International Symposium on Knowledge Acquisition and Modeling*. IEEE, Dec. 2008. [Online]. Available: https://doi.org/10.1109/kam.2008.70

[12] M. Hansen, "Micro-level analysis of airport delay externalities using deterministic queuing models: a case study," *Journal of Air Transport Management*, vol. 8, no. 2, pp. 73–87, Mar. 2002. [Online]. Available: https://doi.org/10.1016/s0969-6997(01)00045-x

[13] K. F. Abdelghany, S. S. Shah, S. Raina, and A. F. Abdelghany, "A model for projecting flight delays during irregular operation conditions," *Journal of Air Transport Management*, vol. 10, no. 6, pp. 385–394, Nov. 2004. [Online]. Available: https://doi.org/10.1016/j.jairtraman.2004.06.008

[14] V. Venkatesh, A. Arya, P. Agarwal, S. Lakshmi, and S. Balana, "Iterative machine and deep learning approach for aviation delay prediction," in *2017 4th IEEE Uttar Pradesh Section International Conference on Electrical, Computer and Electronics (UPCON)*. IEEE, Oct. 2017. [Online]. Available: https://doi.org/10.1109/upcon.2017.8251111

[15] M. Zanin, Y. Zhu, R. Yan, P. Dong, X. Sun, and S. Wandelt, "Characterization and prediction of air transport delays in china," *Applied Sciences*, vol. 10, no. 18, p. 6165, Sep. 2020. [Online]. Available: https://doi.org/10.3390/app10186165

[16] S. M. Al-Tabbakh, H. M. Mohamed, and Z. H. El, "Machine learning techniques for analysis of egyptian flight delay," *International Journal of Data Mining - Knowledge Management Process*, vol. 8, no. 3, pp. 01–14, May 2018. [Online]. Available: https://doi.org/10.5121/ijdkp.2018.8301

[17] Y. J. Kim, S. Choi, S. Briceno, and D. Mavris, "A deep learning approach to flight delay prediction," in *2016 IEEE/AIAA 35th Digital Avionics Systems Conference (DASC)*. IEEE, Sep. 2016. [Online]. Available: https://doi.org/10.1109/dasc.2016.7778092

[18] M. Arikan, "The impact of airline flight schedules on flight delays: An analysis of block-time, delay propagation, and schedule optimization using stochastic models," Ph.D. dissertation, Purdue University, 2011.

[19] S. Deepudev, P. Palanisamy, V. P. Gopi, and M. K. Nelli, "A machine learning based approach for prediction of actual landing time of scheduled flights," in *Proceedings of International Conference on Recent Trends in Machine Learning, IoT, Smart Cities and Applications*. Springer Singapore, Oct. 2020, pp. 755–766. [Online]. Available: https://doi.org/10.1007/978-981-15-7234-0_72

[20] H. Fricke and M. Schultz, "Delay impacts onto turnaround performance optimal time buffering for minimizing delay propagation," in *Proc. of USA/Europe Air Traffic Management Research and Development Seminar*, 2009. [Online]. Available: https://api.semanticscholar.org/CorpusID:28942704

[21] J. O. Rawlings, S. G. Pantula, and D. A. Dickey, *Applied regression analysis: a research tool*. Springer, 1998.

[22] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.

[23] A. P. Varshini, K. A. Kumari, D. Janani, and S. Soundariya, "Comparative analysis of machine learning and deep learning algorithms for software effort estimation," in *Journal of Physics: Conference Series*, vol. 1767, no. 1. IOP Publishing, 2021, p. 012019.

[24] M. Kuhn and K. Johnson, *Feature engineering and selection: A practical approach for predictive models*. Chapman and Hall/CRC, 2019.