Delays analysis in air transportation with ADS-B large dataset

Ivan Ostroumov[1,*,†], Nataliia Kuzmenko[1,†] and Byambajargal Sukhbaatar[1,2,†]

[1] National Aviation University, Liubomyra Huzara Ave., 1, Kyiv, 03058, Ukraine
[2] MIAT Mongolian Airlines, Ulaanbaatar, 17120, Mongolia

Abstract

Civil aviation is an important part of global transportation system. Numerous advantages of civil aviation stimulate a continuous growth in the amount of services provided by air transport except Covid-19 global lockdown. Civil aviation uses a wide network of flight routes to provide main transportation services. Nominal transportation network operation grounds on well-planned air traffic. However multiple factors may affect the system, which significantly degrade network performance and could reduce the safety of air transportation. Low precision of air traffic or action of rare factors results in delays of provided services. In this paper, we provide a statistical analysis of air traffic delays based on probabilities. We use a kernel probability density function to fit punctuality data and estimate risks to be delayed at a particular time. As input data for delay analysis, we use a database of historical flight trajectories archived by Automatic Dependent Surveillance-Broadcast (ADS-B) technology. Air traffic punctuality depends on many factors, however, in trajectory data, we have only the result of their action with no data about reasons. We propose a specially developed software for automatic delay analysis based on ADS-B large datasets. Considered examples of delay analysis for particular flights.

Keywords

big data, ADS-B, airplane trajectory, transportation system, air navigation

1. Introduction

Aviation uses one of the most complicated transportation networks to provide passenger and cargo transfer over the globe [1, 2]. Airplanes use a flight route network to organize air traffic on the basis of flight safety and efficiency [3, 4]. Each participant of air traffic needs to follow certain rules and directions to minimize mid-air collision and risk of collision with a terrain [5]. Air traffic management is provided by numerous automatic systems and air traffic control teams to make sure the airplanes stay safe and don't bump into each other by guiding them over the whole flight.

CMSE'2024: International Workshop on Computational Methods in Systems Engineering, June 17, 2024, Kyiv, Ukraine

* Corresponding author.
† These authors contributed equally.
ostroumov@ukr.net (I. Ostroumov); nataliiakuzmenko@ukr.net (N. Kuzmenko); byambajargal831@gmail.com (J. Sukhbaatar)
0000-0003-2510-9312 (I. Ostroumov); 0000-0002-1482-601X (N. Kuzmenko); 0009-0004-0032-2920 (J. Sukhbaatar)

Over recent years, the growth of air traffic has been remarkable, reflecting the increasing demand for air travel worldwide. This expansion can be attributed to various factors, including globalization, economic development, and advances in aviation technology [6]. As more people and goods seek efficient transportation options, the aviation industry has experienced a surge in both passenger and cargo traffic. Aviation has shown a clear tendency to grow up to 2020. That has been approved by global number of flights represented in Figure 1. However, alongside the opportunities brought by this growth, there are also challenges, such as congestion at airports and in airspace, which necessitate ongoing innovation and collaboration to ensure safe and efficient air travel for all.

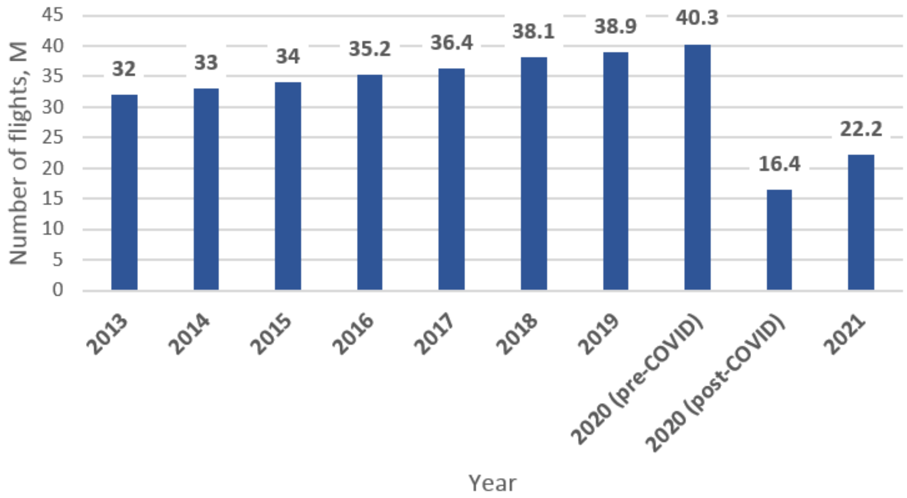

Figure 1: Statistics of number of total flights for the last decade.

The total number of flights experienced consistent growth from 2013 to 2020 (pre-COVID). However, with the onset of the COVID-19 pandemic, there was a notable decline, amounting to a reduction of 23.9 million flights. In 2021, the total number of flights rebounded to 22.2 million [7, 8].

Delay is one of the main parameters of transportation system's effective operation. Delay indicates time actual air traffic is late from planned activities [9, 10]. Thus, delays indicate precision of planning tools and air traffic system resistance to action degradation factors [11]. Airspace volume is limited by existence of flight routes network and system used to meet minimums of safety levels [12, 13]. Thus, further support of continuous growth in air traffic requires significant changes in air space structure and increasing the level of automation in all systems used for air traffic control [14, 15]. A free-route airspace is one of the possible solutions to increase airspace capacity [16, 17]. Development and global implementation of automatic air traffic control systems are other important reasons that limit wide integration of Unmanned aerial vehicles in controlled airspace.

In the paper, we use delay analysis to estimate the probability of an airplane being delayed at a particular time which is an important input data for automatic air traffic control system based on risks. Delays analysis based on probabilities will help to tune transportation system precisely and help improve air traffic prediction.

The organization of the paper is the follows: In the second section, we study reasons for delays and classify factors that have a significant impact on delay forming. In the third section, we describe delay analysis based on probabilities. In the fourth section, a numerical demonstration of developed software for delay analysis is validated with a Large ADS-B dataset of real air traffic data.

2. Delays in air traffic

In recent years, the aviation industry has faced a notable increase in delays due to several factors. The expanding volume of air traffic has led to congestion along flight routes, posing challenges in efficient route planning. Furthermore, limited airspace capacity has strained existing infrastructure, contributing to delays in departure and arrival schedules. The heightened workload on air traffic controllers has necessitated meticulous coordination to ensure the safe and timely flow of flights. Additionally, aging airport infrastructure has struggled to keep pace with the demands of modern aviation, exacerbating delays. If we want to decrease the amount of delays and want to expand the aviation industry, collaboration among aviation stakeholders to enhance airspace capacity through strategic infrastructure upgrades and modernization initiatives is crucial.

In the 2022 edition of the CODA Digest [18], which analyzes all-causes delays and cancellations in European air transport, it was found that the average delay per flight reached a 5-year peak of 17.3 minutes per flight. This represents a notable increase compared to 2021, where the average delay per flight stood at 9.2 minutes (Figure 2). The surge in delays coincided with a significant rise in the number of flights, which increased by 48% compared to the previous year.

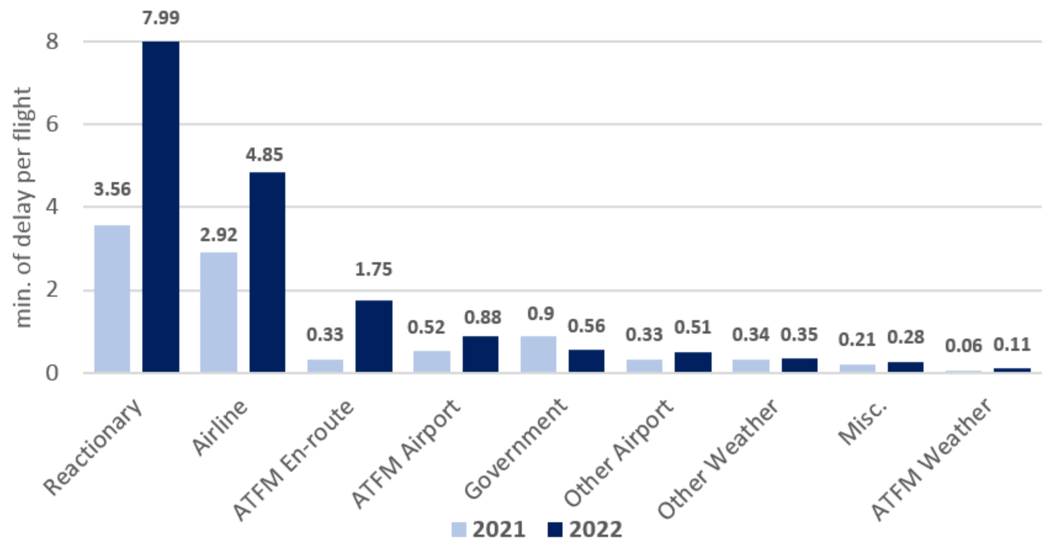

Figure 2: Statistic of airline-reported delays in Europe.

Among the causes attributed to airlines, ground handling and staff shortage-related delays ranked second, accounting for an average delay of 4.9 minutes per flight.

Delays are usually associated with the capacity problems, staff shortage, runway closure, obstruction, night curfew etc.

En-route Air Traffic Flow Management (ATFM) delays showed a sharp escalation, reaching 1.8 minutes per flight. The network experienced a substantial increase in ATFM capacity issues throughout 2022. Additionally, the Ukrainian crisis played a significant role in generating ATFM delays, starting from February 22, 2022, as traffic flows shifted due to the closure of the Ukrainian Flight Information Region [19]. Airport delays are related to labour shortages, parking stands, gate limitations and ramp congestion. Governmental delays are mostly associated to COVID-19 documentation checks, immigration and security.

Understanding the reasons behind flight delays and their various categories is essential for efficient aviation management. Without standardized reporting systems, different systems used by airlines worldwide would lead to confusion and complications. Fortunately, airlines under the International Air Transport Association (IATA) have adopted standardized procedures for reporting flight delays and movement messages electronically, streamlining operations for airline workers, ground handling agents, and airports [20]. The IATA Delay code system comprises nine category sets, described using either a two-digit numeric code or a two-letter alpha code, with numeric formats being more common.

For instance, the delay code 11, associated with late check-in, and code 15, linked to boarding discrepancies and missing passengers at the gate, are the most frequently encountered delays caused by passengers and baggage. This type of delay typically occurs due to various reasons such as passengers arriving late at the airport, long queues at the check-in counters, issues with documentation or baggage handling. For aircraft and ramp handling issues, code 36 relating to fueling or refueling, code 32 relating to loading and unloading, and staff shortages, are commonly observed. Fueling delays, for instance, stem from operational problems like equipment malfunctions or staffing shortages.

Technical issues delays is used for non-scheduled maintenance, entail unplanned aircraft maintenance outside regular schedules, prompted by pre-flight checks or crew reports [21, 22]. These may lead to flight delays or cancellations until maintenance is completed for the safe conduction of flight.

Weather-related delays could be connected with departure or destination station weather, occur due to adverse weather conditions such as thunderstorms, snow, or fog, impacting flight safety and visibility. Air traffic control (ATC) delays are often used for delays related to airport and air traffic control, such as ATC holds and immigration processes. An ATC hold occurs when air traffic control restricts aircraft departure or arrival in specific airspace or at an airport for various reasons. These delays often occur at busy international airports with high air traffic volume, resulting in increased workload, runway demand, and gate capacity constraints. Reactionary delays are used for “Late inbound aircraft," occurs when the aircraft that will be used for the flight arrives late from a previous flight. This can cause a delay in the departure of the subsequent flight, also known as “Aircraft Rotation”.

Regardless of the delay type, the consequences remain consistent, including missed departure slots, extended ground times, potential cascading delays, passenger discomfort, and increased operational costs due to aircraft and crew idle time. Overall, flight delays significantly impact airlines and passengers alike, affecting operational efficiency, customer satisfaction, and financial performance.

3. Delay analysis based on probabilities

Delays and airplane punctuality could be estimated based on trajectory data available via Automatic Dependent Surveillance-Broadcast technology (ADS-B). Each participant in air traffic has to be equipped with a specific transponder of Mode 1090ES to inform all other users and air traffic control authority about airplane identification and exact position. Air traffic management uses ADS-B trajectory data and user's position measured by conventual surveillance equipment like surveillance radars and multilateration systems to provide air traffic control. Also, networks of Software-Defined Radios are used to receive position reports that are transferred over a 1090 MHz data channel. Position of each airspace user is decoded and stored in a global surveillance database. These databases may include trajectories of each airspace user from the day of airplane has been in use.

Big data of ADS-B airplane trajectories could be used for delay analysis in fully automatic mode. Based on rules of surveillance equipment use, on-board ADS-B transponder should be initiated at the moment of starting moving from the gate.

Time of departure (TD) or time of gate left could be obtained by detection of airplane moving initiation. It could be done by calculating the lengths between airplane positions. In case of airplane path is out of two standard deviation errors of primary on-board positioning system the time of gate left could be detected.

Time of take-off (TTO) could be obtained by simple airplane altitude tracking. Point at which airplane starts climbing from runway altitude indicates the time of take-off.

Time of landing (TL) also could be estimated by altitude dataset at the point when altitude became constant to a destination airport runway.

Time of arrival (TA) or time of gate reached is a time when an airplane stops moving at the final leg before ADS-B transponder is shot down.

For general safety reasons publicly available services of ADS-B data sharing limit data set in a time when airplanes are on the ground. Mostly simple filter data by 600 ft altitude is applied to grant permission to the data only when airplanes are in the air.

Air navigation databases provide access to scheduled air traffic data. Scheduled data could be compared to actual and delays could be calculated. Calculation of delays for each scheduled time for historical flights is important and makes it possible to analyze influence of different factors on flight. Result of this analysis identifies the most valuable impact in delay forming and recommendations to airline or air navigation service providers could be issued to minimize delay forming at a particular leg of the flight phase.

Another important step is a statistical analysis of times TD, TTO, TL, and TA. Results of its fluctuations indicate importance to passengers and other users of air transport services. Results of statistical analysis indicate about probability of being delayed in a particular flight connection. These probabilities are important components of risk management in supply chains for cargo delivery.

Risks could be estimated based on a probability density function (PDF). The probability of an airplane being delayed could be estimated as the area below PDF within a particular time frame (Figure 3) [23, 24]. Scheduled time is used as a reference point to get a time frame for a delayed period. Probability of fixing a particular time within defined perils could be estimated as follows:

$$\mathrm{P}[t_a < t < t_b\,] = \int_{t_a}^{t_b} \rho(x)dx. \tag{1}$$

where t_a, t_b are time frame perils for confidence band; $\rho(x)$ is PDF.

The probability of airplane arrival in time could be estimated as a cumulative function:

$$\mathrm{P}[t_s] = \int_{-\infty}^{t_s} \rho(x)dx. \tag{2}$$

where t_s is a scheduled time.

The risk of an airplane being delayed at a particular time could be estimated as follows:

$$\mathrm{R}[d] = \int_{t_s+d}^{+\infty} \rho(x)dx. \tag{3}$$

where d is the time of airplane being delayed.

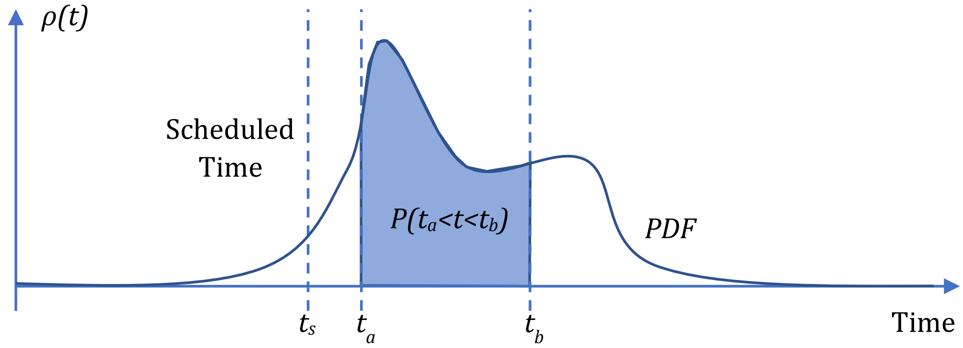

Figure 3: Probability of airplane being delayed.

Simple PDF (Gaussian or Generalized Error) doesn't provide adequate statistical data due to the action of multiple factors which is not Normal. A kernel PDF (KPDF) could provide better fitting in case of a mix of different PDFs [25, 26]. We use a common KPDF with the following model:

$$\rho(t) = \frac{1}{nB^d} \sum_{i=1}^{n} C\left(\frac{t-T_i}{h}\right), \tag{4}$$

$$C(x) = \frac{exp(-x^2/2)}{\int exp(-x^2/2)dx}, \tag{5}$$

where T is a matrix of statistical times, for each flight; i is a flight index; n is the number of flights processed; $C(x)$ is the Gaussian kernel function; B is bandwidth.

Fitting with KPDF required to use effective bandwidth B which could be calculated by minimum of Mean Square Error:

$$B = \left(\frac{4\mu}{n\sigma^4 \int L^2 p(x)^2 dx}\right)^{\frac{1}{d+4}}, \quad (6)$$

where μ is a mean value; σ is a standard deviation; L is a Laplace function.

A special software for delay analysis has been developed for use in MATLAB environment. A structure scheme of software is given in Figure 4.

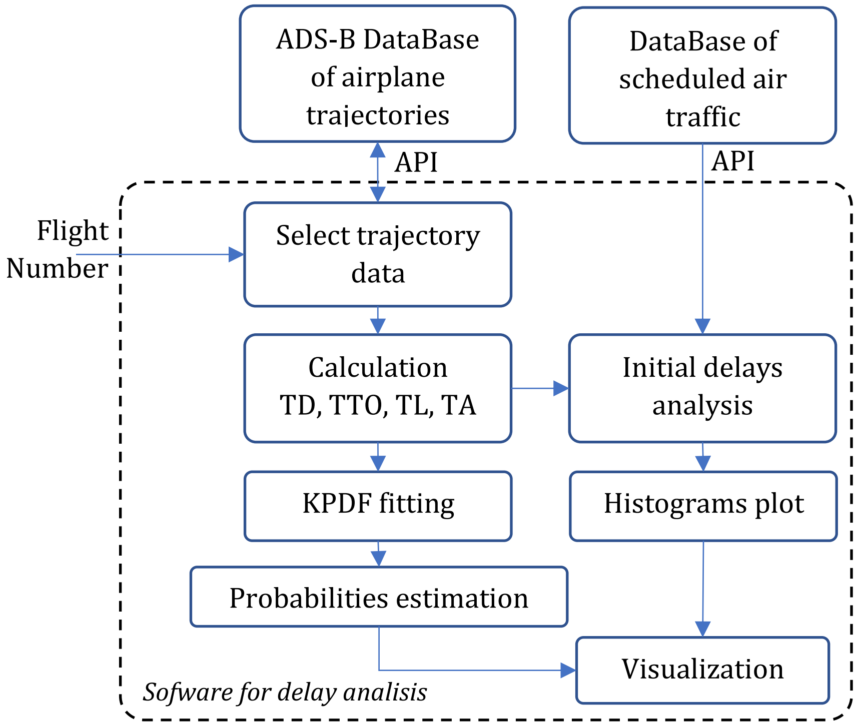

Figure 4: Structure scheme of software for delay analysis.

As input software requires only a particular flight number to be studied. Based on flight number, software requests available trajectory data from ADS-B large-scale database with API. Airplane trajectories are replied in JSON data format which is easily integrated into internal variables of software. Then for each airplane trajectory values of TD, TTO, TL, and TA are calculated. Calculated differences between actual and scheduled times gives delays. Delay values are accumulated in arrays by each flight number and processed. Results of delays frequency distribution estimation are visualized with a histogram plot. A KPDF is applied for each set of times TD, TTO, TL, and TA. Probabilities of airspace users being delayed are estimated by cumulated PDF (2). As a result, software prepares a set of histograms with a result of KPDFs fitting which could be saved in a separate folder titled with flight number.

4. Numerical demonstration

In numerical demonstration, we use developed software to provide delay analysis of flight OM137 (MGL137). Flight OM137 connects international airports of Sergelen Sum (ZMCK, Mongolia) and Frankfurt am Main (EDDF, Germany) operated by airline “Mongol Air”. Airplanes B 787 and B 763 have been used to serve this flight. We use ADS-B trajectory dataset obtained by publicly available API [27]. In the study, we use only recent data for period January-May 2024. This period gives 53 flight realizations. Flight OM137 is a scheduled flight that is repeated on a timetable basis. Therefore, scheduled times of TD, TTO, TL, and TA are still mostly constant excapt summer-winter timetable changes. Duration of ground operations from gate departure to take-off for ZMCK is 10 min. Also, taxing in EDDF from landing to the gate is planned to be 10 min. The scheduled time of ZMCK gate left is 2:30 (2:00 summer time) UTC and scheduled time for EDDF gate arrival is 12:10 (11:00 summer time) UTC. We use calculated delays as the difference between scheduled and actual times for TD, TTO, TL, and TA. Results of statistical analysis in delays for TD, TTO, TL, and TA are presented in Figure 5– 8.

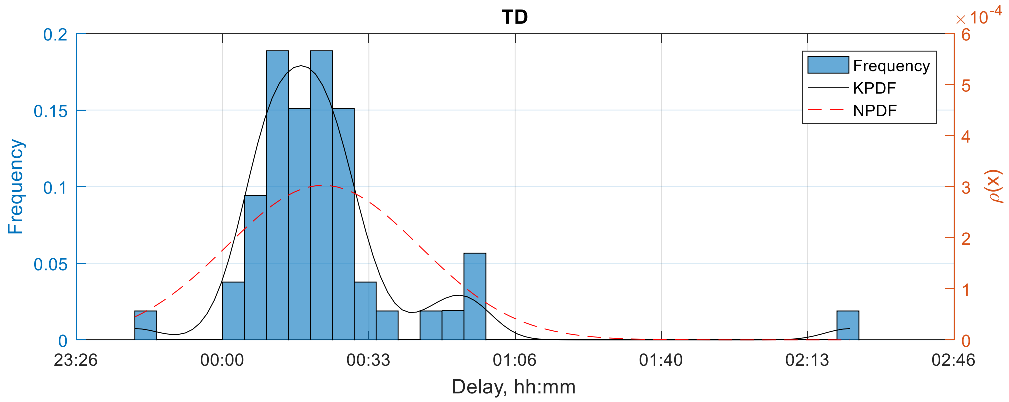

Figure 5: Delay for time of departure.

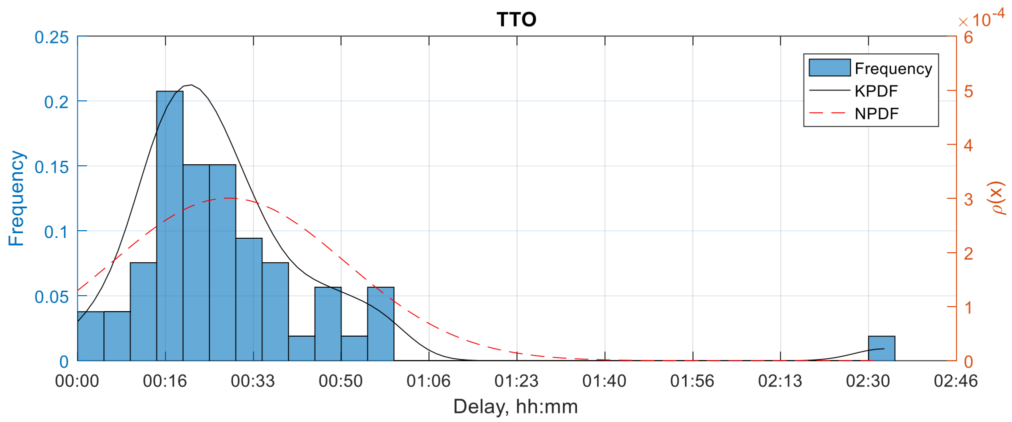

Figure 6: Delay for time of take-off.

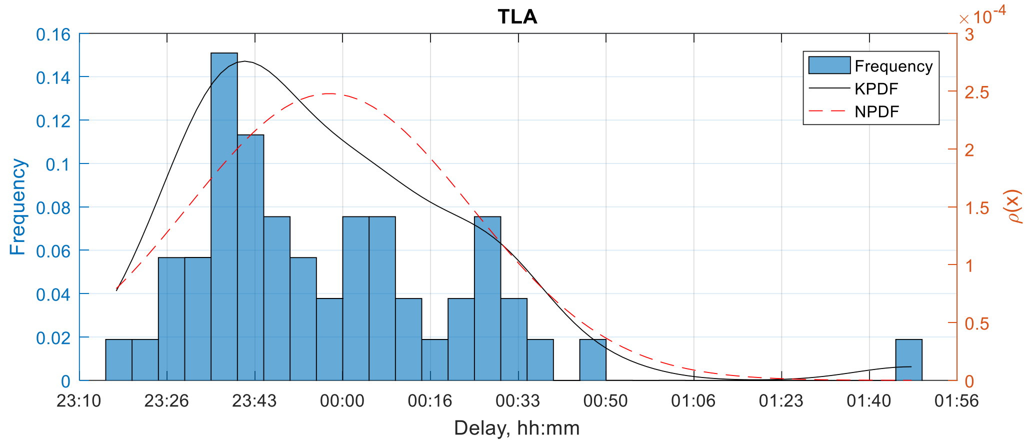

Figure 7: Delay for time of Landing.

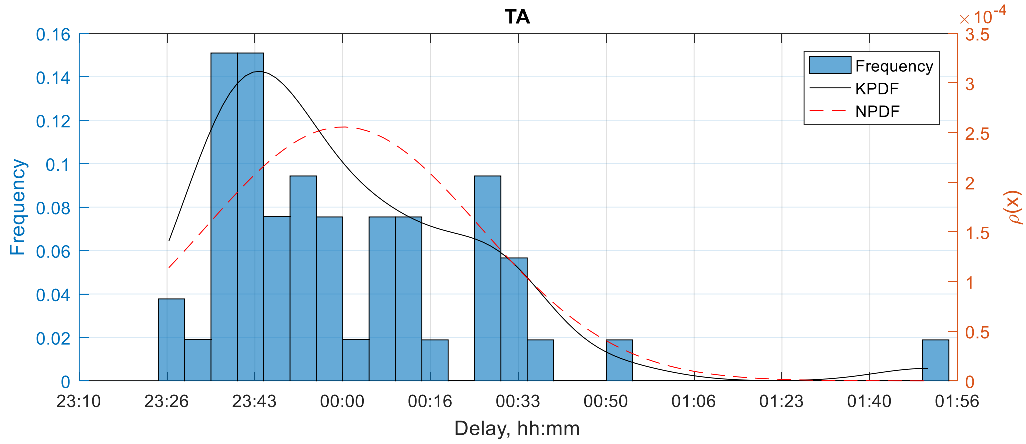

Figure 8: Delay for time of arrival.

Results of delay analysis for OM137 flight are given in Table 1. We calculate parameters of mean and standard deviation of delay for each considered time TD, TTO, TL, and TA. Also, we estimate probabilities of airplanes being early or at least in scheduled time based on (2) for both models NPDF (P_{NPDF}) and KPDF (P_{KPDF}).

Table 1
Results of delay analysis for OM 137 flight

Time	Mean	Standard deviation	P_{NPDF}	P_{KPDF}
Time of departure (TD)	00:22:54	00:21:58	14.86%	4.57%
Time of take-off (TTO)	00:28:39	00:22:06	9.75%	2.10%
Time of Landing (TL)	-0:02:30	00:26:49	53.73%	59.30%
Time of arrival (TA)	00:00:03	00:26:00	49.91%	57.62%

Results in Table 1 show the difference in models. Probability P_{NPDF} is very different from P_{KPDF}. Thus, the simplified model with NPDF gives values with less precision.

Results of cumulated PDF (CPDF) estimation by (2) for KPDF model are shown in Figure 9. CPDF gives a probability of an airplane being delayed at a particular duration.

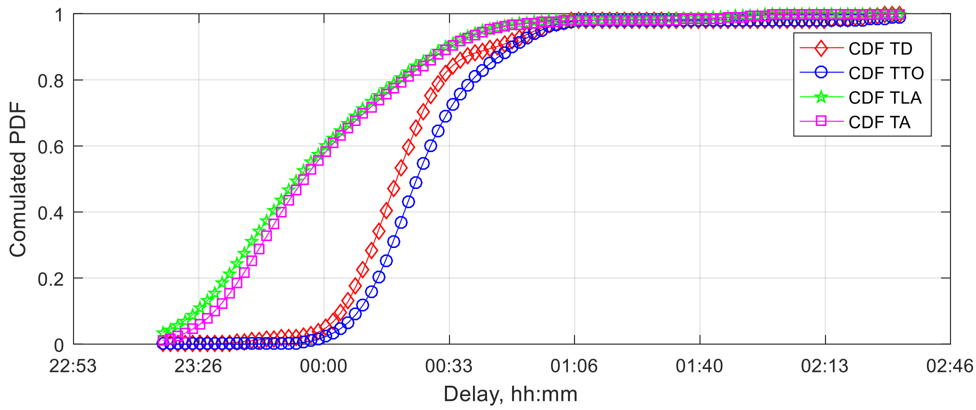

Figure 9: CDF for considered times.

Obtained results indicate that for particular flight connection probability of airplane being delayed in the departure airport (TD and TTO) is high enough (the probability of being delayed in 30 min is more than 70%). However, during the flight airplane corrects its total delayed time to guarantee arrival at the destination airport in scheduled time with a probability of no less than 60%.

5. Conclusions

Delay analysis is an important task performed by each airline to increase performance of provided services. Delay analysis identifies the main reasons for delays and helps to concentrate airline activity to minimize degradation factors action. Usually, delays are associated with direct additional costs for airlines spent on the operation of a particular flight. Thus, delay minimization is one of the primary tasks of airlines. Public available airplane trajectories archive via ADS-B database opens a new possibility for delays analysis. Long-term archives of particular flight connections help to identify the main times for delay analysis: time of gate left, take-off, landing times, and time of arrival at the gate of destination airport.

A KPDF is useful for simply fitting the statistical data for delay analysis. Normal PDF does not fit well due to multiple degradation factors action with different natures. Fitted KPDF makes it easy to estimate probabilities of airplanes being delayed, which are important components in supply chain optimization based on risk models.

Acknowledgements

This project has received funding through the EURIZON project, which is funded by the European Union under grant agreement No. 871072 (Project EU #3035 EURIZON “Research

and development of Ukrainian ground network of navigational aids for increasing the safety of civil aviation”).

References

[1] K. Button, S. Taylor, International air transportation and economic development, Journal of air transport management 6(4) (2000) 209–222. URL: https://hdl.handle.net/10419/114969.

[2] J. Wensveen, Air Transportation: A Global Management Perspective. Routledge. 9th. ed., Taylor, London, 2023. doi: 10.4324/9780429346156.

[3] K. Cherednichenko, V. Ivannikova, O. Sokolova, M. Yanchuk, Model of transport safety assessment in multimodal transportation systems, Transport 38(4) (2023) 204–213. doi: 10.3846/transport.2023.20865.

[4] Air traffic management, Procedures for Air Navigation Services, Doc. 4444, ICAO, 2016.

[5] I. Ostroumov, N. Kuzmenko, Risk Assessment of Mid-air Collision Based on Positioning Performance by Navigational Aids, in: Proceedings of 6th International Conference on Methods and Systems of Navigation and Motion Control (MSNMC), Kyiv, Ukraine, 2020, pp. 34–37. doi: 10.1109/MSNMC50359.2020.9255506.

[6] S. Becken, F. Carmignani, Are the current expectations for growing air travel demand realistic?, Annals of Tourism Research 80 (2020) 102840. doi: 10.1016/j.annals.2019.102840.

[7] X. Sun, S. Wandelt, C. Zheng, A. Zhang, COVID-19 pandemic and air transportation: Successfully navigating the paper hurricane, Journal of Air Transport Management 94 (2021) 102062. doi: 10.1016/j.jairtraman.2021.102062.

[8] G. Andreana, A. Gualini, G. Martini, F. Porta, D. Scotti, The disruptive impact of COVID-19 on air transportation: An ITS econometric analysis, Research in Transportation Economics 90 (2021). 101042. doi: 10.1016/j.retrec.2021.101042.

[9] Q. Li, R. Jing, Characterization of delay propagation in the air traffic network, Journal of Air Transport Management 94 (2021) 102075. doi: 10.1016/j.jairtraman.2021.102075.

[10] I. Ostroumov, N. Kuzmenko, O. Sushchenko, M. Zaliskyi, O. Solomentsev, Y. Averyanova, S. Zhyla, V. Pavlikov, T. Nikitina, A probability estimation of aircraft departures and arrivals delays, in O. Gervasi, et al. (Eds.), Computational Science and Its Applications ICCSA 2021, volume 12950 of Lecture Notes in Computer Science, Springer, Cham, 2021, pp. 363–377. doi:10.1007/978-3-030-86960-1_26.

[11] M. Zaliskyi, R. Odarchenko, S. Gnatyuk, Y. Petrova, A. Chaplits, Method of traffic monitoring for DDoS attacks detection in E-health systems and networks, CEUR Workshop Proceedings, 2255 (2018) 193–204.

[12] Q. Li, R. Jing, Characterization of delay propagation in the air traffic network, Journal of Air Transport Management 94 (2021) 102075. doi: 10.1016/j.jairtraman.2021.102075.

[13] M. Zanin, S. Belkoura, Y. Zhu, Network analysis of Chinese air transport delay propagation, Chinese Journal of Aeronautics 30(2) (2017) 491–499. doi: 10.1016/j.cja.2017.01.012.

[14] I.V. Ostroumov, N.S. Kuzmenko, Compatibility analysis of multi signal processing in APNT with current navigation infrastructure, Telecommunications and Radio Engineering 77(3) (2018) 211–223. doi: 10.1615/TelecomRadEng.v77.i3.30.
[15] I.V. Ostroumov, N.S. Kuzmenko, Accuracy assessment of aircraft positioning by multiple Radio Navigational aids, Telecommunications and Radio Engineering 77(8) (2018) 705–715. doi: 10.1615/TelecomRadEng.v77.i8.40.
[16] H. Hirabayashi, M. Brown, N. Takeichi, Feasibility study of free route airspace over the north pacific. Journal of Air Transportation 30(2) (2022) 59–70. doi: 10.2514/1.D029.
[17] Y. Gencoglu, B. Baspinar, Free Routing in High-Complexity Airspace: Impact Assessment for Environment and Aviation Stakeholders, Journal of Aerospace Information Systems 20(10) (2023) 627–635. doi: 10.2514/1.I011203.
[18] CODA digest. All-Causes Delay and Cancellations to Air Transport in Europe, Eurocontrol (2022).
[19] V. Ivannikova, O. Sokolova, K. Cherednichenko, How the War in Ukraine Impacts Global Air Transportation Ecosystem: Assessment and Forecasting of Consequences. in: O. Prentkovskis, I. Yatskiv (Jackiva), P. Skackauskas, M. Karpenko, M. Stosiak, (Eds.), TRANSBALTICA XIV: Transportation Science and Technology, Lecture Notes in Intelligent Transportation and Infrastructure. Springer, Cham, 2024, pp. 386–401. doi:10.1007/978-3-031-52652-7_38.
[20] Airport Handling Manual (AHM). IATA. (2023).
[21] O. Solomentsev, M. Zaliskyi, O. Zuiev, Radioelectronic equipment availability factor models, in: Proceedings of Signal Processing Symposium 2013 (SPS 2013), Serock, Poland, 2013, pp. 1–4. doi: 10.1109/SPS.2013.6623616.
[22] O.C. Okoro, M. Zaliskyi, S. Dmytriiev, O. Solomentsev, O. Sribna, Optimization of maintenance task interval of aircraft systems, International Journal of Computer Network and Information Security (IJCNIS) 14 (2) (2022) 77–89. doi: 10.5815/ijcnis.2022.02.07.
[23] I. V. Ostroumov, N. S. Kuzmenko, Risk Analysis of Positioning by Navigational Aids, in: Proceedings of Signal Processing Symposium (SPSympo), Krakow, Poland, 2019, pp. 92–95. doi: 10.1109/SPS.2019.8882003.
[24] O. Solomentsev, M. Zaliskyi, Yu. Nemyrovets, M. Asanov, Signal processing in case of radio equipment technical state deterioration, in: Proceedings of Signal Processing Symposium 2015 (SPS 2015), Debe, Poland, 2015, pp. 1–5. doi: 10.1109/SPS.2015.7168312.
[25] C. Yen-Chi, A tutorial on kernel density estimation and recent advances. Biostatistics & Epidemiology 1(1) (2017), 161–187. doi: 10.48550/arXiv.1704.03924.
[26] X. Zhixiao, J. Yan, Kernel density estimation of traffic accidents in a network space, Computers, environment and urban systems 32(5) (2008) 396–406. doi: 10.1016/j.compenvurbsys.2008.05.001.
[27] AeroAPI. Flightaware. 2024. URL: https://www.flightaware.com/commercial/aeroapi.