

UC Berkeley

UC Berkeley Electronic Theses and Dissertations

Title

Air to Ground: Analysis and Evaluation on Air Traffic Management Performance

Permalink

<https://escholarship.org/uc/item/21t5p1kt>

Author

Liu, Ke

Publication Date

2024

Peer reviewed|Thesis/dissertation

Air to Ground:
Analysis and Evaluation on Air Traffic Management Performance

by
Ke Liu

A dissertation submitted in partial satisfaction of the
requirements for the degree of
Doctor of Philosophy
in
Engineering – Civil and Environmental Engineering
in the
Graduate Division
of the
University of California, Berkeley

Committee in charge:

Professor Mark Hansen, Chair
Professor Michael Cassidy
Professor Sandrine Dudoit

Summer 2024

Air to Ground:
Analysis and Evaluation on Air Traffic Management Performance

Copyright 2024
by
Ke Liu

Abstract

Air to Ground:

Analysis and Evaluation on Air Traffic Management Performance

By

Ke Liu

Doctor of Philosophy in Engineering – Civil and Environmental Engineering

University of California, Berkeley

Professor Mark Hansen, Chair

This thesis develops methodologies to evaluate the performance of various components of the National Airspace System (NAS) that are related to air traffic management (ATM). The NAS is a multifaceted network comprising aircraft, airports, air traffic managers and controllers, passengers, and navigational aids. Air traffic management is, in broadest terms, the process of shifting flight schedules to reduce demand capacity imbalances that must to be handled by tactical control and aircraft maneuvers. ATM performance is influenced by a mix of external weather uncertainties and internal operational errors, making the system's management unpredictable and complex. As U.S. air traffic is expected to grow significantly, further complicated by escalating weather uncertainties and operational challenges, there is an urgent need for a comprehensive performance analysis of ATM in the NAS. Such analysis is essential to identify key performance indicators, define metrics for trend assessments, and uncover underlying patterns that can inform future improvements. This thesis is structured into three parts, distributed over four chapters, each focusing on different aspects of NAS operations.

Part I-Chapter 1 focuses on analyzing Actual Airborne Time (AAT) between the U.S. and China. A major goal of ATM is to shift delay in the air to delay on the ground. Since Chinese ATM is in its nascent stages, a comparison of AAT in the two nations sheds light on how more advanced ATM can reduce airborne times. This chapter conducts the first comparative empirical study on AAT, examining how variables such as origin-destination distance, hub status, traffic conditions, and weather influence flight times in both countries. Econometric models reveal that AAT is strongly correlated with distance, with flights in China experiencing longer times under similar distances, suggesting inefficiencies in Chinese airspace management. The chapter also explores the potential for efficiency gains if Chinese airspace operations were aligned with U.S. standards, projecting significant reductions in AAT and associated cost savings.

Part II-Chapter 2 addresses the supply side by investigating runway and terminal airspace capacities, crucial bottlenecks in the NAS. Focusing on JFK Airport, this chapter presents a statistical model of controlled interarrival times (CIT) at landing thresholds, incorporating factors such as weather, runway utilization, and aircraft characteristics. A novel conditional probability model is introduced, capturing CIT behaviors under varying arrival demands. This model aids in understanding and enhancing landing safety and efficiency by depicting the probability distributions of CIT under different operational scenarios.

Part III evaluates the impact of TMIs designed to balance flight demand and airspace capacity across different components of the NAS. **Chapter 3** evaluates Miles-in-Trail (MIT), a common TMI that regulates spacing between aircraft to manage capacity-demand imbalances. A case study at ORD Airport illustrates that eliminating MIT could reduce overall queuing delays but increase arrival delays, indicating a trade-off between local congestion and system-wide throughput. **Chapter 4** delves into the efficiency of Ground Delay Programs (GDPs), which are implemented to address demand-capacity imbalances at arrival airports by converting potential airborne delays into pre-departure ground delays. This chapter introduces a deterministic queuing model to simulate scenarios absent of GDPs, focusing on excess delay and wasted slots. It applies the methodology to 1204 GDPs from 33 airports in 2019. The results indicate that removing GDPs could reduce excess delays by approximately 70%, albeit at the cost of increased airborne holding. The primary inefficiencies in GDPs stem from inaccurate program rate settings and flight time deviations, suggesting that significant improvements in GDP efficiency could be achieved by refining these parameters.

Dedicated to my grandma *Shaolan* and my parents *Wenjie* and *Hongmei*

Table of Contents

TABLE OF CONTENTS	II
LIST OF FIGURES	IV
LIST OF TABLES	VI
ACKNOWLEDGMENTS	VII
INTRODUCTION	1
PART I. EVALUATING AIRBORNE TIME AND FLIGHT DELAY	5
CHAPTER 1 COMPARATIVE STUDY ON AIRBORNE FLIGHT TIMES: INSIGHTS FROM THE US AND CHINESE SYSTEMS	5
1. INTRODUCTION	5
2. MODEL DEVELOPMENT	8
2.1 Operation-related variables	8
2.2 Weather-related variables	12
2.3 Time-related variables	14
2.4 Fixed effects	14
3. DATA	16
3.1 Data sources	16
3.2 Data filtering	17
4. ESTIMATION RESULTS	18
5. SENSITIVITY ANALYSIS	25
5.1 Sensitivity to flight length	25
5.2 Sensitivity to daily flight leg	27
6. COUNTERFACTUAL ANALYSIS OF POTENTIAL EFFICIENCY GAINS IN CHINA	30
6.1 Methodology	31
6.2 Results	31
7. CONCLUSION	35
PART II. EVALUATING TERMINAL AIR TRAFFIC CONTROL AND ARRIVAL SPACING	37
CHAPTER 2 STATISTICAL MODELING OF FLIGHT INTER-ARRIVAL TIMES	37
1. INTRODUCTION	37
1.1 Arrival separation definition	37
2.1 Related study	38
2. DATA AND STUDY AREA: JFK	40
2.1 Data Source	40
2.2 JFK airport	40
3. METHODOLOGY	42
3.1 Mathematic model framework	42
3.2 Feature engineering	45
4. MODEL RESULT: CIT	47
5. CONCLUSION	50
PART III. ENHANCING AIR TRAFFIC MANAGEMENT EFFICIENCY	51

CHAPTER 3: MILES-IN-TRAIL RESTRICTIONS AND AVIATION SYSTEM PERFORMANCE: CHICAGO O'HARE CASE STUDY	52
1. INTRODUCTION	53
1.2 Literature Review	53
1.3 Research question and contributions	54
2. METHODOLOGY	55
2.1 Design of queueing diagrams	56
2.2 Construction of model planned arrival time without MITs	58
2.3 Reallocation models	63
3. CASE STUDY: ORD AIRPORT	64
3.1 Data sources	64
3.2 ORD airport and relevant statistics	65
4. RESULTS AND DISCUSSION	69
4.1 Reallocation results for overall performance	69
4.2 Evaluation on MIT efficiency	70
5. CONCLUSION	74
CHAPTER 4 EXCESS DELAY AND WASTED SLOTS FROM THE GROUND DELAY PROGRAM: MEASUREMENT AND CAUSAL ANALYSIS	75
1. INTRODUCTION	75
2. BACKGROUND	76
2.1 Literature review	76
2.2 GDP process	78
2.3 Improvement opportunities in GDP	80
3. METHODOLOGY	81
3.1 Data description and preparation	81
3.2 GDP flight detection	82
3.3 Counterfactual queueing model	83
3.4 Causal analysis for GDP excess delay	87
4. RESULTS	91
4.1 Summary GDP statistics	91
4.2 GDP excess delay and wasted slots	93
4.3 CAUSAL ANALYSIS ON EXCESS DELAY: STATISTICAL MODELS	97
4.3 Causal analysis on excess delay: factor contribution	102
5 CONCLUSION	103
BIBLIOGRAPHY	105
APPENDIX 1: TOP 30 AIRPORTS IN CHINA AND THE U.S. WITH COMMERCIAL FLIGHTS	110
APPENDIX 2: EXAMPLE OF TRAFFIC DENSITY AT A SPECIFIC TIME IN CHINA AND THE U.S.	111
APPENDIX 3: MAJOR HUB AIRPORTS IN CHINA AND THE U.S.	112
APPENDIX 4: FLIGHT SHARE OF AIRLINE AND AIRCRAFT TYPES OF CORE 30 AIRPORTS IN CHINA & THE U.S.	113
APPENDIX 5: COMPONENTS OF DIRECT OPERATING COST FOR MAJOR AND MINOR CHINESE AIRLINES	114

List of Figures

Figure 1. Workflow overview.....	1
Figure 2. Time components related to a flight.....	5
Figure 3. Air routes among top 30 airports in China and the U.S., ...	7
Figure 4. Example of queueing diagram for demand and actual flight counts	10
Figure 5. Layout of grid across the truncated national airspace	10
Figure 6. Processing flight for the traffic density matrix.....	11
Figure 7. Flight GCD route with its enroute nearby airports within 100 miles (APT: airport).....	12
Figure 8. Distribution of weather stations in China (left) and the U.S.(right) (2016)	13
Figure 9. 2D grid interpolated convective weather grid examples	13
Figure 10. Coefficients with 95% CI for GCD-aircraft-specific variables for China and the U.S.	21
Figure 11. Coefficients with 95% CI for airline-specific variables for China and the U.S.	22
Figure 12. Coefficients with 95% CI for airport-specific variable for China and the U.S.	23
Figure 13. Flight distribution by flight length in China and the U.S.	25
Figure 14. Delay recovery effect variation with different flight length.....	26
Figure 15. Distribution of the number of legs flown by aircraft.....	28
Figure 16. Delay recovery effect variation with NumFltOfAC.....	29
Figure 17. Distribution of AAT change among Chinese flights if AAT behavior in the U.S. were adopted.....	32
Figure 18. Direct operating cost saving for the top 10 airlines in China	34
Figure 19. AAT change among Chinese routes if the AAT behavior in the U.S. were adopted.....	34
Figure 20. In-trail aircrafts along the final approach segment.....	37
Figure 21. Trajectories of Flights Arriving at JFK on 1/1-1/10/2019	41
Figure 22. Timeline for 3 consecutive arrival flights	43
Figure 23. Overall approach to assessing the MIT impacts.....	56
Figure 24 Main decomposition of air travel time	56
Figure 25 Queueing diagram for the actual scenario	57
Figure 26 Queueing diagram for the model scenario without MITs	58
Figure 27 Illustration for Case 1 model planned arrival time.....	59
Figure 28 Illustration for Case 1 model planned arrival time.....	59
Figure 29 Illustration for Case 2 model planned arrival time.....	60
Figure 30 Illustration for Case 3 model planned arrival time.....	60
Figure 31 Bar chart (left) and box chart (right) of TMI behaviors (2018)	65

Figure 32 Annual Counts on MITs based on Restricted Airports for 2018.....	66
Figure 33 Usage frequency of MITs on daily base (left) and quarter-hour base (right).....	67
Figure 34 Histogram of ORD-sourced MIT Spacing (left) and duration (right) (2018)	67
Figure 35. Frequency of direct MIT issued by ORD for surrounding FIX/waypoint (2018).....	68
Figure 36. Change in quarter-hour throughputs for Case 3 under flight-level reallocation	71
Figure 37. Annual counts of throughput change for each quarter hour (2018)	71
Figure 38. Change in flight delay on the day base for Case 3 under flight-level reallocation algorithm	72
Figure 39. Scatter plot of individual MIT performance in terms of change in arrival & queueing delays	73
Figure 40. Distribution of β value for individual MITs (13104 MITs in total)	73
Figure 41. Timeline of GDP planning and execution phases	78
Figure 42. Example of GDP process at EWR on Jan. 12 th , 2018	80
Figure 43. GDP Delay for In-Effect Flight (IEF)	83
Figure 44. GDP Delay for Post-Cancellation Flight (PCF)	83
Figure 45. Flight timings for the actual and counterfactual scenarios	84
Figure 46. Queueing diagram for actual and counterfactual (model) scenarios	86
Figure 47. GDP frequency, duration and causes by airport (2019)	92
Figure 48. GDP affected flights and assigned ground delay by airports (2019)	93
Figure 49 Scenario-based counterfactual queueing model: Example GDP at ORD on Aug 26 th , 2019	94
Figure 50. Excess delay by airport.....	96
Figure 51. Wasted slots and averted holding flights by airport	97
Figure 52. Relative factor contribution in reduction of average GDP excess delay.....	103

List of Tables

Table 1. Description of variables	15
Table 2. Summary statistics of the variables in the specified models after data cleaning (2016)	17
Table 3. Base models for AAT behavior in China and the U.S.....	24
Table 4. Estimation results of AAT for different flight length for China and the U.S.....	26
Table 5. Estimation results of AAT in different NumFltOfAC of China and the U.S.	29
Table 6. Estimates of time, fuel, fuel cost, and direct operating cost saving for Chinese airlines	33
Table 7. IFR- Approach in-trail wake turbulence separations minima (in miles)	38
Table 8. JFK Runway characteristics	41
Table 9. JFK arrival runway configurations (2019)	42
Table 10. Description of variables	46
Table 11. Counts on weight class of arrival pairs at JFK (2019).....	47
Table 12. Summary statistics of the variables	48
Table 13. Estimation results of CIT	49
Table 14. Summary on coefficients of weight class pair	50
Table 15. Description of variables	62
Table 16. Algorithms for two reallocation methods	63
Table 17. ORD-sourced TMIs	66
Table 18. Summary tables for flight delays caused by ORD-sourced MITs (2018)	68
Table 19. Summary tables for MIT restrictions encountered per flight (2018).....	68
Table 20. Summary on difference cases for model planned arrival time	69
Table 21. Change in arrival delays & queueing delays under 15-min aggregated reallocation	69
Table 22. Change in arrival delays & queueing delays under flight-level reallocation.....	69
Table 23. Definition of model explanatory variables	90
Table 24. Summary results on case-based counterfactual queueing model	95
Table 25. Model Performance on Test Set (in min/RF)	98
Table 26. Model estimation results.....	100
Table 27. Descriptive statistics for numerical variables (N=1,204)	101

Acknowledgments

My Master's and Ph.D. studies at Berkeley have been an unforgettable and remarkable journey. I wish it could never end, and that I could go to the NEXTOR office at 107 McLaughlin Hall and have lunch and cups of oat latte with nutmeg and iced americano with team members every day. However, every journey must come to an end, and I will cherish these memories forever.

First and foremost, I would like to express my deepest gratitude to my advisor, Dr. Mark Hansen, for his unwavering support, guidance, mentorship, and enlightenment throughout my academic journey. With the strategic navigation of my advisor, Mark, I have finally landed precisely, despite encountering some GDPs, excess enroute delays, MITs, and even a go-around, but achieving perfect CIT at AAL. He has illuminated the path to the world of aviation for me, and I will follow it for a long time to come. The lessons I have learned from him extend well beyond aviation and air traffic management, shaping my perspective and approach in profound ways.

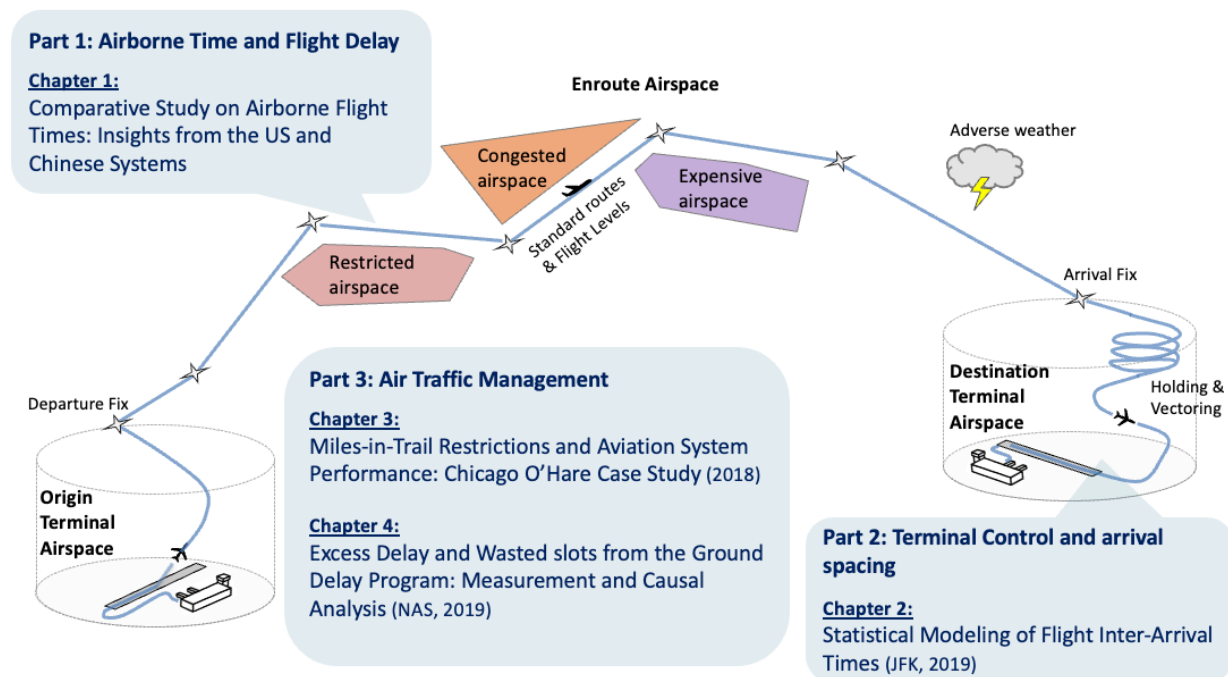
I am deeply grateful to Dave Knorr and Martin Dubin from the FAA for their funding, guidance, and support, which significantly enriched many aspects of this thesis. From the outset of my Ph.D., beginning with my initial research project on MITs at ORD, their influence has been pivotal in my academic and professional development. I also extend my appreciation to John Schade and Kennis Chan from ATAC for their assistance and support in my aviation safety project, which, although not included in this thesis, was crucial to my growth. Special thanks to Dr. Zou Bo from the University of Illinois Chicago for his invaluable help and guidance in my AAT project. I am equally thankful to Dr. Banavar Sridhar, Dr. Raja Sengupta, and Dr. Zvezdelina E. Stankova for their collaboration on CE260D (Aviation Data Science) and my role as a teaching assistant in MATH 10A (Methods of Mathematics: Calculus, Statistics, and Combinatorics) over several semesters. Moreover, I wish to thank the rest of my committee members, Dr. Michael J. Cassidy, Dr. Sandrine Dudoit, and Dr. Susan Shaheen, for their encouragement and insightful comments that greatly contributed to my work.

I am immensely grateful to my colleagues and collaborators: Dr. Kaijing Ding, Dr. Lu Dai, and Dr. Zhe Zheng, who have significantly enriched my early academic career with their research experiences and insights. I owe special thanks to Kaijing, not just as a collaborator but also as the best friend and support in our research group. It has been my great fortune to start and conclude my Ph.D. journey with her. I also extend my gratitude to my colleagues at NEXTOR: Dr. Yulin Liu, Dr. Ivan Tereshchenko, Dr. Ang Li, Tyler Truksa, Albert Cao, Yati Liu, Jing Xu, and Masoud Reihanifar, along with Frank Ketcham from Delta, for their camaraderie and support. Additionally, I would like to thank my friends: Qianhua Luo, TJ Pu, Zoey Zhang, Gaofeng Su, Yuan Jin, Yihong Zhao, Da Yu, Huizhuo Zhou, and Yue Meng, for their enduring friendship and encouragement throughout this journey.

Lastly, my deepest gratitude goes to my family. To my grandmother and parents—Shaolan, Wenjie, and Hongmei—thank you not only for your unconditional love and support but also for instilling in me the passion and courage to explore the world. You are the bedrock of my life. I must also express my profound appreciation to my partner, Xi, whose deep understanding, connection, care, and support have been my steadfast comforts and joys.

Introduction

With U.S. air traffic demand increasing more than tenfold over the past four decades and projected to grow at an average rate of 2.2% over the next twenty years (FAA, 2019; Campbell-Hill, 2017), the strain on National Airspace System (NAS) components is set to escalate. This growth is likely to intensify the workload and complexity for operational planning and system management. The NAS constitutes a complex network of interacting elements, including aircraft, airports, air traffic managers and controllers, passengers, and various navigational facilities. In addition, external uncertainties, such as thunderstorms, low ceilings, and reduced visibility, as well as internal operational errors, such as deficiencies in air traffic control staffing or central computer system breakdowns, can significantly affect system performance. The intricate connections between aircraft, passengers, and crew within different airlines add another layer of complexity, making real-time operations difficult to manage and predict. Extreme cases of system disruptions can lead to significant delays and prolonged recovery times. Therefore, the NAS is a dynamic and complex system where balancing demand and capacity is crucial for maintaining efficiency and safety. With rising demand, more frequent convective weather due to climate change, and increasing system uncertainties, there is a growing need for performance analysis in the NAS to identify key performance indicators, define metrics for trend assessment, and uncover patterns.



Graph Source: Reynolds, T. G. (2008, Sept.) Analysis of lateral flight inefficiency in global air traffic management. 8th AIAA Aviation Technology, Integration and Operations Conference

Figure 1. Workflow overview

As shown in Figure 1, the NAS is segmented into departure and arrival zones in the terminal airspace and the high-altitude enroute airspace. Traffic demand is defined by the number of flights into and out of different airspace components, while supply refers to the capacity of components

such as runways, enroute sectors, and terminal spaces. When demand exceeds the flow constraints of an airport or airspace, air traffic management (ATM) is implemented to shift delays to upstream locations where they can be safely and efficiently absorbed. Air traffic managers have developed various Traffic Management Initiatives (TMIs) with distinct concepts, procedures, automation, and control characteristics. The primary aim of ATM is to shift delays to locations where they can be managed safely and efficiently. Contemporary ATM systems employ various TMIs to alleviate downstream congestion and ensure the safe, orderly movement of air traffic flows. Traditional TMIs, listed in ascending order of restrictiveness, include Altitudes, Miles-in-Trail (MIT), Minutes-in-Trail (MINIT), Fix Balancing, Airborne Holding, Sequencing Programs, Reroutes, Ground Delay Programs (GDP), Airspace Flow Programs (AFP), and Ground Stops (GS) (FAA, 2020). More details on different TMIs can be found in the introduction to Part IV. In response to the rapid evolution of the NAS, technological advancements have led to continual system adjustments and enhancements. A significant development is the design of Trajectory Based Operations (TBO) and Time-Based Flow Management (TBFM), novel TMIs implemented as part of the FAA's NextGen program to address the operational deficiencies of traditional TMIs (FAA, 2018). However, due to their technological immaturity and nascent operational stage, the current NAS management system remains a hybrid of TBFM and traditional TMIs (FAA, 2017; Hampton, 2015).

This thesis aims to measure the performance of different NAS components. Thus, methodologies will be developed to evaluate flight demand behavior, the performance and bottlenecks of airspace component capacities, and the operational outcomes of various TMIs or combinations of TMIs. The analysis is structured into three parts:

- Part I (Chapter 1): Evaluating Airborne Time and Flight Delay – Focusing on the system impacts of ATM, it investigates the differences in airborne flight times between the U.S. and China, assessing the impact of air traffic management practices and operational efficiencies.
- Part II (Chapter 2): Evaluating Terminal Air Traffic Control and Arrival Spacing – From airspace supply side, it develops statistical models for flight inter-arrival times to analyze terminal congestion and runway capacity, extending the analysis to three years of operations at JFK.
- Part III (Chapter 3 & 4): Enhancing Air Traffic Management Efficiency – From ATM operation side, it examines the effectiveness of TMIs like MIT and GDP, providing insights into their impact on air traffic flow and identifying opportunities for improvement.

By addressing these aspects, the thesis provides valuable insights into the performance and management of the NAS, offering recommendations for enhancing operational efficiency and reducing delays amidst increasing air traffic demand. The following contents will give a brief introduction and summary on each chapter.

A major goal of ATM is to shift delay in the air to delay on the ground. Since Chinese ATM is in its nascent stages, a comparison of AAT in the two nations sheds light on how more advanced ATM can reduce airborne times. Therefore, in Chapter 1, we explore the behavior of Actual airborne time (AAT), which is the time between actual wheels-off and actual wheels-on of a flight. Of particular interest is the comparison AAT in different countries with varying air route structures, air traffic management systems, weather, and geography. This chapter performs the first comparative empirical analysis of AAT behavior, focusing on the U.S. and China. The focus is on how AAT is affected by origin-destination (OD) distance, the possible pressure to reduce AAT

from other parts of flight operations, hub status of departure/arrival airports, enroute and terminal traffic conditions, and convective weather. Econometric models are developed to quantify the impacts of factors on AAT behavior in China and the U.S., separately. The estimation results show that in both countries AAT is highly correlated with OD distance. Flight time in China is longer than in the U.S. given the same OD distance which indicates a low effective speed may be the result of a low aircraft speed, or due to the flight experiencing metering, rerouting, holding or vectoring in Chinese airspace. In addition, we find that a flight has limited capability to make up for pre-departure delay. Sensitivity analysis of AAT to flight length and aircraft utilization is further conducted. Given the more abundant civil airspace, flexible routing networks, and efficient air traffic flow management (ATFM) procedures, we performed a counterfactual analysis to investigate how Chinese AATs would change if they were governed by the U.S. model. We find that this would result in significant efficiency gains for the Chinese air traffic system. On average, 13 minutes of AAT per flight would be saved. Systemwide fuel saving would amount to 326 million gallons with CO₂ emission reduction of 2.7 million tons and direct airline operating cost saving of over \$1.3 billion in 2016.

Focusing on the supply side for ATM—the capacities of runways and terminal airspaces—Chapter 2 addresses the pressing issue of runway capacity as a pivotal bottleneck in the NAS. Given the ever-growing civil aviation demand and increasing complexity of the NAS, understanding the behavior of controlled interarrival time (CIT) and its relationship with operational conditions is important. This chapter presents an empirical study of CIT at the landing runway threshold with its influencing factors. A mathematical model is designed to fit the distribution of the observed runway crossing time interval (RCTI) for arrival pairs by using a conditional probability with a uniform distribution for feasible estimated time of arrival (ETA) and a normal distribution for CIT. The behavior of CIT is further explained by a linear model with variables relating to weather, runway, and aircraft. This conditional probability model not only captures the situation when arrival demand is very low and the time gaps between landing aircrafts are constrained by ETAs, but also ideally approximates the left-hand side of RCTI when ETAs are bunched so that short-term arrival demand exceeds capacity. The linear model presents the probability distribution curve of CITs given different operational conditions, which provides important information for improving landing safety and capacity. The methodology is applied to over 194 thousand arrival pairs at JFK airport during the whole year of 2019.

In a relatively tactical traffic management context, Chapter 3 analyzes the application and efficacy of the most common TMI -- Miles-in-Trail (MIT). An MIT places a minimum spacing requirement between aircrafts at a specific location to manage the imbalance between capacity and demand at a given airport, or to support merging streams in the sectors. This paper focuses on the improvement opportunities from reduction of MIT. This chapter designs a model including estimation of planned arrival time and a deterministic model to assess the impacts of a given MIT or a set of MITs. Our model is applied to ORD airport of 2018 as a case study. As results show, if all ORD-sourced MITs were eliminated, the overall queueing delay would increase by 3.3 hours per day on average while arrival delay would decrease by 14.7 hours per day. We also found that ORD during 8:00-9:00, 11:00- 12:00 and 17:30-18:30 tends to experience more throughput reduction when implementing MITs and those delayed flights are allocated into 9:00-10:00, 12:00-13:00 and 18:30-19:30 respectively. Our results strongly suggest that ORD-sourced MITs, while relieving some build-up of flights in the ORD terminal area, do so at a substantial penalty in the form of increased arrival delay and throughput shortfalls.

In the strategic traffic management context, Chapter 4 evaluates the key TMI in the current system -- the Ground Delay Program (GDP). When scheduled demand significantly exceeds forecasted capacity for a substantial period at an arrival airport, a GDP preassigns forecasted arrival delays to incoming flights before their departure, effectively shifting potential airborne delay to the ground. This approach enhances safety, efficiency, and minimizes fuel costs. Despite its intended functionality, the program often faces operational inefficiencies due to weather uncertainties and deviations in flight planning. This study quantifies the performance shortfalls of current GDP operations by focusing on metrics such as "excess delay" and "wasted slots." We developed a flight delay identification algorithm and a deterministic queueing model to measure these metrics across over 1,200 GDPs across multiple airports in 2019. Additionally, we employed regularized regression models to further examine GDP performance and identify key factors that impact GDP-induced delays. Our analysis reveals significant opportunities for improving GDP efficiency through more precise GDP planning and better flight time compliance, potentially aided by advanced technologies like Trajectory Based Operations (TBO) and Time-Based Flow Management (TBFM).

In summary, by systematically analyzing these various aspects of NAS operations, the thesis aims to offer practical insights and identify significant opportunities for improvement. These enhancements are vital for increasing the efficiency, safety, and reliability of air traffic operations. Each chapter contributes to a comprehensive understanding of how various components of air traffic system can be improved or require special attention to meet the escalating demands of global air traffic.

PART I. Evaluating Airborne Time and Flight Delay

Chapter 1

Comparative Study on Airborne Flight Times: Insights from the US and Chinese systems

1. Introduction

With the total number of global air travelers growing tenfold from 1970 to 2017 (World Bank, 2018), air traffic systems worldwide have become increasingly complicated and in many situations congested, causing negative consequences to airlines and the traveling public such as increased airline operating cost, loss of passenger welfare, and added fuel consumption and emissions (see, for example, Ball, et al., 2010; Zou and Hansen, 2012; Dresner et al., 2012; Ryerson et al., 2014). To mitigate these negative consequences, the airline industry and air traffic control (ATC) have made significant efforts. For example, it is common practice for airlines to adjust flight schedules to absorb and prevent delays from being propagated. Integrated disruption management with flight planning has also been considered to help with schedule recovery once delay occurs (e.g., Clausen et al., 2010; Marla et al., 2016). On the ATC side, many traffic management procedures including Ground Delay Programs (GDP), Airspace Flow Programs (AFP), and Time-Based Flow Management (TBFM) have been developed to balance flight demand with terminal and en-route airspace capacity and stream air traffic. The goal of these efforts is to minimize delays and maximize the extent to which aircraft fly according to schedule. To support these efforts, understanding the behavior of actual airborne time (AAT), i.e., the time between the wheels-off at the departure airport and the wheels-on at the arrival airport, which is the most expensive part of a flight, is of critical importance. The ability to predict AAT is essential to effective air traffic management, while interregional comparisons of this metric can shed light on how levels of service provided by different air navigation service providers.

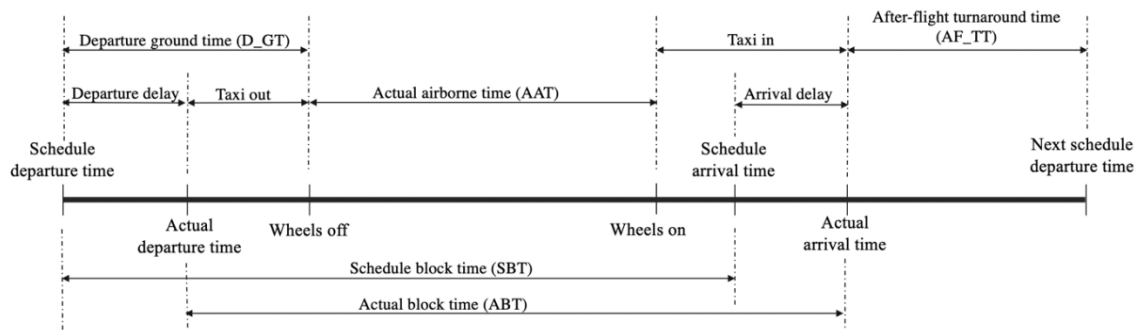


Figure 2. Time components related to a flight

The behavior of AAT is intertwined with other time components related to a flight. As shown in Figure 2, AAT is part of the actual block time (ABT) of a flight which consists of AAT as well as taxi-out and taxi-in times. For a typical flight, taxi times are a small proportion compared to AAT suggesting a high correlation between AAT and ABT. In an ideal situation, ABT exactly

coincides with the scheduled block time (SBT) of a flight and the actual and scheduled departure times match. However, in reality, ABT, which is influenced by taxi, en-route, and terminal delays, usually deviates from SBT. In some cases, flight operators may also try to reduce ABT below SBT to make up for departure delays. Taxi-out time plus gate departure delay forms the ground time at the departure airport (D_GT). After arrival at the gate, the extent of after-flight turnaround time (AF_TT) plus arrival delay relative to the needed time for turnaround determines whether delay will incur to the gate departure of the next flight.

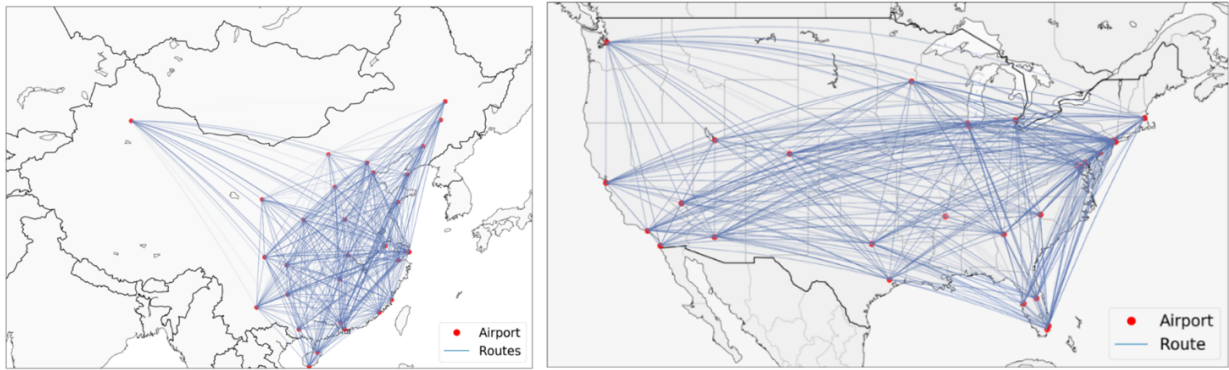
The time components involved in a flight described above provide some useful perspectives on what to consider when modeling AAT. The first important aspect is the extent to which AAT is in line with the flight schedule (OAG), which in turn is influenced by historic ABT of flights operating on the same OD and designed by airline according to their policy and schedule management. A closely related issue is how AAT is aligned with flight physical distance flown and its optimal distance path – great circle distance (GCD) route. Secondly, as on-time performance is a primary concern for airlines, an important factor to consider is the amount of time a flight spends before takeoff and whether pre-takeoff delay, i.e., the difference between the actual and schedule wheels-off time, leads a flight to speed up or take a more direct route to make up for the delay within certain limits. Similarly, the amount of turnaround time between the actual gate arrival of a flight and the scheduled gate departure of the next flight may affect AAT of the flight. If the scheduled turnaround time is short, the ability of a flight to shorten AAT to reduce arrival delay should be explored. Apart from the above, other flight operation-related factors such as aircraft heading direction (eastward vs. westward), hub status of departure/arrival airports, enroute and terminal congestion levels, and convective weather are also expected to exert some influence on AAT. In addition, AAT is likely to be further influenced by time-of-a-day, day-of-a-week, seasonal, and airport-specific factors.

To our knowledge, no prior research has looked into the behavior of AAT. Even the existing efforts to model ABT and SBT are limited. For ABT, Coy (2006) employed a two-stage statistical model with one-year flight records to estimate the block time of U.S. commercial flights. Factors chosen for ABT estimation include historical ABT, flight arrival time, airport utilization, ice, and the interaction of poor weather conditions with traffic. Deshpande and Arikan (2012) used Ordinary Least Square (OLS) to estimate truncated block time, defined as ABT minus delay caused by late arrival at the origin of the departing aircraft, as a function of route, carrier, airports, airport congestion, and aircraft-specific variables. In a series of studies, the question of how the ABT distribution affects airline SBT setting considering flight on-time performance has been examined (Mayer and Sinai, 2003; Hao and Hansen, 2014; Kang and Hansen, 2017, 2018). Further research also looked into a related issue of whether delay as part of ABT is/can be internalized (Brueckner, 2002a, b, 2005; Brueckner and Van Dender, 2008). More recently, Fan (2019) investigated the variation of SBT over the past three decades using quantile and OLS regressions with air traffic condition, flight delays, aircraft type, flight heading, airports slot policy, airport-specific anomalies, airline-specific policies, and changes in crude oil price as explanatory variables. It is found that worldwide, SBT has grown by 0.21-0.33 minutes per year. Belkoura et al. (2018) discusses a computational framework to detect whether the actual trajectory is deviated from the planned trajectory. The airborne delay, the difference between ABT and SBT, can be calculated by the deviations with a hypothesis of a constant velocity.

In view of the existing literature, this chapter aims to make two major contributions. The first one is the explicit modeling of AAT and testing of its relationship with a range of operation-, time-,

weather-related and fixed-effect factors. In particular, we are interested in the extent to which the AAT is determined its OD-pair great circle distance (GCD) and which possible causes can contribute to deviations from that fundamental relationship. Also of interest are the interactions of AAT with predeparture delay and after-landing turnaround time, given that the three time components together determine delay propagation downstream. Large predeparture delay may pressure a flight to reduce AAT to catch up with the schedule within certain limits. Knowing that the turnaround time is small after landing, a flight may also try to reduce AAT to minimize arrival delay and its propagation to connected operations. These conjectures, which are unexplored previously, are part of the hypotheses to be tested in our study. Furthermore, how the AAT behaves differently for flight with longer distance and with more flight legs in a day will be explored.

For the second major contribution, our investigation of AAT behavior is conducted in a comparative manner for flights in the U.S. and China. The air traffic systems of the two countries are the two largest in the world. The two countries are also similar in territory size, both located in the belts of prevailing westerlies and dealing with significant flight delays (Ren and Li, 2018). Yet significant differences exist in their air traffic systems in terms of geographical distributions of air travel demand, airspace availability, air traffic flow management (ATFM), delay management policies by the respective civil aviation authorities, and airline scheduling practices. For the geographical distributions of air travel demand, most of the economic centers in China are concentrated in the eastern part of the country, resulting in the majority of flights being short- and medium-haul. In contrast, the economic centers in the U.S. are more evenly distributed throughout the country, with particular foci on the east and west coasts, leading to longer flight lengths (32). In terms of airspace availability, only 20% of the national airspace in China is made available for civil aviation use while in the U.S. the percentage is 80%. This stark discrepancy means more limited airspace capacity and flight routing flexibility in China (Hsu, 2014), problems that are likely to become worse if forecasted rates of air traffic growth in that country is realized. These differences make the comparison between the two countries worthwhile, with the results possibly informing future flight scheduling and operation strategies in China to improve its system efficiency.



(a) Routes among the top 30 airports in China (b) Routes among the top 30 airports in the U.S.

Figure 3. Air routes among top 30 airports in China and the U.S.,

The remainder of the chapter is organized as follows. In Section 2, we present model development and variable specifications. Section 3 describes the data sources and the data filtering process. Estimation results of the base model are reported and discussed in Section 4. Further sensitivity analysis of the estimation results to flight length and aircraft utilization is performed in

Section 5. As we conjecture a more efficient flying environment in the U.S., Section 6 conducts a counterfactual analysis to assess the potential efficiency gains if the Chinese air traffic system adopted the AAT behavior in the U.S. Findings of the chapter are summarized in Section 7 along with suggestions for future research.

2. Model development

Based on the literature review and conjecture in Section 1, this section specifies the econometric models for AAT to be estimated in the chapter and identify explanatory factors that may influence AAT with related hypotheses. The general form of the AAT models is presented in Eq. (1), in which explanatory variables characterizing the AAT behavior are classified into four categories: operation-related, weather-related, time-related and fixed-effect variables. The variables considered in the model specifications are described in the following subsections and summarized in Table 1.

$$AAT = f(\text{Operation-related variables, Weather-related variables, Time-related variables, Fixed-effect variables}) \quad (1)$$

2.1 Operation-related variables

a) Characteristics of flight OD pair

As mentioned earlier, the most relevant variable is the origin-destination (OD) distance, to represent which we use the great circle distance (**GCD**) between the origin and the destination airports. We consider a **LON_dif** variable, which measures the longitude difference between the departure and the arrival airports, with eastward taking positive values and westward taking negative values. The expectation is that prevailing winds may affect flying speed and thus AAT. We use **HubDep** and **HubArr** to indicate whether a flight takes off from or lands at the airline's hub. We conjecture that a flight from/to a hub airport is likely to receive better ATC services in the terminal airspace, and that an airline usually has more flying experience and a stronger relationship with ATC service providers at its hub airports, which may improve flight operations. On the other hand, prior research on delay internalization (e.g., Brueckner, 2002a) suggests that flight delay, which can be part of AAT, tends to fall with higher airport concentration, as is the case for hub airports. In addition, Mayer and Sinai (2003) mentioned that bunching of departures and arrivals together for short connecting times at hub airports leads to peaks of air traffic that causes congestion and predeparture delays. Thus, we conjecture that air traffic controllers would relieve the pressure from the traffic peaks and flight on-time performance, which could be a factor to reduce the flight AAT. The major hub airports in China and the U.S. are listed in [Appendix 4](#).

b) Flight network effect

Arrival delay of a flight not only increases the operating cost but can propagate to downstream flights of the same aircraft and different connected aircraft through crew rotation and passenger connections (Churchill et al., 2010; Kafle and Zou, 2016). To mitigate the delay propagation, a flight delayed at departure may speed up enroute or take a more direct route to reduce the airborne time (Delgado, et al. 2013). Provided that there is room for reducing airborne time, the larger the

pre-takeoff delay, the higher pressure a flight will have to save more time from AAT. So, D_GT , which is the gate out delay at the origin airport, is designed to capture this effect. We have the hypothesis that the pre-departure delay pressures a flight to reduce AAT so that with everything else being equal, AAT should to some extent, decrease with pre-takeoff delay.

Considering the delay propagation from a different perspective, we also involve the effects from aircraft turnaround time, which is defined as the time between the scheduled arrival of the current flight and the scheduled departure of the next flight. If turnaround time becomes longer, then there is greater capacity to absorb and prevent any arrival delay from being propagated to the next flight (of the same aircraft). On the other hand, turnaround time should not be exceedingly long as it affects the number of legs an aircraft can fly in a day and thus airline productivity. Nonetheless, shorter turnaround time at the destination airport may pressure a flight to fly faster to minimize arrival delay. a variable AF_TT_dif is constructed as the difference between the after-flight turnaround time (AF_TT) and the minimum turnaround time (MTT) at the destination airport. AF_TT is calculated as the difference between the scheduled departure time of the next flight of the same aircraft and the scheduled arrival time of the current flight under study. In China, the Civil Aviation Administration of China (CAAC, 2012) documents values for minimum connection time, which may bear some similarity with MTT, but does not provide details on how the values are calculated. In the U.S., information on MTT is not available from the Federal Aviation Administration (FAA). To make the measurement of MTT consistent in the two countries, we use the 5th percentile value of the actual turnaround time of similar flights at the same airport. Similar flights refer to flights that have the same aircraft type, arrival airport, and airline. The use of the 5th percentile rather than the observed minimum turnaround time follows the idea of existing work (e.g., Britto et al., 2012; Zou and Hansen, 2012) to make the calculation more robust to measurement error and reduces the influence of unusually favorable conditions. The hypothesis is that flights with smaller turnaround time may be more pressured to reduce AAT to minimize arrival delay. Another binary variable, Flt_last , indicating whether a flight is the last leg flown by an aircraft is included, to capture the aircraft utilization and to test whether being the last flight in a day will lead to faster or slower flying.

c) Terminal congestion effect

To investigate the congestion effects at the departure and arrival airports for a flight, we introduce two new variables, $QueueDep/QueueArr$, which indicate airport- and time-specific departure/arrival queueing delays. Inspired by a previous study on airport queueing (Hansen, 2004), we adopt the deterministic queueing modeling to model flight departure and arrival procedures at each airport with quarter-hour time intervals. Taking arrival queueing as an example. At an airport, a flight is counted as arrival demand in each quarter hour between its scheduled wheels-on time and the actual wheels-on time. However, if a flight landed earlier than the schedule time, it is counted into only the earlier time interval of actual arrival as demand. Then, we calculate the cumulative actual arrivals ($A(t)$, $CumActual$) and cumulative arrival demand ($P(t)$, $CumDemand$) for each quarter hour in the course of a day. Figure 4 provides an illustration of a one-day queueing diagram, with the horizontal axis representing quarter-hour periods of the day and the vertical axis representing cumulative flight counts. The area between the two curves indicates the total arrival queueing delay for the airport during the day. Airport-specific departure queueing delay can be similarly constructed based on scheduled and actual flight wheels-off time. For a flight arriving/departing at airport α during quarter hour Q in day j , $QueueDep/QueueArr$

are computed based on Eq. (2). Basically, *QueueDep/QueueArr* are each the sum of airport departure/arrival queueing delays up to 30 min before the quarter hour of the actual flight departure/arrival.

$$\begin{aligned} QueueDep &= \sum_{i=0,1,2} CumDemand_j^\alpha(Q-i) - CumDeparture_j^\alpha(Q-i) \\ QueueArr &= \sum_{i=0,1,2} CumDemand_j^\alpha(Q-i) - CumArrival_j^\alpha(Q-i) \end{aligned} \quad (2)$$

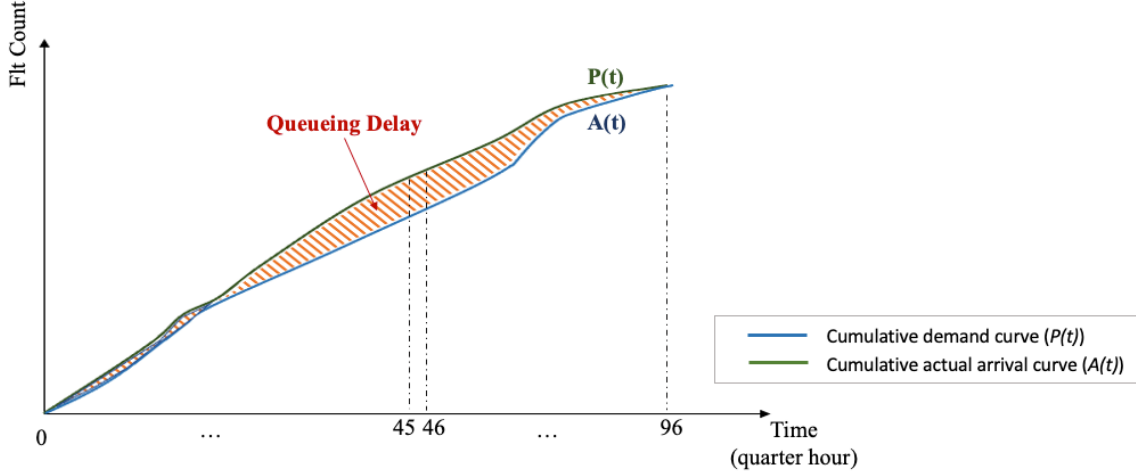
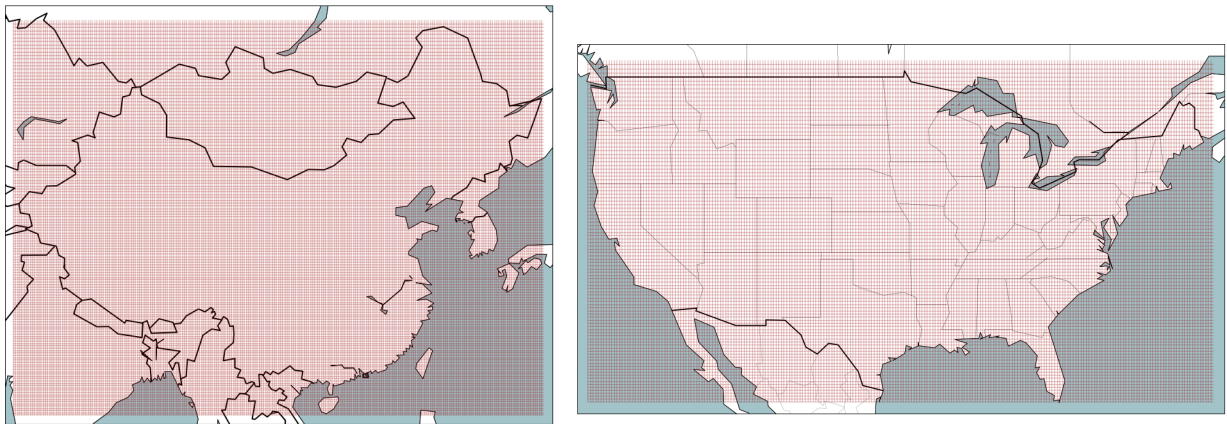


Figure 4. Example of queueing diagram for demand and actual flight counts

d) Enroute congestion effect

Variable, *TdEnroute*, is designed to capture the traffic congestion level of the airspace where a flight schedules to fly along its GCD route. Inspired by the research on wind mapping algorithm (Liu, et al, 2021), we construct the grid system for national airspace and map the flight locations into the grid in chronological order. The U.S. airspace is divided into cells with size of 0.25° by 0.25° , and with latitude from 25°N to 50°N and longitude from 66°W to 125°W , while China airspace is divided into the same size grid with latitude from 18°N to 54°N and longitude from 73°E to 135°E , as showed in Figure 5.



(a) China grid system: 144*248 cells

(b) U.S. grid system: 100*236 cells

Figure 5. Layout of grid across the truncated national airspace

After constructing the airspace grid system, for each domestic flight with its GCD route and schedule departure and arrival time, we apply the distance-based interpolation along the great circle arc on the angle subtended by the GCD route to generate the enroute location of the given flight at quarter-hour intervals during its schedule airborne time with example in Figure 6(a) provided. Then, at each quarter hour of the year, we can identify the locations of concurrent airborne flights and aggregate the traffic density over the entire grid as the number of aircraft located in the cells. The value of a cell indicates the traffic volume passing through a particular region (i.e., the cell) at a given quarter-hour time. To capture the potential impacts of surrounding traffic and flight route deviations, we add weights of traffic density from the adjacent cells by applying convolutional kernel with filtering size of 3 by 3, padding of 1, stride of 1 and activation function as showed in Figure 6(b) to the traffic density matrix over the grid system. To help reader better understand this method, examples on traffic density before and after kernelling at a specific quarter hour in China and the U.S. are provided in [Appendix 2](#).

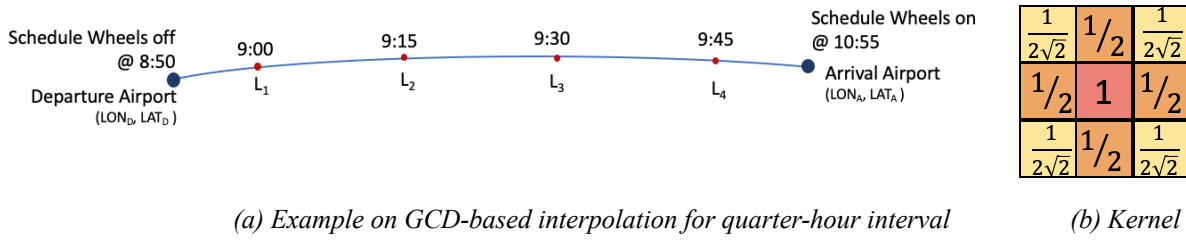


Figure 6. Processing flight for the traffic density matrix

Having the quarter-hour-specific traffic density matrices over the airspace grid along the year, we calculate the en-route traffic congestion for each individual flight, **TdEnroute**, by matching the flight en-route locations at each quarter-hour during its airborne time with the corresponding time-location-specific cells. We then take the average on the traffic density values from these identified cells as in Eq. (3) where N is the number of quarter-hour locations captured for the flight. **TdEnroute** is designed to quantify the schedule traffic density of the region which a flight is flying through along its GCD route. The Larger this value is, the busier regions a flight is overflying.

$$TdEnroute = \frac{1}{N} \sum_{i=1}^N Grid_Traffic_{Location_i}^{QTR_i} \quad (3)$$

e) Nearby terminal effects

Another concern for the impacts of en-route traffic pattern comes from the flight operation and air traffic management of the airports adjacent to the GCD route of a flight. Terminal air traffic controllers not only need to handle the aircrafts out of and into the airport, but also are responsible for the safe separation of flights flying over the busy airspace surrounding the airports. Therefore, air traffic management over and around the airports can be affected by terminal operations. To this end, we design a variable, **NearbyAPT**, to capture the traffic impact from airports near a flight's GCD route. Note that the airspace within a radius of 30-50 miles around an airport and up to 10,000 feet is managed by those terminal air traffic controllers. The nearby airport impact on an overflying flight is expected to vary depending on time of a day and how far an airport is from the route of the overflying flight. For a given flight along its GCD route, we include airports within 100 miles

from the GCD route and identify the actual times when the flight passes by the point on the route that is the closest to each of these airports. Figure 7 provides an illustration. Moreover, as the flight climbing and descending stages may take 30 min or more time (especially for arrival flights), we count traffic volume of a nearby airport based on actual flight departure/arrival time in three consecutive quarter hours. The middle quarter hour contains the time when the flight under study passes by the point on its GCD route that is the closest to the airport. The other two quarter hours immediately precede and follow the middle quarter hour. The counted traffic volume for such a nearby airport j is denoted as T_j . Then, NearbyAPT for a flight is defined as:

$$\text{NearbyAPT} = \sum_{j \in J} \frac{T_j}{D_j} \quad (4)$$

where J is the set of airports within 100 miles away from the GCD route of the flight under study. D_j is the shortest distance from airport j to the flight's GCD route.

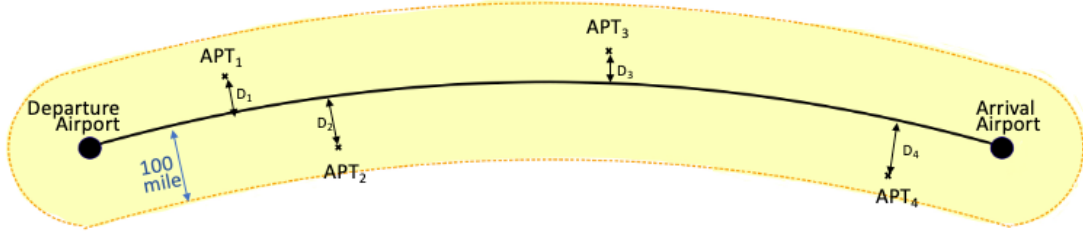


Figure 7. Flight GCD route with its enroute nearby airports within 100 miles (APT: airport)

2.2 Weather-related variables

a) Enroute weather conditions

Convective weather is of great importance in terms of sector capacity and flight en-route inefficiency. As flights may need to reroute or have the backward delay at preceding airspace to avoid convective weather, we conjecture that convective weather, especially thunderstorm, can increase AAT. Inspired by a previous study on normalizing the weather effects (Callahan et al, 2001), we map the convective weather to the same grid systems constructed in Section 2.1-4). The convective weather records come from 2330 weather stations in the U.S. and 2080 stations in China (as shown in Figure 8 where blue dots denote station locations), collected at different frequencies in the two countries. Thus, resampling of the weather records is performed to make the records of the same frequency (which is hourly) in the two countries. In view of the uneven distribution of weather stations and the absence of records in cells without a weather station, we apply a 2D grid interpolation with a linear method to assign convective-weather weights to all the grid matrices, to generate hourly convective-weather matrices over the airspace grid in the year. Figure 9 provides examples on 2D-interpolated convective-weather matrices for a specific hour in China and the U.S.

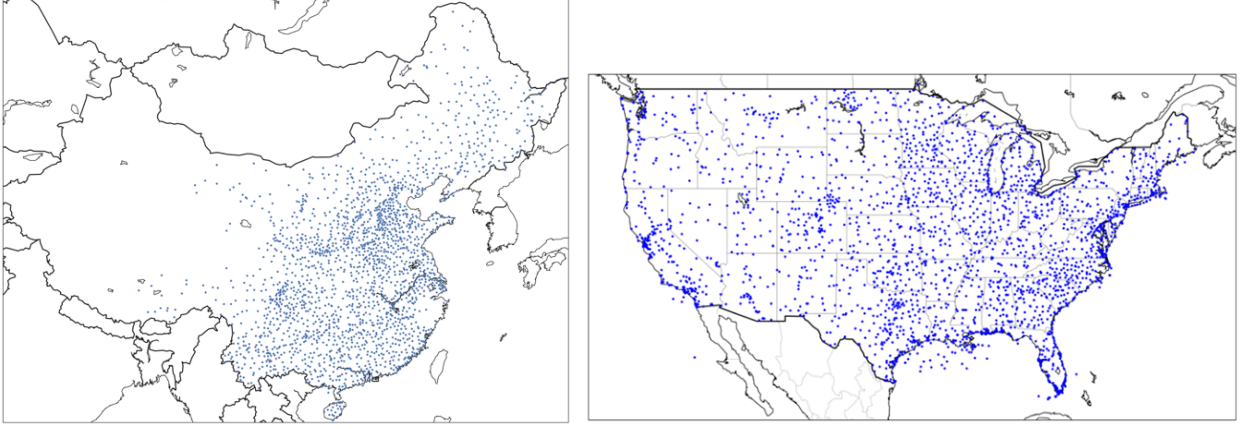


Figure 8. Distribution of weather stations in China (left) and the U.S.(right) (2016)

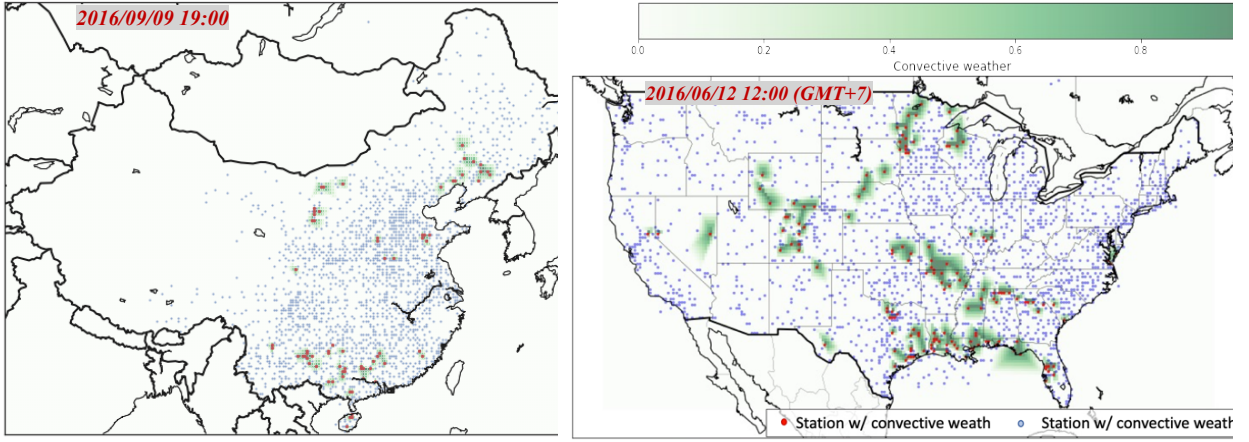


Figure 9. 2D grid interpolated convective weather grid examples

Similarly to the calculation of $TdEnroute$, we calculate the en-route weather conditions for each flight using variable $CWEnroute$, by matching the flight enroute locations in hourly intervals during its airborne time with time-specific cells in the airspace grid. We then take the average on the convective weather weights from the identified cells as shown in Eq. (5), where N is the number of mapped time-specific cells for the flight. In this way, $CWEnroute$ quantifies the proportion of a flight's GCD route with the presence of convective weather.

$$CWEnroute = \frac{1}{N} \sum_{i=1}^N Grid_ConvectiveWeather_{Location_i}^{Hour_i} \quad (5)$$

b) Terminal weather conditions

To capture the impacts of terminal weather on AAT during the descending and climbing procedures, we consider convective weather (i.e., thunderstorm), temperature, crosswind speed, and visibility at the departure and arrival airports of a flight. For temperature, we postulate that both hot and cold temperatures present adverse weather and impact the AAT. Thunderstorm, as the most critical convective weather, can severely reduce the airport capacity and thus affect flight delay and airborne time. So dummy variables, $CWDep$ / $CWArr$ for each flight OD airports, are included to control these effects. Hot temperature negatively affects aircraft engine performance (FAA, 2008).

Cold temperature is often associated with foggy and snowy days which suggest low visibility thus also negatively affecting flying especially during ascent and descent. The absolute difference between the temperature at the actual departure / arrival time of a flight and the average temperature associated with all departure / arrival flights at the same airport is considered (*TempDep* and *TempArr*). The crosswind speed variable, *CWindDep* / *CWindArr*, provides the speed of wind component that is blowing perpendicular to the taking-off / landing runway at the departure / arrival airport during the hour of the actual departure / arrival time of the flight. The visibility variables *VisDep* / *VisArr* are similarly defined but taking logarithmic values to capture the plausible nonlinear relationship between visibility and AAT. For example, visibility decreasing from 1 mile to 0.5 mile impacts flight operations more than decreasing from 8 miles to 7.5 miles. As some records have visibility of zero miles, we add one to the exponent: $VisDep = \ln(\text{Departure Visibility} + 1)$ and $VisArr = \ln(\text{Arrival Visibility} + 1)$.

2.3 Time-related variables

Several variables are considered in time-related category. The first one is *DT*, a binary variable indicating whether a flight operates in the day or night. *DT* equals 1 if the scheduled departure time is between 6:00 and 20:00 at the local departure airport. The hypothesis is that AAT may be higher in the day due to greater traffic and congestion. Second, AAT may vary depending on it is a weekday or in the weekend, again considering the difference in air traffic particularly for business travel. A *Weekday* variable is included which takes value 1 if the scheduled departure day is a weekday and 0 otherwise. Third, traffic pattern differences arising from seasonal changes in air travel demand and consequently airline scheduling strategies may further affect AAT. To this end, three seasonal variables—*Spring*, *Summer* and *Fall*—are introduced.

2.4 Fixed effects

As our econometric model examines the AAT behavior from domestic flights in one country which are operating with different aircraft types and different airlines to connect the airports, the flights can't be considered as independent observation but should be treated a "nested" structure data. Thus, inspired by the research on SBT time model (Fan, 2019), we employ 3 groups of indicator variables on the aircraft type, airline and airport to account for the idiosyncratic effects and to control the unobserved sources of variation at the group levels. First group is aircraft type. Its effect is modeled with its mathematical product with flight GCD (as *GCD*AC_XXX*). Aircraft type with annual flight share over 1% are converted into GCD-dummy variables for China and the U.S. respectively, while the remaining aircraft types are grouped as "others". The U.S. model includes B738, A320, A321, B737, B739, A319, E170, B752, MD88, B712, E190, MD90, CRJ7, B733, R75L, B753, CRJ9 and B763, and China model includes B738, A320, A321, A319, B737, A333, A332 and E190. B738 is excluded for both countries to avoid multicollinearity problems. Second group is airline. We select the airlines with annual flight share over 1% for dummy variables for in China and the U.S and group the remaining airline as "others". The U.S. model includes AAL, DAL, UAL, SWA, JBU, NKS, SKW, ASA, VRD and FFT, and China model includes CSN, CSC, CDG, CSH, DKH, GCR, CBJ, CQH, CHB, LKE and UEA. AAL is excluded in the U.S. model and CSN is excluded in the China model. The statistics on flight shares for airlines and aircrafts are provided in [Appendix 4](#). The last group of fixed-effect dummy variable is airport specific,

including both departure and arrival airports. Guangzhou Baiyun International Airport (CAN) is considered the base airport in the China model and Los Angeles International Airport (LAX) is considered the base airport in the U.S. model. The airport dummies are expected to capture the systematic variations on efficiency of terminal traffic management and traffic congestion for flight departure and arrival.

Overall, the following Eq.(6) presents the full specification of AAT behavior model specifications. In the specifications, β 's are the coefficients to be estimated and ε is the error term.

$$\begin{aligned}
 AAT = & \beta_0 + \beta_1 GCD + \beta_2 Lon_dif + \beta_3 Spring + \beta_4 Summer + \beta_5 Fall + \beta_6 DayTime + \beta_7 Weekday \\
 & + \beta_8 HubDep + \beta_9 HubArr + \beta_{10} D_GT + \beta_{11} AF_TT_dif + \beta_{12} Flt_last \\
 & + \beta_{13} QueueDep + \beta_{14} QueueArr + \beta_{15} TdEnroute + \beta_{16} NearbyApt \\
 & + \beta_{17} CwEnroute + \beta_{18} CwDep + \beta_{19} CwArr + \beta_{20} VisDep + \beta_{21} VisArr \\
 & + \beta_{22} TempDep + \beta_{23} TempArr + \beta_{24} CWindDep + \beta_{25} CWindArr \\
 & + \sum_{i=1}^{N_{aircraft}} \beta_i^{AC} GCD * Aircraft_t_i + \sum_{j=1}^{N_{airline}} \beta_j^{AL} Airline_j + \sum_{k=1}^{N_{Dep}} \beta_k^{Dep} DepApt_k + \sum_{l=1}^{N_{Arr}} \beta_l^{Arr} ArrApt_l + \varepsilon
 \end{aligned} \tag{6}$$

Table 1. Description of variables

Variables		Description
Y-variables		
	<i>AAT</i> (min)	Actual airborne time, which is the time difference between wheels-on and wheels-off time
X-variables		
Operation related	<i>GCD</i> (mile)	Great circle distance between departure and arrival airports
	<i>Lon_dif</i> (degree)	Longitude difference between departure and arrival airports, positive if flying eastward and negative if flying westward
	<i>D_GT</i> (hour)	Pre-departure delay, calculated as max (0, actual gate-out time – scheduled gate-out time)
	<i>AF_TT_dif</i> (hour)	Difference between after flight turnaround time (AF_TT), calculated as scheduled gate departure time of the next flight minus scheduled gate arrival time of the current flight, and the minimum turnaround time (MTT, defined as 5th percentile value of the actual turnaround time according to the given arrival airport and aircraft type): AF_TT – MTT
	<i>Flt_last</i>	Takes value 1 if the flight is the last leg flown by an aircraft in a day
	<i>HubDep / HubArr</i>	Takes value 1 if the flight departs from / arrives at a hub airport
	<i>TdEnroute</i> (flight counts)	Enroute traffic density along the flight GCD route, calculated as $TdEnroute = \frac{1}{N} \sum_{i=1}^N Grid_Traffic_{Location_i}^{QTR_i}$
	<i>NearbyAPT</i> (100 flight counts)	Traffic impact from airports near a flight's GCD route, calculated as, $NearbyAPT = \sum_{j \in J} \frac{T_j}{D_j}$

	<i>QueueDep/ QueueArr</i> (hour)	Cumulative queuing delay at departure/ arrival airport , calculated as, $QueueDep/QueueArr = \sum_{i=0,1,2} CumDemand_j^a(Q-i) - CumActual_j^a(Q-i)$
Time related	<i>Weekday</i>	Takes value 1 if the flight operates on a weekday, and 0 otherwise
	<i>Springg / Summer / Fall</i>	Takes value 1 if the flight is scheduled to depart in Spring / Summer / Fall, and 0 otherwise
	<i>DayTime</i>	Takes value 1 if the flight operates in the day (scheduled departure between 6:00 and 20:00), and 0 otherwise
Weather related	<i>CwEnroute</i>	Enroute convective weather along the flight GCD route, calculated as, $CWEnroute = \frac{1}{N} \sum_{i=1}^N Grid_ConvectiveWeather_{Location_i}^{Hour_i}$ which is range from 0 to 1.
	<i>CWDep/ CWArr</i>	Take value 1 if there is convective weather during the actual departure / arrival hour at of a flight at the corresponding airport
	<i>VisDep/ VisArr</i> (ln mile)	Log of visibility (in miles) plus one mile during the actual departure / arrival hour at of a flight at the corresponding airport
	<i>TempDep/ TempArr</i> (Fahrenheit)	Absolute difference between the temperature at the actual departure / arrival time of a flight and the average temperature associated with all departure / arrival flights at the same airport
	<i>CWindDep/ CWindArr</i> (knots)	The speed of wind component that is blowing perpendicular to the taking-off / landing runway at the actual departure / arrival time of a flight
Fixed effects	<i>GCD*AC_XXX</i>	Aircraft dummies multiplying with its flight distance in mile (B738 is the benchmark and excluded for both countries)
	<i>AL_XXX</i>	Take value 1 if the flight airline is XXX (China Southern Airlines (CSN) / American Airline (AAL) is the benchmark and excluded in the China/U.S. model.)
	<i>DepApt_XXX / ArrApt_XXX</i>	Take value 1 if the flight departure/ arrival airport is XXX (Guangzhou Baiyun International Airport (CAN) / Los Angeles International Airport (LAX) is the benchmark and excluded in the China/U.S. model.)

3. Data

3.1 Data sources

The data needed for model estimation come from several sources. For operation- and time-related variables, flight-level records are collected for the year 2016 from the Operation Supervisory Center of CAAC on the China side and the Aviation System Performance Metrics (ASPM) of FAA on the U.S. side. This results in 5,814,061 and 6,883,674 flight records collected in China and the U.S. respectively. On the China side, the flight records include commercial and general aviation flights. On the U.S. side, the flight records consist of commercial flights, freights, and air taxi. Each record documents the departure and arrival airports, departure and arrival dates, airline to which the flight belongs, aircraft type of the flight, tail number of the aircraft, the scheduled gate departure time, the actual wheels-off time, the actual wheels-on time, and the scheduled gate arrival time. The time information allows us to construct *AAT*, *D_GT*, *AF_TT*, *AF_TT_dif*,

Flt_last, *DT*, *Wkd*, and the seasonal dummy variables. In addition, *GCD* between any two airports is calculated based on the latitudes and longitudes of the airports, the information of which is collected from OpenFlight (2019). The longitude information also allows for construction of the *Lon_dif* variable. *TdEnroute* and *NearbyAPT* are derived based on flight time, GCD route and airport locations.

For weather-related variables, the Meteorological Terminal Aviation Routine (METAR) weather report provides weather information including temperature, wind speed, and visibility at each of the airports covered in the data on the China side, at 60-minute intervals. On the US side, the ASPM data offers similar weather records, at a higher frequency of every 15 minutes. As for the convective weather (thunderstorm), we collected weather records of 2330 stations from Global Hourly – Integrated Surface Database (ISD) for the U.S. model, and derived hourly data from 2080 surface meteorological stations in China from China Meteorological Administration, both for the entire year of 2016.

3.2 Data filtering

Multiple data filtering steps are performed to ensure data consistency. First, as we are interested in AAT of commercial flights, other types of flights are removed. Second, we are interested in flights between airports that have high levels of traffic and are thus subject to congestion. In this study, we focus on flights that flew between the top 30 airports in terms of total flight operations in each country in 2016. These airports are shown in Figure 2 and also listed in *Appendix I*. This filtering reduces the number of flight records to 1,815,416 for China and 2,531,949 in the U.S, which accounts for 31.2% and 36.8% of all the flights records in China and the U.S. respectively.

In addition to the above, records with missing or abnormal variable values are eliminated from the data. Flights with GCD less than 10 miles between the departure and the arrival airports are deleted, so that flights with extremely short distance and unusual flight operations with the same origin and destination are not included for our regression analysis. For each OD pair, we delete flight records with AAT larger than the 85th percentile value plus 1.5 times the difference between the 85th and the 15th percentile values in the AAT distribution, and flight records with AAT less than the 15th percentile value minus 1.5 times the difference between the 85th and the 15th percentile values in the AAT distribution. These deleted flights are viewed as outliers as they have exceedingly long or short AAT. Similar filtering of flight records with extremely long and short *D_GT* is also performed. After filtering, the China dataset has 1,704,232 flights, while the U.S. dataset has 2,468,941 records. The summary statistics for some of the variables are reported in Table 2.

Table 2. Summary statistics of the variables in the specified models after data cleaning (2016)

Variable	Countries	Mean	Std. Dev.	Min.	Max.
AAT	China	120.045	43.733	8.000	378.000
(min)	US	138.422	76.053	4.000	569.000
GCD	China	782.161	333.626	27.221	2184.177
(mile)	US	1030.977	640.988	11.023	2721.314
D_GT	China	0.436	0.511	0.000	19.217
(hour)	US	0.164	0.404	0.000	8.667
AF_TT_dif	China	0.383	0.925	0.000	19.417

(hour)	US	2.096	2.292	0.000	23.100
LON_dif	China	-0.040	9.833	-38.213	34.318
(degree)	US	-0.043	20.285	-51.589	51.589
CwDep	China	0.005	0.072	0.000	1.000
[0,1]	US	0.011	0.101	0.000	1.000
TempDep	China	15.795	11.299	0.343	93.257
(Fahrenheit)	US	12.766	9.777	0.000	71.975
VisDep	China	1.713	0.395	0.000	1.976
(ln mile)	US	2.328	0.268	0.000	2.398
CWindDep	China	3.414	3.074	0.000	29.228
(knots)	US	4.942	4.499	0.000	42.347
CwArr	China	0.006	0.077	0.000	1.000
{0,1}	US	0.011	0.103	0.000	1.000
TempArr	China	15.640	11.194	0.008	91.808
(Fahrenheit)	US	12.718	9.748	0.000	72.807
VisArr	China	1.729	0.381	0.000	1.976
(ln mile)	US	2.335	0.251	0.000	2.398
CWindArr	China	3.428	3.061	0.000	30.631
(knots)	US	5.033	4.530	0.000	42.347
TdEnroute	China	2.405	0.724	0.000	10.756
(flight counts)	US	2.794	1.414	0.000	20.828
CwEnroute	China	0.010	0.039	0.000	1.000
[0,1]	US	0.012	0.040	0.000	1.000
QueueDep	China	5.432	4.637	0.000	72.500
(hour)	US	5.617	5.701	0.000	105.000
QueueArr	China	2.668	3.354	0.000	58.750
(hour)	US	5.308	5.216	0.000	127.250
NearbyAPT	China	0.023	0.082	0.000	4.128
(100 flights)	US	0.158	0.529	0.000	15.009

4. Estimation results

This section presents the estimation results of the AAT model specified in Section 2 Eq. (6). Models are estimated using Ordinary Least Squares regression (OLS). To reduce the heteroscedasticity, all standard errors are clustered by OD pairs. Model results for the U.S. and China are presented in Table 4. The China and U.S. coefficients are compared and discussed as followings.

Looking at the individual coefficients, the **GCD** coefficients are 0.129 for China and 0.117 for the U.S. both with $p < 0.001$. Due to the lack of flight plan and actual trajectory data for both China and the U.S., in this research we focus on flight GCD routes between departure and arrival airports to capture the effective flight speed along the GCD routes. The effective speed describes how fast a flight is traveling along its GCD route. A low effective speed may be the result of a low aircraft speed, or due to the flight experiencing rerouting, holding or vectoring in the airspace. The GCD coefficients can be used to infer the effective flight speed along the GCD routes, by multiplying the inverse of the GCD coefficients by 60 (one hour has 60 minutes), which leads to an average effective speed of 475 mph in China and 513 mph in the U.S. The smaller effective speed in China suggests that flights fly more slowly or longer the actual distance than flights in the U.S. to traverse the same GCD. One reason is that China has a complexed airspace with many restricted and prohibited areas, leading to only 20% en-route airspace available and “zigzag” flight paths in the real-world operation, while the civil aviation use in the U.S. is over 80%. Besides, compared to the more spread-out air traffic network in the U.S., the traffic congestion concentrates more in the

eastern areas in China due to more closely-spaced airports which can affect the air traffic management over and around the regions. Thus, flight might have more detouring or vectoring in the eastern part of China.

The predeparture, *D_GT* coefficient (China: -0.912, US: -1.082) is negative with statistically highly significance for both countries. It indicates that more flight delay before takeoff has a negative effect on AAT. This confirms our hypothesis that predeparture delay will pressure a flight to fly faster or take a shorter route if operation conditions allow. However, the magnitude of this effect is quite small: if a Chinese flight had an additional predeparture delay of 20 minutes, on average it would compress AAT by less than a minute. The number is somewhat higher for a U.S. flight (1.08 minutes). The estimate suggests limited capability for flights in both countries to speed up to recover pre-takeoff delays. Possible constraining factors include airspace route structure and ATC procedures, as well as greater fuel burn if flying faster.

The next variable of interest is *AF_TT_dif*, whose coefficients are positive for both countries, but not statistically significant with very small values (China: 0.103; US: 0.055). Thus, our a priori expectation that flights with smaller turnaround time will have more pressure to reduce AAT is rejected. In other words, aircraft turnaround time has a minimum effect on AAT and can be neglected.

For the remaining operation-related variables, the *Lon_dif* variable has a highly significant coefficient in both regions (China: -0.680, US: -0.447). As China and the U.S. are on the similar latitudes of the belts of prevailing westerlies, flights flying eastward are expected to spend less time than flights flying westward. This effect is clearly observed in our model. For a given longitude difference between an OD pair, the more negative coefficient in China may imply a greater heading effect than in the US. However, this effect on effective flight speed along the GCD route is offset by the larger positive GCD coefficient in China that may arise from shorter flying distances and greater constraints by surrounding air traffic.

For the hub airport variables, China and the U.S. models report the statistically significant opposing effects (China: *HubDep* 0.720 & *HubArr* 0.247, US: *HubDep* -0.373 & *HubArr* -0.338) except *HubDep* in China. Everything else being equal, flights at hubs are likely to incur extra airborne time in China. This suggests that a higher concentration of flights from an airline and flight bunching effects for short connecting time at a hub airport could lead to longer delays and thus longer AAT. Moreover, the different estimates for departure and arrival hubs are a sign that the traffic congestion impacts at the two ends of the flight spectrum are different. On the other hand, flights into and out of hubs in the U.S. are associated with AAT reduction. One explanation is that flights in the U.S. may receive better ATC support and service in and around hub terminal airspace than in China. Another possible explanation is that an airline typically has more flights and slots at their hubs, and thus more opportunities under Traffic Management Initiatives (TMIs) for flight substitution and swap in the presence of empty slots or flight cancellation due to rescheduling or unexpected en-route emergency.

The *Flt_last* coefficient is negative for China and the U.S. (China: -0.950 $p < 0.001$, US: -0.576 $p < 0.1$). A possible explanation is that a flight, as its last flight leg of the aircraft, is typically late in a day with high aircraft utilization, the crew have the motivation to get off work earlier by flying faster. Also, flights with the last flight leg may face less congestion both on enroute and terminal airspace, as many aircraft have already ended their operation of a day.

TdEnroute and *NearbyAPT* are designed to capture impacts from airspace traffic volume on flight AATs, with the former focusing on the enroute congestion and the latter on route-adjacent terminal congestion. The coefficient of *TdEnroute* is positive in the U.S. (0.078, $p < 0.1$) as opposed to a negative coefficient in China (-0.318, $p < 0.1$). The US coefficient reports a modest positive relation between enroute traffic and AAT which is consistent with our hypothesis that flights overflying busy airspace can have excessive airborne time. However, coefficients of both models are not statistically significant and their confidence interval both include the value of zero. Hence, *TdEnroute* is not a good estimator in explaining AAT. This could be because the traffic volume in the model is calculated based flight GCD route and schedule flight time, which can deviate from the actual trajectory and real-time flight locations. Coefficient of *NearbyAPT* also reports low significance for China and insignificance for the U.S. (China: -3.054, US: 0.174) with only the sign of U.S. model consistent with our expectation. As the range and mean of *NearbyAPT* in China is rather small as in Table 2, the impacts of this variable are expected to be neglected. The bad performance of the variable can also be addressed by using the flight GCD route to identify the route nearby airports.

The last operation-related variables are related to terminal congestion effect. The arrival queueing delay variable, *QueueArr*, has a negative and statistically highly significant coefficient for both countries (China: 0.387, US: 0.307). The estimate reveals that a long queue to the arrival airport will increase flight AAT by 0.3-0.4 min per hour of cumulative queueing delay at the arrival airport for a half hour before flight actual arrival. The result is reasonable as a larger queueing delay indicates that the arrival traffic demand exceeds airport capacity leading to a high congestion level in the terminal airspace. So, a flight needs to be held in the terminal airspace or might be delayed further upstream of the terminal airspace resulting in extra airborne time. However, the departure queueing delay variable, *QueueDep*, reports negative coefficient estimates which are also statistically significant for both countries (China: -0.062, US: -0.040). The absolute values, however, are much smaller than those for *QueueArr*. The estimates suggest that departure queueing delay, which can lead to an increase in pre-takeoff ground time, has a very small effect on AAT, or flight time in the air.

Turning now to the time-related variables, the coefficients for *DayTime* are both positive in two regions (China: 1.325, US: 1.078) with highly statistical significance. The results reflect that AATs tend to be longer in the daytime as a result of increased traffic levels. Higher traffic during weekdays also leads to greater AAT. Similarly, the positive coefficients on *Weekday* (China: 0.279, US: 0.288, both $p < 0.001$) also suggests higher traffic volume during weekdays. With regard to the seasonal effects, our estimates reveal that summer and fall have shorter flight times on average than spring and winter in both countries. Seasonal effects vary more in China, with a 3.7-minute difference between the “fastest” and the “slowest” seasons, while in the U.S. the difference is 2.7 minutes.

Weather-related variables also influence AAT greatly. The most significant effect occurs when flights meeting the convective weather. The variable of enroute convective weather, *CwEnroute*, reports a positive and highly statistically significant coefficients (China: 5.888, US: 15.115) on both countries. It confirms with our expectation that flights need to detour or be delayed at the backward/upstream sectors to avoid the hazardous airspace, which leads to a much longer flight airborne time. In the U.S., a flight will have over 15 minutes more when encountering thunderstorm over all the proportion of its GCD route whereas the effect is only 5.89 minutes on the China side. One reason is that the *CwEnroute* for GCD route is a better estimator for the U.S.

practice as flights have less deviations from their GCD route compared to the situation in China route structure. The terminal convective weather not only has adverse effects on descending operation but can severely reduce terminal arrival capacity leads flights to holding patterns to avoid unsafety operation environment. So, flight intending to arrive during the convection-related period at terminal can have extra airborne time from holding or backward delay assigned by controllers. Accordingly, in our model, the **CwArr** variable presents highly significant coefficients (China: 4.865 US: 2.484). As for the convective weather at departure airport, **CwDep** generates a significant positive coefficient (China: 1.061) while the same variable is not significant on the U.S. side. Apart from the convective weather, other terminal weather factors show relatively small effects. The arrival visibility (**VisArr**, China: -0.481, US: -1.684) has a statistically significant and negative coefficients for both countries. **VisArr** presents a much larger coefficient in the U.S. but smaller for China, which can be a sign that flights in the U.S. tend to be more susceptible to the lower visibility. As for the effect of departure visibility, **VisDep** has modest effects with insignificant coefficient (China: 0.204 US: -0.287). The difference is not surprising, as convective weather or low visibility at the departure airport mainly affect ground holding time before takeoff, which is not included in AAT. In contrast, clear sky and better visibility at the arrival end help decrease separation while forming the landing queue which contributes to reducing AAT.

Model coefficients of fixed-effect variables, including aircraft, airline, and airport-specific effects, are visualized with their point estimate and 95% confidence interval (Figure 10, Figure 11, and Figure 12). First for aircraft types, Figure 10 shows that different aircraft types lead to generate different impacts on AAT behavior. Both China and the U.S. model adopt B738 as the baseline aircraft type. Overall, wide body aircraft tend to have higher speed and thus shorter AATs. A332 and A333 in the China side and B763 in the U.S. side both reports relatively smaller and significant coefficients among the major aircraft-specific variables. It is noted that the exact magnitude of aircraft-specific effect for a flight is related to its GCD distance, and for example a flight operating from PEK to SHA with GCD of 668.147 mile will have 2.67 minutes faster if flying by A332 compared to B738 in China.

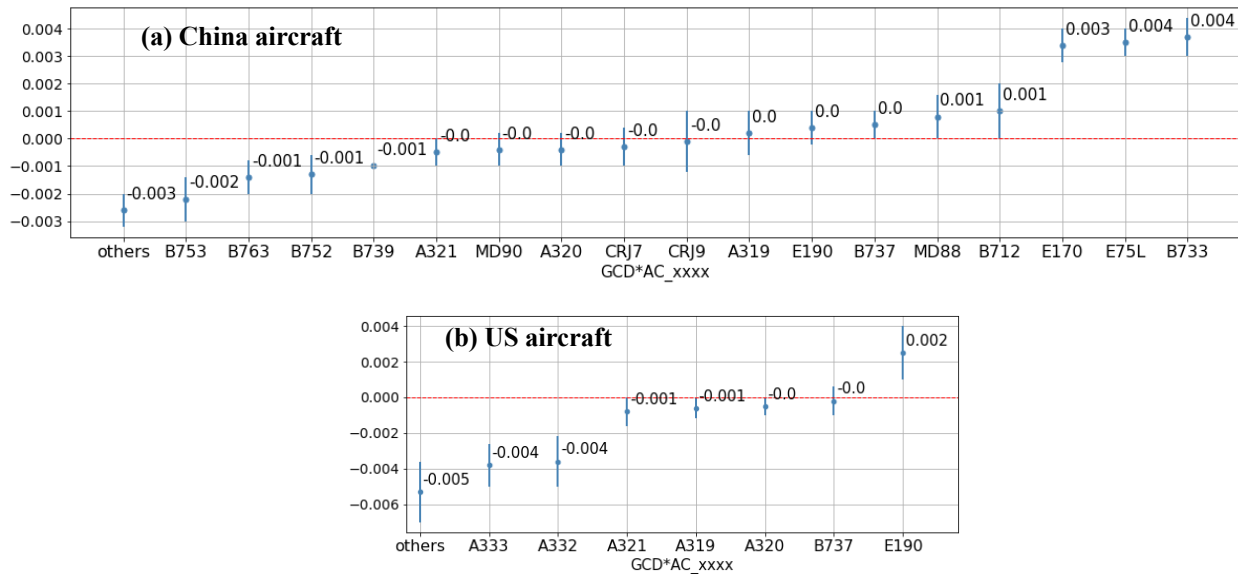


Figure 10. Coefficients with 95% CI for GCD-aircraft-specific variables for China and the U.S.

Secondly, we expect airline dummies to describe how AAT can be impacted by the airline policies, especially their schedule management. In the model, China Southern Airlines (CSN) and American Airlines (AAL) are set as baselines for China and the U.S. model, respectively. As shown in Figure 11, the coefficient difference among airline specific variables is up to 5.3 min in China while the scale is much smaller in the U.S. with range of 2 min. On the China side, flights of Shenzhen Airlines (CSZ) and Shanghai Airlines (CSH) have the relatively longer AAT on average, while flights of Sichuan Airlines (CSC) have the shorter AAT. On the U.S. side, flights of Alaska Airlines (ASA) have the relatively longer AAT on average, while flights from Southwest Airlines (SWA), United Airlines (UAL) and Delta Airlines (DAL) can perform slightly better on their AAT.

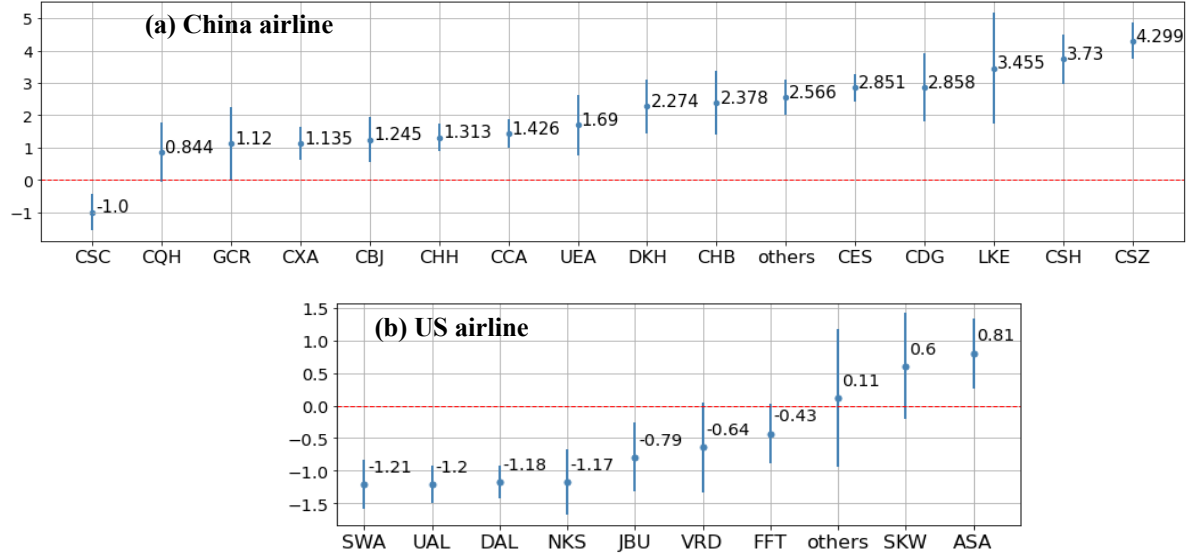


Figure 11. Coefficients with 95% CI for airline-specific variables for China and the U.S.

Third, for the airport dummies, the U.S. has more airports with significant coefficients. For both countries, arrival airports have more significant variables. On the China side, the largest AAT difference between two departure airports is almost 14 min (between SHE and NNG) as shown in Figure 12(a), while only 8.5 min (between JFK and DEN) in the U.S. model in Figure 12(c). For Chinese arrival airports in Figure 12(b), the largest AAT difference is between CAN and URC, at 9.3 min, which is shorter than the difference at the departure end. Similarly, on the U.S. side in Figure 12(d), the arrival airports only present a maximum difference of 7.3 minutes between SEA and ATL. We also find that airports around the New York area are more likely to experience extra airborne time, probably because flights can face with large traffic volume and complex ATC environment near the New York area. On the China side, airports in the northeastern area in China such as SHE, TAO and DLC tend to have more airborne time. Flights between northeast and eastern areas of China (typically, Shanghai, Nanjing, Hangzhou, Qingdao) need to cross the Bohai sea which has a very limited airspace (for example, only one main enroute segment from DLC to TAO). Therefore, these flights are usually congested enroute over the sea, leading to a longer airborne time. This also causes the flights at TAO located at the end of this en-route segment, to be affected severely during their landing and climbing procedure.

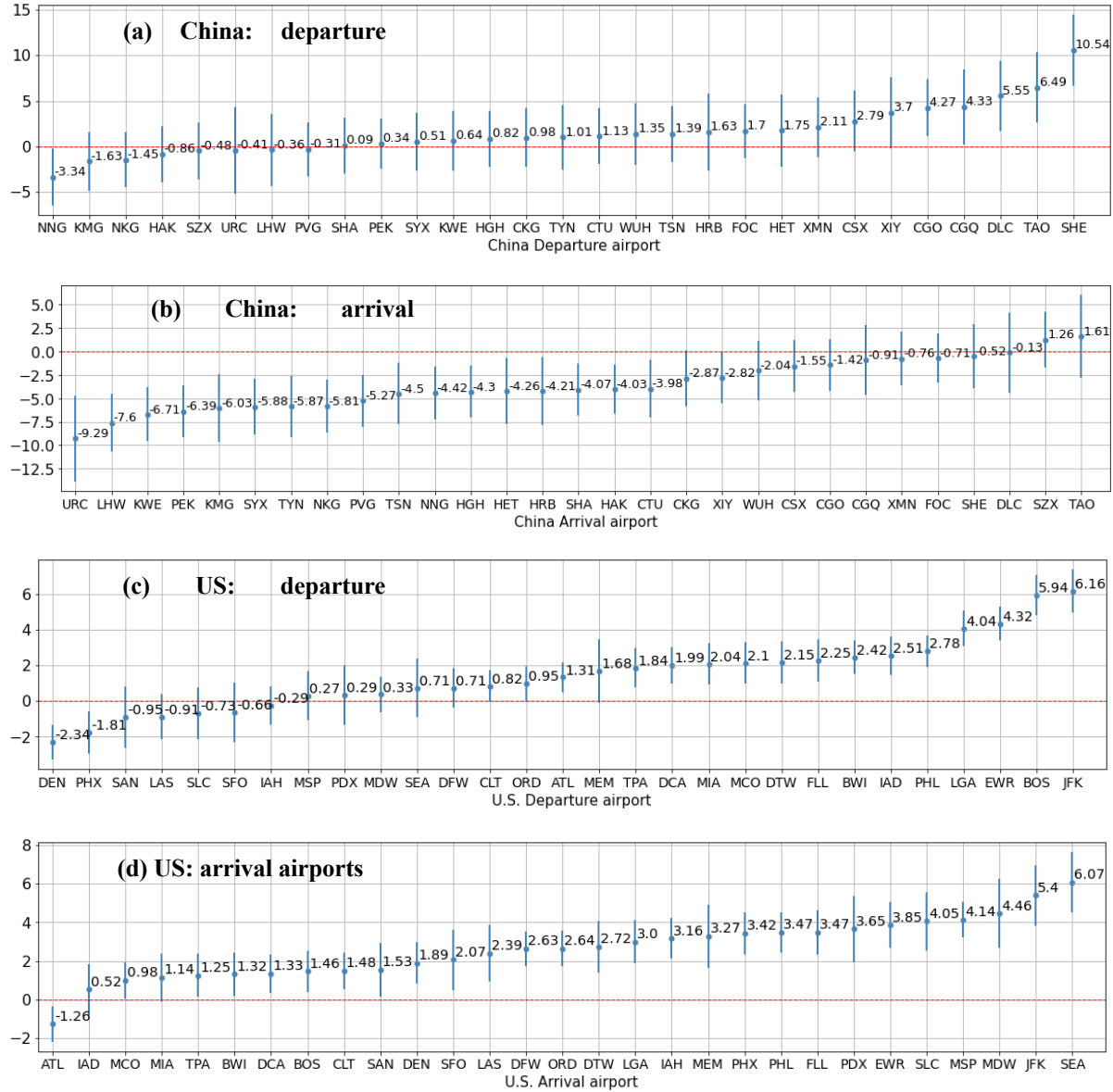


Figure 12. Coefficients with 95% CI for airport-specific variable for China and the U.S.

Besides the above coefficients, the estimates for the constant term are also worth discussions. The meaning of the constant not only reflects the additional time a flight spends on acceleration and deceleration, but also quantifies the average value of the model fixed effects for the benchmark airline and airport. The estimates are 19.4 minutes for Chinese flights and 18.1 minutes for U.S. flights.

A last point to notice is R^2 's of the models. Models in both regions report good performance, and the U.S. side has a slightly greater goodness-of-fit than the China side. The larger R^2 on the U.S. side can be attributed to two facts. First, the higher airspace availability in the U.S. provides higher en-route capacity and allows flights to deviate less from GCD routes, which also contributes to better explanatory power of the independent variables (e.g., *GCD*, *TdEnroute*, *NearbyAPT*, and *CWEroute*). Second, the U.S. has an effective and comprehensive air traffic management system

which includes and integrates strategical and tactical initiatives such as Ground Delay Program (GDP), Airspace Flow Program (AFP), Mile-in-Trail (MIT), and Time-Based Flow Management (TBFM). These TMIs with Collaborative Decision Making (CDM) can efficiently assign more proactive delays to the flights even before departure and allows airline to swap, substitute and cancel flights to meet their schedule and help reduce AAT fluctuations. In contrast, GDP was not implemented in China in 2016. Chinese air traffic management held incoming flights at their departure airports if there was foreseen airport or en-route congestion, but flights were not assigned with a Control Time of Departure when being held at the airports. As a result, AATs are somewhat less predictable in China.

Table 3. Base models for AAT behavior in China and the U.S.

	China: OLS model				The U.S.: OLS model			
	coef	std err	[0.025	0.975]	coef	std err	[0.025	0.975]
<i>Cons</i>	19.433 ***	2.034	15.446	23.419	18.138 ***	0.964	16.249	20.027
<i>GCD</i>	0.129 ***	0.001	0.128	0.131	0.117 ***	0.000	0.117	0.118
<i>LON_dif</i>	-0.680 ***	0.024	-0.727	-0.633	-0.447 ***	0.009	-0.465	-0.403
<i>Spring</i>	0.977 ***	0.199	0.587	1.368	-0.561 ***	0.209	-0.970	-0.152
<i>Summer</i>	-1.507 ***	0.175	-1.850	-1.164	-2.722 ***	0.381	-3.469	-1.974
<i>Fall</i>	-2.678 ***	0.286	-3.24	-2.117	-1.883 ***	0.274	-2.42	-1.347
<i>DayTime</i>	1.325 ***	0.195	0.943	1.706	1.078 ***	0.088	0.906	1.250
<i>Weekday</i>	0.279 ***	0.029	0.221	0.337	0.288 ***	0.025	0.239	0.337
<i>D_GT</i>	-0.912 ***	0.097	-1.102	-0.723	-1.082 ***	0.055	-1.189	-0.974
<i>AF_TT_dif</i>	0.103 **	0.047	0.010	0.196	0.055	0.063	-0.068	0.179
<i>Flt_last</i>	-0.950 ***	0.139	-1.222	-0.678	-0.576 *	0.348	-1.258	0.106
<i>TdEnroute</i>	-0.318 *	0.189	-0.688	0.051	0.078 *	0.046	-0.012	0.168
<i>CwEnroute</i>	5.888 ***	0.675	4.565	7.211	15.115 ***	0.808	13.532	16.698
<i>NearbyAPT</i>	-3.054 **	1.221	-5.448	-0.659	0.174	0.154	-0.128	0.476
<i>HubDep</i>	0.720 ***	0.208	0.313	1.127	-0.373 **	0.158	-0.683	-0.062
<i>HubArr</i>	0.247	0.210	-0.166	0.659	-0.338 *	0.174	-0.680	0.004
<i>QueueDep</i>	-0.062 ***	0.011	-0.084	-0.04	-0.040 ***	0.006	-0.051	-0.028
<i>QueueArr</i>	0.387 ***	0.017	0.353	0.421	0.307 ***	0.011	0.285	0.328
<i>CwDep</i>	1.061 ***	0.159	0.750	1.373	-0.287 *	0.164	-0.609	0.034
<i>CwArr</i>	4.865 ***	0.204	4.465	5.265	2.484 ***	0.180	2.131	2.836
<i>VisDep</i>	0.204 **	0.091	0.026	0.382	-0.189 *	0.099	-0.384	0.005
<i>VisArr</i>	-0.481 ***	0.132	-0.738	-0.223	-1.684 ***	0.122	-1.924	-1.444
<i>TempDep</i>	0.028 ***	0.008	0.012	0.045	0.001	0.004	-0.007	0.009
<i>TempArr</i>	0.010	0.010	-0.010	0.03	-0.019 ***	0.004	-0.027	-0.011
<i>CWindDep</i>	0.003	0.011	-0.019	0.024	0.018 ***	0.006	0.007	0.029
<i>CWindArr</i>	0.068 ***	0.010	0.048	0.088	0.037 ***	0.006	0.026	0.049
<i>Fixed effects (estimates for airline, aircraft and airport dummies are not presented here to save space)</i>								
Num. Obs.	1704232				2468688			
R ²	0.957				0.986			

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$

5. Sensitivity analysis

Recalling the discussion in Section 4, the GCD coefficient in our model can reflect the effective speed about how fast a flight is traveling along its GCD route. Flights with longer GCD are likely to be affected by more interactions with the en-route airspace environment, which can lead to excess airborne time. We expect that this phenomenon can differ significantly in the U.S. and China, based on the coefficient of GCD derived in the general model as shown in Table 3. For aircraft utilization, propensity of delay propagation from an earlier flight to a later flight for higher aircraft utilization can be more significant for on time performance and passenger connection problem, so we conjecture that airlines and pilot might take different strategies on AAT behaviors, such as taking more direct. Thus, in this section, we investigate the sensitivity of the estimation results on flight length and aircraft utilization which is measured as the number of flights flown by the same aircraft per day.

5.1 Sensitivity to flight length

We divide the flight records in each country into three groups based on GCD values: 1) less than 800 miles as short-haul flights; 2) between 800 and 1,600 miles as medium-haul flights; 3) greater than 1,600 miles as long-haul flights. Figure 13 shows the number of flights in each group in the two countries. As the top 30 airports in China are concentrated in the eastern part of the country (see Figure 2), most flights are short haul. The busiest routes, notably Beijing-Shanghai Pudong (PEK-PVG), Shanghai Hongqiao-Shenzhen (SHA-SZX), and Beijing-Guangzhou (PEK-CAN), all fall into the short-haul category. Long-haul flights are a very small proportion of the total. In contrast, the top 30 airports in the U.S. are more dispersed across the nation; flights are more evenly distributed among the three flight groups. Among the busiest routes, New York (JFK)-Los Angeles (LAX) is a long-haul route, Los Angeles (LAX)-San Francisco (SFO) is a short-haul route, and Los Angeles (LGA)-Chicago (ORD) is a medium-haul route. The results are reported in Table 4 with standard errors clustered by OD pair.

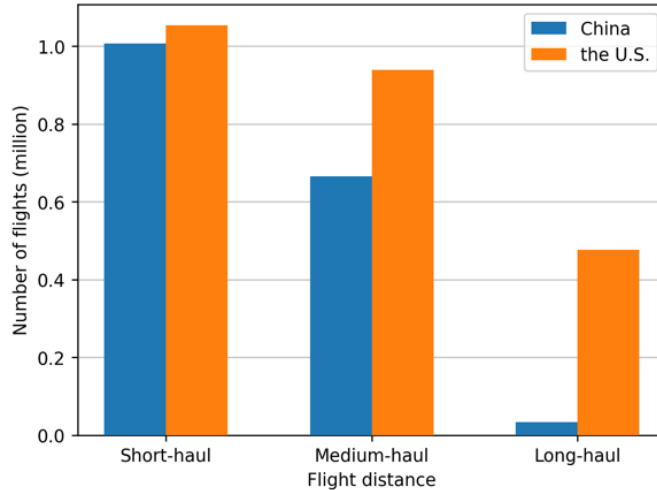


Figure 13. Flight distribution by flight length in China and the U.S.

For AAT behavior, we focus on comparing the **GCD** and **D_{GT}** coefficients in these models. Figure 14 plots the point estimates of **GCD** and **D_{GT}** coefficients for short-, medium-, and long-

haul flights in the two countries along with the 95% confidence intervals. Figure14(a) shows that the point estimates of the **GCD** coefficients remain unchanged with flight length increasing both in China and the U.S. The effective GCD speed of Chinese flight is slower and with wider confidence interval compared to the U.S. flights in any flight length group. It suggests that the effective GCD speed is not affected by the flight length in the U.S. However, the confidence interval for China, especially for long-haul flights, presents a rather wide distribution, suggesting the GCD speed range from 395 mph to 556 mph. It indicates that long-haul flights in China have more variant operation environments in terms of enroute traffic density, nearby airport impacts and airspace constraints. For example, flights at Urumqi in the West or with long North-South routes linking can be less affected by the busy airports in China they fly over.

Figure14(b) shows different trends in the relationships between the **D_GT** coefficient and flight length for China and the U.S. The **D_GT** coefficient on the U.S. side slightly increase in absolute value with flight length. A possible explanation is that longer flights in the U.S. are more likely to go through low-density airspace in which direct routings are more likely to be permitted. In China, the **D_GT** coefficients increase from short- to median-haul and decrease from median- to long-haul. Similar to GCD coefficient variation, the confidence interval become wide in the long-haul. Specifically, the lower bound of the confidence interval present a decrease trend. It indicates that **D_GT** effect in China grows with flight length as U.S. flights in some cases. However, **D_GT** effect is not stable on the China side, due to its complex airspace and operations that long-haul flights may face. Overall, we conclude that it is the distinct nature of the en-route flying environment in China and the U.S. that results in the different changing trends of the coefficients for **D_GT** as a flight fly longer, especially for long-haul.

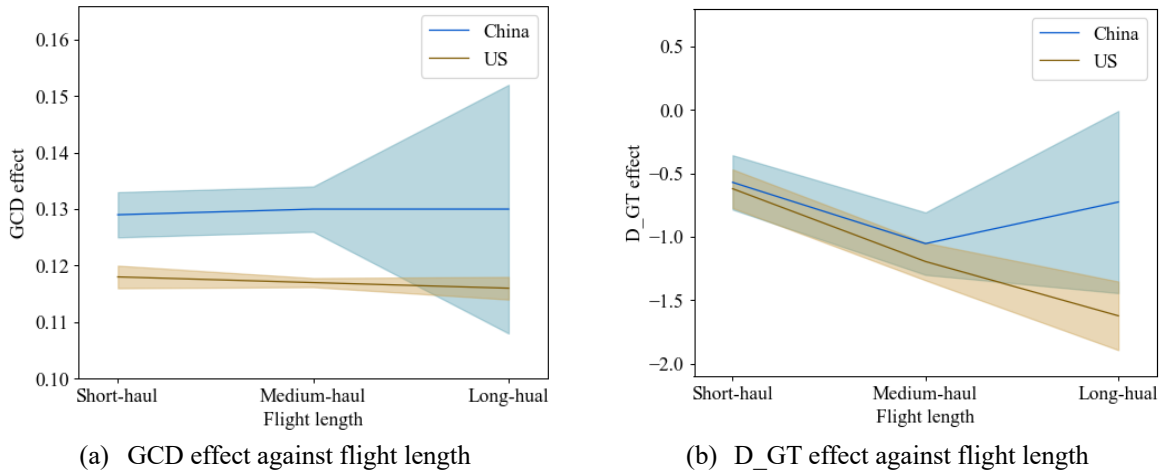


Figure 14. Delay recovery effect variation with different flight length

Table 4. Estimation results of AAT for different flight length for China and the U.S.

	Short-haul		Medium-haul		Long-haul	
	China	U.S.	China	U.S.	China	U.S.
<i>Cons</i>	17.503***	15.856***	18.552***	17.374***	2.533	23.410***
<i>GCD</i>	0.129***	0.118***	0.130***	0.117***	0.130***	0.116***
<i>LON_dif</i>	-0.701***	-0.418***	-0.689***	-0.467***	-0.783***	-0.422***

<i>Spring</i>	0.687***	-0.555***	1.453***	-0.663**	0.618	-0.755
<i>Summer</i>	-1.285***	-1.644***	-1.825***	-3.092***	-3.524***	-4.848***
<i>Fall</i>	-1.885***	-0.965***	-3.702***	-2.098***	-6.040**	-3.810***
<i>DayTime</i>	1.154***	0.964***	1.895***	0.642**	2.540*	0.927***
<i>Weekday</i>	0.286***	0.139***	0.290***	0.342***	0.237	0.314***
<i>D_GT</i>	-0.570***	-0.620***	-1.053***	-1.194***	-0.725**	-1.622***
<i>AF_TT_dif</i>	0.028	-0.090	0.104	0.156***	0.557*	0.261***
<i>Flt_last</i>	-0.619***	0.285	-1.377***	-1.294***	-0.204	-1.205***
<i>TdEnroute</i>	-0.280	0.170***	-0.369	0.318**	0.547	0.603***
<i>CwEnroute</i>	5.855***	9.231***	7.701***	21.136***	7.455	28.978***
<i>NearbyAPT</i>	-2.427**	0.048	-4.058	-0.036	3.119	0.135
<i>HubDep</i>	0.573***	0.137	0.203	-0.168	-0.345	-0.071
<i>HubArr</i>	-0.178	0.189	-0.089	-0.273	-2.059*	-0.557*
<i>QueueDep</i>	-0.041***	-0.014**	-0.084***	-0.047***	0.081	-0.076***
<i>QueueArr</i>	0.280***	0.224***	0.484***	0.362***	0.600***	0.363***
<i>CwDep</i>	0.643***	0.376**	1.578***	-0.054	-3.083**	-3.138***
<i>CwArr</i>	4.430***	1.115***	5.160***	2.314***	10.795***	6.373***
<i>VisDep</i>	0.168*	0.115	0.195	-0.530***	1.797***	-0.393
<i>VisArr</i>	0.034	-1.967***	-0.708***	-1.475**	-4.378***	-1.389***
<i>TempDep</i>	0.014	-0.013***	0.048***	-0.017***	-0.02	0.046***
<i>TempArr</i>	0.005	-0.001	-0.002	-0.002	0.083*	-0.083***
<i>CWindDep</i>	0	0.014**	0.004	0.003	0.063	0.057***
<i>CWindArr</i>	0.077***	0.040***	0.061***	0.062***	0.005	-0.009
<i>Fixed effects (estimates for airline, aircraft and airport dummies are not presented here to save space)</i>						
Num. Obs.	1,006,238	1,053,470	664,746	938,218	33,248	477,000
R ²	0.815	0.870	0.820	0.870	0.745	0.918

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

5.2 Sensitivity to daily flight leg

We measure aircraft utilization as the number of flight legs the aircraft associated with the flight under study flies in a day and denote the number as *NumFltOfAC*. As mentioned at the beginning of this section, flights may increase the effective GCD speed in order to mitigate the increased propensity for delay propagation from an earlier flight to a later flight. Figure 15 illustrates the distribution of flights in seven groups based on *NumFltOfAC*: 1, 2, 3, 4, 5, 6, and >6. Over 70% flights are associated with aircraft with 4-6 legs per day in both countries.

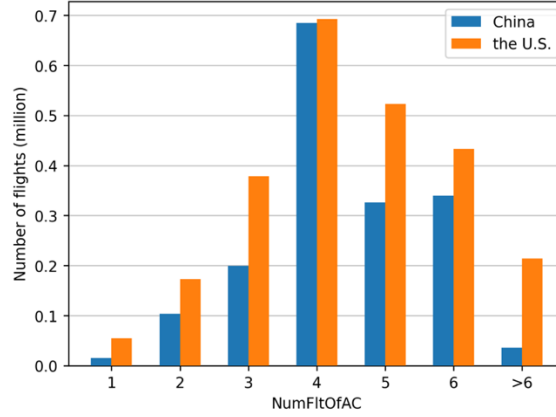


Figure 15. Distribution of the number of legs flown by aircraft

Table 5 reports the OLS estimation results of the AAT model for each of the seven flight groups for each country, with all standard errors clustered by OD pair. As in subsection 5.1, the focus of our analysis is on the coefficients for GCD and D_GT , for which point estimates and 95% confidence intervals are plotted in Figure 16. In Figure 16(a), the GCD coefficient remains almost unchanged in any $NumFltOfAC$, which means that flight effective GCD speed for flights is not significantly impacted by the $NumFltOfAC$. It implies that flights may not be able to keep on-time performance by accelerating their effective speed in general. One possible explanation is that the flight speed and route is limited by the air traffic controller. To keep the traffic flow stable and decrease the ATC workload, large adjustments in speed and trajectory may be not allowed in air traffic networks.

Figure 16(b) shows the D_GT coefficient estimates as a function of $NumFltOfAC$. Consistent with the base model estimates in Section 4, the ability of a flight to recover departure delay is limited in both countries. The overall trend is an increase in the absolute value of the D_GT coefficient until $NumFltOfAC$ equals to 4, then a decrease with aircraft utilization increasing. There is a tradeoff between $NumFltOfAC$ and flight distance as with the increase on $NumFltOfAC$, flights tend to be shorter and to have less opportunity to take a more direct route or speed up enroute. Overall, flight operators are able to respond albeit to a modest extent, when faced with a greater need to recover flight time because of the threat of delay propagation. But this effect will reduce if an aircraft is under a higher utilization.

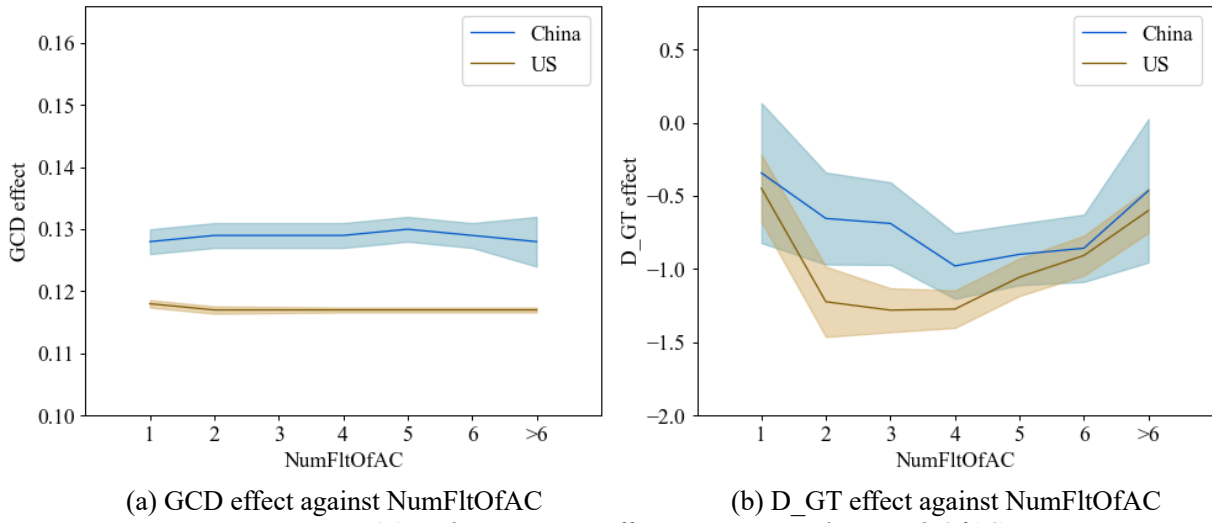


Figure 16. Delay recovery effect variation with NumFltOfAC

Table 5. Estimation results of AAT in different NumFltOfAC of China and the U.S.

China	1	2	3	4	5	6	>6
Cons	8.812***	16.377***	16.387***	20.425***	19.373***	21.317***	20.783***
GCD	0.128***	0.129***	0.129***	0.129***	0.130***	0.129***	0.128***
LON_dif	-0.710***	-0.696***	-0.674***	-0.700***	-0.650***	-0.625***	-0.607***
Spring	0.827*	1.086***	1.156***	1.241***	0.865***	0.528***	0.172
Summer	-1.364***	-1.640***	-1.378***	-1.639***	-1.497***	-1.372***	-1.167***
Fall	-2.504***	-3.164***	-3.053***	-2.976***	-2.510***	-2.036***	-1.473***
DayTime	1.291**	1.461***	1.394***	1.090***	0.849**	1.374***	1.662**
Weekday	0.119	0.256***	0.247***	0.300***	0.276***	0.263***	0.214**
D_GT	-0.343	-0.654***	-0.688***	-0.978***	-0.899***	-0.857***	-0.464*
AF_TT_dif	0	-0.068	0.030	0.160*	0.212**	0.31	-0.069
Flt_last	8.812***	-1.069***	-0.857***	-1.286***	-0.981***	-0.859**	-0.849
TdEnroute	-0.005	0.234	0.097	-0.354*	-0.469**	-0.180	0.035
CwEnroute	7.957***	8.755***	6.356***	6.205***	5.948***	4.931***	2.607
NearbyAPT	-14.129	-4.773	-3.603	-4.603**	-3.927***	-1.733**	-1.852
HubDep	1.202*	0.553	0.49	0.800***	0.738***	0.452	-1.067
HubArr	0.247	0.724	0.443	0.564**	0.014	-0.532*	-1.02
QueueDep	-0.058*	-0.046***	-0.066***	-0.076***	-0.067***	-0.018	-0.038
QueueArr	0.320***	0.382***	0.369***	0.438***	0.348***	0.329***	0.302***
CwDep	0.693	0.935*	1.154***	1.355***	1.069***	0.542**	0.598
CwArr	5.385***	5.648***	5.559***	4.901***	4.733***	4.285***	3.714***
VisDep	0.540*	0.770***	0.579***	0.157	0.235**	-0.013	-0.044
VisArr	-1.351***	-1.083***	-0.823***	-0.614***	-0.224	0.011	0.036
TempDep	0.047**	0.041***	0.047***	0.032***	0.025**	0	-0.029*
TempArr	0.002	0.004	-0.003	0.011	0.004	0.025***	0.052***
CWindDep	-0.048	-0.008	0.010	-0.007	-0.005	0.019	0.027
CWindArr	0.085**	0.082***	0.083***	0.067***	0.064***	0.052***	0.050*

Fixed effects (estimates for airline, aircraft and airport dummies are not presented here to save space)

Num. Obs.	15,653	103,328	199,378	684,636	326,087	339,714	35,436
R ²	0.966	0.963	0.964	0.952	0.948	0.936	0.937

U.S.	1	2	3	4	5	6	>6
Cons	16.490***	17.241***	18.950***	17.574***	17.970***	16.785***	15.893***
GCD	0.118***	0.117***	0.117***	0.117***	0.117***	0.117***	0.117***
LON_dif	-0.503***	-0.441***	-0.436***	-0.444***	-0.449***	-0.451***	-0.435***
Spring	0.066	-0.216	-0.232	-0.539**	-0.697***	-0.772***	-0.601***
Summer	-2.119**	-3.206***	-3.139***	-2.943***	-2.768***	-2.243***	-1.824***
Fall	-1.682***	-2.678***	-2.320***	-1.982***	-1.771***	-1.475***	-1.031***
DayTime	1.138***	1.368***	0.807***	0.703***	1.014***	0.895***	0.685***
Weekday	0.727***	0.483***	0.343***	0.323***	0.275***	0.220***	0.122**
D_GT	-0.448***	-1.223***	-1.280***	-1.273***	-1.055***	-0.907***	-0.600***
AF_TT_dif	-0.28	0.026	0.077**	0.083**	0.072*	0.071	0.033
Flt_last	2.931	-0.279	-0.834***	-0.925***	-0.975***	-0.851**	-0.563
TdEnroute	0.352**	0.179**	0.171**	0.155***	0.094**	0.118***	0.072*
CwEnroute	12.160***	18.060***	21.539***	17.884***	14.447***	11.286***	7.497***
NearbyAPT	0.717*	0.579**	0.537**	0.007	-0.055	-0.033	0.11
HubDep	-1.072**	0.268	-0.246	-0.173	-0.112	-0.195	0.165
HubArr	-0.599	0.028	-0.694***	-0.11	-0.234	-0.210	0.043
QueueDep	-0.002	-0.079***	-0.058***	-0.034***	-0.030***	-0.027***	-0.014*
QueueArr	0.261***	0.353***	0.341***	0.318***	0.294***	0.262***	0.244***
CwDep	-0.515	-1.312***	-0.604**	-0.322	-0.097	0.109	0.1
CwArr	2.482***	3.922***	3.162***	2.853***	2.221***	1.592***	1.292***
VisDep	-0.247	0.05	-0.19	-0.337***	-0.297***	0.06	-0.04
VisArr	-2.389***	-2.154***	-1.658***	-1.457***	-1.614***	-1.818***	-1.498***
TempDep	0.021*	0.035**	0.017**	0.001	-0.010***	-0.013***	-0.009*
TempArr	0.004	-0.048***	-0.022***	-0.017***	-0.020***	-0.014***	-0.008
CWindDep	0.006	0.033**	0.009	0.016**	0.018***	0.014**	0.030***
CWindArr	0.080***	0.02	0.038***	0.037***	0.035***	0.039***	0.039***

Fixed effects (estimates for airline, aircraft and airport dummies are not presented here to save space)

Num. Obs.	54,822	172,903	378,320	692,675	523,077	433,196	213,695
R ²	0.985	0.985	0.984	0.982	0.982	0.981	0.984

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

6. Counterfactual analysis of potential efficiency gains in China

Finally, as our study attempts to compare the AAT behavior in China and the U.S., recognizing the operational differences between the two countries is important. It is conjectured that the U.S. has a more efficient air traffic system due to higher airspace availability and a more flexible routing network which allow aircraft to adjust speed and route. Additionally, in the U.S. strategic traffic management initiatives, such as GDPs and AFPs, are used to shift delay from air to the airport surface, reducing the need to absorb delay while airborne. This chapter examines whether and to

what extent the AAT behavior is different between the two countries. Counterfactual analysis is conducted to understand the implications for systemwide efficiency gains if the AAT behavior in China became the same as in the U.S. In addition, given the difference in the behavior of AAT between China and the U.S., this section also investigates how AAT would change and the implications for airline fuel consumption, CO_2 emission and operating cost if AATs in China followed the U.S. model. This brings the last hypothesis of the chapter that provided that the air traffic system in China had the same operation environment as in the U.S., Chinese flights would have overall shorter AATs leading to operating cost and fuel savings.

6.1 Methodology

The basic idea for performing the counterfactual analysis is to apply the variable coefficient estimates from the U.S. model to the China flight data. However, it is worth noting that some variables, including the airline dummies, airport dummies, *Lon_dif* and constant, may not be transferable from one country to the other. This is because the inherent characteristics of Chinese airlines and their scheduling strategies would likely persist even if the air operation environment in China became the same as that of the U.S. Also, departure and arrival airport dummies capture airport-specific characteristics and their impacts on terminal traffic management and traffic congestion level. The *Lon_dif* effect is location specific. Thus, its coefficient for China should be kept. Last but not least, the constant term quantifies the average value of the model fixed effects, which includes the effects of base airline – CSC, base departure airport – CAN and base arrival airport – CAN while constructing the airline and airport dummies in China. Additionally, most of the aircraft types (e.g., B738, A320, A321, A319, and B737) selected as dummies in the China model can be matched in the U.S. model, and A333, A332 and “others” in China model will be treated as “others” in the U.S. model.

6.2 Results

Figure 17 provides the distribution of AAT change over all Chinese flights if the AAT behavior in the U.S. were adopted. Consistent with *a priori* expectation, most of the flights would experience a reduction in AAT. The average AAT change is a 13.05-minute reduction with a standard deviation of 9.98 minutes. From the qualitative perspectives in flight operation conditions, the counterfactual analysis simulates these Chinese flights in the U.S. air traffic system. First, over 80% airspace in the U.S. is available for civil aviation, compared with 20% in China. Chinese airports are mostly concentrated in the eastern of the country as shown in Figure 3(a). Although the U.S. network has some similar situations such as New York area and Florida area, the flight density of Chinese eastern area should be higher during peak hours. Second, the U.S. airlines have more control over flight rescheduling and route planning. Moreover, compared with the air traffic management system in China, the U.S. side not only has more comprehensive management functions among air traffic managers, controllers, dispatchers, and pilots, but has also developed a much more effective set of TMIs including GDP, MIT and TBFM to balance demand and capacity and to shift foreseen delays from the air to the ground. Airlines can also respond by cancelling or swapping flights to meet the schedule needs. In contrast, GDP was not systematically implemented in China as of 2016. Although Chinese air traffic management could hold incoming flights at departure airports if there was predicated congestion en route or around the arrival terminal airspace, flights were not

assigned with a Control Time of Departure when being held at the airports, which limited the airlines' capability of responding to changing conditions and adjusting their flight schedule to avoid double delay penalty. Therefore, the AAT reduction is not necessarily caused higher cruising speed, but can be resulted from more abundant airspace, more flexible and direct route planning, less vectoring and holding delay due to imbalance between demand and capacity, less unnecessary enroute delay, better communication among airline, pilot and controller, and resilience against weather or system emergency.

The average AAT reduction of 13.05 minutes per flight should not be directly interpreted as the pure result of a higher average cruise speed but also attributed more to greater system-level operation efficiency. For the former, we compare the cost index for all aircraft types of U.S. and Chinese airlines under Star Alliance as found in Star Alliance Virtual (2021). The cost index is a setting on the airline flight management system that specifies the ratio of time cost (in \$/hour) and fuel cost (in cents per pound), and effectively determines the target cruise speed. We find that the cost index is the same among all Star Alliance U.S. and Chinese airlines for 56 out of 58 aircraft types. On the other hand, as mentioned in Section 1, underlying the AAT differences are several fundamental differences in the two countries' air traffic systems. First is the airspace availability. In the U.S., 80% of the airspace is available for civil aviation use while the percentage is only 20% in China. Second is greater control by airlines of flight rescheduling and route planning by U.S. airlines. Third, the U.S. system has developed a more comprehensive and mature suite of TMIs than China, including GDP, AFP, and TBFM, to balance flight demand and airspace capacity and to shift airborne delay to the surface. The result is more flexible and direct routing, less vectoring and airborne holding, and possibly greater resilience against weather and other system uncertainties.

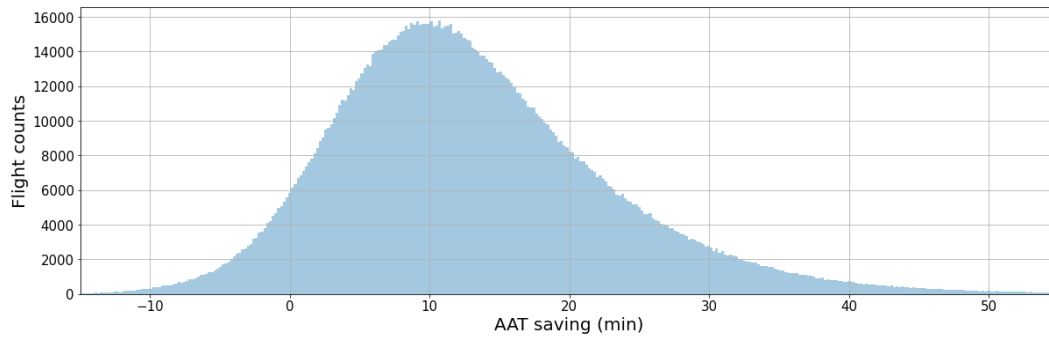


Figure 17. Distribution of AAT change among Chinese flights if AAT behavior in the U.S. were adopted

The above estimates of AAT reduction can be used to assess the potential savings in airline fuel, CO_2 emission, fuel cost, and direct operating cost. Fuel consumption for each flight is calculated based on aircraft types according to FAA's aircraft operating costs documentation (2016) and Statistical Data on Civil Aviation of China (2017), while CO_2 emission is calculated according to Greenhouse Gas Inventory Guidance: Direct Emissions from Stationary Combustion Sources (2016). Table 7 reports the calculated fuel savings per flight and in total for the four largest airlines (Air China, China Eastern, China Southern, and Hainan Airlines) and the rest combined named as minor airlines. On average, 191.08 gallons of fuel would be saved per flight, with a systemwide saving of 325.64 million gallons during 2016. Correspondingly, 2.7 million tons of CO_2 would be reduced. The largest benefiting airline would be China Southern which has the largest number of flights in the dataset (over 20% in the total), whereas Air China would save the most on a per-

flight basis. Using the average fuel price of \$1.45 per gallon in China in 2016, the fuel saving translates to \$472.18 million cost savings. Note that this does not include savings from the ability of U.S. airlines to fly at more fuel-efficient altitudes as airspace congestion in China forces some flights to fly at lower altitudes where they must burn more fuel.

The reduction in AAT saves not only fuel consumption and CO_2 emission, but also other cost related to flight operations, as the related flight operation costs, such as fuel cost, aircraft depreciation, crew working time, and aircraft maintenance and rentals, should decrease with the flight time. We name these costs as direct operating cost and assume the direct operating cost and fuel cost are both perfectly positive correlated with the flight time. So, to estimate the direct operating cost, we multiply fuel cost saving by the ratio of fuel cost to the total direct operating cost of the airlines based on their financial statements. For Air China, China Eastern, China Southern, and Hainan Airlines, the relevant cost information can be obtained from their annual income statements (Air China LLC., 2016; China Eastern LLC., 2016; China Southern Airlines LLC., 2016; Hainan Airlines LLC., 2016). For the other airlines, the ratio is based on the average of the ratios of Juneyao Airlines and Spring Airlines, the only two airlines for which cost information is available to us (Juneyao Airlines LLC., 2016; Spring Airlines LLC., 2016). The detail calculation on the direct operating cost ratio is provided in *Appendix 5*. As shown in Table 7, the ratios are quite consistent across airlines, in the range of 0.33-0.37. The resulting total direct operating cost saving amounts to around \$1.32 billion, again with the largest saving reaped by China Southern, close to \$283 million in 2016. Figure 18 shows the top 10 airlines in terms of the total direct operating cost saving. Finally, recalling that the flight data are filtered within the top 30 commercial airports and do not include all flights in China, the original data contain 3,135,075 domestic commercial flights. As a first-order extrapolation, the nationwide saving of total direct operating cost would amount to about \$4.4 billion for the year 2016.

Table 6. Estimates of time, fuel, fuel cost, and direct operating cost saving for Chinese airlines

Airline	Time saving (min)		Fuel saving (gallons)		CO_2 emission reduction (m kg)	Fuel cost saving (\$m)	Ratio	Direct operating cost saving (\$m)
	Per flight	Total (million)	Per flight	Total (million)				
Air China	13.89	2.69	225.18	43.63	362.61	63.27	0.331	191.15
China Eastern	12.60	3.22	187.10	47.85	397.60	69.38	0.340	204.05
China Southern	13.55	4.80	204.07	72.24	600.33	104.75	0.370	283.11
Hainan Airline	13.67	1.94	223.47	31.69	263.38	45.96	0.346	132.82
Minor Airlines (184 airlines)	13.00	9.59	172.43	130.23	1082.13	188.83	0.359	525.98
Total	13.05	22.25	191.08	325.64	2706.04	472.18	/	1315.27

Notes:

1. The ratio for the minor airlines is derived from the average value of Juneyao Airlines and Spring Airlines.
2. As the major airlines hold direct or indirect interests in their subsidiary airlines, the same ratio of fuel cost to direct operating cost is adopted. Below is the correspondence between major and subsidiary airlines:
 - a) Air China: Shenzhen Airline, Kunming Airline, Air Macau, Dalian Airlines and Air China Inner Mongolia, Shandong Airline, Tibet Airlines;
 - b) China Eastern: Shanghai Airlines, China United Airlines;
 - c) China Southern: Xiamen Airlines, Chongqing Airlines, Sichuan Airlines;

- d) Hainan Airline: Air Chang'an, Beijing Capital Airlines, Fuzhou Airlines, Grand China Air, Lucky Air, Tianjin Airlines, Urumqi Airlines.

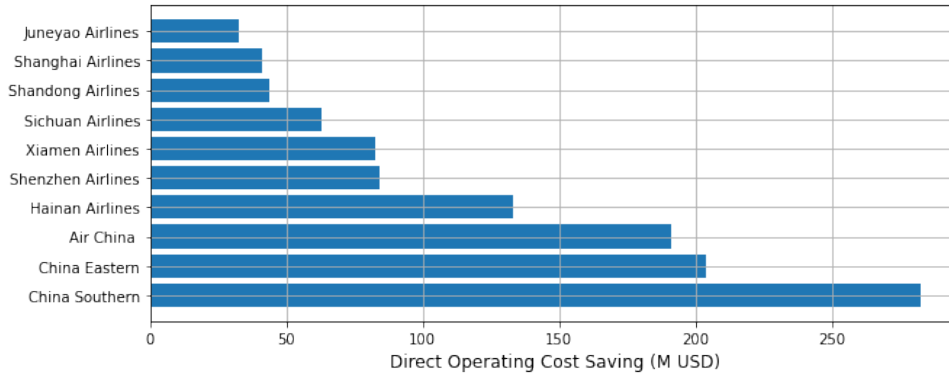


Figure 18. Direct operating cost saving for the top 10 airlines in China

We further compute the average AAT change for routes that had at least 200 flights in 2016 which reduces the number of OD pairs from 823 to 711. Routes with too few flights may have been scheduled on a more *ad hoc* basis, and thus less representative for the purpose of route-level analysis. Figure 19 visualizes the average AAT change by route. The lines in Figure 19(a) indicate routes with a decrease in AAT, whereas the lines in Figure 19(b) depict the routes with an increase in AAT. The thickness of each line corresponds to the extent of AAT change. In total, 704 routes would experience reduced AAT. The largest amount of AAT reduction would occur to Ürümqi Diwopu International Airport (URC) → Shenzhen Bao'an International Airport (SZX), Sanya Phoenix International Airport (SYX) → Harbin Taiping International Airport (HRB) and Ürümqi Diwopu International Airport (URC) → Shanghai Pudong International Airport (PVG), all with an average flight AAT saving of 29 minutes. In contrast, only 7 routes would experience increased AAT. The largest amount of AAT increase would occur to Dalian Zhoushuizi International Airport (DLA) → Qingdao International Airport (TAO) route, by 5 minutes.

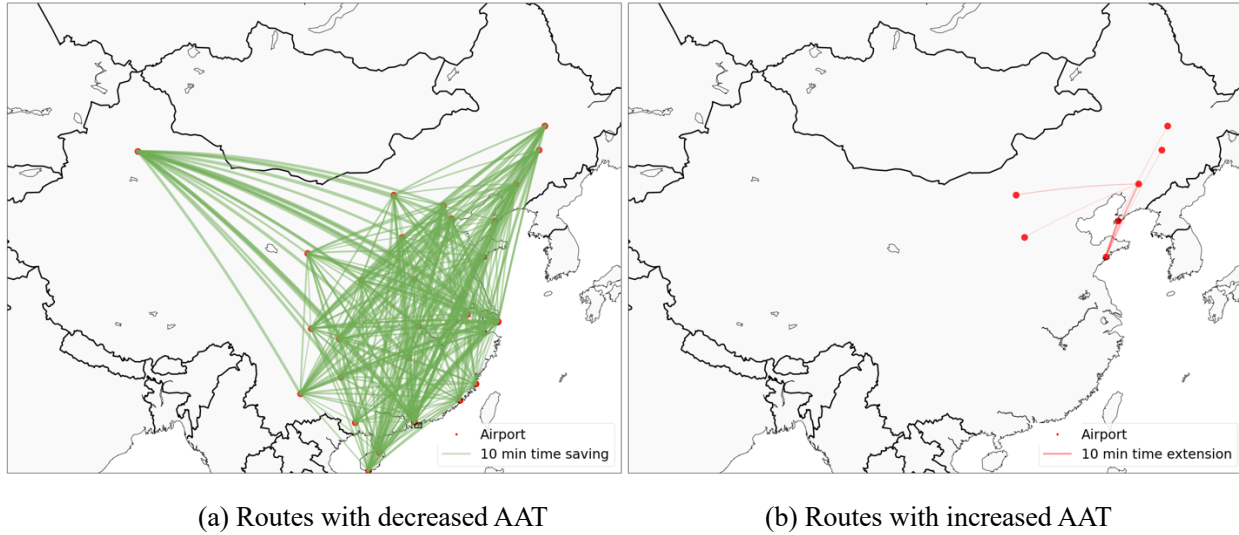


Figure 19. AAT change among Chinese routes if the AAT behavior in the U.S. were adopted

7. Conclusion

Understanding the behavior of flight actual airborne time is important given the ever-growing air travel demand and increasing flight delays. For the first time in the literature, this chapter performs an empirical investigation of AAT and its influencing factors, particularly how AAT is affected by OD distance, enroute and terminal congestion, convective weather, the possible pressure to reduce AAT from other parts of flight operations and airline/airport-specific effects. Unique to our investigation is its comparative nature: the modeling of AAT is side-by-side for the U.S. and China, which have the two largest air traffic systems in the world with distinct characteristics in terms of geographical distributions of air travel demand, airspace availability, air traffic flow management, delay management policies by civil aviation authorities, and airline scheduling practices.

The different model specifications and empirical estimation yield several interesting insights, with the most important ones reflected in the testing of the four hypotheses. Specifically, AAT is highly correlated with OD distance. The effective GCD speed is higher in the U.S. Chinese low effective speed indicates the flight might have lower aircraft speed, or experience rerouting, holding or vectoring enroute due to limited airspace availability and inefficient traffic management. Facing with ground delay before takeoff, a flight has limited capability to make up for the delay due to possible constraints from airspace route structure and ATC procedures, as well as greater fuel burn if flying faster. The pressure from short turnaround time after landing to reduce AAT is found immaterial. These findings suggest that existing ATC procedures do not allow much flexibility for an individual flight to adjust AAT to mitigate delay. The question of how fast a flight can fly thus hinges mostly on ATC as well as airline scheduling. Since airline scheduling depends not only on OD distance but also on ATC (e.g., an airline will likely have longer flight schedules if constantly experiencing delays under ATC), ATC is an ultimate critical factor affecting AAT. Given that OD distance cannot be altered, if one wants to improve aircraft flying time a systemwide change in the ATC would be the most relevant.

Taking the above argument to a counterfactual analysis for the Chinese air traffic system to adopt the AAT behavior in the U.S., results suggest that significant efficiency gains would be achieved for Chinese flights in terms of AAT, fuel consumption, and direct airline operating cost. In view of the ATC systems in the two countries, such adoption of the AAT behavior would mean sweeping changes in airspace abundance, air routing flexibility, and efficiency of TMIs in the Chinese system. The benefit estimates, therefore, should be viewed as an upper bound of the potential benefits moving toward a more constraint-relaxed and efficiency-enhancing ATC in China. To achieve these benefits, the modeling and analysis presented in the chapter will be only part of the considerations. Other factors, such as legacy system configuration (e.g., most of the airspace in China is still controlled by the military) and ease of regulatory adaptation, also need to be taken into account and warrant further examination. Thus, Chinese air traffic system could learn the refined management procedures and techniques from the U.S. sides, such as a more resilient route network, advanced traffic management initiatives such as GDP, MIT, and TBFM to assign the foreseen airborne delay to the ground, and if possible, more abundant airspace. Chinese air traffic system could also improve the efficiency of management function among controllers, traffic managers, dispatchers, and pilots, and give airlines more control over flight rescheduling and route planning.

This research presents a beginning towards understanding the behavior of actual flight airborne time. Future research can be extended in a few directions. First, as en-route traffic and

weather conditions, and ATC all affect AAT behavior, collecting and incorporating the flight plan/actual trajectories and TMIs may help further enhance model fit and interpretability. Second, provided that data become available in China, we could analyze gate departure delay and taxi-out delay and explore whether the two types of pre-takeoff delays have the same influence on AAT. Third, as flights flown by the same aircraft in a day are connected, it would be interesting to model AATs for a series of flight legs to further understand the behavior of AAT of not just a single flight but multiple ones and their interdependence.

PART II. Evaluating Terminal Air Traffic Control and Arrival Spacing

Chapter 2

Statistical Modeling of Flight Inter-arrival Times

1. Introduction

Airports, as the nodes of an air transport network, are the major bottlenecks in the National Air System (NAS), leading to significant delays at many major airports (Hansen, 2002; Yin, et al., 2019). The runway system has been identified as the critical airport bottleneck (Sölveling and Clarke, 2014). In practice, runway capacity is determined by how air traffic controllers (ATC) sequence and space arrivals and departures in real time. This study focuses on aircraft arrival separation with the aim of isolating the roles of required separations imposed by the controller or pilot and the “natural” separations resulting from upstream scheduling and air traffic controller actions.

1.1 Arrival separation definition

To avoid the wake vortex hazard and simultaneous runway occupancy (SRO), a trailing aircraft is required to maintain a certain safe separation distance from its leading aircraft during approaches. The strength of wake turbulence is related to the weight of aircraft, speed, wingspan, and wing shape (FAA 2002). FAA previously mandated a minimum wake vortex separation standard for different weight classifications of lead and trail aircraft, as outlined in Table 7. This legacy standard, although using a rough classification system, was straightforward in identifying patterns. However, since 2019, it has been superseded by the Consolidated Wake Turbulence (CWT) Separation Standards. These updated standards now categorize aircraft into nine distinct types, ranging from A-Super to I-Lower Small. Detailed information on these standards is available in FAA Order([ORDER JO 7110.126A](#)). The given spacing criteria are in terms of distance and are applied to arrival flight pairs landing on the same runway along the final approach segment which usually begins at the final approach fix (FAF) and ends at the runway threshold (FAA, 2011; Ballin and Erzberger, 1996).

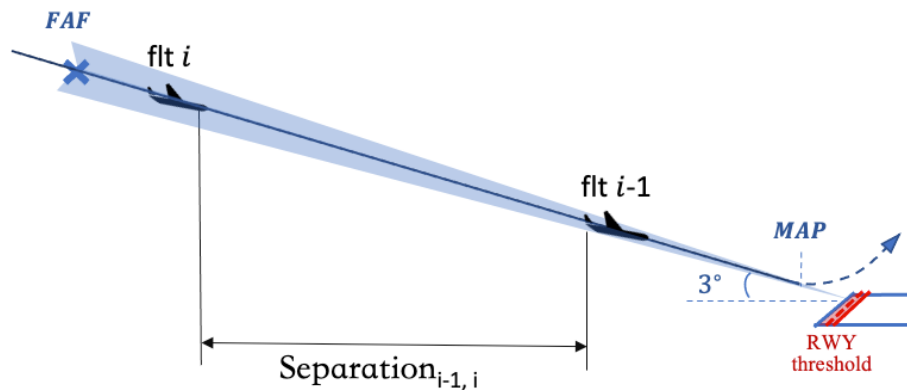


Figure 20. In-trail aircrafts along the final approach segment

Controllers and pilots monitor horizontal distance between arrival pairs to comply with the separation minima from trailing aircraft passing FAF up to the time that leading aircraft crosses the runway threshold as shown in Figure 20. While the FAF locations do not move for a given runway, the crossing speeds at these fixes vary. A slower trailing aircraft must have a longer distance behind the fast-trailing aircraft when crossing the runway threshold to maintain the minimum separation. This is known as the ‘opening case’ (Hansen, 2002), as separations between arrival pairs are increasing during the final approach. In contrast, the minimum separation happens at touchdown if a fast aircraft trails a slower one. In addition to the above separations between aircraft landing on the same runway, there exist more complicated cases including separation between arrivals on parallel dependent runways, intersecting runways and mix usage runway with departure flights (FAA, 2022). During the arrival, ATC will establish the sequence of landing aircrafts given their estimated time of arrival (ETA) and thus request pilots to adjust their airspeeds to achieve the sequence and maintain proper spacing (FAA, 2017).

Table 7. IFR- Approach in-trail wake turbulence separations minima (in miles)

Leading Aircraft	Trailing Aircraft			
	Small	Large	B-757	Heavy
Small	3.0	3.0	3.0	3.0
Large	4.0	3.0	3.0	3.0
B-757	5.0	4.0	4.0	4.0
Heavy	6.0	5.0	5.0	4.0

Source: FAA ORDER JO 7110.65T 5-5-4. MINIMA (FAA, 2011); updated standard (2019) is in (ORDER JO 7110.126A).

The required minimum arrival separation can also be described in terms of time. In theory, distance-based separation can be translated into time-based one given the arrival pair airspeeds. Accordingly, the time-based separation can be used to derive the average minimum interarrival time for a given period of time and fleet mix, and then to calculate runway arrival capacity as time divided by the average minimum interarrival time (Hansen, 2002). In this research, we focus on the **controlled interarrival time** between arrival pairs, as **CIT**, which is the interarrival time controllers (or pilots) achieve when they are actively maintaining separation, as opposed to when upstream scheduling and ATC actions result in “natural separation” that does not require active control. In general, **CIT** is not a single and static value for an aircraft pair, but depends on aircraft weight class, aircraft speed, runway configuration, controller performance and meteorological conditions.

2.1 Related study

In view of the existing literature, there are many different methods and statistical models to estimate the arrival separation and investigate its behavior. Some research (Hansen, 2002; Ballin and Erzberger, 1996) works directly with the minimum separation distances (shown in Table 7) and the approach speeds of different planes to calculate the minimum interarrival time, which make the separation time fixed for a given set of lead-trail aircraft types. This is straightforward and useful if the study aims to approximate the runway capacity or optimize the landing sequence but does not account for stochastic or systematic variability in **CIT** for a given lead-trail type combination.

Other researchers have introduced different probability distributions to fit the density of interarrival time. Jeddi et al. (2009) used log-normal and log-logistic distributions to fit the interarrival time at FAF and runway threshold. The work employed one week of arrival data of DTW in 2003. To account for the capacity constraint issue, it only kept sample pairs during peak periods. The data was divided by meteorologic conditions and follow-trail aircraft weight class to estimate the distribution parameters respectively. An early but innovative study by Vandevenne and Lippert (2000) decomposed actual separation time into target time separation that the controller attempted to achieve, imprecision error and inevitable gap due to the level of available control. It further introduced a joint distribution model with a negative exponential distribution for gap and a normal distribution for error. This work addressed the concerns about large separations during low demand, but the gap needs to be better defined and a constant arrival rate for the study period is assumed. Building upon Vandevenne model, a more recent study (Rakas and Yin, 2005) proposed a double normal model to capture the distribution of landing time intervals as a mixture of pre-control gaps and post-control landing intervals. It better approximates the left-hand side for smaller landing time intervals. This chapter also used the same one-month PDARS data of LAX during June of 2003 for a hypothesis test and concluded that dominant airline has shorter landing time intervals for large aircraft pairs under VFR. An innovation of this chapter is that it assumes the probability distribution of observed landing time intervals to be a segment function conditioned on pre-controlled landing time intervals. However, its limited analysis scope of only large-large aircraft pairs under VFR, and its use of arrivals on different landing runways to calibrate a single model, may reduce the accuracy and interpretability of the results.

Aside from the above studies, work on behaviors of flight interarrival time and controlled interarrival time is relatively limited. In particular, the relationship between the probability distribution of CIT and system conditions has received little study. To fill this gap in literature and practice, this chapter aims to provide a model to investigate the behavior of interarrival time between landing aircraft crossing runway threshold, by analyzing historical operation data. The separation time can be modelled with a conditional probability based on the convolution of uniform distribution for estimated time of arrival and a normal distribution, whose mean is a linear model of relevant explanatory variables, of a controlled arrival time. The model is estimated using the maximum likelihood estimation (MLE) method. We applied this methodology to over 194 thousand arrival pairs at JFK airport during the whole year of 2019.

The contributions of this chapter are two-fold. First, while most existing study is limited to one week up to one month arrival data, our research use an annual sample of arrivals at JFK, leading to results of greater comprehensiveness. Moreover, instead of picking operations during peak hour, our study adopts a conditional probability and MLE model to address the right-hand-side distribution of interarrival time which is constrained by upstream separation decisions. For the second major contribution, our explicit modeling on controlled interarrival time and testing of its relationship with a range of weather-, runway-, and aircraft-related factors. This can help identify importance of such factors determining interarrival times.

This chapter is organized as follows. Section 2 describes the data sources and relevant statistics. Section 3 presents our methodology, including formulation of the mathematical model on controlled interarrival time and feature engineering. Estimation results of model are reported and discussed in Section 4. Section 5 offers conclusions and discusses future directions for future research.

2. Data and study area: JFK

In this chapter, we focus on flight arrivals at John F. Kennedy International Airport (JFK) for the whole year of 2019. JFK airport is selected for testing based on its multi-runway configuration and complex operational conditions.

2.1 Data Source

We collected four datasets for the calendar year of 2019 for JFK airport.

- The first dataset is retrieved from the *Events (EV) dataset* and *Reduced Data (RD) summary* of the NASA Sherlock Data Warehouse, which are gathered from 76 FAA facilities and formatted by ATAC corporation. The EV files record the derived parameters of the aircraft at the time of an event and RD files is a one-line summary for each flight flown in the facility in the last 24 hours. We use EV data to obtain runway threshold crossing time and RD for weight class information of each flight.
- *Airport Configuration Data (APTC)* from National Traffic Management Log records the time of each runway configuration change, the new arrival and departure runway configuration, and corresponding aircraft arrival rates and departure rates of airport. We use this data to identify mixed-use runways.
- *ASPM Airport Quarter Hour data* captures the operational conditions such as acceptance rate, and weather information including temperature, ceiling, windspeed, and visibility at top 77 commercial airports, at 15-minute intervals.
- *Global Hourly - Integrated Surface Database (ISD)* provides the surface observations and numerous meteorological parameters such as thunderstorm, precipitation amounts, snow depth, wind speed and direction, temperature, dew point, cloud data, visibility, and various other elements as observed by each station.

2.2 JFK airport

JFK is the major international airport serving New York City and the 6th-busiest airport in the United States with more than 460K arrivals and departures in 2019 (BTS, 2020). It has two pairs of parallel runways, all equipped with Instrument Landing System (ILS) except for runway 13R. Based on 2019 data, runway 22L is the most heavily used arrival runway, accounting for more than 40% of all arrivals; 4R the second, accounting for 30%; and 31R the third, accounting for 10%. It should be noted that runway 13L/31R was closed from April 1 to Nov. 16, 2019, to receive a major reconstruction to “ensure the long-term safety and viability of the runway” (International Airport Review, 2018). Therefore, these two runways were used less in 2019 compared to previous years when they handled approximately one-third of arrivals.

The eight runways have quite different arrival routes and varying final approach lengths, as shown in Figure 21, which plots the below-5000ft trajectories of arrival flights at JFK during 1/1-1/10/2019. Due to its proximity to other airports (LaGuardia, Newark Liberty International, and Teterboro) in the northwest direction, the flight arrival routes toward JFK approach from other directions. Table 8 displays the distance of final approach segment from Final Approach Fix (FAF) to Missed Approach Point (MAP) in instrument approach, or to runway threshold in visual approach, under different navigation conditions (NAVAID) for each runway (FlightAware, 2023).

Runway 13L, due to the limited airspace, has the shortest final approach of 2.6nm under VOR navigation instruments.

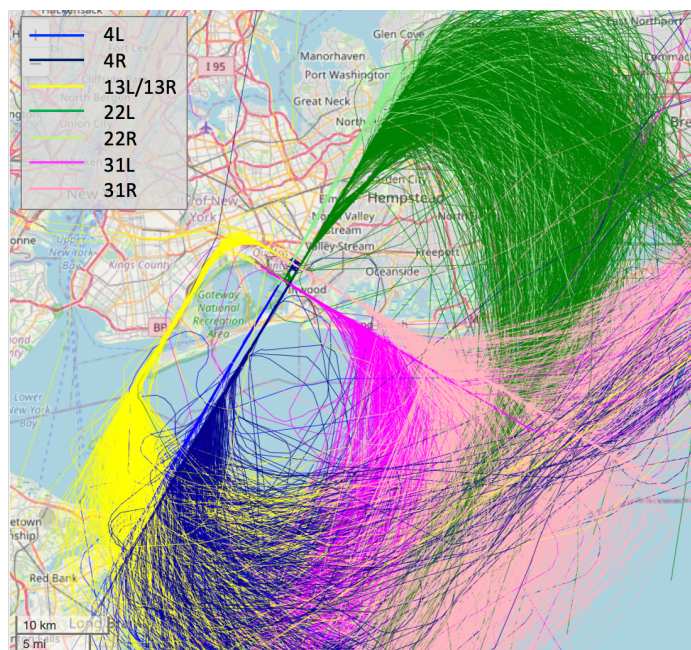


Figure 21. Trajectories of Flights Arriving at JFK on 1/1-1/10/2019

Table 8. JFK Runway characteristics

RWY	Usage Percentage in terms of Arrival Operations	Distance from FAF to MAP, in nm		
		ILS/LOC	VOR	RNAV(GPS)
04L	4.7%	4.5	4.5	5.5
04R	30.5%	4.5	4.4	5.5
13L	3.1%	4.5	2.6	6.1
13R	0.1%	/	2.6	6.1
22L	41.1%	5.5	3.9	5.1
22R	5.2%	5.7	/	5.5
31L	5.0%	5.4	4.7	5.3
31R	10.2%	5.7	/	5.8

Table 9 summarizes the most frequently used arrival runway configurations and their corresponding departure runway configurations based on APTC data. Parallel runways are used for arrivals for more than 80% of time. Usually, one runway in the pair is used as the main arrival runway and the other (in red) is mix-used for both arrival and departure. We also obtained the Aircraft Arrival Rates (AAR) information from APTC for each arrival runway configuration. AAR can be as low as 20 when ILS 31L is in use and can triple that number when 13L and 22L are used together, leading to a pretty wide range.

Table 9. JFK arrival runway configurations (2019)

Arrival RWY Configuration	Use Percent (time based)	AAR	Departure RWY Configuration
ILS 22L/22R	40.0%	40-44	[22R , 22R /31L, 13R, 31L]
ILS 04R/04L	35.3%	42-48	[04L /31L, 04L , 31L, 22R/31L, 04L /31L]
ILS 31R VAP 31L	11.6%	52-58	[31L]
VOR 13L/22L	2.7%	54-60	[13R]
ILS 31L	1.9%	20-40	[31L]
VOR 22L	1.6%	36	[22R/31L, 22R, 31L, 13R, 22L /31L]
STAGGERED 31R/31L	1.6%	46-54	[31L , 22R]
other	5.3%	/	/

3. Methodology

This section includes construction of mathematic model for the controlled interarrival time (*CIT*), feature engineering, and delay analysis with queueing model.

3.1 Mathematic model framework

During arrivals, air traffic controllers (ATC) will establish the sequence of arriving aircrafts and arrival aircraft will be requested to adjust their airspeed or vector to achieve proper sequencing and separation (FAA IPH, 2022). In other cases, the pilot of the trail aircraft will take responsibility for maintaining separation from the lead aircraft. When arrival demand is lower than capacity, there will often be a sufficient time gap between scheduled flight arrival times, which allows aircrafts to land around their initial estimated time of arrival (ETA) while fulfilling the minimum separation requirement. On the other hand, when the arrival demand is heavy, especially beyond the expected runway capacity, air traffic controllers will need to monitor the aircrafts closely and sequence arrivals by adjusting airspeed of landing aircrafts tentatively at fixes along their approaches. In this case, arrival flights during the congested period may receive en route delays based on speed control or vectoring. The primary goal of arrival control is to guarantee safe spacing to avoid wake vortices from leading aircrafts and to prevent simultaneous runway occupancy. Therefore, the observed RCTI should depend on pre-control ETA and controlled separation and which can be modeled using probability distributions.

First, we will define the distribution of ETA. Though the exact ETA information is not available to us, we know the final landing sequence is assigned based on initial ETA. Therefore, the estimated arrival time of aircraft i , ETA_i given the event that its leading aircraft ($i+1$) has landed, should be later than ETA_{i-1} and earlier than ETA_{i+1} . Without ETA data, we use observed runway threshold crossing time of the leading aircraft, RCT_{i-1} , as an approximation of ETA_{i-1} , and RCT_{i+1} as an approximation of ETA_{i+1} . ETA_i is thus assumed to be a random variable uniformly distributed over the time interval between the runway threshold crossing time of its leading aircraft, RCT_{i-1} , and that of its trailing aircraft, RCT_{i+1} , which is

$$ETA_i \sim Uniform(RCT_{i-1}, RCT_{i+1}) \quad (1)$$

Define the time interval between RCT_{i-1} and RCT_{i+1} as T_i ,

$$P_i = (ETA_i - RCT_{i-1}) \sim Uniform(0, T_i), \quad (2)$$

where $T_i = RCT_{i+1} - RCT_{i-1}$

Then we introduce the distribution of the observed runway crossing time interval, $RCTI_i$. It is determined by two variables, P_i and CIT_i , which is the controlled inter-arrival time between the landing aircraft i and its leading aircraft $i-1$ when crossing the runway threshold. There are two scenarios for the observed $RCTI_i$ of aircraft i given the values of RIT_i and P_i ,

$$RCTI_i = \begin{cases} P_i, & \text{if } CIT_i < P_i \text{ (i.e., } RCT_{i-1} + CIT_i < ETA_i) \\ CIT_i, & \text{if } CIT_i \geq P_i \text{ (i.e., } RCT_{i-1} + CIT_i \geq ETA_i) \end{cases} \quad (3)$$

The first scenario, $P_i > CIT_i$, indicates that aircraft i is planned to arrive with a sufficiently large time gap after its leading aircraft has landed. So, its landing time (RCT_{i+1}) is not constrained by the minimum separation requirement. For the second scenario, $P_i \leq CIT_i$, aircraft i could have landed earlier without pilot or controller intervention. But its actual landing time is constrained by the controlled inter-arrival separation CIT_i , with an arrival delay as,

$$\Delta delay = RCT_{i-1} + CIT_i - ETA_i = CIT_i - P_i \quad (4)$$

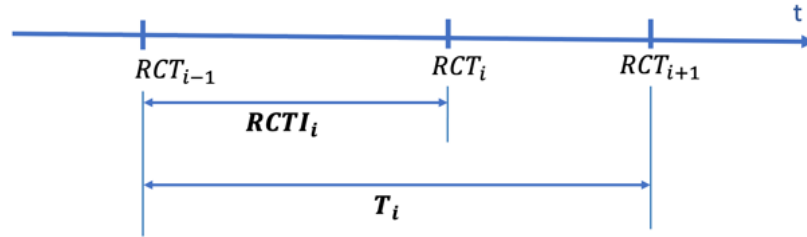


Figure 22. Timeline for 3 consecutive arrival flights

Finally, we turn to the variable of our most interest, **controlled inter-arrival time (CIT)**. Since the value of CIT depends on aircraft type, weather, runway configuration and other operational factors, we assume that for landing aircraft i , CIT_i follows a normal linear model with X_i as the explanatory variable vector to capture the impacts from operational conditions and unknown parameters (β, σ):

$$\begin{aligned} CIT_i &\sim Normal(\mu_i, \sigma^2), \text{ where } \mu_i = X_i^T \beta \\ \text{or } CIT_i &= X_i^T \beta + \varepsilon_i, \text{ where } \varepsilon_i \sim N(0, \sigma^2) \end{aligned} \quad (5)$$

Let $f_{P_i}(p_i)$ and $F_{P_i}(p_i)$ denote the probability density function (PDF) and cumulative distribution function (CDF) of P_i , which follows $Uniform(0, T_i)$, as shown in Eq. (2); $\varphi_{CIT_i}(rit_i)$ and $\Phi_{CIT_i}(rit_i)$ denote the PDF and CDF of CIT_i , which follows $N(X_i^T \beta, \sigma^2)$, as shown in Eq. (5). According to the function of $RCTI_i$ in Eq. (3), we can calculate the cumulative probability for the observed $RCTI_i$ for aircraft i , as

$$Prob(RCTI_i < rcti_i) = F_{P_i}(rcti_i) \Phi_{CIT_i}(rcti_i) \quad (6)$$

By taking the derivative of CDF in Eq. (5), we derive the PDF for $RCTI_i$ as,

$$\begin{aligned}
\text{Prob}(RCTI_i = rcti_i) &= \frac{d\text{Prob}(RCTI_i < rcti_i)}{drcti_i} \\
&= \frac{dF_{P_i}(rcti_i)\phi_{CIT_i}(rcti_i)}{drcti_i} \\
&= f_{P_i}(rcti_i)\phi_{CIT_i}(rcti_i) + \phi_{CIT_i}(rcti_i)F_{P_i}(rcti_i) \\
&= \frac{1}{T_i}\phi_{CIT_i}(rcti_i) + \frac{rcti_i}{T_i}\phi_{CIT_i}(rcti_i)
\end{aligned} \tag{7}$$

For each observation (aircraft i) in the dataset, $RCTI_i$, T_i and X_i are known, which is denoted as $data_i$. We will use maximum likelihood estimation (MLE) (King, 1998) to derive the values of β and σ for the distribution of controlled interval arrival time RIT_i in Eq. (6). The posterior density for the unknown parameters β and σ is then:

$$\text{Prob}(\beta, \sigma | data) = \frac{\text{Prob}(data | \beta, \sigma)p(\beta, \sigma)}{\text{Prob}(data)} \propto \text{Prob}(data | \beta, \sigma) \tag{8}$$

Assuming observations are all independent, the likelihood of the whole sample can be written as the product of the individual likelihoods over all observations. Plug in Eq. (7), then the final likelihood function can be constructed as:

$$\begin{aligned}
L &= \prod_{i=1}^N \text{Prob}(data_i | \hat{\beta}, \hat{\sigma}) \\
&= \prod_{i=1}^N \left(\frac{1}{T_i}\phi_{CIT_i}(rcti_i) + \frac{RCTI_i}{T_i}\phi_{CIT_i}(rcti_i) \right)
\end{aligned} \tag{9}$$

Then we can write down its log-likelihood function:

$$\begin{aligned}
LL &= \sum_{i=1}^N \log \left(\frac{1}{T_i}\phi_{CIT_i}(rcti_i) + \frac{RCTI_i}{T_i}\phi_{CIT_i}(rcti_i) \right) \\
&= \sum_{i=1}^N \log \frac{1}{T_i} + \sum_{i=1}^N \log \left(\phi_{CIT_i}(rcti_i) + rcti_i \times \phi_{CIT_i}(rcti_i) \right)
\end{aligned} \tag{10}$$

Given the data, T_i is a constant value so $\sum_{i=1}^N \log \frac{1}{T_i}$ will not impact the optimization of the likelihood function in Eq. (10). This indicates that the assumption that ETA follows a uniform distribution between RCT_{i-1} and RCT_{i+1} indirectly implies that the optimal value of parameters will not depend on $T_i = RCT_{i+1} - RCT_{i-1}$, which is the observed time interval between the runway threshold crossing time of aircraft $i + 1$ and aircraft $i - 1$.

$$\uparrow LL \leftrightarrow \uparrow \sum_{i=1}^N \log \left(\phi_{CIT_i}(rcti_i) + rcti_i \times \phi_{CIT_i}(rcti_i) \right) \tag{11}$$

Since we have constructed the log-likelihood function, we can derive the optimal values for parameters β and σ by maximizing Eq. (11) numerically.

3.2 Feature engineering

In Eq. (5), we assume the controlled inter-arrival time of adjacent aircrafts crossing runway threshold, CIT , to be a random variable following a normal linear model. In this section, we will introduce three categories of variables to try to account for the variation of CIT , including airport-related, aircraft-related, and weather-related variables. The specific variables considered in the model are listed in Table 10 and the general form of linear model is as follows:

$$RIT = f(\text{weather variables, airport variables, aircraft variables}) \quad (12)$$

Several weather variables are considered to affect the controlled interarrival time. First, we have the meteorological condition (MC) of Visual / Instrument Meteorologic Conditions (VMC/IMC), which refers to the actual weather conditions. ICAO defines the VMC minima with requirements of visibility, cloud distance, cloud ceilings, and weather phenomena (ICAO, 2005), which is a boundary criterion between VMC and IMC. An IMC variable is included which takes value 1 if MC is instrument during the given arrival and 0 otherwise. Second, the variables of runway crosswind and headwind speed, $CWind_RWY$ and $HWind_RWY$, represent the wind velocity components perpendicular to, and away from respectively, the target arrival runway of the given aircraft during the quarter hour of landing. Note that both runway angle and wind angle are with respect to magnetic north. Third, the terminal visibility variable, log_vis , is calculated with a logarithm transformation with base 10 to capture the nonlinear relationship between visibility and minimum separations. For example, we expect that a decrease of visibility from 2 miles to 1 mile have a greater impact than a decrease of visibility from 10 miles to 9 miles. As there exists 0-mile visibility in the records, we add a constant term, i.e., $log_vis = \log(\text{terminal visibility} + 1)$. Similarly, the terminal ceiling variable, log_ceil , is defined as $log_ceil = \log(\text{terminal ceiling} + 1)$. Lastly, thunderstorm is the most critical convective weather event which can severely affect arrival control and reduce airport capacity. Therefore, we use ISD synoptic surface observations at John F Kennedy International Airport surface station (WBANZ: 92789), which reports the presence of thunderstorms, to generate the dummy variable, TS . It represents the occurrence of thunderstorm in the airport vicinity when the given aircraft is landing.

For airport-related factors, we focus on the operational conditions related to runway configuration. First, we consider the shared runway usage, Mix_RWY , by identifying whether the landing runway of the object aircraft is also used for departure during the same time period. For the arrival-priority scenario, departure aircrafts are usually squeezed between two landing aircrafts without affecting the arrival separation time. However, when departure demand is higher, the time gaps between arrivals will be adjusted to accommodate the increasing number of departure operations (Glaab, et al., 2016). Second, as JFK airport has two types of arrival runway configuration: single-runway operations and two-runway operations, we create another variable, Dep_RWY , to represent the dependency between arrival runways for the latter configuration group. Two-runway operations at JFK include 22L/22R, 4L/4R, 31L/31R and 13L/22L as partially shown in Table 8. The independent parallel approach operations require a 4300-ft distance between runway centerlines and an individual monitor controller for each runway according to FAA policy (FAA, 2015). Runways 4L/22R and 4R/22L are 3000 ft apart and fail to meet the criteria, and thus their parallel operation requires a 1.5 NM stagger. Although runways 31L and 31R are 6700 ft apart that meets the geometric criteria, the operation situation is complex due to staffing considerations (Glaab, et al., 2016). If staffing resource and traffic volume allows, it will be “ILS 31R VAP 31R” which is viewed to be operated without dependency. Otherwise, it will be “staggered 31R/31L” operating with dependency requiring a 2-mile stagger. For configuration 13L/22L, the two runways do not

intersect and have approximately 1.75NM of distance between their runway ends, thus we regard them as independent arrival runways. Last, we include the length of final approach segment of landing runway under given NAVAID (ILS, LOC or RNAV), as *FAF_MAP*. It is the distance between final approach fix (FAF) and missed approach point (MAP) as listed in Table 7. This variable is an approximation on the length of final approach segment of given runway and designed to explain the impacts of opening/ closing arrival cases.

To capture the aircraft and fleet mix impacts on controlled inter-arrival time, we introduce fixed-effect dummy variables on the weight class of lead-trail aircraft pair as *WC_A-B* where A and B represent the weight class of lead and trail aircrafts respectively, and they take values from weight class set $\Omega = \{H: \text{heavy}, L: \text{large}, F: B757, S: \text{small}\}$. Weight class pair of L-L is set as reference group to avoid multicollinearity, so there is 15 weight class of lead-trail aircraft pair remained in the model.

Table 10. Description of variables

Variable	Description
Y-variables	
CIT (in second)	Controlled interarrival time, which is the interarrival time controllers (or pilots) achieve when they are actively maintaining
X-variables	
Weather-related	
IMC (dummy)	Takes value of 1 if the Meteorological Condition Flag is I-instrument during the quarter hour when the aircraft is landing; 0 otherwise (V-visual)
log_ceil (in log 100ft)	Log of ceiling (in 100 ft) plus one hundred ft during the quarter hour when the aircraft is landing
log_vis (in log mile)	Log of visibility (in miles) plus one mile during the quarter hour when the aircraft is landing
CWind_RWY (in knots)	The speed of wind component to the landing runway during the quarter hour when the aircraft is landing
HWind_RWY (in knots)	The speed of wind component that away from the landing runway during the quarter hour when the aircraft is landing (take negative if tailwind)
TS (dummy)	Takes value of 1 if there is thunderstorm in the airport vicinity when the aircraft is landing on arrival runway; 0, otherwise
Airport-related	
Mix_RWY (dummy)	Takes value of 1 if the arrival runway is also used for departure when the aircraft is landing; 0, otherwise
Dep_RWY	Takes value of 1 if the arrival runway configuration is 22L/22R, 4R/4L or staggered 31L/31R when the aircraft is landing; 0, otherwise
FAF_MAP	Distance (in NM), which is the distance between FAF and MAP of landing runway with given NAVAID when the aircraft is landing
Aircraft-related	
WC_A-B	Takes value of 1 if the weight class of lead-trail aircraft weight class pair A-B, where A and B represent weight class L, H, F(B757) and S (L-L is the reference in the model).

By combining linear model in Eq. (5), revised log-likelihood function in Eq. (11) and explanatory variables discussed above, we can now construct the full specifications of the negative

log-likelihood function for our problem as follows. In the specifications, β 's and σ are the coefficients to be estimated, while N represents the total number of samples and i indicates observed aircraft i .

$$\begin{aligned}
\min_{\beta, \sigma} & - \sum_{i=1}^N \log \left(\phi \left(\frac{RCTI_i - RIT_i}{\sigma} \right) + RCTI_i \times \varphi \left(\frac{RCTI_i - RIT_i}{\sigma} \right) \right) \\
s. t. & RIT_i = \beta_0 + \beta_1 IMC_i + \beta_2 HWind_{RWY_i} + \beta_3 CWind_{RWY_i} \\
& + \beta_2 HWind_{RWY_i} + \beta_3 \log_ceil_i + \beta_4 \log_vis_i + \beta_5 TS_i \\
& + \beta_6 Mix_{RWY_i} + \beta_7 Dep_{RWY_i} + \beta_8 FAF_MAP_i \\
& + \sum_{A, B \in \Omega} \beta_{WC_A-B} (WC_A - B)_i
\end{aligned} \tag{13}$$

where $\Omega = \{H: \text{heavy}, L: \text{large}, F: B757, S: \text{small}\}$ and weight class pair of L-L is set as the reference group to avoid multicollinearity problem in the model.

4. Model result: CIT

This section presents the estimation results of the CIT model specified in Section 3 (Eq. 13) using the arrival pairs at JFK airport from January 1, 2019, through December 31, 2019. There are 220,970 arrivals at JFK in 2019 from EV dataset, including 196,833 pairs of arrival trailing arrival and 24,137 pairs of arrival trailing departure. We calculate the RCTI for each arrival aircraft. As our model focus on the required separation between arrival-arrival pairs the departure-arrival observations are deleted. We then merge aircraft weight class information from RD data, including heavy (H), large (L), B757 (F), small (S) and unknown. There are 1.44% arrival aircrafts detected with unknown weight class, which are deleted after calculating RCTI. In the end, there are 194,325 observations in our training model with lead-trail weight class counts as shown in Table 11. Weather information from ASPM quarter hour data and ISD data, and runway configuration from APTC data are first cleaned, filtered, and unified with time zones (UTC) before being merged into arrival-pair-based master datasets for modeling and performance assessment. The summary statistics for some of the variables are reported in Table 12. The median of RCTI is 117s with a standard deviation of 7481s and maximum value of 985,544s, which exhibits a right skewness and extremely long right tail.

Table 11. Counts on weight class of arrival pairs at JFK (2019)

lead \ trail	H	L	F	S
H	17476	32702	2320	403
L	32870	89241	6408	1465
F	2331	6491	505	79
S	607	1301	79	53

Table 12. Summary statistics of the variables

Variables	mean	std	min	median	max
RCTI	412.807	7481.404	23.000	117.000	985544.00
IMC	0.135	0.342	0.000	0.000	1.000
HWind_RWY	6.373	5.738	-14.381	5.985	32.020
CWind_RWY	6.035	4.511	0.000	5.260	27.932
log_ceil	1.796	0.380	0.000	2.004	2.004
log_vis	0.991	0.162	0.000	1.041	1.041
TS	0.004	0.062	0.000	0.000	1.000
Mix_RWY	0.045	0.208	0.000	0.000	1.000
Dep_RWY	0.801	0.399	0.000	1.000	1.000
FAF_MAP	5.031	0.698	2.600	5.500	6.100

The sample data are inputted in the model specified in Eq. (13) and the unknown parameters (β, σ^2) are estimated by minimizing the negative log-likelihood function. We use the `scipy.optimize.minimize` package in Python with Nelder-Mead search algorithm and maximum iterations of 75,000 (SciPy User Guide, 2023) to derive the optimal solutions. Then we numerically calculate the second derivative of the log-likelihood function in Eq. (11) as the Hessian matrix $H(\beta, \sigma^2)$. Having the samples and estimated value of (β, σ^2) , we calculate the expected value of the Hessian matrix and derive the variance-covariance matrix: $\text{var}(\beta, \sigma^2) = (-E[H(\beta, \sigma^2)])^{-1}$ (Pawitan, 2013). In the end, the standard errors of the estimators are the square roots of the diagonal terms in matrix $\text{var}(\beta, \sigma^2)$. Model results are presented in Table 13 including the minimum negative log-likelihood value.

Looking at the individual coefficients, the estimate for constant term is 91.046 seconds. It quantifies the average value of *CIT* for the benchmark aircraft weight class pair (L-L) under VMC. The magnitude of constant is reasonable given the minimum separation requirement in Table 7 and regular approaching airspeed. For the weather-related variables, the meteorological condition, **IMC**, has a highly significant coefficient indicating that aircraft under IMC need on average 2.590 more seconds on *CIT*, compared to that under VMC. The variable of terminal ceiling, **log_ceil**, presents a negative coefficient with small magnitude and statistical insignificance, but its negative sign still suggests that arrival aircrafts tend to have a longer *CIT* if the ceiling is lower. The insignificance of this coefficient could be due to the collinearity between MC and ceiling variables in our data. **log_vis** presents a statistically significant coefficient with relatively large value. It indicates that if the **log_vis** decreases by 1 (or equivalently if visibility decreases by a factor of -), the landing aircraft needs 7.453 more seconds for *CIT*. With regard to the wind effects, our estimate on **CWind_RWY** is statistically significant and reveals that stronger cross wind leads to larger *CIT* with 0.115 extra second per knot, while the effect of **HWind_RWY** is insignificant. Last, the variable of thunderstorm, **TS**, reports a positive and highly statistically significant coefficient on *CIT*. It confirms with our expectation that aircrafts need extra *CIT* (=2.119s) when thunderstorms are present in the terminal area.

Turning now to the runway-related variables, the coefficient of mixed runway use, **Mix_RWY**, reports a statistically insignificance with a small positive value of 0.067. This suggests that flights landing after another arrival on a mixed-use runway have about the same controlled separation as

is the case on a dedicated arrival runway. However, our sample only contains a small proportion (4.5%) of flights that lands on a mixed-use runway and those arrival flights trailing departures are deleted during data processing, so this result is not definitive.

Landing on a dependent runway (*Dep_RWY*), compared with an independent one, will require 6.970 less seconds of *CIT*. Further research is required to verify and explain this surprising result. One possibility is that simultaneous operations on dependent runways occur only under favorable operating conditions. Another guess is that dependent runways might be used more frequently when the overall airport demand is higher. This may create pressure for flights need to be squeezed in and arranged more densely to make the most of limited available usage time, which will lead to a shorter *CIT*.

The coefficient of the length of final approach segment (*FAF-MAP*) is estimated to be 2.350 with statistical significance, which indicates that one additional nautical mile of final approach segment can lead to 2.350 additional seconds of *CIT* on average. This might be a result of the minimum separation requirement in the scenario of opening lead-trail aircraft pair. When a flight is entering the final approach and its speed is less than its leading aircraft, the minimum separation criteria must be satisfied at the entrance of final approach, which is FAF, and this separation will become larger as it approaches the airport. Then with a longer final approach is, the gain in separation is likely to be larger, leading to the result that *CIT* also becomes larger.

Table 13. Estimation results of *CIT*

Var	Coef.	std err	Z	P> Z	[0.025	0.975]
cons	91.046	0.907	100.409	0.000***	89.268	92.823
IMC	2.590	0.422	6.142	0.000***	1.763	3.416
log_ceil	-0.500	0.397	-1.261	0.207	-1.277	0.277
log_vis	-7.453	0.561	-13.293	0.000***	-8.552	-6.354
Cwind_RWY	0.115	0.012	9.642	0.000***	0.091	0.138
Hwind_RWY	-0.006	0.014	-0.386	0.700	-0.034	0.023
TS	2.119	1.003	2.113	0.035**	0.153	4.085
Mix_RWY	0.067	0.468	0.143	0.887	-0.851	0.985
Dep_RWY	-6.970	0.179	-38.956	0.000***	-7.320	-6.619
FAF-MAP	2.350	0.094	24.992	0.000***	2.166	2.534
WC_H-L	48.299	0.178	271.687	0.000***	47.952	48.648
WC_H-H	12.287	0.236	51.981	0.000***	11.824	12.752
WC_H-F	53.155	0.563	94.407	0.000***	52.051	54.258
WC_H-S	81.468	2.206	36.927	0.000***	77.144	85.792
WC_L-H	-3.173	0.176	-18.008	0.000***	-3.519	-2.828
WC_L-F	3.564	0.355	10.043	0.000***	2.869	4.260
WC_L-S	22.629	0.790	28.650	0.000***	21.081	24.177
WC_F-L	-2.064	0.338	-6.111	0.000***	-2.726	-1.402
WC_F-H	-6.260	0.566	-11.056	0.000***	-7.370	-5.151
WC_F-F	-3.009	1.274	-2.362	0.018**	-5.505	-0.512
WC_F-S	30.265	2.742	11.040	0.000***	24.892	35.639
WC_S-L	-6.472	0.872	-7.423	0.000***	-8.181	-4.763
WC_S-H	-9.842	1.250	-7.875	0.000***	-12.291	-7.392
WC_S-F	2.520	3.426	0.736	0.462	-4.195	9.235
WC_S-S	5.442	4.648	1.171	0.242	-3.668	14.553
Sigma, σ	16.141	0.050	326.097	0.000***	16.044	16.238
Num. Obs.			194325			
Neg log-likelihood			-98503.241			

AIC	197058.482
BIC	197323.091

***p < 0.01, **p < 0.05, *p < 0.10.

As for aircraft weight class pair, coefficients for smaller aircrafts trailing larger aircrafts appear to be positive while parameters for larger aircrafts trailing smaller aircrafts tend to be negative. All coefficients except for that of small-B757 (WC_S-F) and of small-small (WC_S-S) are statistically significant. Coefficient of heavy-small pair (WC_H-S) is the largest among all weight class pairs. A small aircraft trailing a heavy one will require 81s of more time separation than the baseline large-large pair due to wake vortex. Besides, it appears that the larger the weight difference between the pairs is, the larger the absolute value of the parameter tends to be. As shown in Table 14, values at the top and right are larger than values at the bottom and left, which is in the same trend as FAA arrival-arrival separation standard (shown in Table 7).

Table 14. Summary on coefficients of weight class pair

(second)	trail			
lead	H	L	F	S
H	12.287	48.299	53.155	81.468
L	-3.173	0 ^a	3.564	22.629
F	-6.260	-2.064	-3.009	30.265
S	-9.842	-6.472	2.520	5.442

Note: a. L-L is set as reference in the model

5. Conclusion

Understanding the behavior of controlled interarrival time and its relationship with operational conditions is important given the ever-growing civil aviation demand and increasing complexity in NAS. This work documents an empirical study of *CIT* and its influencing factors. A mathematical model is designed to fit the distribution of observed runway crossing time interval by using a conditional probability with uniform distribution for feasible planned arrival time and normal distribution for *CIT*. The behavior of *CIT* is further explained by a normal linear model with weather-, runway- and aircraft-related variables. This conditional probability model not only captures the situation when arrival demand is very low and the time gaps between landing aircraft is constrained by ETA, but also approximates the left-hand side of *RCTI* when ETAs are bunched with short-term arrival demand exceeding capacity. The normal linear model of *CIT* reveals the probability distribution curve of *CIT* with given operational conditions which provides information for better landing safety and capacity. We applied the methodology to 194 thousand arrival pairs at JFK airport during the whole year of 2019.

Based upon our work, future study can be extended in several directions. First, the behaviors of terminal traffic management, controllers, airlines, and pilots could be considered by developing new variables into the *CIT* model. Instead of normal linear model, other statistical models, such as random forest, can be tested. Second, for the conditional probability for *RCTI*, we make assumption with uniform distribution for feasible planned arrival time and normal distribution for *CIT*. Different distributions, such as negative exponential and log-normal, can be assumed and examined to compare with the existing work. Additionally, more complicated scenarios with departures as well as arrivals can be studied as well.

PART III. Enhancing Air Traffic Management Efficiency

This Part is about analysis on the usage of Traffic Management Initiatives (TMIs) within the NAS which is controlled by FAA. In this introduction, some background, including several important concepts related to the air traffic management, and motivation is introduced.

As Traffic Flow Management (TFM) focuses on the traffic demand management given the flow constraints of airport capacity and airspace capacity, its objective is shifting delay to the locations where it can be absorbed safely and efficiently. The air traffic managers have evolved several types of TMIs over the past decades to manage downstream congestion and to ensure the safety and orderly movement of traffic flow within the NAS. Each of them has its own concepts, procedures, automation, and control characteristics. According to Chapter 18 of the FAA's Facility Operation and Administration manual (2020), common TMIs are summarized and provided in order of increasing restrictiveness as follows.

(1) Altitude:

This program is designed to separate different traffic flows, or to distribute the number of flights heading to a specific geographic region. In case of severe weather conditions, Low Altitude Arrival/Departure Routing (LAADR) will be adopted.

(2) Miles-in-Trail (MIT):

This program places spacing requirement (multiple of 5 nm) between aircrafts arriving at an airport, over a fix/waypoint, through a sector, or on a specific route.

(3) Minutes-in-Trail (MINIT):

This program is similar to MIT, but it uses time to space the flights instead of distance. It is usually adopted for non-radar situations.

(4) Fix Balancing:

This program is designed for arrival or departure procedure to balance the demand by assigning a flight to a fix which is not in the filed flight plan.

(5) Airborne Holding:

This program can hold aircrafts to fill the airport capacity when holding is allowed by the operating environment and the weather condition is predicted to improve soon.

(6) Sequencing Programs:

These programs include Departure Sequencing Program, Enroute Sequencing Program, Arrival Sequencing Program and Time-Based Metering. They are used to accommodate different phases of flights at some fix.

(7) Reroutes:

This program assigns routes to aircrafts, which are not the original flight plan.

(8) Ground Delay Program (GDP):

This TM is determined by ATCSCC to hold the aircrafts on the ground at their departure airports to balance the demand and capacity at their arrival airports. Flight restricted by GDP will be assigned an Expected Departure Clearance Time (EDCT).

(9) Airspace Flow Program (AFP):

This program is similar to GDP, but it is used to manage the traffic demands filed into or through the constrained areas instead of arrival slots.

(10) Ground Stop (GS):

This program is the most restrictive. It holds aircrafts which meet specific criteria (i.e. airport, airspace, or equipment specific) on the ground at their departure airports.

In 2007, in order to improve the safety, efficiency and predictability of operations in the NAS, FAA proposed to develop the Next Generation Air Transportation System (NextGen) on Trajectory Based Operations (TBO) by 2025 (FAA, 2017). At the current stage of the program, Time-Based Flow Management (TBFM) is adopted as a decision support system and automation tool to help controllers and local facilities to manage traffic with sequencing and de-conflicting aircrafts in a dynamic time-based environment as opposed to the static MIT restrictions. In theory, it could take the place of the traditional distance-based managements and resolve their potential deficiencies (FAA, 2018), but the current TBFM has not been implemented effectively, nor met controllers' needs, due to airspace complexity, lack of clear guidance on operation, insufficient technical support and reliability issues (Hampton, 2015). Hence, the current management of NAS is hybrid system of TBFM and traditional TMIs.

As growth in demand will generate further strain on the NAS components such as airports and sectors, and more workload for traffic controller and managers, more efficient TMIs and more reliable decision tools are required. However, several research and reports have pointed out the deficiencies and problems in the current management system and questioned the potential flight delay and throughput shortfall when implementing TMIs. Against this background, the goal of this part is to develop and apply a method for assessing the impact of TMIs on operational performance in the NAS. Specifically, while ideally a TMI merely shifts delay without sacrificing throughput or increasing delay, in practice TMIs may cause unnecessary delays or even exacerbate demand-capacity imbalances. We want to find the extent to which real-world TMI's meet their intended purpose and the degree to which induce unnecessary delays or otherwise negatively impact system performance.

Chapter 3:

Miles-in-Trail Restrictions and Aviation System Performance: Chicago O'Hare Case Study

1. Introduction

Among all the TMIs, MITs has been the most frequently used program in the past couple of decades (Rios, 2010; NBAA, 2020). When demand is predicted to be greater than the capacity at certain NAS elements, such as sectors or terminal areas, the corresponding control centers can institute MIT restrictions that impose minimum spacing requirements (normally multiples of 5 nautical miles) between aircraft arriving at an airport, flying over a fix/waypoint, traversing a sector, or flying a specific route. Therefore, an MIT restriction can be regarded as a distance-based metering restriction to meter traffic flows. Facing an MIT, flights can experience departure delay, speed control, detour (vector), or airborne holding (FAA, 2009). Moreover, when a control center could not absorb the required delays within its operational air space, the MIT restrictions can be propagated upstream from one facility to one or more facilities. This procedure is called passback (Sheth, Gutierrez-Nolasco, & Petersen, Development of Miles-in-Trail Passback Restrictions for Air Traffic, 2014). In the current management system, several deficiencies exist in the implementation of MITs. Firstly, an MIT could be set incorrectly. Frequent errors include having the MIT be active for a longer period of time or setting larger spacing values than actually required. Second, the value of spacing requirement is a multiple of five nautical miles, limiting the ability to precisely calibrate the spacing to available capacity. Third, the setting of MIT highly depends on the prediction of weather conditions in the NAS and airport capacity, so the uncertainties of weather and inaccurate predictions of traffic demand and airport or airspace capacity can result in MIT parameters that are not well-suited for actual conditions.

In 2007, in order to improve the safety, efficiency and predictability of the NAS, FAA proposed to develop the Next Generation Air Transportation System (NextGen) with focus on Trajectory Based Operations (TBO) by 2025 (FAA, 2017). At the current stage of this plan, Time-Based Flow Management (TBFM) is adopted as a new TMI to help controllers and local facilities to manage traffic by sequencing and de-conflicting aircrafts in a dynamic manner that can respond quickly to changing conditions (weather, winds, departure delays, etc.), as opposed to the static and blunt MIT restrictions. In theory, TBFM could resolve the operational deficiencies of the traditional distance-based management (FAA, 2018), but the current TBFM has not reached operational maturity, nor met controllers' needs, due to airspace complexity, lack of clear guidance on operation, insufficient technical support and reliability issues (Hampton, 2015). Therefore, the current management of NAS is hybrid system of TBFM and traditional MITs. Still, it is expected that TBFM system and other TBO-related initiatives can gradually replace MITs in the future and correspondingly improve the system effectiveness. Therefore, the inefficiency of the current operation of MITs is worth investigating to better estimate the improvement opportunities from reduced MITs as the deployment of TBFM progresses.

1.2 Literature Review

Most of the research on MIT restrictions has focused on operational statistics and simulation modeling to provide decision support. Rios (2010) provided a very detailed description of the various TMIs, especially for MITs, in the NAS with focus on causes and usage frequency at different airports or facilities. Myers (2005) developed a preliminary algorithm by comparing the actual progress (from ETMS dataset) with MIT parameters to identify whether a flight was involved in a MIT and tested the algorithm for two-week samples in May of 2004 to assess MIT impacts. It quantified the MIT delays by comparing the flight delays of MIT-restricted flights and

the remaining flights, and concluded that no significant impacts existed in en route delays. This research also assessed the necessity of MITs based on the number of restricted flights and concluded that 39% MITs during the study period may not be needed as they involved less than 5 flights. The assessment criteria is this research is simplistic and not aligned with the purpose of MIT restrictions. Ostwald & Topiwala (2006) stated that the use of MIT was highly dependent on the experience of controllers, and consequently may not generate the intended effects on traffic flows. They developed a prototype of MIT Impact Assessment (MIA) capability to assess MIT impacts prior to implementation. They also compared the flight times of MIT and non-MIT flights for two airports to quantify the MIT-induced delays. They proposed the idea of “periods of opportunity” in which airport throughput is below capacity that can be identified based on the comparison of demand, actual arrivals and target arrival rate of the airport during the given time period given different frequencies of MITs.

Several research studies have been carried out on modeling the MIT restriction parameters. Evans (2004) developed a model to describe the use of MITs which involves a restriction generation model based on acceptance rate and arrival demand at the airport, and a delay flow model which takes the predicted restrictions and converts them into spacing separations according to historical data. This research is focused on prediction of MIT parameters based on historical data but does not evaluate MIT performance, but its ideas on matching the ASPM demand and the MIT frequency over arrival fix can be used as reference. More recent research done by Sheth (2013) implemented the FACET software developed by NASA as a simulation tools to measure the delays from MIT-impacted flights and the corresponding sector congestion. It simulated the MITs on three playbook routes in the NAS and compared the flight delays and sector congestion with varying MIT spacing values. This chapter emphasized on the importance of MIT passback to the upstream facilities and suggested that the environment for passback MIT is complicated so a more advanced simulation is needed. Other researchers have proposed machine learning algorithms including basic support vector machine, random forest and decision tree algorithms to assist managers for setting MIT restrictions. The algorithms were evaluated with traffic flows at EWR for 5 months and they concluded that the decision tree performed the best but further testing was required to confirm the conclusion (Bloem, Hattaway, & Bambos, 2012).

1.3 Research question and contributions

The aim of our work is to assess the impact of TMIs on airport throughput, arrival delay, and en route queueing delay. While ideally a MIT merely shifts delay without sacrificing throughput or increasing delay, in practice MITs may cause unnecessary delays or even exacerbate demand-capacity imbalances. However, to our knowledge, no research has looked into the system impacts of MITs in terms of unnecessary flight delays and throughput shortfalls at the airports over a long study period. Accordingly, this chapter aims to measure the improvement opportunities from the better setting of MITs or the reduction of MIT usage under the full implementation of TBFM system in the near future. More specifically, we will develop models to quantify the deficiencies of MITs with emphasis on flight delays and airport throughput reduction. In order to assess MIT performance, we will develop methodologies that enable us to predict how operational outcomes would be different in the absence of a given MITs or set of MITs. These methodologies must allow us to predict (1) when a given MIT-impacted flight would have arrived at a potential bottleneck (i.e. airport) in the absence of the MIT; (2) the actual capacities of the bottleneck; and (3) the delays

and throughput that would have resulted based on the new arrival times and capacities. We apply these methodologies to MITs originating from ORD airport, but they can eventually be used to estimate the improvements opportunities from efficient traffic management at the NAS level.

In view of the current hybrid system of TBFM and MITs and the existing literature on MIT evaluations, there is an ongoing need for better assessing MIT restrictions at both flight level and NAS level in order to detect and reduce their inefficiencies. One contribution of our work is quantitatively measuring the impacts on delay and throughputs for the use of MITs at a given airport during a long period of time. As elaborated in Section 2, our methodology includes a deterministic queueing model, and various models for predicting planned arrival times that enable us to predict how the performance and behavior of the airport queueing system would be changes in the absence of a specified set of historical MITs . Instead of calculating the flight delays by computing the flight time difference between restricted flights and non-restricted fights as in several previous works, this model is capable of identifying the flight performance with and without MIT restrictions directly. Moreover, the change in queueing delay and arrival delay for both restricted and non-restricted flights under different operation conditions can be identified. Besides the flight delay, this model can also evaluate the change in throughputs at the given airport. Another contribution of this work is the design of planned arrival time as input for the reallocation model. To generate this input, we employ several approaches, including the en route time included in the flight plan as well as en route time predicted by different machine learning models, to calculate the planned arrival time, defined as the time the aircraft would touch down in the absence of terminal capacity constraints. Knowing the planned arrival times for MIT-impacted flights in the absence of the MITs, and using simple but plausible assumptions about airport arrival capacity, it is possible to predict the impacts of of MITs on operational performance, including the additional queueing delay or flights arriving at the ORD terminal area sooner, as well as the reduced arrival delay when the airport is able to accept these earlier arrivals. As similar approach can eventually be used assess the efficiency of other TMIs, including MITs at other airports, as well as GDPs, over the entire system.

The remainder of this chapter is organized as follows. In Section 2, the methodology is presented which includes the queueing diagram, prediction on planned arrival time and reallocation model. Then Section 3 describes the data sources and introduce the background for the case study. Estimation results of the models are reported and discussed in Section 4, with special focus on the evaluation of MIT performance. Finally, Section 6 offers the conclusions and discusses the limitations of the current study.

2. Methodology

A high-level framework of how our work quantifies the impacts of MIT restriction on flight delay and airport throughput is illustrated in Figure 23. This diagram includes the input data, models and outcomes. The boxes labeled [1], [7] and [11] are the required input information for flight records, MIT restriction parameters and airport operational conditions. The boxes labeled [4], [9], and [12] represents the models we adopted during the assessment, while the remaining green boxes depicts the intermediate results from the corresponding models. The red box labeled [14] are the final outcomes of assessment. As our datasets directly include the records of TMI-restricted flights with the corresponding induced delay which are provided by MITRE, whose methodology we will discuss briefly prior to discussing the models of boxes [4], [9] and [12] in the following subsections.

MITRE adopted its own algorithm to match the NTML restrictions with the impacted flights and to estimate the induced flight delays. Flights subject to MITs could experience airborne and/or ground delays depending on the location of the restrictions. The airborne MIT delays including holding and vectoring delay which is obtained from matching the holding and vectoring database reported by MITRE's in-house algorithm with MIT parameters. The ground MIT delays are measured by matching flights with MIT restrictions based on MIT providing facilities, flight departure time and flight destination airport. It includes delays from directly restricted flights, and delays from indirectly impacted flights which were not subject to the MITs but had similar scheduled departure time with the MIT restricted flights and were thus delayed in the departure queue.

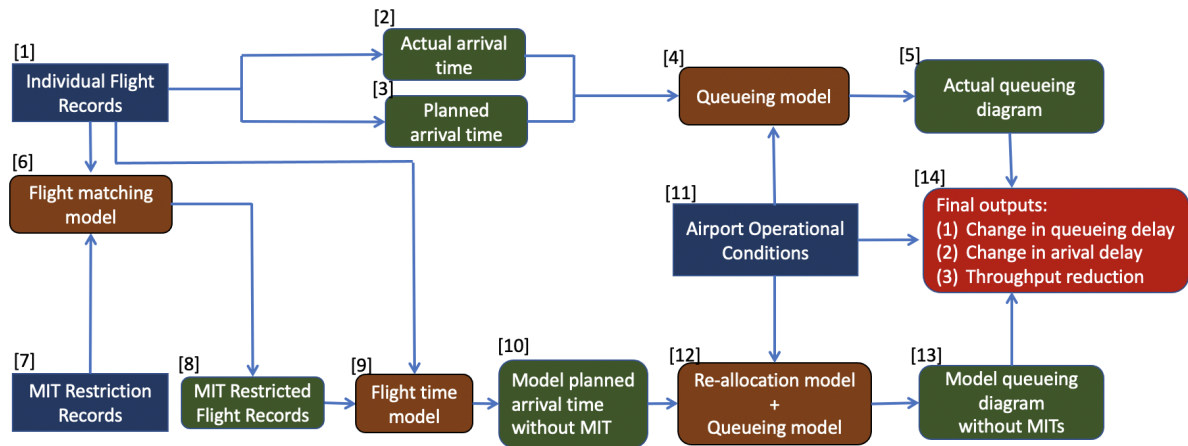


Figure 23. Overall approach to assessing the MIT impacts

2.1 Design of queueing diagrams

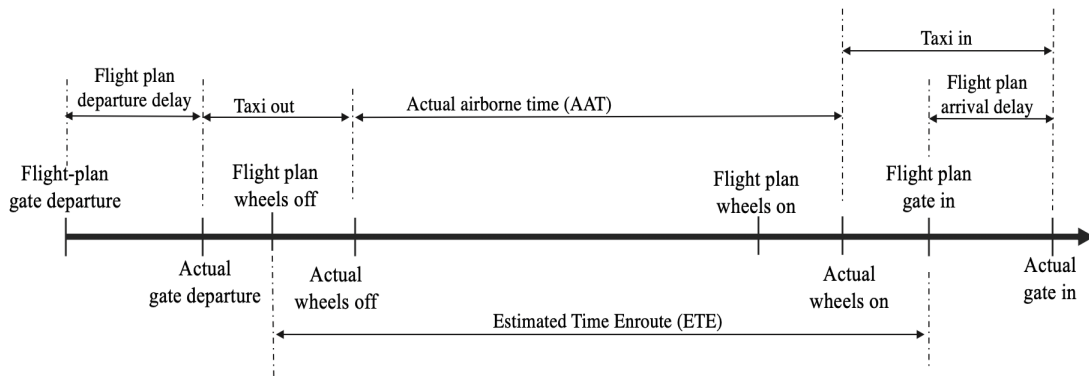


Figure 24 Main decomposition of air travel time

Queueing theory has been applied for a vast range of problems, from improving traffic flow, scheduling customer service to design of telecommunications systems. Especially inspired by the work done by Hansen (2004) and Ding (2018) on evaluation of the flight delays of different events using queueing systems, the deterministic queueing diagram is adopted in our study to quantify the flight delay and throughput impacts from MIT restrictions at a given airport. To construct the queueing diagram at an airport during a certain time period of length T (e.g. a day), we need to

derive the cumulative planned arrivals which is the cumulative count of flights joining the queue to wait for a runway from the beginning of the time period until T , and the cumulative actual arrivals which is the cumulative count of flights landing on the runway also from beginning of cycle starts until T , or some time after T when this is required for all the planned arrivals to have landed. Accordingly, information on the actual and planned arrival time of each flight during T is required. As shown in Figure 24, *actual arrival time* can be obtained directly from the *wheels-on time*, while the *planned arrival time* can, in theory, be calculated from the *actual wheels-off time* plus the *estimated time en route*. The *estimated time en route* is available from the flight plan, which is determined a few hours before the flight departure based on the prediction of en route weather conditions (especially for wind speed and direction) and route planning. However, for some flights in the available dataset, this yields a time later than the *actual wheels-on time*, which implies that the en route time was shorter than estimated. So, we take the minimum value of *actual* and *estimated time en route* for the planned time en route. The two flight times are defined as follows,

$$\text{Actual arrival time} = \text{actual wheels on time} \quad (1)$$

$$\text{Planned arrival time} = \text{actual wheels off} + \min \{ \text{estimated time en route}, \text{actual time en route} \} \quad (2)$$

Based on Eq (1) and (2), the cumulative planned arrival curve ($P(t)$, the red curve) and cumulative actual arrival curve ($A(t)$, the orange curve) can be derived for any given airport during a specified time interval as shown in Figure 25(a). The orange hatched area between $P(t)$ and $A(t)$ represents the induced queueing delay during time period T . Figure 25(b) shows a real example of actual queueing diagram for arrival flights at ORD airport during the whole day of Nov. 26th, 2018. There were 692 arrivals during this time period with 71 ORD-sourced MIT restricted flights. We can observe significant congestion occurring from 6:30–14:30, and there were 161 hours of queueing delay generated in total.

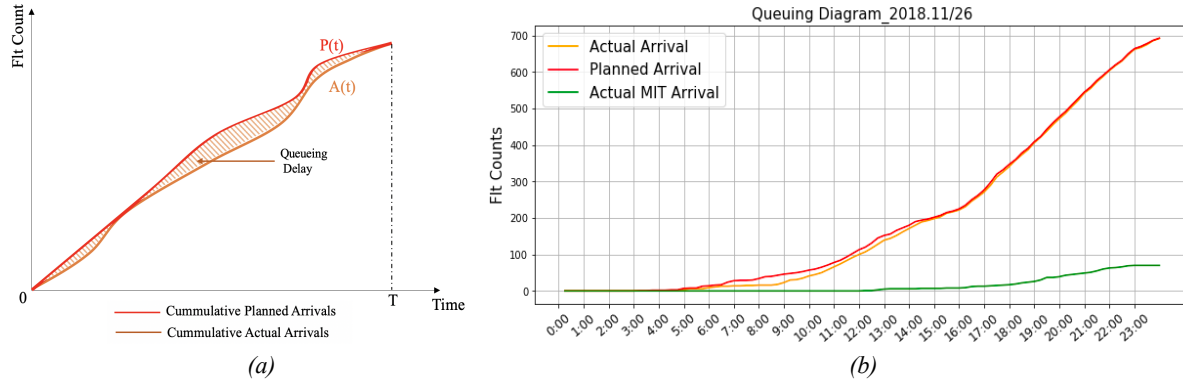


Figure 25 Queueing diagram for the actual scenario

After deriving the actual queueing diagram for arrival flights at a given airport, we now wish to estimate how these curves would have behaved differently in the absence of MIT delays, so the actual operation is compared with the model scenario without the airport-sourced MITs. As the MIT restrictions are removed from the system, we assume the corresponding MIT delays will be eliminated from the restricted flights and these flights will have a new *model planned arrival time*

which is supposed to be earlier than the original *planned arrival time*. Then the flights can join the queue earlier to land on the runway, at a time that can be estimated based airport capacity constraint and the first-come-first-served (FIFO) assumption. Accordingly, a cumulative model planned arrival curve ($P'(t)$, the green curve) and a cumulative model actual arrival curve ($A'(t)$, the blue curve) can be derived using the deterministic queueing diagram, as shown in Figure 26(a). The red hatched area between $P'(t)$ and $A'(t)$ indicates the new queueing delay generated from the model scenario without the airport-sourced MITs. Comparing the actual queueing delay in Figure 25(a) and new queueing delay in Figure 26(a), the delay difference is the impacts on queueing delay from removal of MITs. Similarly, if comparing $A(t)$ and $A'(t)$ as shown in the green hatched area in Figure 26(b), the change in arrival delay which can be calculated as well.

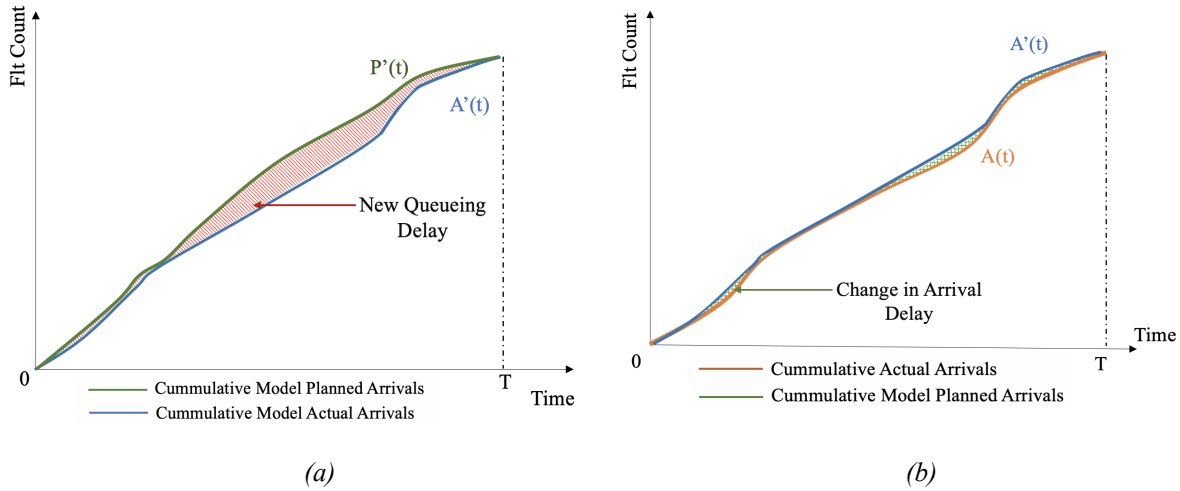


Figure 26 Queueing diagram for the model scenario without MITs

2.2 Construction of model planned arrival time without MITs

In section 2.1, we adopt the queueing diagram to capture the queueing and arrival delays when removing MITs from a given airport and defined the actual arrival time and planned arrival time for each flight based on Eq (1) and (2). The remaining question is how to calculate the *model planned arrival time* and *model actual arrival time* when the airport-sourced MITs were removed. From the data sets available to us, only schedule time, flight plan time and actual flight time as shown in Figure 24 are available and there is no information on the flight en route performance such as trajectories. From the available data, we must estimate the *planned arrival time* for a MIT-impacted flight if the MIT did not occur.

For illustration and anticipating the case study to be used in our analysis, let us assume and ORD-sourced MIT. The timeline in Figure 27 represents the actual flight time for an aircraft arriving at ORD airport on a specific day. During its flight, it experienced three delays due to MIT restrictions. Delay 1 is ground delay from an ORD-sourced MIT, while Delay 2 and Delay 3 are en route delays from JFK-sourced and ORD-sourced MITs respectively. The yellow box represents flight's queueing delay which is the time difference between its *planned arrival time* and *actual arrival time* according to Eq (1) and (2).

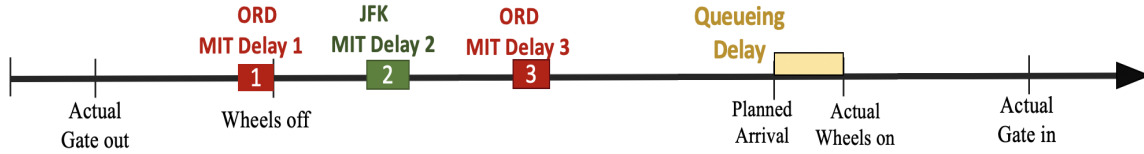


Figure 27 Illustration for Case 1 model planned arrival time

The different settings for the *model planned arrival time* are discussed as follow. The corresponding examples in the below are based on the example flight in Figure 27.

[Case 1]

Model planned arrival time

$= \text{planned arrival time} - \text{ORD-MIT delays}$

$= \text{actual wheels off time} + \min \{ \text{estimated time en route}, \text{actual en route time} \} - \text{ORD-MIT delays} \quad (3)$

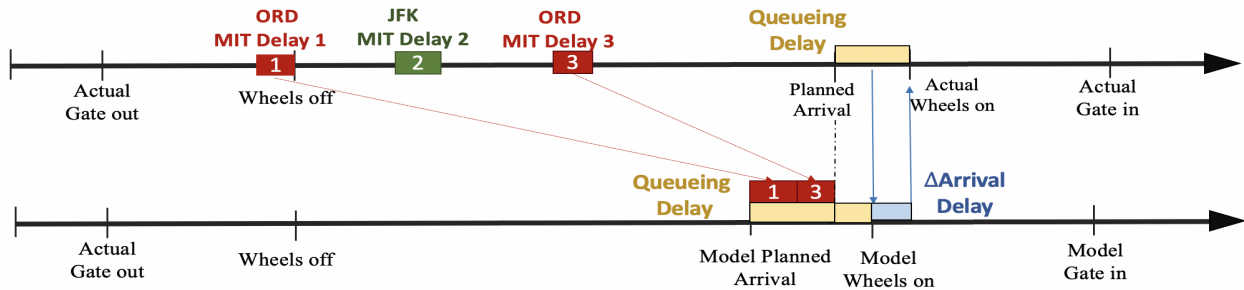


Figure 28 Illustration for Case 1 model planned arrival time

Case 1 removes the ORD-sourced MIT delays directly from the planned arrival time. In Figure 28, the upper timeline represents the actual flight time and the bottom one is the model flight time. As we remove the ORD-sourced MITs, only the 1st and 3rd delays are removed from the flight. Then we have a model planned arrival time which is (delay 1 + delay 3) time earlier than the planned arrival time. With earlier planned arrival time, the flight will land earlier with a new model wheel on time. The blue box depicts the change in arrival delays for the given flight when the ORD MITs were removed. However, as MIT delays are not considered when setting the flight plan time (i.e. estimated time en route), Case 1 will lead to the planned arrival time earlier than it should be.

[Case 2]

$\text{Model planned arrival time} = \text{actual wheels on time} - \text{ORD-MIT delays} \quad (4)$

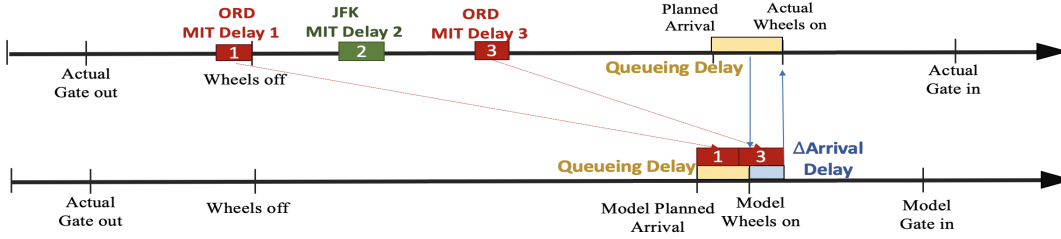


Figure 29 Illustration for Case 2 model planned arrival time

Case 2 removes the ORD-sourced MIT delays directly from the actual wheel on time. It does not remove en route and terminal delays that are caused by the MIT. Therefore, the model planned arrival time is likely to be biased toward the late side.

[Case 3]

Model planned arrival time

$$= \text{actual wheels off} - \text{ORD-MIT ground delays} + \min \{ \text{estimated time en route} + \text{other TMI en route delays, actual en route time} - \text{ORD-MIT en route delays} \} \quad (5)$$

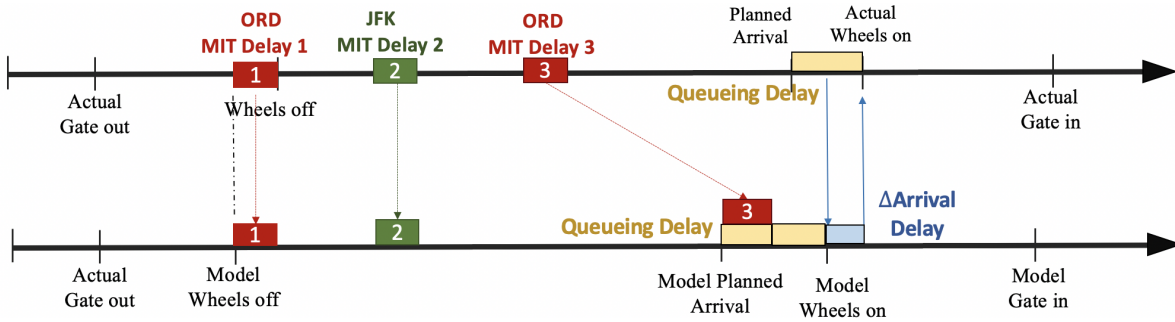


Figure 30 Illustration for Case 3 model planned arrival time

Case 3 separates ORD-sourced MIT delays into ground delays and en route delays. We calculate a new *model wheels off time* as actual wheels off time minus ORD-MIT ground delays as we assume when removing MIT and its corresponding ground delay, the restricted flight can take off earlier by the ground delay time. Since we only consider the impacts of MITs sourced from the target ORD airport, the delays of MITs which are not sourced from ORD airport should be excluded from the computation. So, we calculate the *model en route time* as the smaller value between estimated time en route plus other TMI en route delays and actual en route time minus ORD-MIT en route delays. In the end, the *model planned arrival time* is derived by adding *model en route time* to *model wheels off time*. This definition can, on some level, resolve the issues in case 1 and case 2.

[Case 4]

Model planned arrival time

$$= \text{actual wheels off} - \text{ORD-MIT ground delays} + \min \{ \text{predicted en route time} \\ + \text{other TMI en route delays, actual en route time} - \text{ORD-MIT en route delays} \} \quad (6)$$

Case 4 is very similar to Case 3, but instead of using the flight plan time (i.e. estimated time en route), different predictive models are applied to predict the en route time for flights as discussed in the following content.

As the predictive model used for Case 4 aims to estimate the flight en route time without the constraints of airport capacity, we need to exclude the delay impacts from a capacity-constrained airport. At the current stage of research, we focus on the single airport MIT problem so flights arriving at a specific airport (e.g. ORD airport) are taken into consideration. We assume that when removing the airport capacity constraint, flights would not need to wait in the terminal queue nor be impacted by any airport-sourced MIT delay. To exclude the queueing effect and MIT effect, a filtering procedure is applied for individual flights. First, we calculate the ratio α as follows,

$$\alpha = \frac{\text{actual arrivals}}{\text{capacity}} \quad (7)$$

for each quarter hour for the given airport. Then, for each flight, we check the α values for its wheels-on quarter and one quarter before (i.e. α_t, α_{t-1}). If one of the two values exceed certain threshold (i.e. $\max \{ \alpha_t, \alpha_{t-1} \}$, e.g. the third quartile value of α), the flight is defined as in the queue. Next, we check whether a flight experiences any airport-sourced MIT delays. In the end, we collect the flights which are not in the queue nor with any airport-sourced MIT delay, as input data of our prediction model.

The general form of the en route time model is presented in Eq. (8), in which explanatory variables affecting en route time are classified into three categories: operation related, time related, and weather related. The specific variables considered in the model specifications are listed in Table 15.

$$\text{En route time} = f(\text{operational variables, temporal variables, meteorological variables}) \quad (8)$$

Several operational variables are considered to affect the en route flight time. The most relevant variable is flight-plan estimated en route time because it is determined few hours before departure according to the prediction of en route weather conditions (especially for wind speed and direction) and route planning. Other operational variables include the airline type, flight type, and operational runway number at the arrival airport. Gate departure delay is also included because the pre-takeoff delays may pressure a flight to reduce actual airborne time as a flight delayed at departure may speed up en route or takes a shorter air route to reduce airborne time to mitigate delay propagation (Deshpande & Mazhar, 2012; Liu et al., 2020).

For time-related factors, we consider how different time periods could affect the traffic patterns. One factor is daytime, a binary variable indicating whether a flight wheels on in the day between 7:00 and 20:00. It is assumed that the en route flight time may be longer in the daytime due to greater traffic volume and congestion. Second, considering the difference in air traffic particularly for business travel, the en route flight time may vary depending on it is a weekday or in the weekend. Third, seasonal changes can also lead to the traffic pattern variation due to seasonal air travel demand.

To capture the meteorological impacts on en route time, we include ceiling, temperature, wind direction, visibility, nearby en route impact, nearby thunderstorm impact and operational network impact at the arrival airport (i.e. ORD) only due to the limitation of available datasets. The en route weather conditions are neglected, especially for the wind impact along the route. However, the high-altitude wind speed can range up to 100-120 knots, and wind information is very important for predicting the aircraft en route location and airborne time. For example, if an aircraft is flying at certain airspeed and get tailwind, then the ground speed will be much faster. Convective weather data is also of great importance for sector capacity measurement and reroute of aircrafts, which usually is handled image process techniques. In such case, there exist limitations on the capability and accuracy in the prediction model considering its variables.

In the models, we performed multiple data filtering steps for our designed variables and used one hot encoding to transform categorical variables to dummies. 80% of samples is used as training set and 20% is for test set. In the end, three methods – ordinary least squares regression, random forest and neural network—are adopted to do the prediction.

Table 15. Description of variables

Variables	Description
Y-variables	
Flight enroute time	
XEn route time (in min), which is the time difference between actual wheels-on time and actual wheels-off time but excluding the queuing time near terminal area and MIT en route delay.-variables	
Operational variables	
FILED_T	Estimated Minutes En route set in Flight Plan.
Gate_Delay	Gate delay time (in min) at departure airport, calculated as: actual gate-out time - scheduled gate departure time.
LCC	Takes value 1 if the airline is low-cost carrier.
FT_A, FT_N	Describe the flight type (i.e. N = Domestic departure and arrival; A = Foreign departure, domestic arrival), which takes value 1 if the flight is N/ A, and 0 otherwise
AC_C, AC_T, AC_F, AC_O	Describe the aircraft type (i.e. C = Commercial; T = Air Taxi; F = Freight; O = Other), which takes value 1 if the aircraft type of the flight is C/ T/ F/ O, and 0 otherwise.
Num_Runway	Count the number of arrival runways in operation based on the actual wheels on time of the flights.
Temporal variables	
Daytime	Takes value 1 if the flight is in the day (departure between 7:00 and 20:00), and 0 otherwise.
Weekend	Takes value 1 if the flight operates on during weekend, and 0 otherwise
Summer / Fall/ Winter	Takes value 1 if the flight is scheduled to depart in Summer / Fall/ Winter and 0 otherwise.
Meteorological variables	
Ceiling	Ceiling (In hundreds of feet) of ORD at the flight-plan arrival time.
Temperature	Temperature (in Fahrenheit) of ORD at the flight-plan arrival time.
Visibility	Visibility (in miles) of ORD at the flight-plan arrival time.
Coef_Runway_Wind	Measure the effects of windspeed and wind angle, calculated as, (Angle between arrival runway and wind)/180 × wind speed.
En route_Impact	Measure the nearby en route condition of ORD at the flight-plan arrival time, which is a categorical variable including values of None/ Minor/ Moderate/ Severe.
Near_By_TS_Impact	Measure the impact of thunderstorm of ORD at the flight-plan arrival time, which is a categorical variable including values of None/ Minor/ Moderate/ Severe.
OPSNET_Impact	Measure the impact of traffic control condition of ORD at the flight-plan arrival time, which is a categorical variable including values of None/ Minor/ Moderate/ Severe.

2.3 Reallocation models

Having the *model planned arrival time* for the restricted flights from previous section, we can reallocate the flights to land on the runways. We propose two reallocation algorithms.

The first one is a 15-min aggregated reallocation algorithm. We use the time interval of 15 minutes (i.e. quarter hour which is the time interval adopted by FAA to measure the airport operational conditions in ASPM) as the unit of discrete time period. Then we will carry out the reallocation by aggregating the available waiting flights at the end of each quarter hour and locating them into the airport within its quarter-hour capacity. In this case, the changes in queueing delay and arrival delay are calculated by taking the difference between the counts in each quarter hour and multiplying by 15 minutes. This reallocation algorithm is illustrated in Table 16: *Reallocation Algorithm 1*.

The second algorithm is a flight-level reallocation. It regards the airport as a multi-server system to serve a maximum of one customer (i.e. aircraft) at any given server (i.e. available runway). The number of servers during a quarter hour equals to the number of operational runways during the same time period. According to first-come-first-served assumption, we can allocate flights in sequence based on its *model planned arrival time*, the *model actual arrival time* of its previous flights onto each available server and the required headway. The required headway (in seconds) functions as service time which assumed to be deterministic but time-varying, and it is calculated by dividing 900 seconds (= 1 quarter hour) by the airport capacity and multiplying by the number of operational runways for each quarter hour. The details of this algorithm are presented in Table 16: *Reallocation Algorithm 2*.

Table 16. Algorithms for two reallocation methods

Reallocation Algorithm 1_ [15-min aggregated reallocation]
- Inputs: $CP = \{CP_1, \dots, CP_i, \dots, CP_T\}$ – cumulative planned arrivals w/o MIT in quarter hour i $CA = \{CA_1, \dots, CA_i, \dots, CA_T\}$ – cumulative actual arrivals in quarter hour i $AR = \{AR_1, \dots, AR_i, \dots, AR_T\}$ – airport capacity per quarter hour i $AQ = \{AQ_1, \dots, AQ_i, \dots, AQ_T\}$ – total actual queueing delays (in min) in quarter hour i - Outputs: R_AA – reallocation actual arrivals w/o MIT for each quarter hour R_CA – reallocation cumulative actual arrivals w/o MIT for each quarter hour $\Delta Queue_Delay$ – total change on queueing delays (in min) when MITs were removed $\Delta Arrival_Delay$ – total change on arrival delays (in min) when MITs were removed <pre> R_AA[1]=min{AR[1], CP[1]} R_CA[1]=0 + R_AA[1] for i from 2 to T, plan_arrival= CP[i] – R_CA[i-1] if AR[i] > plan_arrival: R_AA[i] = plan_arrival else: R_AA[i] = AR[i] R_CA[i] = R_CA[i-1] + R_AA[i] ΔQueue_Delay = sum(AQ - (CP - R_CA) × 15) ΔArrival_Delay = sum (R_CA - CA) × 15 </pre>
Re-allocation algorithm 2 _ [flight-level algorithm]

- Inputs:
 - $AQ = \{AQ_1, \dots, AQ_k, \dots, AQ_N\}$ – actual queueing delays (in min) for flight k
 - $PAT = \{PAT_1, \dots, PAT_k, \dots, PAT_N\}$ – model planned arrival time w/o MITs for flight k
 - $AAT = \{AAT_1, \dots, AAT_k, \dots, AAT_N\}$ – actual arrival time for flight k
 - $AR = \{AR_1, \dots, AR_i, \dots, AR_T\}$ –ARR_RATE from ASPM quarter hour data (15-min interval)
 - $RN = \{RN_1, \dots, RN_i, \dots, RN_T\}$ –the number of operational runways from ASPM quarter hour data (15-min interval)
- Outputs:
 - $MAT = \{MAT_1, \dots, MAT, \dots, MAT_N\}$ – Model arrival time for flight k
 - $\Delta Queue_Delay$ – total change on queueing delays (in min) when MITs were removed
 - $\Delta Arrival_Delay$ – total change on arrival delays (in min) when MITs were removed

```

MAT[1] = PTA[1]
Headway = 15×60 / AR[i] × RN[i] for i in range (1, 365×24×4)
T=900    #note: time interval = 15 min × 60 sec/min = 900 sec
Runway_time = zeros(RN1)    #note: initialize the available time runway for runway
h = Headway[0]    #note: find the headway for 1st quarter
for k from 2 to N,
    if PAT[k] >= T:
        #note: check if flight plans to arrive into a new quarter, then runway and headway need to be updated
        T=T+900    #note: update the time threshold
        Runway_time = zeros(RN2) + T    #note: update the available time for runways
        h = Headway [T÷900]    #note: update the headway requirement
    MAT[k] = max {min { Runway_time } + h, PTA[k]}
    Runway_time [argmin{ Runway_time }] = MAT[k]
ΔArrival_Delay = sum (MAT - AAT) ÷ 60    #note: convert second into minute
ΔQueue_Delay = sum (AQ - (PAT - MAT) ÷ 60)    #note: convert second into minute

```

3. Case study: ORD airport

After reviewing the existing research and checking the availability of datasets, we focus on the MIT problem at Chicago O'Hare International Airport (ORD) for the whole year of 2018. ORD airport is selected for testing based on its extensive use of MIT restrictions and the conditions for which the MITs are typically implemented, which will be discussed in Section 3.2.

3.1 Data sources

We collected four datasets from two data sources, the MITER corporation and FAA's Aviation System Performance Metrics (ASPM) for calendar year of 2018.

a) ASPM Flight Level Data

This dataset contains records of commercial, general aviation, air taxi, and military flights which departure or arrive at 77 major airports in the U.S. Each record includes scheduled, flight-plan and actual flight times and the corresponding delays (e.g. gate-out delay, gate-in delay) for a flight. The time decomposition of flight time, which is relevant to our study is shown in Figure 24.

b) ASPM Airport Quarter Hour Data

This data set captures the operational conditions such as acceptance rate, and weather information including temperature, ceiling, windspeed, and visibility at top 77 commercial

airports, at 15-minute intervals. From ASPM, ARR_RATE (Airport supplied Arrival Rate for capacity) and AAR_EFF (Efficiency Airport Arrival Rate) are available during each 15-min time period, but some quarter hours have actual arrivals larger than ARR_RATE and AAR_EFF. So, we make a simple and straightforward assumption on airport capacity as,

$$\text{Airport capacity} = \max \{ \min \{ \text{ARR_RATE}, \text{ARR_EFF} \}, \text{actual arrival} \} \quad (9)$$

c) *Post-processing National Traffic Management Log (NTML) Restriction Data (MITER)*

This dataset was generated from the raw NTML records by correcting data entry errors, improving identification of restriction information, and filtering and merging the updated restriction records. It includes all TMIs in the NAS in 2018. Information on restriction type, sourced airport, issued/start/end time, duration, requiring/ providing facility and pass-back are all available.

d) *Layered TMIs Metrics for ORD arrivals (MITER)*

This data set includes flight delays from TMIs (i.e. GDP, GS, AFP, MINT and MIT, together with TBFM) of TMI-restricted flights arriving at ORD in 2018, but the data doesn't include flight delays of GDP/GS/AFP for Sept.-Dec. 2018. Each record represents a flight delay per TMI event, so there may be multiple records for the same flight. Information on flight ID, arrival/departure time, TMI type, delay type and delay duration are available.

3.2 ORD airport and relevant statistics

In Post-processing NTML data, TMIs including MIT, TBFM, APREQ, STOP, MINIT, GS, GDP, and DSP were recorded for 2018. In total, over 345 thousand TMIs were implemented in the NAS and the annual frequency of each TMI is shown in Figure 31 (left). MIT restrictions count for about 59%, making MITs the most commonly used TMI in 2018. The side by side box chart in Figure 31 (right) shows the duration of each TMI. GDP has the longest average duration about 7.37 hours with standard deviation of 12.17 hours and skewness of 1.117, while MIT has duration of 2.56 hours with standard deviation of 3.46 hours and skewness of 3.158.

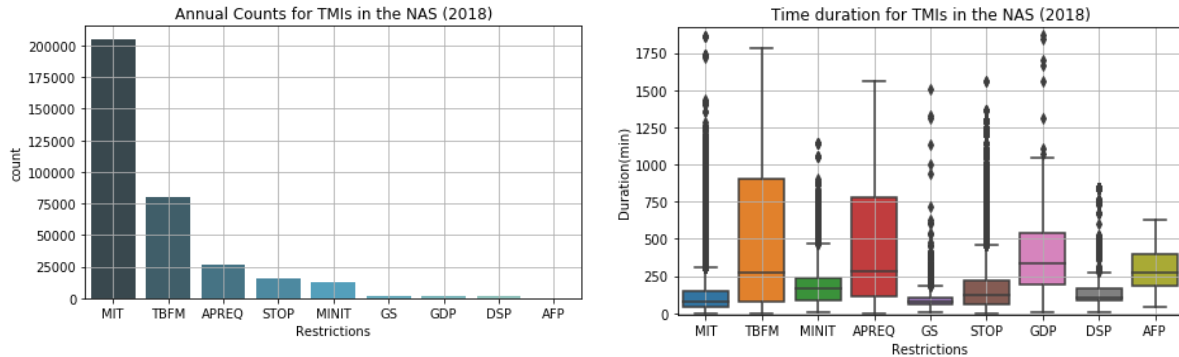


Figure 31 Bar chart (left) and box chart (right) of TMI behaviors (2018)

As discussed in Section 1, MIT is adopted to manage the imbalance between capacity and demand at a given airport, or to provide space in flows to support merging streams in the sectors. In the datasets, 4% MIT restrictions were used for traffic imbalance at terminal areas, and the bar chart in Figure 32 presents the annual frequency of MITs sourced from the top 20 airports in the U.S. in 2018. We can observe that ORD airport was the source the most MITs, 24,337 in total,

which counts over 10% of the overall MIT records, followed by DCA (15,101 MITs), JFK (10,520 MITs) and EWR (10,209 MITs).

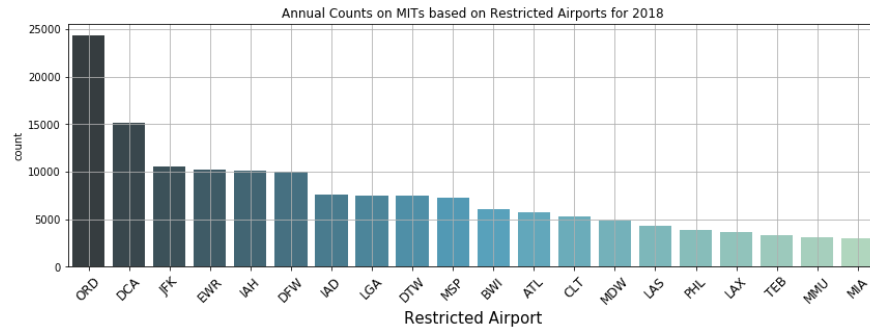


Figure 32 Annual Counts on MITs based on Restricted Airports for 2018

More in-depth study of TMI programs at ORD airport, especially for MIT restrictions, was performed to reveal the basic operations and usage patterns. ORD airport generated 28,012 TMIs in 2018, and Table 17 shows the frequency for each TMI. MITs with the highest frequency led to 471,341min (= 7,855 hr) delays while GDP with relatively low frequency generated the largest total flight delay of 1,615,621 min (= 26,927 hr).

Table 17. ORD-sourced TMIs

TMI sourced from ORD (2018)	Annual frequency	Total delay counts	Total delay duration (min)
MIT	24,341	57,987	471,341
APREQ	2,297	2,686	29,847
STOP	565	458	10,811
TBFM	389	44,880	314,766
MINIT	193	288	5,096
GS	133	3,893	257,043
GDP	62	18,844	1,615,621
DSP	32	8	203

On average, there were around 67 MIT restrictions sourced from ORD airport each day and the distribution of daily MIT counts is plotted in Figure 33(a), with maximum number of 104 MITs on July 20th, 2018. The usage patterns of MITs can be further decomposed into each quarter hour and the side-by-side box plot in Figure 33(b) presents the distribution of usage frequency for each quarter hour along the day over the whole year of 2018. We can observe relatively high volume of MIT use during morning peak around 7:00 am and afternoon peak from 4:30 pm to 6:00 pm, while in the midnight and the morning, there is almost no use of MITs.

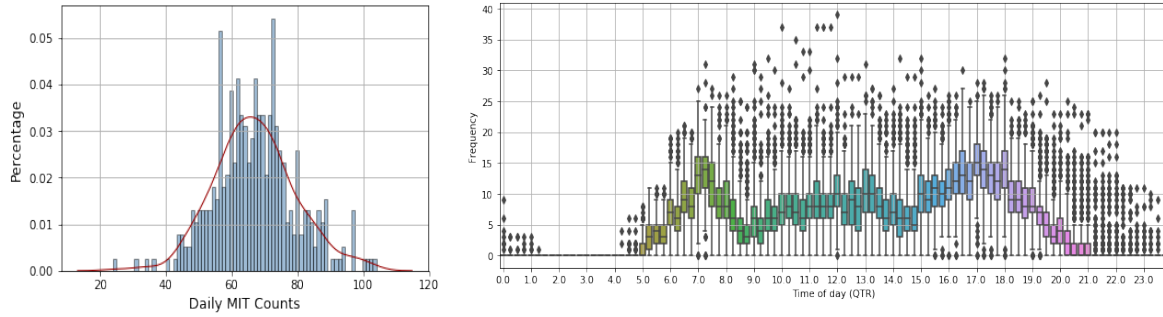


Figure 33 Usage frequency of MITs on daily base (left) and quarter-hour base (right)

Besides use frequency, other parameters of MIT restrictions, such as spacing and duration, are also analyzed. ORD-sourced MITs in 2018 have spacing ranges from 10 nm to 250 nm, while most spacing ranges from 10 nm to 30 nm. The histogram in Figure 34 shows the MIT spacing frequency distribution, and in the data samples, the most frequent spacing value is 20 nm which occurs 43.78% of the MIT restrictions. The distribution of MIT duration is right skewed with average of 106 min and standard deviation of 98.75 min.

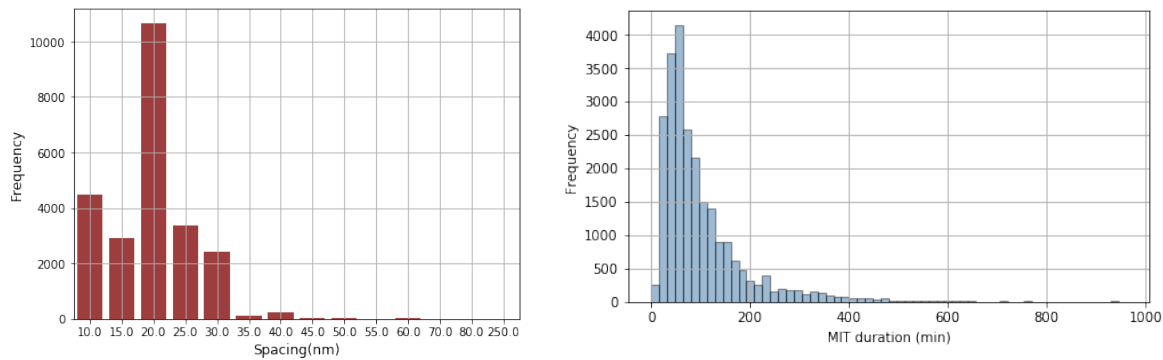


Figure 34 Histogram of ORD-sourced MIT Spacing (left) and duration (right) (2018)

The MITs sourced from ORD airport are either direct issued from its control centers or issued from the backward control centers as passback restrictions. In 2018, 57% of MITs (14,424) are the direct restrictions and the remaining 43% (14,424) are pass-back restrictions. Direct MIT restrictions have a longer average duration of 117.14 min with standard deviation of 111.15 min and a smaller required spacing of 17.36 nm with standard deviation of 5.94 nm, compared to the passback restrictions (duration: mean = 90.28 min, std = 74.40 min; spacing: mean = 23.04 nm, std = 7.03nm). Considering the direct MIT issued from ORD, the map in Figure 35 shows the MIT frequency on the surrounding FIX/waypoint. The blue dot is ORD airport, while the red dots represent the FIX/waypoint and the size of the red dot indicates the frequency. Traffic flows from ZOB experienced the most restrictions (4,667 MITs), and followed by ZMP (3,433 MITs), ZKC (2,922 MITs) and ZID (2,864 MITs).

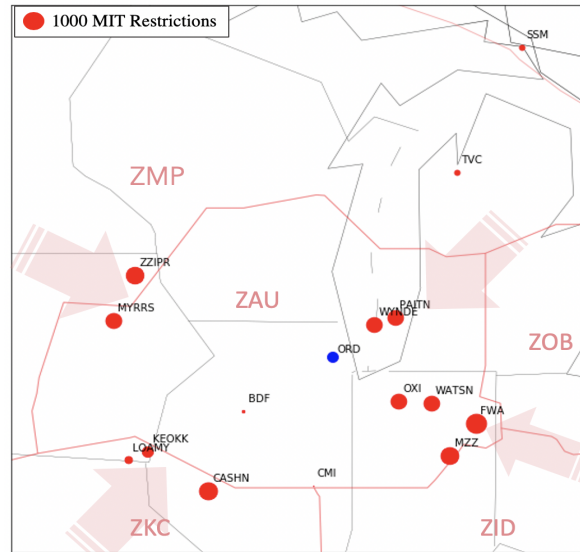


Figure 35. Frequency of direct MIT issued by ORD for surrounding FIX/waypoint (2018)

Among arrival flights at ORD, 57,987 flight delays of 471,341 min were generated by ORD-sourced MITs as shown in Table 17. The MIT flight delays can be categorized into departure, vectoring and holding delays and Table 18 presents the counts, average duration and standard deviation of different delay types for flight delays resulted from ORD-sourced MITs. Checking the flight ID and arrival time, the flight delays can be grouped for the same flight, which indicates that a flight might experience several delays from different MIT restrictions. Table 19 shows the number of MIT restriction per ORD arrival flight. It is found that 54,217 arrival flights, which counts for 12.18% of total ORD arrival flights were restricted by ORD-sourced MITs during 2018 with a daily average of 148 flights.

Table 18. Summary tables for flight delays caused by ORD-sourced MITs (2018)

Delay type	Counts	Average flight delay (min)	Std. of flight delay (min)
Departure	40153	10.03	16.28
Holding	739	18.30	12.88
Vectoring	17095	3.23	4.02

Table 19. Summary tables for MIT restrictions encountered per flight (2018)

MIT restrictions encountered per flight	Counts (flights)
1	50,540
2	3,594
3	74
4	8
5	1

4. Results and Discussion

Section 3 introduces the methodology of queueing diagram and the reallocation algorithm to quantify the impacts of MIT at a given airport. In this section, we apply these methods to ORD airport for 2018 and the results are presented in the following contents.

4.1 Reallocation results for overall performance

We have calculated the *actual arrival time*, *planned arrival time* and *model planned arrival time* for each arrival flights at ORD airport in 2018. Table 20 summarizes the corresponding 6 settings mentioned in Section 2.2 for *model planned arrival time* under different hypotheses in the absence of ORD-sourced MITs. In particular, Case 4 includes 3 similar settings but with different predictive models on flight en route en route – linear regression (LR), random forest (RF) and neural network (NR).

Table 20. Summary on difference cases for model planned arrival time

No	Case	Description
1	1	= actual wheels off time + min {estimated time en route, actual time en route} – ORD-MIT delay
2	2	= actual wheels on time – ORD-MIT delay
3	3	= actual wheels off – ORD-MIT ground delays + min {estimated time en route + other TMI en route delays, actual en route time – ORD-MIT en route delay}
4	4 (LR)	= actual wheels off – ORD-MIT ground delays + min { OLS regression model time en route + other TMI en route delays, actual en route time – ORD-MIT en route delay}
5	4 (RF)	= actual wheels off – ORD-MIT ground delays + min { random forest model time en route + other TMI en route delays, actual en route time – ORD-MIT en route delay}
6	4 (NR)	= actual wheels off – ORD-MIT ground delays+ min { neural network model time en route + other TMI en route delays, actual en route time – ORD-MIT en route delay}

The reallocation is carried out to quantify the delay impacts and following Table 21 and Table 22 summarize the results of change in queueing delays and arrival delays for difference cases in the absence of ORD-sourced MITs under 15-min aggregated reallocation and flight-level reallocation.

Table 21. Change in arrival delays & queueing delays under 15-min aggregated reallocation

	Case 1	Case 2	Case 3	Case 4 (LR)	Case 4 (RF)	Case 4 (NR)
Saving on arrival delay per day (hr)	18.584	14.727	14.958	15.796	16.269	15.419
Saving on arrival delay per ORD-MIT flt (min)	8.350	6.617	6.720	7.097	7.310	6.928
Increase on queueing delay per day (hr)	-0.643	3.214	2.982	2.145	1.672	2.522
Increase on queueing delay per ORD-MIT flt (min)	-0.289	1.444	1.340	0.964	0.751	1.133

Table 22. Change in arrival delays & queueing delays under flight-level reallocation

	Case 1	Case 2	Case 3	Case 4 (LR)	Case 4 (RF)	Case 4 (NR)
Saving on arrival delay per day (hr)	20.795	13.551	14.662	15.609	16.158	15.251
Saving on Arrival delay per flt (min) [non-MIT flt, MIT flt]	-0.335, 12.229	-0.192, 7.754	-0.242, 8.665	-0.241, 9.093	-0.241, 9.339	-0.240, 8.923
Increase on queue delay per day (hr)	-2.853	4.391	3.279	2.332	1.783	2.690
Increase on queue delay per flt (min) [non-MIT flt, MIT flt]	0.335, -4.066	0.192, 0.408	0.242, -0.501	0.241, -0.930	0.241, -1.176	0.241, -0.760

In Table 21, the change in arrival and queueing delays on the daily average and ORD-MIT-flight average is calculated. As the reallocation adopted here is on a quarter-hour basis, the queueing and arrival delay is not assigned to individual flights but quantified at a macro level. If a flight's planned or actual arrival time change is within the same quarter hour, we can't capture it from the model. And if a flight's planned or actual arrival time is across a quarter hour, it is counted as 15-min change. As a result, the individual flight performance cannot be obtained. Similarly, Table 22 summarized change in delays under flight-level reallocation algorithm, and in this model, the queueing and arrival delay change for each flight can be calculated at a more granular level. So, the performance of MIT restricted flights and non-MIT flights can be compared.

As discussed in Section 2.2, Case 1 yields a planned arrival time that is unrealistically early. Consequently, Case 1 generates larger arrival delay saving estimates and small or even negative queueing delay increases in both Table 21 and 22. Case 2 defines the time later than it should be and we discover it has the smallest arrival delay saving and the largest queueing delay increase compared to other cases when using the same reallocation algorithm. Considering the problems in Case 1 and Case 2, we believe the remaining two cases provide more reasonable and accurate results. For the 15-min aggregated reallocation in Table 21, we can see Case 3 exhibit similar values with Case 4. There are 15.5-16.5 hours in arrival delay saving per day and 1.5-2.5 hour increase on queueing delay. For the flight-level reallocation in Table 22, the values for Case 3 and Case 4 are even closer with saving on arrival delay per day ranging from 11.1 to 11.6 hours and increase on queueing delay per day ranging from 6.3 to 6.8 hours per day. In additions, non-MIT flights experience more arrival delay which means they tend to arrive later when removing MITs with longer queueing time, while MIT flights, on average, tend to arrive earlier and experience shorter queueing time if MITs were removed.

Comparing tables 21 and 22, we find that that the aggregated reallocation presents similar results with flight-level reallocation. The 15-min aggregated reallocation model assumes that the flights planning to arrive during the same quarter are located to arrive at a uniform rate, while the flight-level reallocation model, in contrast, can reflect the unevenness of the actual arrival, and is thus expected to predict somewhat larger saving on arrival delays than the aggregated model which is consistent with Table 21 and Table 22 except for Case 1.

4.2 Evaluation on MIT efficiency

Two assessment measures are introduced in this subsection, and a pilot test on individual MIT performance is presented in the end. It should be noted that the assessments are all evaluated for Case 3 under the flight-level reallocation algorithm.

The first measurement focuses on the change in throughputs for each quarter hour. When we carry out the reallocation for flights without MIT delays, the arrivals in each quarter could increase, decrease or remain the same. Figure 36 presents the change in throughputs of quarters (=35,040 in total) in 2018. The change in throughput for a given quarter hour is calculated as its reallocation throughput subtracted by its actual throughputs. Overall, 18.03% quarters experience throughput reduction when implementing MITs, and 17.73% have throughput gain. In detail, the side by side bar chart in Figure 37 shows the total throughput gain (the orange bar) and loss (the blue bar) for each quarter hour during 2018 when implementing MITs. For example, for the quarter hour of 6:00~6:15, there is around 102-flight loss in throughput and 5-flight gain in throughput during 2018. The bar chart shows that ORD airport during 8:00-9:00, 11:00-12:00 and 17:30-18:30 tends to experience more throughput reduction when implementing MITs and those corresponding delayed flights seems to be allocated into 9:00-10:00, 12:00-13:00 and 18:30-19:30 respectively. This conclusion is drawn from macro level and the individual flight reallocation can be studied in the future.

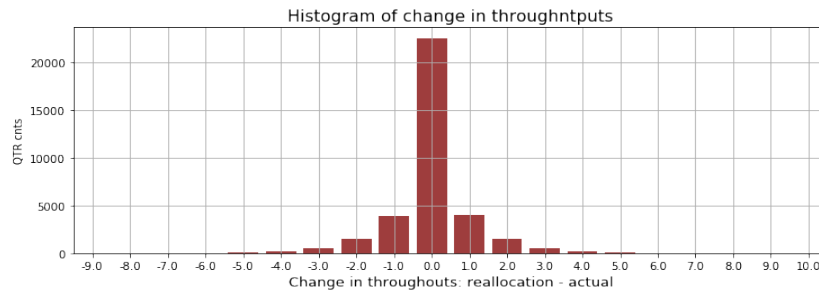


Figure 36. Change in quarter-hour throughputs for Case 3 under flight-level reallocation

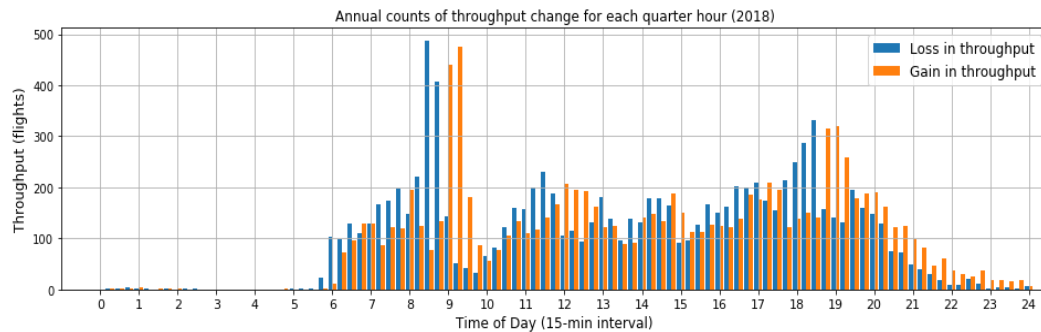


Figure 37. Annual counts of throughput change for each quarter hour (2018)

Second measurement is the change in queueing and arrival delay. Both delays for each flight can be grouped into day base and Figure 38 plots these two delays along the day of year. The change in flight delay is calculated as: *model flight delays without MITs* – *actual flight delays*. The yellow area represents the increase of queueing delay if the MIT were removed, while the red area indicates the saving from arrival delay. As we can see, for some days, removal of MITs saves substantial arrival delay with minimal increase in queueing delay, while there are also days queueing delay increases and arrival delay decreases are of comparable magnitudes. The change in arrival delay from removing MIT restrictions is always negative. If MIT restrictions are eliminated, the restricted flights would arrive no later than their original arrival time, while some of the non-restricted flights may arrive later than their original times because restricted flights join

the queue ahead of them. The next impact, however, is inevitably to reduce—or at worst leave unchanged--arrival delays. As for the change in queueing delay, 91.8% days of the year have positive change in queueing delay while the remaining 8.2% days have decrease in queueing delay when removing MITs.

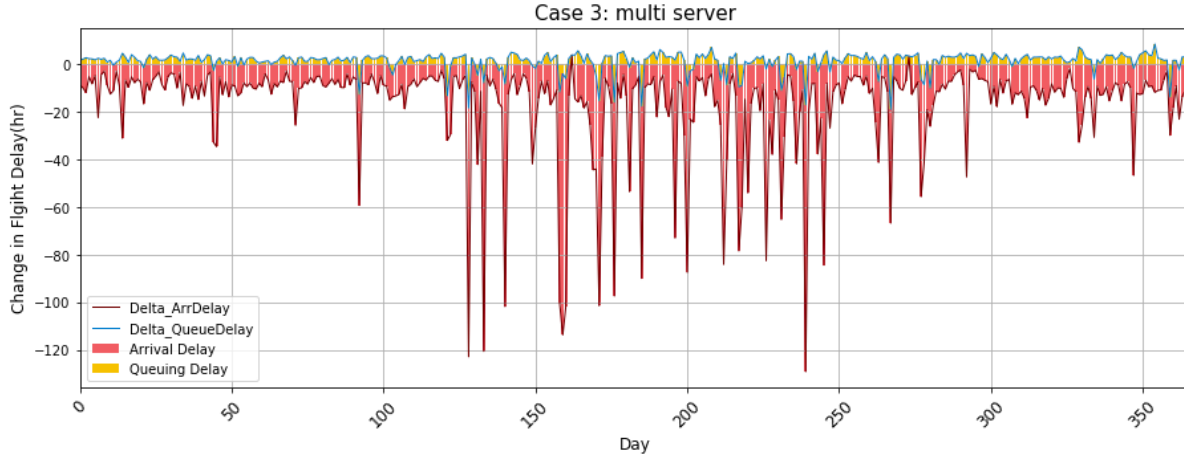


Figure 38. Change in flight delay on the day base for Case 3 under flight-level reallocation algorithm

This suggests a performance metric, which we denote β , based on the changes in queueing and arrival delay when MITs were removed to measure the performance of MITs as follows:

$$\beta = \frac{\Delta \text{Queueing Delay}}{\Delta \text{Arrival Delay}} = \frac{\text{Increase on queueing Delay}}{\text{Saving on Arrival Delay}} = \frac{\text{Benefits of implementing MIT}}{\text{Costs of implementing MIT}} \quad (10)$$

When implementing a MIT, the ideal outcome is that there will be no increase on arrival delay but a decrease in queueing delay compared to the situation without MIT. So, in the reallocation model, we can regard the decrease in arrival delay as the cost of implementing MIT and increase on queueing delay as the benefit. Because the total saving on arrival delay is non-negative, then the value of β can be evaluated as follow,

- (1) If $\beta < 0$, there is reduction on both queueing delay and arrival delay if MITs are removed. MITs in the category appear to have no apparent redeeming value.
- (2) If $\beta \geq 0$, there is reduction on arrival delay but an increase in queueing delay if MITs are removed. The larger the value of β is, the better the performance of the MIT.

Under this setting, we can calculate the β for each day, but the day-based evaluation describes the operations of multiple MIT restrictions in that day, which is not detailed enough. However, Eq. 10 can be applied to the measurement of individual MIT performance. Similar ideas on trade-off between terminal delay and en route/ground delay are presented in Christien (2019).

Last, a pilot test on the individual MIT performance has been studied. For a given MIT issued from ORD, we detect its restricted flights and match the MIT delays. Next, the model planned arrival time is derived and the flight-level reallocation is carried out if the given MIT is removed. So, the corresponding change in queueing delay and arrival delay can be calculated. There are 13,104 MITs sourced from ORD which involve restricted flight delay, and Figure 39 plots the

performance of these MIT restrictions in term of change in total flight delays. Each dot in the graph represents a MIT removal, so increase in queueing delay and saving from arrival delays can be tracked if this MIT is removed. From the scatter plot, we can divide it into three parts:

- (1) part 1 (90.84%): flights tend to arrive earlier with longer queueing time when MIT in part 1 is removed.
- (2) part 2 (9.16%): flights tend to arrive earlier with shorter queueing time when MIT in part 2 is removed.

Accordingly, the β value in Eq. (10) can be calculated for individual MITs to quantify their performance. Figure 40 presents the distribution of β values for these 13,104 MITs. The last bin in the graph with $\beta = 10$ includes the count of MIT with $\beta \geq 10$ for easier visualization.

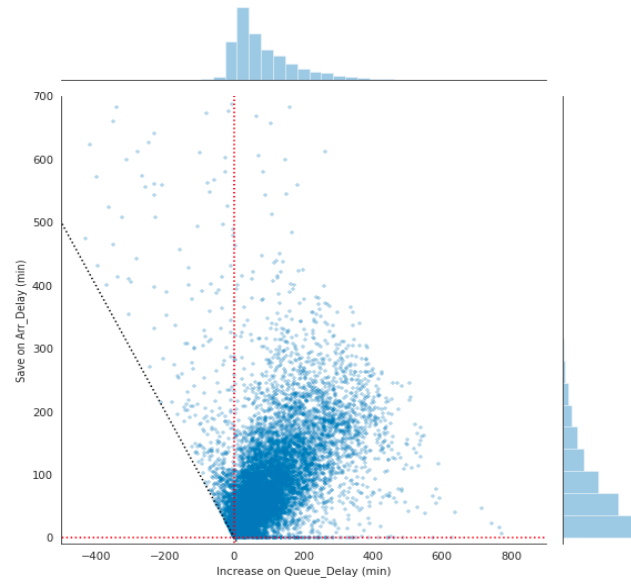


Figure 39. Scatter plot of individual MIT performance in terms of change in arrival & queueing delays

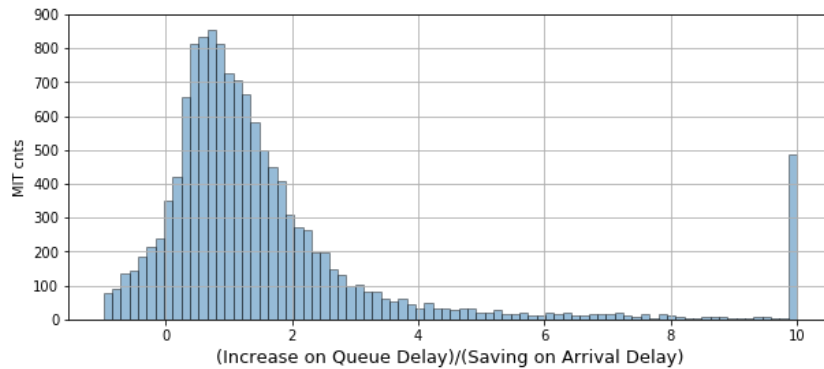


Figure 40. Distribution of β value for individual MITs (13104 MITs in total)

5. Conclusion

Our work proposed and designed a model including estimation on planned arrival time, deterministic queueing diagram and reallocation algorithm to assess the impacts of MIT restrictions on flight delays and throughput. In detail, the change in queueing and arrival delay if the MIT restrictions were eliminated can be calculated based on different methods for estimating planned arrival time. The change in throughputs at the given airport can be also computed after the reallocation algorithm and the quarter hour with high possibility of throughput loss due to MIT can be identified. This method is developed to assess the current efficiency of MITs and opportunities to improve NAS operational performance through curtailing the use of MITs.

In particular, we use ORD airport as the case study to test our model because of its high frequency of MIT usage in 2018. Statistics for the MIT use pattern for ORD in 2018 is provided. More importantly, we have tested the deterministic queueing diagram and reallocation model to estimate the operation of arrival flights at ORD if the ORD-sourced MIT restrictions were eliminated. Our results show that current operation of MITs leads to certain inefficiencies in the NAS. It is found that as we remove the MIT restrictions, there will be saving from arrival delays with a certain degree of queueing delay increase as trade-off effects. Also, unnecessary throughputs reduction is discovered base on quarter hour operation. Although the results vary depending upon the different assumptions on the airport capacity and en route flight time, the findings suggest that the MIT retractions, while saving some queueing delay, do so at sizable cost in the form of greater arrival delay. On average, ORD-sourced MITs cause about 15 hours of arrival delay per day, while saving queueing delay in the terminal area of about 2-3 hours.

There exist some limitations in the study and further work towards refining the methods is needed. First, the model only focuses on the capacity constraint at airport runways and the sector capacity or fix capacity is neglected. Although airport capacity always functions as the primary bottleneck in the NAS, the en route constraints are still of great importance when we try to remove the MIT delays from flights and reallocate their wheels on time. For example, capacities of the four arrival fixes: KRENA, PLANO, KUBBS, and BEARZ could be considered when we convert the queueing system from single server into multi servers. Second, our research results are highly dependent on the accuracy of MITRE datasets as they provided the most important information on matching flight delays with TMIs. However, when processing the MITRE datasets, we observed some shortcomings. One example is the speed reduction of flights caused by MITs could not be captured by MITRE algorithms. Another example is that ORD issued 24,341 MIT restrictions during 2018 but only 13,104 MITs have records on restricted flights from MITRE data. This suggests that there could be some restricted flights which were not matched.

These issues notwithstanding, our results strongly suggest that ORD-sourced MIT restrictions, while relieving some build-up of flights in the ORD terminal area, do so at a substantial penalty in the form of increased arrival delay. The ability of curtail the use of MIT restrictions through increased use of TBFM can therefore be expected to yield substantial operational benefit.

Chapter 4

Excess Delay and Wasted slots from the Ground Delay Program: Measurement and Causal Analysis

1. Introduction

The Ground Delay Program (GDP) has played significant role in TFM over the past three decades. When scheduled demand significantly exceeds forecasted capacity for a substantial period at an arrival airport, a GDP preassigns forecasted arrival delays to incoming flights before their departure, effectively shifting potential airborne delay to the ground. This approach enhances safety, efficiency, and minimizes fuel costs. The major triggers for a GDP include adverse weather conditions such as thunderstorms, snow and low ceiling, and airport runway/taxiway maintenance. In an ideal scenario, GDP would transfer all necessary flight airborne delays due to arrival capacity constraints to the ground before departure without incurring any extra arrival delay or throughput loss. However, challenges such as uncertainties in weather forecasting, flight time planning, and airspace resource availability often undermine its effectiveness. To address these challenges, advancements like Trajectory Based Operations (TBO) and Time-Based Flow Management (TBFM) are being developed and integrated to support air traffic controllers in managing flows and reducing system deviations (Durbin, Noonan, Knorr, Bonn, & Lloyd, 2021). These innovations are expected to lead to a substantial improvement in TFM and more precise design and execution of GDPs, potentially enabling greater throughput and lesser delays during GDP operation.

This chapter is designed to quantify operational shortfalls within the context of Ground Delay Programs (GDPs). We aim to identify opportunities for improvement by analyzing the performance of current GDP executions. While GDP performance is multifaceted, we primarily focus on two specific GDP metrics, which we term “excess delay” and “wasted slots”. Excess delay is the additional arrival delay incurred by flights under GDP, while wasted slot gauges the unutilized arrival capacity in a scenario where these flights would have otherwise taken off without a GDP-induced ground delay. To estimate these metrics, we devise a flight detection algorithm and utilize a deterministic queueing model to measure excess delay and wasted slots from an individual GDP. This methodology was applied to over 1200 GDPs implemented in 2019. The excess delay metric is then used as an outcome variable in a causal analysis, with features related to GDP planning and execution. The results reveal the potential for improvements in GDP planning and execution to reduce excess delay.

This chapter makes two key contributions. Firstly, while most existing GDP performance research is confined to one or two airports, our approach spans a large sample of GDPs across numerous airports in the NAS, yielding findings of broader generalization and comprehensiveness. In addition, our method offers a novel perspective on how airport arrivals would differ in the absence of a specific GDP, thus providing a more precise and accurate measurement of the impacts. Another key contribution of our study is the exploration of how various planning features—such as GDP scope, duration, and the accuracy of flight time estimates—impact GDP excess delays. This analysis not only highlights potential improvements in GDP execution but also examines the trade-offs between increased throughput and reduced airborne delays, further supported by advancements in TBFM and TBO.

The remainder of this chapter is organized as follows. Section 2 provides the background including existing research on GDP, introduction on GDP process and its improvement opportunities. Section 3 introduce our methodology, including the use of counterfactual queueing model for excess delay and wasted slot measurement, and causal analysis on GDP excess delay. Section 4 presents and discuss the model results. Finally, Section 5 offers conclusions and future directions.

2. Background

2.1 Literature review

GDPs have been the subject of extensive research, primarily focused on their planning, execution, and performance outcomes. Ball's early study (2001) utilized a stochastic integer program and a simulation model to examine arrivals under varied demand uncertainties during GDP operations, including flight cancellations, pop-ups, and flight arrival time variations. The study offered insights into demand uncertainty effects and the trade-offs between airborne holding and different arrival rate strategies. However, its simplified assumptions about GDP procedures and basic statistical representations of demand uncertainty limited the depth of its findings. More recent research by Bilimoria (2016) examined the additional delays imposed on flights regulated through GDP or Air Flow Program (AFP) across five U.S. airports—LGA, SFO, JFK, PHL, and EWR—in 2015. The study revealed that flights subjected to ground holding experienced significantly more gate-arrival delays, roughly two to three times more on average, compared to non-ground-held arrivals, primarily due to taxi-out phases, with airborne and taxi-in delays playing minor roles. However, the study's methodology, which relied on simple comparisons rather than rigorous statistical testing, was somewhat basic. Moreover, by combining GDP-restricted and AFP-restricted flights in its analysis, despite their different causes, the study risked confounding the assessment of factors contributing to delays. Durbin's 2021 study evaluated the under-delivery of aircraft during GDPs at NAS airports, highlighting uncertainties in EDCT compliance, taxi congestion, and variances in airline flight planning as key sources of uncertainty. By using historical data and statistical distributions, it concluded that uncertainties in departure and estimated enroute times led to arrival uncertainty, causing unused slots and inefficiencies in the NAS. However, the study's methodology had significant limitations as it treated different operational factors independently, overlooking their interrelated nature. Furthermore, its reliance on histogram distribution plots for analysis without robust statistical models or tests limited the reliability of its findings.

Other studies focused on using defined metrics to evaluate the performance of historical GDPs. An early work (Liu & Hansen, 2014) defined comprehensive GDP performance metrics, focusing on five pivotal objectives: capacity utilization, efficiency, predictability, equity, and flexibility. These metrics, calculated using available flight-level and airport data, were articulated through mathematical formulations on a uniform 0-to-1 scale, thereby allowing for both absolute and relative evaluations of GDP performance. It offered FAA traffic managers and flight operators an insightful retrospective tool for GDP performance analysis, thereby assisting in discerning trends across varying airports and time periods. This methodology was subsequently employed for evaluating GDPs issued at SFO and EWR in 2006 and 2011. A subsequent study by Ren and Kim (2018) utilized the Partitioning Around Medoids (PAM) clustering method to categorize GDPs

based on forecasted adverse weather, weather severity and stability over time, as well as GDP plan parameters at EWR in 2010 and 2014. Using Liu and Hansen's (2004) retrospective metrics, it discovered higher efficiency in GDPs with low-severity weather, frequent revisions, and larger scopes. Additionally, GDPs with high program rates, narrow scopes, and moderate durations showed higher airport utilization. However, the granularity of the clusters could potentially compromise the stability of GDP performance interpretation. (Manley & Sherry, 2010) study the trade-offs between passenger delays, excess surface fuel burn, airline equity, and passenger equity in GDP slot allocation by analyzing various rationing rules. The study utilized an official GDP Rationing Rule Simulator (GDP-RRS) to evaluate performance and equity metrics across six alternative rules. The case study conducted at EWR in 2007 revealed the efficacy of the Ration-by-Passengers rule in significantly reducing passenger delay and excess fuel burn without increasing flight delay, compared to the traditional Ration-by-Schedule. Nonetheless, when prioritizing equity for airlines or passengers, Ration-by-Schedule was favored.

Another stream of research focuses on the application of optimization techniques and machine learning methodologies to enhance GDP design parameters, planning algorithms, and execution. Such studies typically employ objective functions involving a weighted sum of ground and air delay and airline equity, while considering constraints related to uncertainties in flight delays, weather conditions, and airport capacity (Estes & Ball, 2017; Ball et al., 2010; Cox & Kochenderfer, 2016; Glover & Ball, 2013; Mukherjee et al., 2014; Bloem & Bambos, 2014). Various solution strategies, ranging from conventional machine learning approaches, dynamic programming, stochastic optimization, to reinforcement learning, have been employed to determine optimal GDP parameters or operational strategies. However, owing to the high complexity of the NAS and existing technological limitations, these models have not yet achieved a stage of practical implementation. Other research has recently leveraged machine learning models by using historical data to predict the GDP events and their decisions (Wang & Kulkarni, 2011; Kuhn, 2013; Mukherjee et al., 2014; Bloem & Bambos, 2014; Mangortey et al., 2019; Liu et al., 2019). For instance, Wang and Kulkarni (2011) utilized decision trees and neural networks for predicting GDP revisions, while Mukherjee et al. (2014) employed logistic regression to anticipate GDP incidences. Bloem and Bambos (2015) used random forest and inverse reinforcement learning (IRL) to predict GDP initiation and parameters. However, both models struggled with accurately predicting GDP initiation and cancellation, partly due to the exclusion of convective weather variables and the simplification of not allowing for GDP modifications or extensions. Concurrently, Kuhn (2016) and Mangortey et al. (2019) applied random forest and a diverse set of machine learning techniques to forecast GDP incidences, highlighting the vital role of weather variables. Lastly, Liu (2019) harnessed SVM for the regional weather-GDP incidence interplay and trained LR and RF for GDP incidence prediction. These investigations emphasize the multiplicity of factors that must be considered in GDP prediction models.

Overall, the breadth of extant literature offers many insights into the planning, prediction, and performance outcomes of GDP, which have guided this work. Despite these advancements, there is a lack of consensus on how to measure GDP performance. Moreover, while previous investigations have identified individual factors and uncertainties affecting the delivery accuracy of GDP operations, none has undertaken a comprehensive, quantitative assessment of the relative importance of these factors. While feature engineering techniques have proven useful in isolating relevant variables and improving the accuracy of predictive models, the literature lacks an integrated analysis of the potential interplay among these factors. In light of these gaps in both

literature and practice, the objective of our research is to measure GDP performance and statistically model how a wide range of planning and execution features drive performance variation. Ultimately, this analysis will reveal the potential of different improvement opportunities—from more accurate program rates to great compliance with controlled departure and arrival times—to improve GDP efficiency.

2.2 GDP process

The primary objective of a GDP is to transfer anticipated airborne delays—originating from a significant imbalance between demand and capacity at the arrival airport—to the ground. FAA utilizes the Flight Schedule Monitor (FSM) as a central decision support tool to dynamically model, implement, and monitor the GDP operation (FAA, 2021). The FSM allows traffic management specialists to set GDP parameters such as start/end times, the Program Acceptance Rate (PAR), and the departure airport scope. Specialists use the Ration-by-Schedule (RBS) algorithm to assign arrival slots, subsequently issuing a Controlled Time of Arrival (CTA) and Departure (CTD) for flights within the scope, based on their Scheduled Runway Time of Arrival (SRTA). The CTD is derived by subtracting the Estimated Time Enroute (ETE), which is the projected flight en-route time. Ground delays, represented as Expected Departure Clearance Time (EDCT) HOLD, can also arise from other TMIs like AFP and GS. Once a GDP is activated, airlines can adjust their operations through the Collaborative Decision Making (CDM) system, allowing for flight cancellations or swaps to accommodate scheduling needs. This process is followed by a compression step performed by the ATC to fill any open slots.

The GDP retains a degree of flexibility, allowing program parameters to be modified in response to factors such as EDCT compliance, changing weather conditions, and variations in air traffic. Flights are notified with updated EDCT as GDP event changes occur. So, a GDP can encompass multiple GDP events, each representing the initiation or modification of GDP plan. The GDP operation concludes when it reaches the predetermined end time, or in the event the program is cancelled earlier. The timeline for GDP planning and execution is shown in Figure 41.

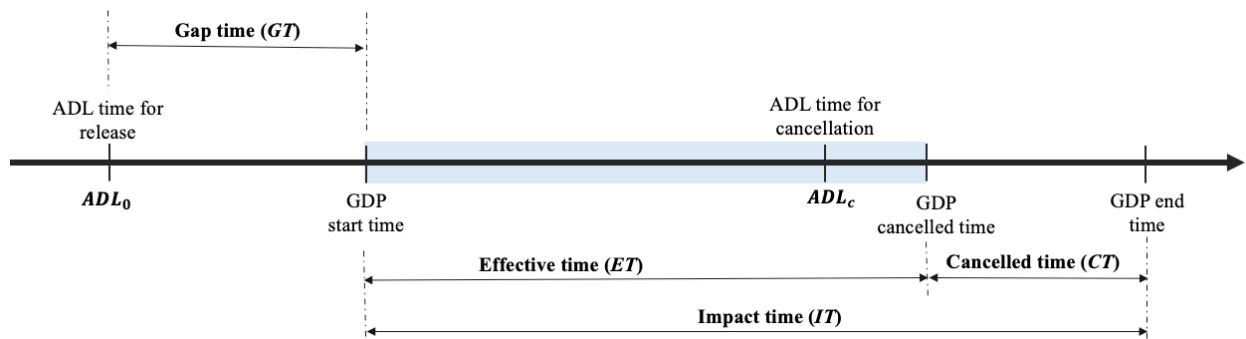


Figure 41. Timeline of GDP planning and execution phases

The terminologies relevant to GDP within this study are outlined as follows:

- **Scheduled Runway Time of Arrival (SRTA):** Defined as the Scheduled Gate Time of Arrival mins standard taxi time, typically assumed to be 10 minutes.
- **ADL Time:** The timestamp when a GDP event is modeled and released, with revisions noted as ADL_i (where i represents the i^{th} revision), and cancellations as ADLc.
- **Scope:** The set of originating airports included in a GDP, typically delineated by parameters such as the NAS sector, distance, or tier-based measures.
- **GDP Start/End Time:** Time markers for the start and end of a GDP event.
- **Canceled Time (CT):** Duration when a GDP is terminated before its planned end.
- **Gap Time (GT):** The duration between when a GDP is released and when it starts.
- **Effective Time (ET):** Duration between the GDP start time and when the GDP is canceled.
- **Impacted Time (IT):** Duration between when the GDP starts and its planned end time.
- **Program rate (PAR)** and **Arrival rate (AAR):** Represent the planned and actual hourly acceptance rates at the airport, respectively.

To illustrate the GDP process, an example from Newark Liberty International Airport (EWR) on January 12, 2018, is detailed in Figure 42. This GDP was initially modeled and released at 10:28, with a start time of 13:00 and a 12-hour duration. It underwent a revision at 16:15 and was prematurely terminated at 23:52. In the plots, the x-axis represents time, divided into quarter-hour increments, and the y-axis signifies the number of flights. The bars quantify the number of flights for each quarter-hour, whereas the horizontal line signifies the airport's declared acceptance rate per quarter hour.

Figure 42(a) displays the scheduled demand as orange bars against the initial GDP PAR estimated by air traffic control, shown as a red line, highlighting a significant demand-capacity imbalance. This led to the design and modeling of the GDP, as depicted in Figure 42(b), where blue bars indicate exempt flights not subject to GDP restrictions, either due to being outside the GDP scope or airborne at the time of issuance. The RBS algorithm was utilized to allocate slots and assign EDCT delays to GDP-regulated flights. Figure 42(c) shows a subsequent revision of the GDP, reducing the program rate to about one slot per quarter hour after three hours of execution, necessitating updates to EDCT delays for affected flights. This adjustment reflects the dynamic response to operational conditions. The program concluded earlier than scheduled, as illustrated in Figure 42(d). The actual AAR is shown by the blue line, sourced from ASPM quarter-hour data, contrasting with the final PAR (red line), evidencing a discrepancy between the planned and actual throughput. This divergence underscores the challenges in GDP planning and execution, likely exacerbated by planning uncertainties and flight time variability.

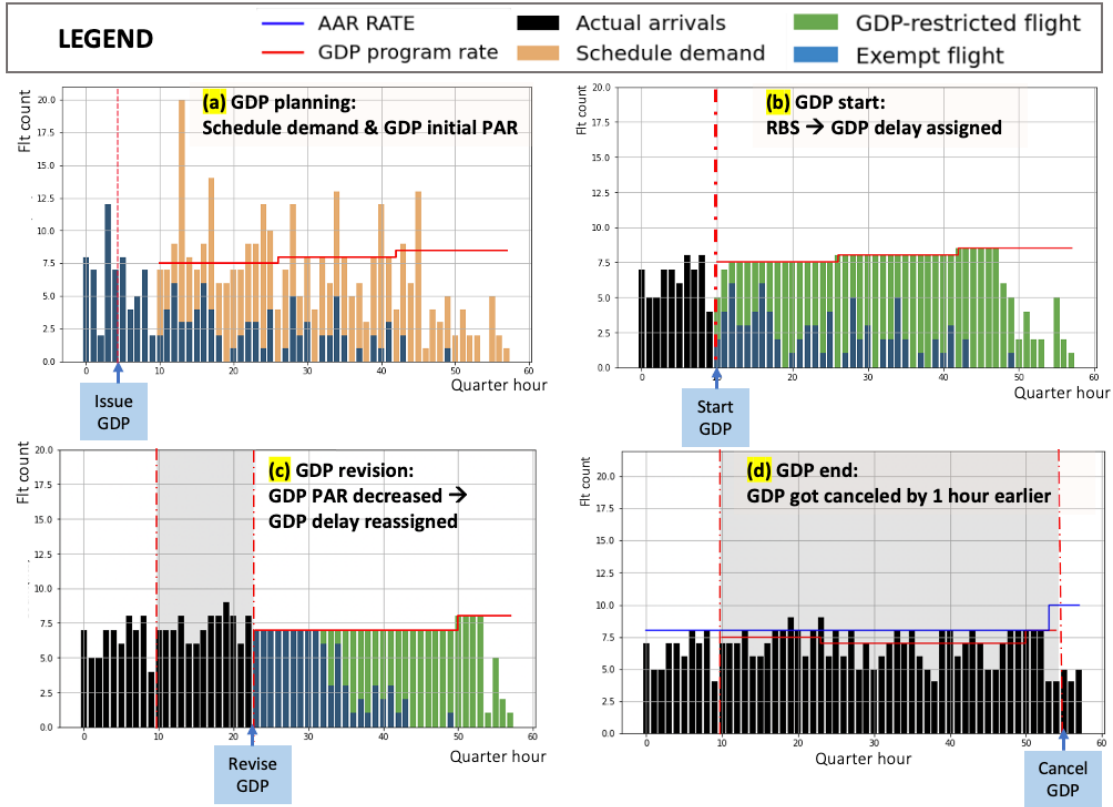


Figure 42. Example of GDP process at EWR on Jan. 12th, 2018

2.3 Improvement opportunities in GDP

GDPs are preemptively designed to shift anticipated airborne delays due to arrival capacity constraints to the ground, yet often falls short of this ideal due to unpredictable operational and environmental factors. These deviations present potential opportunities for system improvements, broadly under two areas: forecast uncertainties during the GDP design phase and operational challenges arising during GDP execution. Firstly, as the planning horizon for GDPs can span several hours, extending over a day, the inherent uncertainties in predicting future weather, airport capacity, and arrival demand often necessitate GDP revisions or cancellation to accommodate the changing traffic conditions. Furthermore, FSM functions primarily as a validation instrument rather than an optimization tool, guiding TFM actions that significantly depend on the expertise and intuition of the ATC specialists (Schummer & Vohra, 2013). This subjective involvement can introduce potential inaccuracies. Overestimation of the airport's acceptance rate may yield fewer ground delays, albeit with a consequential surge in airborne delays. Conversely, underestimation could initiate conservative ground delay assignments, culminating in heightened arrival delays and underutilized airport capacity.

The second challenge is predominantly related to the accurate delivery of flights to the GDP airport. This necessitates that flights under restriction comply with their assigned CTA. To achieve this, a high degree of precision in adherence to departure times (within ± 5 minutes of EDCT) is required to meet the CTD, while maintaining airborne times closely aligned with flight plan

enroute time ETE to ensure CTA compliance. However, variations can occur due to constraints such as departure delay from the previous flight leg, aircraft maintenance, crew availability, surface movement and terminal area limitations, weather disturbances, congested airspace, and special use airspace (Durbin, Noonan, Knorr, Bonn, & Lloyd, 2021). Unforeseen events such pop-up traffic and airline flight substitution and cancellation can further complicate accurate delivery during a GDP. In light of these uncertainties and challenges, this study quantifies the losses from these uncertainties and estimates contributions of a wide range of design and operational factors to these losses, providing valuable insights for future improvements.

3. Methodology

This section outlines our methodology for estimating and assessing excess delay attributed to GDPs. In overview, our methodology is as follows. Initially, we pinpoint crucial GDP timings—release, start/end, revision, and cancellation—as illustrated in Figure 41. Subsequently, we classify flights affected by the GDP into in-effect flights, exempt flights, post-cancellation flights and cancelled flight, followed by determining GDP delays for each completed flight. Next, we implement a deterministic queuing model to estimate the delay that would have occurred in the absence of the GDP. Comparing these hypothetical delays with the actual GDP delays facilitates the estimation of excess delay. Lastly, we construct statistical models to gauge the influence of specified features on the excess delay, thereby identifying those features which exert the most significant impact.

3.1 Data description and preparation

Our study is based on GDP operations across the entire NAS over the span of one calendar year, from January 1, 2019, through December 31, 2019. To execute a comprehensive assessment, we have assembled and processed four key datasets derived from three principal data sources: the Traffic Management Initiative (TMI) Advisories Database, the Federal Aviation Administration's (FAA's) Aviation System Performance Metrics (ASPM) and the U.S. Bureau of Transportation Statistics (BTS).

- **ASPM Flight Level Data:** This dataset embodies a compilation of individual flight records, including flight plans, schedules, Expected Departure Clearance Times (EDCT), and actual timings for flights arriving at 77 major U.S. airports. Note that this dataset exclusively comprises records of flights that were successfully executed, thereby omitting any flights that were cancelled.
- **ASPM Quarter Hour Data:** This data provides a granular insight into the operational conditions at 15-minute intervals, encompassing elements such as airport capacity (AAR or ARR RATE), runway layout, and weather conditions at the same 77 major U.S. airports.
- **ATCSCC GDP Advisory Data:** This dataset comprises records of all GDP operations within the NAS. Each record signifies an event related to GDP proposal, release, revision, or cancellation and includes pertinent details regarding scope, release timing, start/end timing, causes, program rate, and cancellation time.

- **BTS Marketing Carrier On-Time Performance Data:** This dataset was reported by US certified air carriers. It provides extra information on cancelled flights.

The collected datasets underwent an extensive data preprocessing process to ensure their reliability and consistency for subsequent analyses. This included merging and updating GDP events occurring within the same time frame, accurately matching each flight to its corresponding GDP event, aligning the scope of each GDP with the correct airports in our dataset, and standardizing all time data to UTC to preclude inaccuracies due to time zone differences. Then, the datasets were integrated into three master datasets—flight-based, airport-quarter-hour-based, and GDP-based—to facilitate a comprehensive performance evaluation and causality analysis.

3.2 GDP flight detection

In our study, we define a "GDP-affected" flight as one scheduled to land at the GDP-regulated airport during the initial GDP impact window (see 4.1), which consists of two phases: the "GDP effective phase" (ET) and the "GDP cancellation phase" (CT). We distinguish these flights into four distinct and mutually exclusive groups, with each group experiencing GDP delays in a unique way.

- a) *In-Effect Flights (IEFs)*: An IEF meets the following criteria:
- Its SRTA falls within the GDP effective time window (ET).
 - Its departure airport lies within the GDP control scope.
 - It has not taken off when the program is released.

During a GDP, we presume that *IEFs* consistently experience positive delay, implying that the assigned EDCT wheels off (*EDCT_OFF*) time should be no earlier than the flight plan off time. The calculation of the GDP delay for *IEFs* varies depending on the *Flight Plan wheels off time (FP_OFF)* relative to the GDP initial release time (*ADL₀*). If *FP_OFF* is later than *ADL₀* as Figure 43 (left),

$$GDP\ delay = EDCT_OFF - FP_OFF = EDCT\ HOLD \quad (1)$$

If the *FP_OFF* is earlier than *ADL₀* as depicted in Figure 43 (right), but *EDCT_OFF* is later the *ADL₀*, there must be some averted holding delay that postponed its departure until after *ADL₀*. This additional delay can result from several sources, including the delay propagated from previous flight segments or a GS. Therefore, in this scenario, the calculated GDP delay incorporates this pre-existing delay, as follows:

$$GDP\ delay = EDCT_OFF - ADL_0 \quad (2)$$

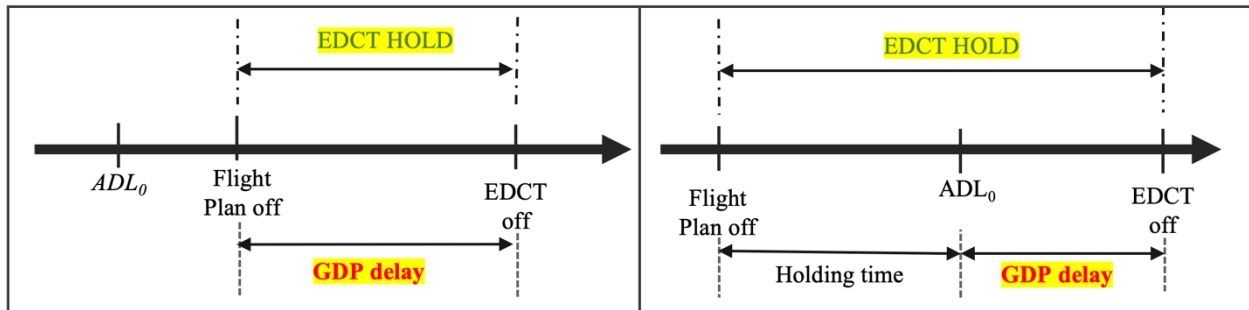


Figure 43. GDP Delay for In-Effect Flight (IEF)

- b) **Post-Cancellation Flights (PCFs)**: A PCF is defined by:
- Its SRTA falls within the GDP cancellation time window (CT).
 - Its departure airport lies within the GDP control scope.
 - Its flight plan wheels off time precedes the GDP cancellation time (ADL_c).

We assume that when a GDP is cancelled, PCFs on the ground are immediately cleared for departure. If the $EDCT_OFF$ is earlier than ADL_c , the flight has already executed the complete delay before the GDP was canceled, indicating a full realization of the planned GDP delay, as shown in Figure 44 (left). However, if the $EDCT_OFF$ is later than the GDP Cancellation window, as depicted in Figure 44 (right), the GDP-induced delay is less than what was initially anticipated. This happens because the cancellation of the GDP allows flights to depart sooner than their assigned $EDCT_OFF$, thus reducing the actual delay experienced. The formula for these situations is expressed as follows:

$$GDP\ delay = \min(EDCT_OFF - FP_OFF, ADL_c - FP_OFF) \quad (3)$$

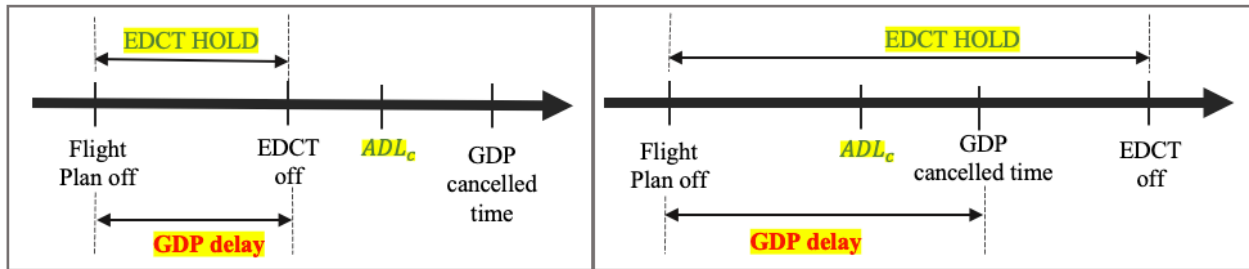


Figure 44. GDP Delay for Post-Cancellation Flight (PCF)

- c) **Exempt Flights (EFs)**: An EF is characterized as such:
- Its SRTA is within the GDP effective time window (ET).
 - Its departure airport is out of the restricted scope, or the flight has already taken off when the program is released.
- Therefore, EFs do not experience any GDP delay as they are immune to the regulations imposed by the GDP.
- d) **Cancelled Flights (CFs)**: These are flights that are cancelled by the airlines due to the imposition of a GDP. This allows the airlines to optimize their remaining schedules in light of the prevailing constraints.

For ease of reference, we use the term "*GDP-restricted flights*" (**RFs**) to collectively refer to *IEFs* and *PCFs*—those flights that are delayed as a result of a GDP.

3.3 Counterfactual queueing model

This subsection answers the question on how to measure the excess delay and wasted slots caused by a particular GDP at an airport. We construct a counterfactual scenario to model the airport arrival performance if a given GDP was not implemented. Absent the GDP, the incoming flights

that were originally under its control could potentially depart sooner, reach their terminal areas earlier, and if capacity allowed, land earlier. Therefore, we define the excess delay of a GDP as the aggregate arrival delay savings in a non-GDP scenario, while *wasted slots* represent the sum throughput shortfalls relative to capacity.

3.3.1 Modeling airport arrivals in the absence of GDP

To assess the airport arrival demand without the implication of GDPs, we followed the flight detection and GDP delay calculation methods described in Section 3.2. We identified three types of flights impacted by a GDP: IEFs, PCFs, and EFs, each subjected to calculated GDP delays as detailed in equations (1)-(3). It's important to clarify that flight departure delays can be caused by factors other than GDPs, such as delays from previous flight legs, departure congestion or aircraft maintenance. Therefore, we solely remove the GDP delay from the flights, while incorporating other ground delays, as well as airborne time deviations that are not induced by GDP. This allows us to calculate the model planned wheels-on time (MP_ON) for each flight as defined in Eq. (4):

$$MP_ON = (AGO - GDP\ Delay) + ATO + AAT \quad (4)$$

where AGO is the actual gate-out time, ATO is the actual taxi-out time, and AAT is the actual airborne time. This equation assumes that GDP delays are primarily accounted for before aircraft pushback. However, since GDP delays can affect both pre- and post-pushback phases, Eq. (4) is applicable regardless of how the GDP delay is distributed between the gate and taxi-out phases.

Figure 45 illustrates Eq. (4) for a single GDP-restricted flight. The upper timeline depicts the actual flight time, including a GDP delay. Conversely, the lower timeline represents a modeled scenario hypothetical scenario where the GDP is removed. Under this scenario, the flight is planned to reach the terminal area earlier. However, due to airport congestion, it may need to incur an additional queueing delay or holding time before it can land, effectively translating to *airborne delay saving* denoted as $\Delta\text{airborne delay saving}$. Consequently, the flight ultimately arrives earlier than in the GDP scenario, with the time saved indicated as $\Delta\text{excess delay}$ from non-GDP scenario.

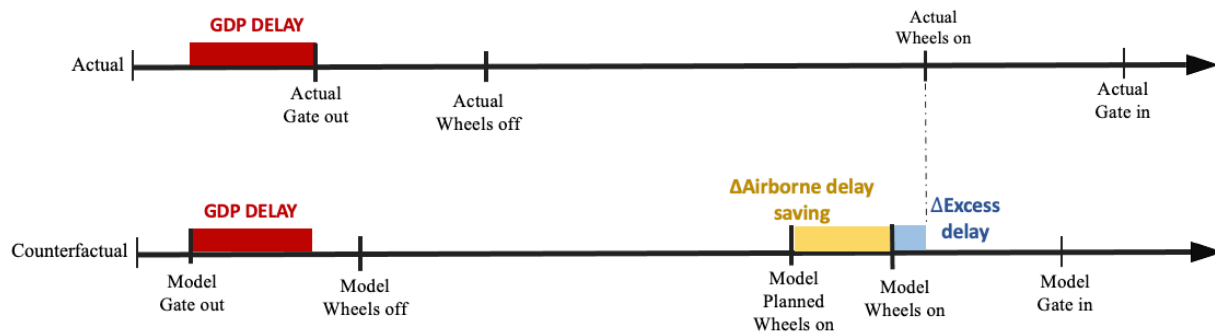


Figure 45. Flight timings for the actual and counterfactual scenarios

To estimate $\Delta\text{airborne delay saving}$ and $\Delta\text{excess delay}$ for each GDP, we construct a model based on the following parameters: airport arrival capacity, observed throughput, and arrival demand in the absence of GDPs. We analyze GDP operations by discretizing each period into

quarter-hour intervals, using a deterministic queueing model to allocate arrival slots according to modeled demand and the airport's quarter-hour capacity constraints.

We calculate the actual arrival curve, $A_{(t)}$, using the wheels-on times at the GDP-constrained airport. To predict the arrival curve in the absence of GDP, denoted as $A'_{(t)}$, we assume removing the GDP eliminates related ground delays, enabling earlier departures and potentially earlier arrivals at the model planned wheels-on time, MP_ON , with demand represented by the cumulative model planned arrival curve, $P_{(t)}$. Then, we apply a queueing model with runway capacity constraints to model the wheels-on time of flights and thus obtain cumulative model arrival curve, $A'_{(t)}$. For each quarter-hour, arrival throughput is the minimum value between the capacity ($C_{(t)}$) and the total landing demand ($TLD_{(t)}$) in that period. Total landing demand consists of the unsatisfied model demand from the previous quarter-hour and any new demand as $p_{(t)}$ within the current quarter-hour:

$$TLD_{(t)} = TLD_{(t-1)} - A'_{(t-1)} + p_{(t)} \quad (5)$$

$$A'_{(t)} = A'_{(t-1)} + \min(TLD_{(t)}, C_{(t)}) \quad (6)$$

Figure 46 visualizes these calculations. The green hatched area between $A_{(t)}$ and $A'_{(t)}$ represents the saving on arrival delay if the GDP was removed, termed *excess delay (ED)*, from removing the GDP. The red hatched area between $P_{(t)}$ and $A'_{(t)}$ in Figure 46 represents the increase in airborne delay that would occur if the GDP were removed, or equivalently the *airborne delay saving (ADS)* resulting from the GDP. In discretized form, assuming 15-minute time periods, we calculate these metrics as:

$$ED_{(T)} = 15 \cdot \sum_{t=1}^T ed_{(t)} = 15 \cdot \sum_{t=1}^T A'_{(t)} - A_{(t)} \quad (7)$$

$$ADS_{(T)} = 15 \cdot \sum_{t=1}^T ads_{(t)} = 15 \cdot \sum_{t=1}^T P_{(t)} - A'_{(t)} \quad (8)$$

Note that total ground delay, GD , imposed by the GDP is the sum of ED and ADS , and that for a “perfect” GDP $ED = 0$ and $GD = ADS$.

In addition to the delay metrics, we also measure the number of averted holding flights (AHF) and wasted slots (WS) for each quarter-hour during GDP operations. As illustrated in Figure 46, AHF at quarter hour t is determined as the difference between $P_{(t)}$ and $A'_{(t)}$. On the other hand, the number of WS measures capacity that is underutilized even though traffic demand is present. This is calculated by taking the minimum value between unused slots (calculated as the difference between arrival capacity and observed throughput for the quarter-hour) and the new served flights in the absence of GDP (the difference between $a'_{(t)}$ - $a_{(t)}$). The calculation is represented in Eq. (9) and Eq. (10):

$$AHF_{(t)} = P_{(t)} - A'_{(t)} \quad (9)$$

$$WS_{(t)} = \min\{ (C_{(t)} - a_{(t)}), \Delta_+(a'_{(t)} - a_{(t)}) \} \quad (10)$$

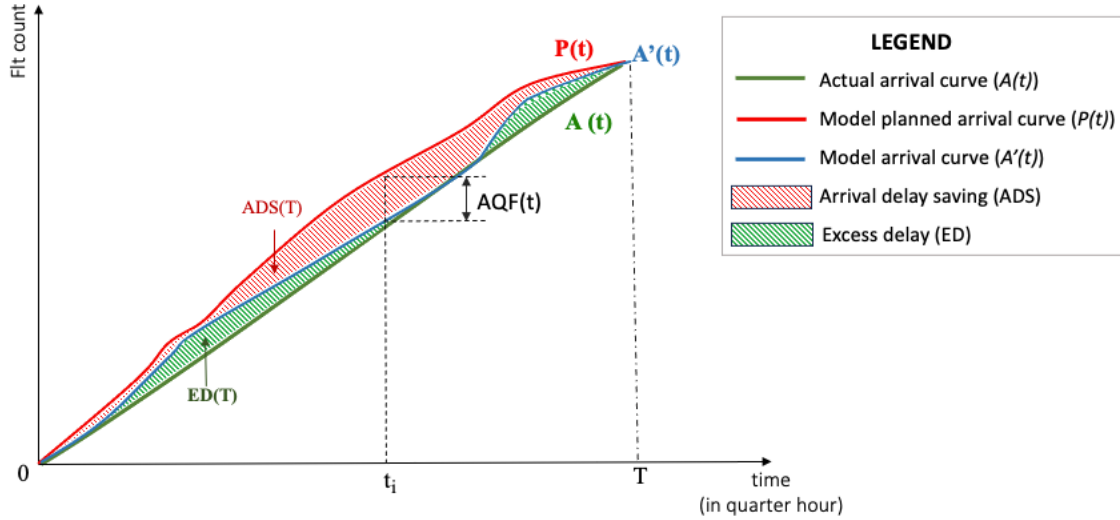


Figure 46. Queueing diagram for actual and counterfactual (model) scenarios

3.3.2 Airport capacity constraint

The queueing model includes as an input a time series of airport arrival capacities $C_{(t)}$. The capacity is, by definition, the maximum possible arrival throughput in a given time period. As a maximum, it is not directly observable. Moreover, there is a conceptual question about the true capacity variable to use when measuring GDP performance. We therefore perform our analysis using four different methods for determining $C_{(t)}$. The methods are:

a. ASPM Arrival Rate

In the ASPM quarter-hour data, the airport-supplied arrival rate, termed AAR , is an estimate of airport arrival capacity for each quarter-hour, recorded by terminal controllers based on visibility conditions, runway configuration, and arrival fleet mix. This is the most straightforward approach for determining $C_{(t)}$.

b. Throughput Adjusted Arrival Rate

The actual quarter-hourly arrival capacity is a given quarter-hour is does not precisely match the ASPM AAR , with differences resulting from arrival timing, winds, fleet mix, and human factors. This can lead to the terminal controllers to be able to handle more flights than AAR in certain quarter hours but less flights in others. Prior studies (Durbin et al., 2021) have remarked on the occasional unreliability of the AAR as an estimate of airport capacity, particularly during periods of low demand. Our investigation of the study period revealed that 19% of quarter-hours had an AAR lower than the observed throughput during the study period over GDP airports. The majority of these quarter-hours had rates shifted by one or two flights, and most of them had their preceding and successive quarter-hours with observed throughput below AAR . This inconsistency might be attributed to demand shifts or operational variations. However, in a small portion of cases there is

a significant mismatch between the rate and observed throughput, possibly as a result of data recording errors.

In light of the above, our second method for determining $C_{(t)}$ includes an upward correction to the ASPM rate when it is consistently exceeded by observed throughput. Referring to the work by Durbin et al. (2021), we implemented a tri-quarterly clamped moving average on observed throughputs ($MA_{A_{(t)}}$) for Case 2, incorporating the given quarter-hour and the quarter-hours both preceding and following it. When the average throughput across the three quarter-hours exceeds the recorded AAR in the given quarter hour, we use the average throughput instead of the ASPM rate.

c. Initial GDP Program Rate

The first two methods are based on realized capacity, while GDPs are planned based on forecast capacity. The results of the first two methods are thus strongly affected by the accuracy of the capacity predictions. To eliminate this factor, in our third method we used the initial GDP program rate (PAR_0) to determine $C_{(t)}$.

d. Final GDP Program Rate

As explained above GDP program rates are often adjusted as the program proceeds. Thus, in our fourth and final method, we determine $C_{(t)}$ from the final program rate (PAR_f). The justification is essentially the same as that for the third method.

To summarize, the capacity settings for these cases can be represented as follows:

$$\text{Case 1: } C_{(t)} = AAR_{(t)} \quad (11)$$

$$\text{Case 2: } C_{(t)} = \max (AAR_{(t)} , MA_{A_{(t)}}) \quad (12)$$

$$\text{Case 3: } C_{(t)} = GDP \ PAR_0 \quad (13)$$

$$\text{Case 4: } C_{(t)} = GDP \ PAR_f \quad (14)$$

3.4 Causal analysis for GDP excess delay

We explore how factors related to GDP planning and execution influence excess delays through a two-stage analytical framework. Initially, statistical models are developed to assess the impact of specific variables on delays caused by Ground Delay Programs. Subsequently, these models are used to simulate counterfactual scenarios to predict potential reductions in delays under various theoretical adjustments. Detailed analyses of these stages are provided in the subsequent sections, offering insights into optimizing GDP operations.

3.4.1 Statistical model development

We establish quantitative statistical models in our first stage, focusing on how multiple factors, particularly those related to GDP planning and execution stages, can influence GDP excess delay. Every observation in our set is specific to a single GDP. The response variable is the excess delay per GDP restricted flight (minutes /restricted-flight), based on the measurements from Section 3.3.

In alignment with the literature review and discussion in Section 2.3, we specify econometric models of GDP excess delay. The variables that delineate excess delay variation are categorized as planning-related, execution-related and contextual variables. The subsequent subsections provide a comprehensive description of the variables included the model specifications, which are summarized in Table 23.

a. Planning-related Features

First subset of features relates to the characteristics of the GDP planning phase. We consider effective time (*EF*), gap time (*GT*), and cancelled time (*CT*) (as depicted in Figure 41) to characterize the GDP planning horizon. A longer *EF* implies an extended planning horizon which, by default, involves more uncertainties due to its temporal breadth. The *GT* represents the period from when the GDP is first modeled and released to when it commences. An earlier release of the GDP provides increases the number of flights that can be controlled, allowing for less severe delays per controlled flight. On the other hand, weather and demand forecasts become increasingly uncertain with the expansion of this time window. The *CT*, on the other hand, signals an improvement in weather or traffic conditions which obviate the need for the GDP. In such instances, the previous GDP delay assignment might be seen as excessive, leading to an over-allocation of delay time. Furthermore, we account for the number of revisions a GDP undergoes during its execution (*n_r*). The act of revising a GDP, which may involve changes in its duration or program rate, often signals alterations in weather conditions, arrival capacity, or traffic demand. However, such revisions inevitably face limitations due to the pre-existing ground delays and active flights en route, which are set by the earlier GDP parameter settings and cannot be altered by subsequent revisions.

Additional planning phase variables considered in our model include the quantity of in-effect flights (*n_IEF*), post-cancellation flights (*n_PCF*), exempt flights (*n_EF*) and cancelled flight (*n_CF*). The *n_IEF* captures the flights that are directly controlled by the GDP, providing insights into the scale and immediate impact of the delay assignment. The *n_PCF* reflects the number of flights that are abruptly able to depart as the result of a cancellation; these flights may incur additional delay in responding to this early release. The *n_EF* captures flights that are not controlled by the GDP even though they are scheduled to arrive while the GDP is in effect. These reduce the capacity available to GDP controlled flights and may increase their vulnerability to forecast uncertainties. Airline flight cancellation is one of the direct consequences from GDP. It can lead to substantial costs to airlines and air passengers, whereas cancelling a flight sometimes can relieve the capacity-demand imbalance at arrival airports and reduce delays of some other flights during airport congestion period, especially during GDP time. The *n_CF* can help to capture the severity of disruptions in airline flight schedule that result from the GDP implementation and how the reshuffling of flights to use the slots of the cancelled flights affects GDP performance.

Longer flights have greater flight time variability and require a longer GDP planning horizon, which may degrade GDP efficiency. Moreover, as GDPs at a U.S. airport result in ground delays to flights originating from both domestic and Canadian airports, we allow for the possibility that differences in implementation between the two countries may affect excess delay. Accordingly, we introduce the average projected enroute time for U.S. and Canadian originating flights (*sc_us* for U.S. departures, *sc_ca* for Canadian departures) as model features.

A fourth set of variables related to GDP planning involve rates, including the observed airport called rate (AAR) and the GDP program rate (PAR). First, to measure the capacity shortfall severity, we calculate the average quarter-hour rate difference between AAR during the given GDP and the annual average AAR at the arrival airport, denoted as ΔAAR . This metric provides a standardized and comparative measure of how significantly the AAR during a GDP deviates from its annual average, indicating the relative strain on airport capacity. Additionally, we calculate the mean and standard deviation of the quarter-hour rate discrepancy between the observed called rate (AAR) and GDP PAR at both the initial GDP release and the final stages ($uPAR_0 / uPAR_f$, $\sigma PAR_0 / \sigma PAR_f$). The average rate difference gauges the overall bias of PARs, while the standard deviation of rate difference measures the variability of differences between PARs and realized capacities. The variables capture the accuracy GDP program rates. Overestimating the acceptance rate might yield fewer ground delays, but at the cost of considerable airborne delay increase. Conversely, underestimating could lead to a conservative strategy for ground delay assignments, potentially resulting in unnecessary arrival delays and underutilization of airport capacity.

b. execution-related features

Execution-related features concern the adherence of flights to the GDP plan so that they arrive at the GDP airport at the appropriate time. We consider fluctuations in flight times starting from gate departure, through taxi-out, and finally, during the enroute phase for both in-effect and exempt flights during a GDP. Both the mean and standard deviation of the time difference between actual and planned timings for both flight categories across the three flight stages are computed. For in-effect flights, the EDCT gate-out time is calculated as the difference between EDCT wheels-off time and unimpeded taxi-out time, using data from the ASPM flight level. Similarly, the EDCT enroute time is computed as the duration between the EDCT wheels-off and wheels-on times. Then, we can derive the mean and standard deviation of time differences between actual gate-out time and EDCT gate-out time (ΔGO_IEF and σGO_IEF), between actual taxi-out time and unimpeded taxi-out time (ΔTO_IEF and σTO_IEF), and between actual enroute time and EDCT enroute time (ΔETE_IEF and σETE_IEF). For exempt flights, we use Flight Plan wheels-off time, Flight Plan wheels-on time, and unimpeded taxi-out time to calculate the features (ΔGO_EF and σGO_EF , ΔTO_EF and σTO_IEF , ΔETE_EF and σETE_EF) following a similar methodology. Together, these variables capture how well GDP flight times match those assumed in the GDP plan.

c. Contextual Variables

Last, to account for systematic variations in GDP excess delay across different airports, we include fixed-effect dummy variables for each arrival airport and for GDP causes. The variables account for systematic and otherwise unexplained variation in GDP excess delay between airports and the conditions that give rise to the GDP. Causes resulting in GDP include a range of conditions such as snow-ice, low ceiling, thunderstorms, runway construction, and wind, among others, which can also vary significantly by airport. In the interest of maintaining robust statistical representation, airports with less than 52 annual GDPs are grouped under the category of "others". This consolidation preserves degrees-of-freedom in our data set and assures that the estimated airport fixed effects are statistically reliable. Including the fixed-effect dummy variables prevents

potentially spurious conclusions that could result airport-to-airport and cause-to-cause differences in the planning- and execution-related features discussed above.

As for the model selection, we employ and compare three types of interpretable models: OLS, Lasso, and Ridge regressions. Following several rounds of data filtering for the constructed variables, we partition the entire dataset randomly into a training set (80%) and a testing set (20%). All numerical features in the training set are standardized, and the testing set is rescaled using the same scaling parameters, including mean and standard deviation for each feature from the training set. We then apply five-fold cross-validation on the training sample for hyperparameter tuning in order to find the combination of hyperparameters that minimizes error. The use of OLS, Lasso, and Ridge regressions allows us to have different levels of complexity and regularization in our models. The Ridge and Lasso regressions, in particular, help control overfitting through their built-in penalty term. Finally, we train the models with the optimal hyperparameters on the entire training set and evaluate them on the testing set, using root mean square error (RMSE) and mean absolute error (MAE) as performance metrics.

3.4.2 Factor contribution

In the second stage of our causal analysis, we construct counterfactual scenarios to estimate the changes in excess delay of GDPs when a particular effect is removed. We achieve this by adjusting the values of corresponding features to zero. We process this altered feature space and quantify the resultant variations in excess delays from the base scenario by utilizing the statistical model trained in the prior stage. For example, suppose we assume all the in-effect flights follow their EDCT gate-out time precisely. To construct the counterfactual scenario for this perfect EDCT compliance scenario, we set the ΔGO_IEF and σGO_IEF for each observation in the data set as zero, while leaving all other values unchanged. Summing the resulting counterfactual predictions yields the second term in the numerator of Eq. (15), while the original predictions appear in the first term. The fractional reduction in excess delay is thus given by:

$$\delta = \frac{\sum_i^N \hat{f}(X_i) - \sum_i^N \hat{f}(X_{i \setminus c})}{\sum_i^N \hat{f}(X_i)} \quad (15)$$

where $\hat{f}(X)$ is the predicted excess delay generated by the selected model with feature matrix X , X_i is the original feature matrix for GDP i , and $X_{i \setminus c}$ is a new counterfactual feature matrix with the values of feature set c changed to specific counterfactual values.

Table 23. Definition of model explanatory variables

Notation	Description
<i>ET</i>	Duration when GDP is in effect (in hr)
<i>GT</i>	Duration between when GDP was release and started (in hr)
<i>CT</i>	Duration between when GDP was canceled and its initially planned to end (in hr)
<i>n_r</i>	The number of revisions during GDP execution (Cancellation is not included)
<i>sc_{us}</i>	Mean estimated time enroute between in-effect departure airport in the U.S. and GDP airport, weighted by in-effect flight counts (in hr)

<i>sc_ca</i>	Mean estimated time enroute between in-effect departure airport in Canada and GDP airport, weighted by in-effect flight counts (in hr)
<i>n_EF</i>	The number of exempt flights during a GDP
<i>n_IEF</i>	The number of in-effect flights during a GDP
<i>n_PCF</i>	The number of post-cancellation flights during a GDP
<i>n_CF</i>	The number of cancelled flights during a GDP
<i>ph</i>	Ground holding delay received by in-effect flights prior to GDP start (in hr)
<i>al_hhi</i>	Airline Herfindahl-Hirschman Index (HHI) of incoming flights during GDP at the airport
<i>AAAR</i>	Rate difference between <i>AAR</i> and airport annual average <i>AAR</i> during GDP effective time averaged by quarter hour
<i>uPAR₀</i>	Mean rate difference during GDP by quarter hour: <i>AAR</i> – initial <i>PAR</i>
<i>σPAR₀</i>	Standard deviation of rate difference during GDP by quarter hour: <i>AAR</i> – initial <i>PAR</i>
<i>uPAR_f</i>	Mean rate difference during GDP by quarter hour: <i>AAR</i> – last-updated <i>PAR</i>
<i>σPAR_f</i>	Standard deviation of rate difference during GDP by quarter hour: <i>AAR</i> – last-updated <i>PAR</i>
<i>uPAR_r</i>	Mean rate difference during GDP by quarter hour: initial <i>PAR</i> – last-updated <i>PAR</i>
<i>σPAR_r</i>	Standard deviation of rate difference during GDP by quarter hour: initial <i>PAR</i> – last-updated <i>PAR</i>
<i>ΔGO_IF</i>	Mean gate out time difference per in- effect flight: actual gate out – EDCT gate out (in min/ft)
<i>σGO_IF</i>	Standard deviation of gate out time difference per in- effect flight: actual gate out – EDCT gate out (in min/ft)
<i>ΔTO_IF</i>	Mean taxi out time difference per in-effect flight: actual taxi out – unimpeded taxi out (in min/ft)
<i>σTO_IF</i>	Standard deviation of taxi out time difference per in-effect flight: actual taxi out – unimpeded taxi out (in min/ft)
<i>ΔETE_IF</i>	Mean enroute time difference per in- effect flight: actual enroute time – ETE (in min/ft)
<i>σETE_IF</i>	Standard deviation of enroute time difference per in-effect flight: actual enroute time – ETE (in min/ft)
<i>ΔGO_EF</i>	Mean of gate out time difference per exempt flight: actual gate out – Flight Plan gate out (in min/ft)
<i>σGO_EF</i>	Standard deviation of gate out time difference per exempt flight: actual gate out – Flight Plan gate out (in min/ft)
<i>ΔTO_EF</i>	Mean taxi out time difference per exempt flight: actual taxi out – unimpeded taxi out (in min/ft)
<i>σTO_EF</i>	Standard deviation of taxi out time difference per exempt flight: actual taxi out – unimpeded taxi out (in min/ft)
<i>ΔETE_EF</i>	Mean enroute time difference per exempt flight: actual enroute time – ETE (in min/ft)
<i>σETE_EF</i>	Standard deviation of enroute time difference per exempt flight: actual enroute time – ETE (in min/ft)
<i>c_RWY/LC/ SNOW/WIN D/others</i>	Take value 1 if the cause of GDP is snow-ice / low ceiling/ wind/ runway construction and maintenance; take value 0 otherwise (cause of GDP as thunderstorm is the reference in the model)
<i>apt_XXX/ others</i>	Take value 1 if GDP airport = BOS/ JFK/ LGA/ ORD/ PHL/ SEA/ SFO/ other airports; take value 0 otherwise (Newark Liberty International Airport (EWR) is the reference in the model)

4. Results

We applied the counterfactual queueing model and causal analysis model to assess 1204 GDPs implemented for arrivals into 33 U.S. airports during 2019. This covers all GDPs for that year, with the exception of 24 GDPs that had incomplete or anomalous information, were canceled before GDP start time.

4.1 Summary GDP statistics

Figure 47 and 48 present summary statistics for the GDPs we analyzed. Figure 47(a) shows that GDPs are highly concentrated among a few airports, including EWR (212 GDPs), SFO (181), LGA (140), BOS (98), and ORD (81). These five airports account for 59% of total GDPs. There exist 12 airports which issued over 24 GDPs throughout the year, averaging approximately 2 GDPs per month. Figure 47(b) presents the actual duration between the initiation and termination of a

GDP, with vertical bar lines denoting standard deviation. The mean duration of GDPs across the NAS stands at 6.3 hours. LGA manifests the longest average duration of 9.2 hours, while SEA exhibits a relatively shorter duration of 3.2 hours among the top issuing airports. A single GDP can undergo multiple revisions during its operational period, featuring alterations in the PAR, scope, or planned end time. It can also be canceled prior to the initial end time due to the volatility of weather conditions and flight deliveries. Approximately 40% of GDPs undergo more than one revision during their lifespan, and, on average, a GDP is canceled 2.5 hours before its initial planned end time. Figure 47(c) shows that most GDPs are caused by adverse weather conditions including thunderstorms, low ceiling, wind, and snow/ ice, while other reasons such as airport outage, runway problems and special events can also lead to GDPs. For EWR and LGA, over 70% GDPs are triggered by convective weather – winds and thunderstorms. For SFO and BOS, low ceiling due to marine stratus is one of the leading causes for GDPs.

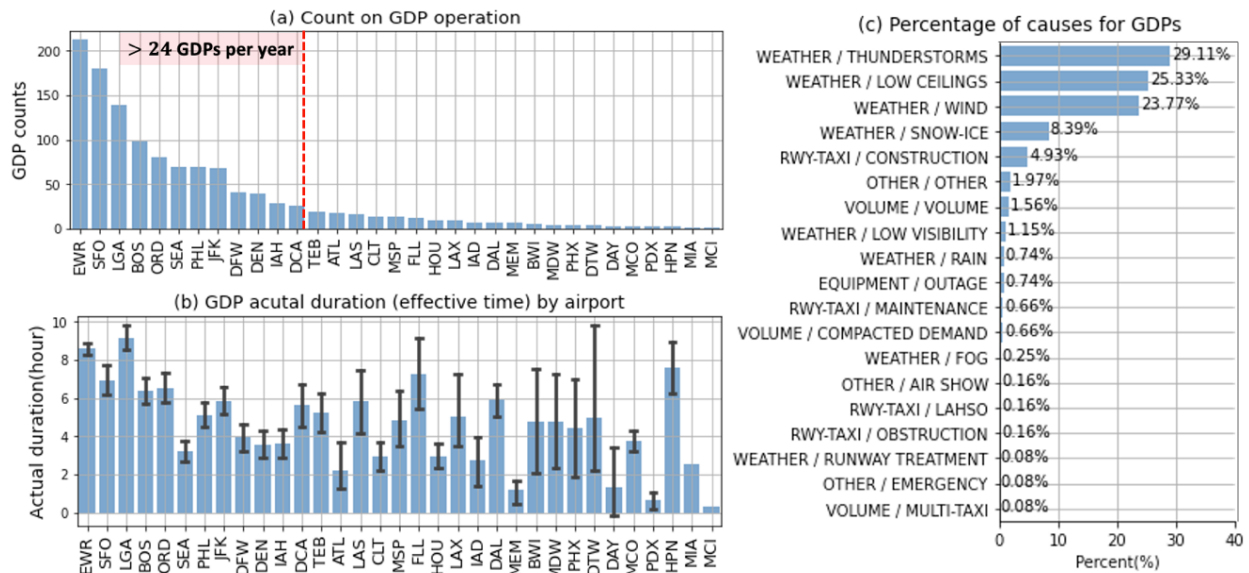


Figure 47. GDP frequency, duration and causes by airport (2019)

As explained in section 3.2, we categorize arrivals with their original scheduled arrival time within the GDP timeline into four distinctive categories: in-effect, post-cancellation, exempt and cancelled flights. For each flight, the corresponding GDP ground delay was also calculated. Figure 48(a) illustrates the mean counts of in-scope, exempt, and cancel-delay flights for each GDP, by arrival airport. On average, we observe 168.7 in-effect flights, 22.2 post-cancellation flights, 68.3 exempt flights and 19.5 cancelled flight across all GDPs. For the corresponding GDP delay shown in Figure 48(b), the vertical bars signify the aggregate GDP delays originating from their restricted flights, accumulated for each airport over the year, while the overlaid red dots represent the average GDP delays, normalized by the total number of GDPs. The average total delay associated with a GDP is 143.4 hours, with a standard deviation of 135.2 hours. GDP delay imposed per restricted flight averaged 34.4 minutes, with a standard deviation of 18.6 minutes per restricted flight (min/RF). ORD GDPs featured the highest average imposed GDP delay of 395 hours, while EWR GDPs accounted highest total GDP delays of 40.5k hours.

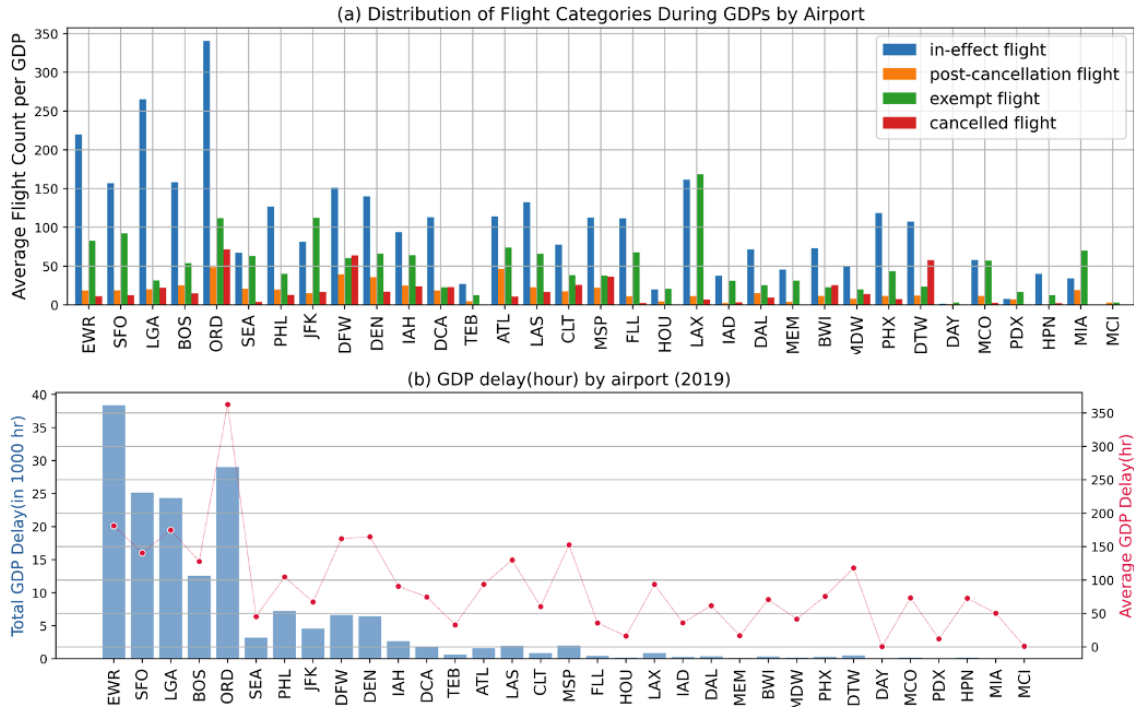


Figure 48. GDP affected flights and assigned ground delay by airports (2019)

4.2 GDP excess delay and wasted slots

This section presents the results of our GDP excess delay and wasted slot analysis. As explained in Section 3.3.2, these calculations involve assumptions about airport capacity and four different approaches to determining capacity are used in our analysis. To illustrate the role of the capacity assumptions in our analysis, we first apply them to a specific GDP and then broaden the scope to encompass all GDPs in our dataset.

The example GDP controlled the arrivals into Chicago O'Hare International Airport (ORD) on August 26th, 2019, due to demand-capacity imbalances caused by the low ceiling condition. It was modeled and initially released at 5:50 with a designated control window scheduled from 7:00 to 21:59. However, it underwent revisions to its program rate at 11:00 and 16:39 and was eventually cancelled at 18:53. Thus, the actual duration of this GDP was 11.9 hours. During this operational window, the GDP encompassed 857 in-scope flights, 113 cancel-delay flights, and 129 exempt flights. The total imposed delay was 735 hours, averaging 45.5 minutes per affected flight.. Based on our queueing model, we estimate the excess delay from the GDP under the four capacity calculation methods. Figure 49 depicts the results with the x-axis denoting quarter-hourly intervals and the y-axis showing flight counts. Each bar represents flight counts for a given quarter-hour, and the horizontal line signifies the called rate. Figure 49(a) illustrates the observed throughput and AAR set by ATC. Figure 49(b) represents the flight demand based on model planned arrival time with GDP and its corresponding delays removed. Arrival demand is served at rates based on capacity cases 1-4 during the GDP operation. Throughputs under the queueing model are displayed in Figure 49(c)-(f). Cases 1 and 2 are similar, except that the latter has slightly more throughput between periods 50 and 60 when observed throughput exceeded the called rate. The differences

between Cases 1 and 2 on the one hand and Cases 3 and 4 on the other are more pronounced, since called rates substantially exceeded planned rates, beginning around period 60, allowing for a commensurate increase in throughput in the former cases.

Excess delay is quantified by comparing the arrivals between observed and counterfactual scenarios. The calculated excess delay per GDP-restricted flight for Cases 1 through 4 was 26.4, 31.1, 10.8 and 16.4 minutes, respectively. The lower excess delays for Cases 3 and 4 reflect that called rates exceed planned rates during the latter part of the GDP.

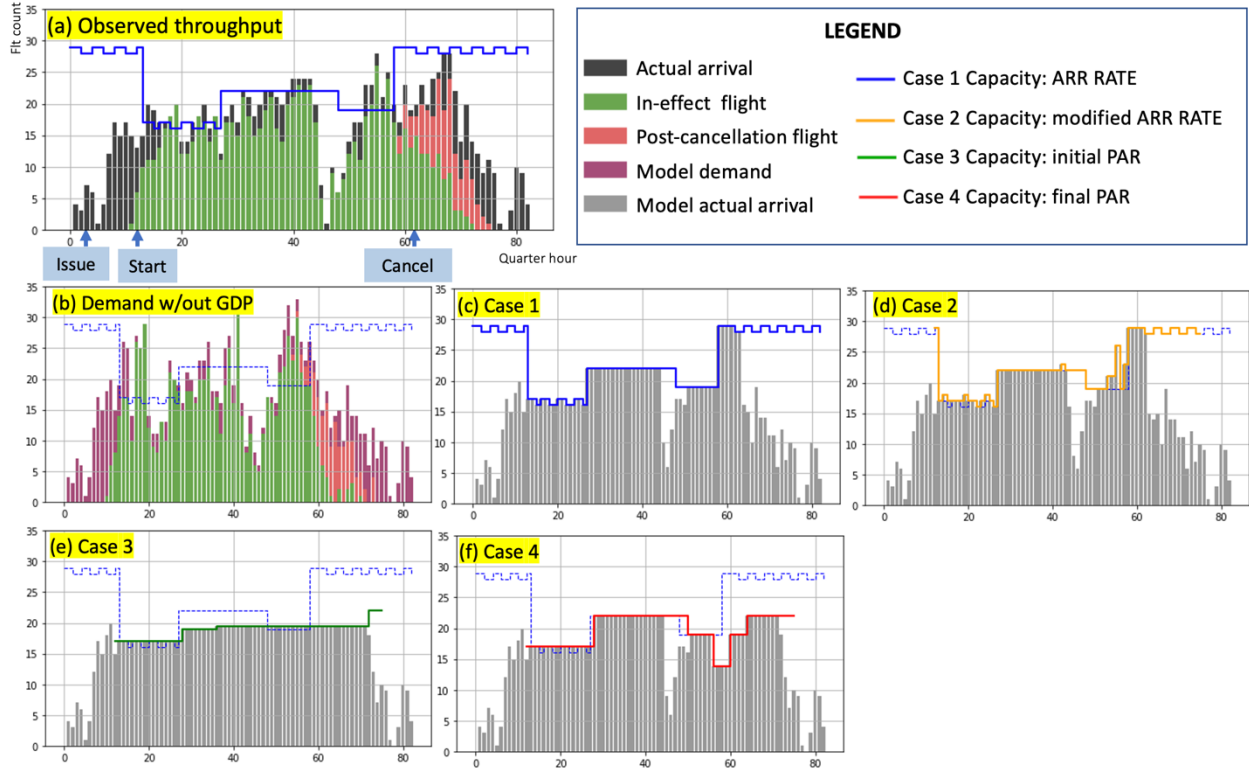


Figure 49 Scenario-based counterfactual queueing model: Example GDP at ORD on Aug 26th, 2019

Expanding to all 1,204 GDPs, deploying the queuing model and four different methods for calculating capacity. The results are summarized in Table 24. The excess delay for each GDP was calculated as the average delay incurred by its corresponding GDP-restricted flights (RF). Additionally, metrics for wasted slots and averted holding flights were averaged across the *effective quarter-hours within the scope of GDPs*. Two variables, β and α , are introduced to quantify, respectively, the proportion of aggregate excess delay to the total GDP delay, and the proportion of the total number of wasted slots to averted holding flights, for all GDPs under consideration. These variables are formally defined in Equations (16) and (17), where N denotes the total number of GDPs,

$$\beta = \frac{\sum_{i=1}^N \text{excess delay}_i}{\sum_{i=1}^N \text{GDP delay}_i} \quad (16)$$

$$\alpha = \frac{\sum_{i=1}^N WS_i}{\sum_{i=1}^N AHF_i} \quad (17)$$

Table 24. Summary results on case-based counterfactual queueing model

Case	Capacity ($C_{(t)}$, by quarter hour)	Excess Delay (min per RF)		Delay Ratio β (%)	Wasted slots (/15 min)	Averted Holding flights (/15 min)	Waste -to- Holding Ratio α (%)
		mean	std.				
1	$AAR_{(t)}$	24.31	36.37	63.05	0.62	6.40	9.69
2	$\max (AAR_{(t)}, MA_A_{(t)})$	29.03	31.07	69.62	0.64	5.61	11.41
3	$GDP\ PAR_{0(t)}$	17.40	39.48	50.62	0.57	8.26	6.90
4	$GDP\ PAR_{f(t)}$	16.13	38.14	48.04	0.54	8.23	6.56

In Case 1, where the input airport capacity is the rate estimated by ATC, we observe that the "efficiency" of a Ground Delay Program (GDP) varies depending on the metric used. On the one hand, over 60% of the delay caused by GDPs is "unnecessary," suggesting that GDPs are inefficient. This translates into an average excess delay of 24.31 minutes per *RF*. On the other hand, eliminating GDPs would allow an increase in slot utilization of just 0.62 per quarter hour while requiring 6.4 aircraft to be holding in the air. This leads to greater airspace congestion, higher energy consumption, increased safety risks, and additional workload for air traffic controllers, thereby casting GDP performance in a more favorable light.

As there exists mismatch between *AAR* and observed throughput (refer to section 3.3.2), arrival capacity is adjusted in Case 2 to account for periods when observed throughput exceeds the called rate. As expected, this more optimistic capacity assumption results in less delay in the absence of the GDP and hence a higher estimate—by about 20%—of excess delay caused by the GDP. Wasted slots are commensurately more and averted holding flights without the GDP commensurately less in this scenario.

For Cases 3 and 4, where we assume the airport called rate is aligned with GDP planned rates – initial released PAR and last-updated PAR, we observe a decrease in the excess delay ratio to 50.62% and 48.04%, with 0.57 and 0.54 wasted slots per quarter hour respectively. The reduced estimates for excess delay and wasted slots in these cases reflect the fact that rates used in planning the GDP tend to be somewhat lower than realized rates. Thus, if there were no GDP but flights could only land at the GDP planned rates, there would more airborne delay and fewer unused slots than under Cases 1 and 2, result in a 30-40% reduction in GDP excess delay and 12-15% fewer wasted slots. While these differences are sizable, the results also show that even a perfect match between planned and realized capacity does not result in a "perfect" GDP; deviations between planned and actual flight times also impair GDP performance.

In assessing overall GDP performance based on post-operation data, we consider Case 2, which derives the actual capacity rate from ATC estimates but relaxes the upper bound using the 3-quarter-hour moving average throughput to account for periods sustained over-delivery, as the most plausible capacity assumption for the queueing model. We therefore focus on Case 2 results for further analysis in subsequent sections. Figure 50 and Figure 51 present an in-depth depiction of GDP excess delay, wasted slots, and averted holding flights, analyzed by airport. In Figure 50(a), we use side-by-side bars to represent the total excess delay resulting from GDP implementation at various airports and the total executed GDP delay. These are arranged in descending order based on the frequency of GDP issuance. EWR and ORD stand out with the highest excess delays, amounting to 26.3k hours (out of a 38.4k hour GDP delay) and 23.4k hours (out of a 29.1k hour GDP delay), respectively. Figure 50(b) presents the average excess delay per restricted flight for

each airport, illustrated as bars, with the proportion of excess delay over GDP delay (β) represented as red dots. While the average β across all GDPs is 69.62%, it's noteworthy that SEA and SFO exhibit lower excess delay ratios, while JFK, DFW, IAH, and ORD have relatively higher excess delay ratios. This pattern of delay ratio β is consistent with the trend observed in the average excess delay per restricted flight.

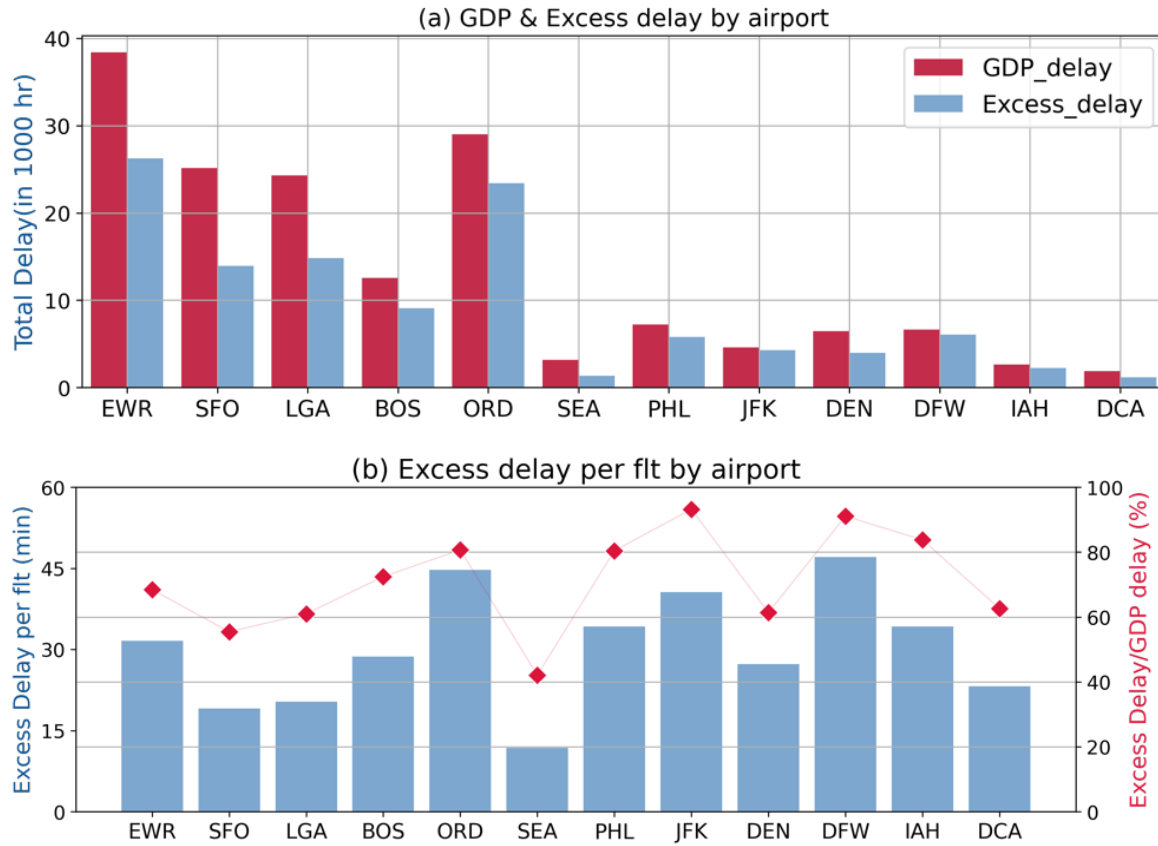


Figure 50. Excess delay by airport

Figure 51(a) presents the average wasted slots per quarter hour during GDPs by airport, depicted as bars, with the proportion of wasted slots over available capacity represented as red dots. Newark Liberty (EWR), San Francisco (SFO), LaGuardia (LGA), Boston Logan (BOS) and Seattle-Tacoma (SEA) airport all show, on average, less than one wasted slot per quarter hour, translating to less than 10% of their capacity. In contrast, O'Hare (ORD), Dallas/Fort Worth (DFW) and George Bush (IAH) airports have a relatively large number of wasted slots and a high wasted-slot-capacity ratio, mirroring the excess delay ratio measurement.

Figure 51(b) presents a scatter plot analysis of β , the ratio of aggregated excess delay per flight to total GDP delay, against α , the ratio of total GDP wasted slots to averted holding flights, for each GDP. β is plotted on the y-axis and α , on a logarithmic scale due to its wide range, is plotted on the x-axis. As expected, there is positive correlation between slot inefficiency (α) and delay inefficiency (β): wasted slots leads to more unnecessary delay. The scatter plot reveals an asymptotic sigmoidal pattern of performance for individual GDPs, as depicted by the blue dots. Blue dots in the lower left corner suggest GDPs with better planning and execution, where both excess delay and wasted slots are minimized. The red crosses, representing airport averages, show

considerable inter-airport variation in excess delay and wasted slots. Seattle (SEA) and San Francisco (SFO) show the least inefficiency, which excess delay representing around half of total GDP delay, and less than one wasted slot for every 10 flights spared airborne holding by the GDP. In contrast, New York Kennedy (JFK) and Dallas-Fort Worth (DFW) airports have the least efficient GDPs, with 90% of GDP delay unnecessary and roughly a 1:1 ratio of wasted slots to averted airborne holding.

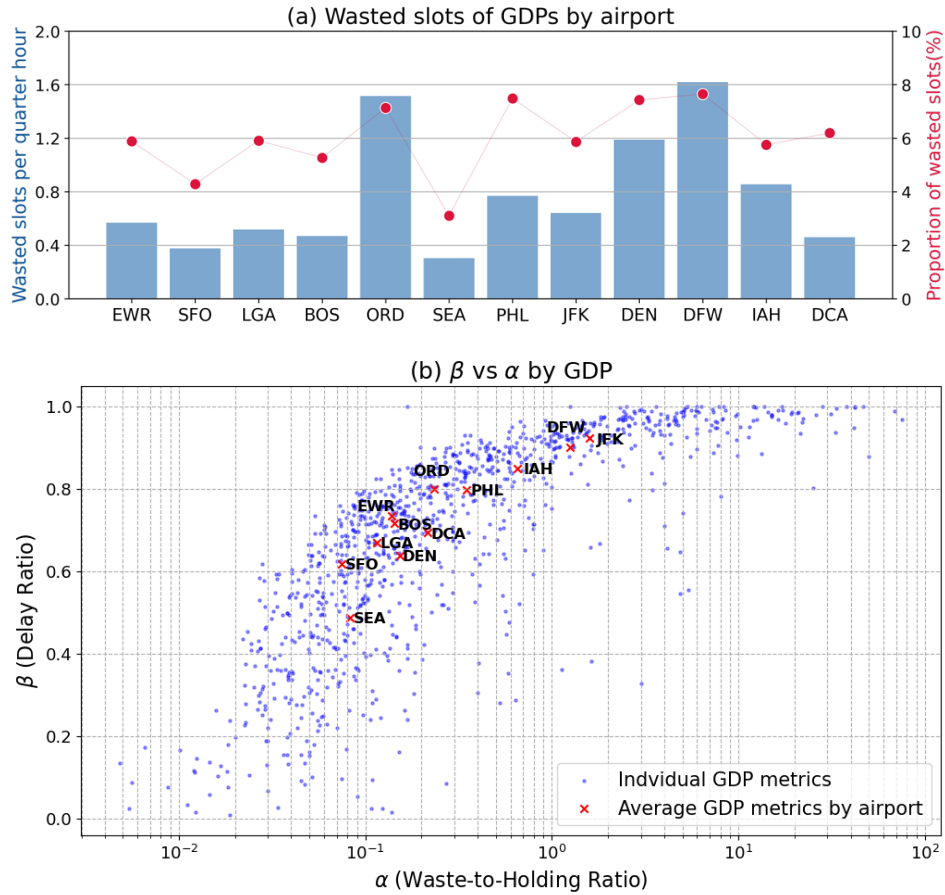


Figure 51. Wasted slots and averted holding flights by airport

4.3 Causal analysis on excess delay: statistical models

We now investigate the factors that drive variation in performance across the GDPs. For the outcome variable, we use excess delay per restricted flight (in min/ RF) measured from Case 2 for each GDP. The explanatory variables follow the model specifications in Section 3.4. To determine hyperparameters for the Lasso and Ridge regressions we use cross-validation. Subsequently, we trained the three candidate models (Lasso, Ridge, and OLS) using the entire training set. A comparative evaluation of the models' performances, in terms of Mean Absolute Error (MAE) and Root Mean Square Error ($RMSE$), is outlined in Table 25. The MAE scores for the models range between 8.6 and 9.3 minutes per flight, while the $RMSE$ s are also close--13.1 to 14.1 minutes per flight. These narrow ranges suggest that the performances of the three models are not substantially different. However, the Ridge regression exhibits a slight edge over its counterparts. This superior performance of the Ridge regression can be attributed to its regularization process, which serves to prevent overfitting by reducing model variance without shrinking coefficients to exactly zero as

Lasso does. This capability ensures a balance between model simplicity and precision, preventing overcomplication while retaining significant variables in the model.

Table 25. Model Performance on Test Set (in min/RF)

	OLS	Lasso	Ridge
RMSE	13.762	14.030	13.183
MAE	9.182	9.370	8.643

In the regression models employed, we standardized all numerical predictors (original feature matrix $\mathbf{X}$) to facilitate uniform interpretation across different scales. The standardization formula used is:

$$\mathbf{X}_{std} = \frac{\mathbf{X} - \boldsymbol{\mu}}{\boldsymbol{\sigma}} \quad (18)$$

where $\boldsymbol{\mu}$ is the mean vector and $\boldsymbol{\sigma}$ is the standard deviation vector of predictors. Dummy variables were left in their original state, assuming a default mean of 0 and standard deviation of 1.

This standardization process allowed us to interpret the standardized coefficients as the change in the response variable per one standard deviation increase in a numerical predictor, while holding other predictors constant. For enhanced interpretability, we rescaled these coefficients back to the original scale of the predictors. For Ridge regression, given the coefficient $\boldsymbol{\beta}_{rdg}$ and the rescaled coefficient $\boldsymbol{\beta}_{scale}$, the model equation is:

$$\mathbf{Y} = \boldsymbol{\beta}_{0,rdg} + \mathbf{X}_{std}^T \boldsymbol{\beta}_{x,rdg} + \boldsymbol{\varepsilon} = \left(\boldsymbol{\beta}_{0,rdg} - \frac{\boldsymbol{\mu}^T \boldsymbol{\beta}_{x,rdg}}{\boldsymbol{\sigma}} \right) + \frac{\mathbf{X}^T \boldsymbol{\beta}_{x,rdg}}{\boldsymbol{\sigma}} + \boldsymbol{\varepsilon} \quad (19)$$

where $\boldsymbol{\varepsilon}$ is the normally distributed error vector; $\boldsymbol{\beta}_{0,rdg}$ and $\boldsymbol{\beta}_{x,rdg}$ are the constant and slope coefficients of $\boldsymbol{\beta}_{rdg}$, respectively. Therefore, the rescaled $\boldsymbol{\beta}_{scale}$ can be calculated as follows: for the rescaled constant, use $\left(\boldsymbol{\beta}_{0,rdg} - \frac{\boldsymbol{\mu}^T \boldsymbol{\beta}_{x,rdg}}{\boldsymbol{\sigma}} \right)$, and for the slope coefficients, use $\left(\frac{\mathbf{X}^T \boldsymbol{\beta}_{x,rdg}}{\boldsymbol{\sigma}} \right)$.

Table 26 presents the results of OLS, Lasso, and Ridge regression applied to the entire dataset, with Ridge being identified as the optimal model. Subsequent analyses and discussions are based on the results from the Ridge regression. First for the GDP planning aspects, **ET** coefficient reveal that for every additional hour of effective duration, the excess delay per RF increases by 1.56 min on average, assuming all other variables remain constant. This could be explained by the fact that longer GDP time window requires a longer planning horizon, which could introduce greater uncertainties and thus possibly higher excess delays. Conversely, **GT** coefficient shows that each additional hour of gap time is associated with a decrease in excess delay by 2.91 min/RF. A likely explanation for this result is that the longer gap time allows more flights to be restricted by the GDP, thus reducing excess delay per flight. It may also be that GDPs have longer gap times when more reliable information is available. However, it's crucial to acknowledge a trade-off between the quantity of controlled flights and the reliability of information over time. On the other hand, planning GDPs further ahead of time is unlikely to increase efficiency, since early GDP planning will result in greater uncertainty in weather and traffic conditions. The **CT** coefficient demonstrates that an additional hour in the cancellation window correlates with an increase in excess delay by

0.37 minutes per RF. This outcome might be explained by the fact that the cancellation of a GDP, indicating an improvement in weather or traffic conditions, can lead to unrecoverable excess delays as the earlier GDP design becomes unnecessarily stringent. For the variable n_r , it shows that, on average, every additional revision of a GDP during its execution is associated with an increase of 3.28 min/RF in excess delay. This outcome aligns with expectations as when a GDP is revised, there are unrecoverable delays to flights imposed by the previous GDP setting. The estimate for the sc_{us} variable implies that excess delay increases with the average duration of restricted flights, which must depart earlier and thus are subject to more unrecoverable delay when there are GDP revisions. Lastly, the variable n_{CF} exhibits a positive coefficient, indicating that a larger number of flights canceled by airlines during a GDP results in a larger difference between realized flight demand and demand assumed when planning the GDP. This ultimately leads to an increase in excess delays.

The next group of variables of interest are related to AAR and GDP PAR, which account for capacity uncertainties and offer insights into the degree of conservatism or aggression in GDP design. The variables $uPAR_0$ and $uPAR_f$ both carry positive coefficients of 1.52 and 1.79, indicating that if the GDP initial or last-updated PAR is smaller than AAR by one slot per quarter hour during, the excess delay increases by approximately 1.52 or 1.79 minutes per RF, respectively. This result is not surprising since deficits between GDP planned, and actual capacities will necessarily result in excess ground holding and delay. The ΔAAR coefficient shows that for every unit decrease in the quarter-hour rate (meaning the AAR is one unit lower than its annual average), the excess delay per RF increases by approximately 1.47 minutes. It intuitively suggests that the severity of the capacity shortfall plays a significant role in exacerbating excess delays and GDP inefficiency.

Turning now to the demand uncertainties related to flight time deviations, our estimation results indicate that the mean time differences between EDCT/Flight Plan time and the actual flight time among in-effect flights or exempt flights during gate out ($\Delta GO_IEF/EF$), taxi out ($\Delta TO_IEF/EF$), and enroute ($\Delta ETE_IEF/EF$) have considerable impacts on excess delay. The negative coefficients for ΔGO_IEF and ΔETE_IEF (-0.284 and -0.425 respectively) indicate that GDPs in which gate-out times and en-route times of non-exempt flights are late relative to plan have less excess delay. On the other hand, positive coefficients for ΔTO_IEF , ΔGO_EF , ΔTO_EF and ΔETE_EF (0.301, 0.160, 0.358 and 0.099 respectively) suggest that GDPs in which non-exempt flights have longer taxi-out delays, or in which exempt flights are delayed at the gate, during taxi-out, or en route have greater excess delay. These complex impacts could be due to the potential balancing out of under- and over-delivery of flights during the same arrival window on a macro level. When we look at the dispersion of these time deviations from the EDCT/Flight Plan (σ_GO_IEF/EF , $\sigma TO_IEF/EF$, $\sigma ETE_IEF/EF$), all variables demonstrate a significant positive impact on excess delay, with σETE_IEF being the most influential (rescaled coefficient of 0.396). This suggests that larger variability in actual flight times relative to their planned times, especially during enroute for in-effect flights, leads to more excess delay, underlying the importance of reliable flight time predictions. Overall, these results highlight the importance of minimizing deviations from planned times, especially during the enroute phase for in-effect flights, in order to reduce excess delay and improve GDP delivery.

Lastly, we direct our focus to the model coefficients of fixed-effect variables, namely, GDP cause, GDP-issued airport and the constant term. These variables were incorporated into the model to encapsulate the distinct characteristics associated with airport operations—especially those

pertinent to arrival management—and to capture the overarching patterns of GDP operational conditions, respectively. For the GDP cause variables, thunderstorms were set as the reference category. The model coefficients corresponding to other causes—low ceiling, wind, runway maintenance, and others—all exhibit considerable negative values (-2.440, -0.528, -8.701, and -4.234, respectively), while the ice-snow cause yields a positive coefficient of 2.382. This pattern implies that GDPs prompted by thunderstorms and icy or snowy conditions tend to engender larger excess delay. This result is in line with the recognized understanding that convective weather phenomena, inclusive of thunderstorms and snow, often foster increased regional unpredictability. This unpredictability can impinge on both airport and enroute sector capacities, potentially leading to a less efficient implementation of GDPs.

In terms of the airport dummy variables, EWR was denoted as the reference category. The model revealed that LGA and SFO are associated with substantially lower excess delay (coefficients of -12.573 and -4.132 min/RF, respectively). Conversely, ORD and PHL are associated with higher excess delay (coefficients of 3.773 and 5.038 min/RF, respectively). These findings are consistent with the counterfactual queueing model depicted in Figure 50(b), underscoring the variability in performance among different airports under GDP conditions. These differences may reflect variations in operational efficiency, arrival management strategies, or the magnitude of weather-related disruptions at these airports.

For the constant term, the original Ridge regression intercept of 30.919 indicates the predicted excess delay per RF is 30.919 minutes when all standardized predictors (X_{std}) are set to zero, corresponding to the original predictor values (X) being at their mean. This aligns with the measurements of true values in Case 2 presented in Table 22. The rescaled intercept of -15.328 minutes, on the other hand, represents the predicted excess delay per RF when the original predictor values (X) are at zero.

Table 26. Model estimation results.

Variable	Ridge		Lasso	OLS				
	Coef.	rescaled Coef.	Coef.	Coef.	std err	P> t	[0.025	0.975]
<i>const</i>	30.919	-15.328	33.042	32.497	3.013	0.000 ***	26.585	38.410
<i>ET</i>	6.041	1.562	6.922	7.773	2.251	0.001 ***	3.356	12.189
<i>GT</i>	-3.391	-2.909	-3.338	-3.459	1.165	0.003 ***	-5.745	-1.172
<i>CT</i>	0.824	0.375	0.888	1.129	0.982	0.251	-0.798	3.056
<i>n_r</i>	3.153	3.282	3.110	3.058	1.211	0.012 **	0.682	5.434
<i>sc_us</i>	3.592	8.349	3.208	3.705	1.579	0.019 **	0.607	6.802
<i>sc_ca</i>	-0.694	-0.861	-0.670	-0.783	1.050	0.456	-2.843	1.278
<i>n_IEF</i>	-2.019	-0.016	-1.971	-2.593	1.862	0.164	-6.245	1.059
<i>n_EF</i>	-0.548	-0.024	-0.391	-0.517	1.002	0.606	-2.483	1.449
<i>n_PCF</i>	-0.770	-0.018	-1.091	-1.372	1.154	0.235	-3.635	0.892
<i>n_CF</i>	3.527	0.083	3.668	3.765	1.259	0.003 ***	1.294	6.235
<i>ph</i>	-1.329	-0.168	-1.297	-1.298	0.951	0.172	-3.163	0.567
<i>al_hhi</i>	1.201	0.001	1.019	1.051	0.926	0.257	-0.767	2.868
<i>AAAR</i>	4.711	1.474	4.869	5.639	1.327	0.000 ***	3.036	8.242
<i>uPAR₀</i>	4.248	1.520	0.000	3.877	0.584	0.000 ***	2.731	5.023
<i>σPAR₀</i>	-0.221	-0.134	0.000	-0.327	2.603	0.900	-5.433	4.780
<i>uPAR_f</i>	4.868	1.793	8.921	4.504	0.617	0.000 ***	3.295	5.714
<i>σPAR_f</i>	-0.756	-0.485	-0.902	-0.814	2.512	0.746	-5.742	4.114
<i>uPAR_r</i>	-1.598	-1.907	-0.322	-1.657	0.944	0.080 *	-3.510	0.196
<i>σPAR_r</i>	2.173	2.246	2.065	2.005	1.250	0.109	-0.447	4.458
<i>ΔGO_IEF</i>	-2.927	-0.284	-2.921	-3.149	0.905	0.001 ***	-4.924	-1.374

σGO_IEF	2.547	0.103	2.358	2.065	1.042	0.048 **	0.019	4.110
ΔTO_IEF	1.368	0.301	1.488	2.278	1.677	0.174	-1.011	5.568
σTO_IEF	0.034	0.006	0.000	-0.694	1.708	0.685	-4.045	2.657
ΔETE_IEF	-2.324	-0.425	-2.174	-2.527	1.123	0.025 **	-4.731	-0.323
σETE_IEF	3.290	0.396	3.202	3.399	1.283	0.008 ***	0.882	5.916
ΔGO_EF	1.178	0.160	1.830	1.970	1.208	0.103	-0.400	4.341
σGO_EF	1.746	0.283	1.362	1.186	1.023	0.246	-0.821	3.193
ΔTO_EF	0.834	0.358	0.533	1.050	1.289	0.416	-1.480	3.579
σTO_EF	0.813	0.254	0.537	0.831	1.214	0.494	-1.552	3.213
ΔETE_EF	1.182	0.099	1.590	1.921	1.291	0.137	-0.612	4.454
σETE_EF	1.637	0.152	1.218	1.155	1.122	0.304	-1.046	3.355
c_SNOW	2.382	2.382	1.454	1.041	3.347	0.756	-5.525	7.607
c_LC	-2.440	-2.440	-3.301	-4.344	2.749	0.114	-9.737	1.049
c_WIND	-0.528	-0.528	0.000	-1.037	-2.527	0.681	-3.920	5.995
c_RWY	-8.701	-8.701	-11.343	-13.837	3.833	0.000 ***	-21.358	-6.317
c_others	-4.234	-4.234	-5.343	-6.868	3.200	0.032 **	-13.146	-0.591
apt_BOS	1.371	1.371	0.000	1.654	3.637	0.649	-5.483	8.791
apt_JFK	1.295	1.295	0.525	2.753	4.132	0.505	-5.354	10.859
apt_LGA	-12.573	-12.573	-17.915	-18.668	4.270	0.000 ***	-27.047	-10.290
apt_ORD	3.773	3.773	2.934	7.959	5.448	0.144	-2.731	18.648
apt_PHL	5.038	5.038	4.371	6.761	4.238	0.111	-1.555	15.076
apt_SEA	-1.692	-1.692	-2.434	-2.329	4.277	0.586	-10.720	6.062
apt_SFO	-4.132	-4.132	-5.495	-5.267	3.700	0.155	-12.526	1.993
apt_others	3.461	3.461	1.970	4.730	4.184	0.259	-3.479	12.940
$R-sqr$	0.592		0.525	0.562				

(* p < 0.10, ** p<0.05, *** p<0.01).

Table 27. Descriptive statistics for numerical variables (N=1,204)

Variable	Mean	Std.	Min	Median	Max
ET (hr)	6.40	3.87	0.02	8.95	17.97
GT (hr)	1.47	1.17	0.00	2.33	5.82
CT (hr)	2.54	2.20	0.00	3.33	13.40
n_r	0.62	0.96	0.00	1.00	9.00
sc_us (hr)	1.63	0.43	0.00	1.95	3.00
sc_ca (hr)	1.28	0.81	0.00	1.56	3.81
n_IEF (flt)	168.69	129.82	0.00	247.00	904.00
n_EF (flt)	22.24	22.41	0.00	33.00	126.00
n_PCF (flt)	68.27	43.03	0.00	91.00	335.00
n_CF (flt)	19.55	42.47	0.00	18.00	479.00
ph (hr)	5.51	7.89	0.00	6.81	66.90
al_hhi	1857.74	909.06	645.56	1975.16	10000.00
ΔAAR (slot)	-1.81	3.20	-20.52	0.17	4.25
$uPAR_0$ (slot)	1.25	2.80	-7.82	1.74	17.31
σPAR_0 (slot)	1.42	1.65	0.00	1.74	11.46
$uPAR_f$ (slot)	1.45	2.71	-7.00	1.95	17.31
σPAR_f (slot)	1.37	1.56	0.00	1.66	11.46
$uPAR_r$ (slot)	-0.20	0.84	-8.64	0.00	5.15
σPAR_r (slot)	0.48	0.97	0.00	0.60	7.58
ΔGO_IEF (min/flt)	-11.69	10.32	-81.50	-7.41	38.87
σGO_IEF (min/flt)	35.56	24.71	0.00	41.48	302.79
ΔTO_IEF (min/flt)	13.93	4.54	-3.50	15.98	51.65
σTO_IEF (min/flt)	15.67	5.53	0.00	18.44	44.62
ΔETE_IEF (min/flt)	3.97	5.47	-17.37	6.22	47.33
σETE_IEF (min/flt)	13.86	8.30	0.00	15.49	63.92
ΔGO_EF (min/flt)	1.02	7.34	-22.00	4.28	54.65
σGO_EF (min/flt)	18.62	6.18	0.00	22.22	58.55
ΔTO_EF (min/flt)	6.39	2.33	-1.00	7.45	25.58
σTO_EF (min/flt)	8.30	3.20	0.00	9.40	38.82
ΔETE_EF (min/flt)	1.20	11.88	-20.80	5.26	89.83
σETE_EF (min/flt)	18.70	10.74	0.00	21.77	76.39

4.3 Causal analysis on excess delay: factor contribution

In this section, we construct counterfactual scenarios to assess the potential changes in the excess delay of GDPs upon the removal of specific effect categories. This is achieved by adjusting the feature values before standardization of the corresponding features to zero. By processing this modified feature space, we were able to quantify the resulting changes in excess delay compared to the baseline scenario. This process involved utilizing the previously trained ridge regression model, in conjunction with Eq. (15). We structured three counterfactual scenarios, focusing on the accuracy of the program rate, the deviation of in-effect flight time, and the deviation of exempt flight time. The relative contributions of these factor effects on the reduction of excess delay are depicted in Figure 52. On the y-axis, the groups of factors altered for model performance testing are displayed, whereas the x-axis denotes the percentage change in excess delay, thereby reflecting each factor's contribution.

Within the GDP planning stage, we primarily concentrated on program rate accuracy. We chose the set of features which capture the discrepancy between the GDP program rate and the actual airport capacity and set their values such that there was no slot difference across all observations. In a scenario where the final updated program rate was precisely accurate (i.e., $uPAR_f$ and σPAR_f equal zero), we observed that the excess delay would decrease by an average of 6.1%. Moreover, if the initial program rate was also in perfect alignment with the actual capacity and remained unaltered during the execution (i.e., $uPAR_0$, σPAR_0 , $uPAR_r$, and σPAR_r equal zero), it could result in a total reduction in excess delay by 13%. While these counterfactual scenarios show that there is some value in improving the ability to accurately set arrival rates, such improvement alone will only have modest impact on excess GDP delays.

During the GDP execution stage, we evaluate two groups of effects related to the deviations in flight times from the gate out (*GO*), taxi out (*TO*), and enroute (*ETE*) time for both in-effect flights (*IEF*) and exempt flights (*EF*). For each flight stage, we establish counterfactual scenarios by initially setting the bias of flight time uncertainty to zero (mean time difference: $\Delta GO/TO/ETE_IEF/EF = 0$), followed by setting its variance to zero (standard deviation of time difference: $\sigma GO/TO/ETE_IEF/EF = 0$), and ultimately, having both bias and variance set to zero to simulate a scenario of perfect time adherence. Our findings indicate that the conformity of in-effect flights (*IEF*) to EDCT flight times contributed significantly to the reduction of excess delay. Furthermore, the variability in flight time compliance is a more potent driver of delay reduction than the bias in flight time. When both bias and variance of flight time deviation are zeroed out, we observe a reduction in excess delay by 23.9% for gate out time, 14.7% for taxi out time, and 13.0% for enroute time for *IEF*. Overall, precise in-effect flight times account for a 51.6% reduction in excess delay. For exempt flights (*EF*), the contributions of flight time deviations are slightly smaller, but nonetheless accounted for a 29.2% reduction in total, with gate-out delay the most important component. Most strikingly, if both in-scope and exempt flights had their actual flight times perfectly align with the plan, the excess delay could be reduced by as much as 80.9%. These findings underscore the vital role of accurate flight time estimation and adherence to planned times in reducing GDP inefficiencies. In short, precise execution of GDPs has far greater potential to reduce excess delays from GDPs than does more accurate rate setting.

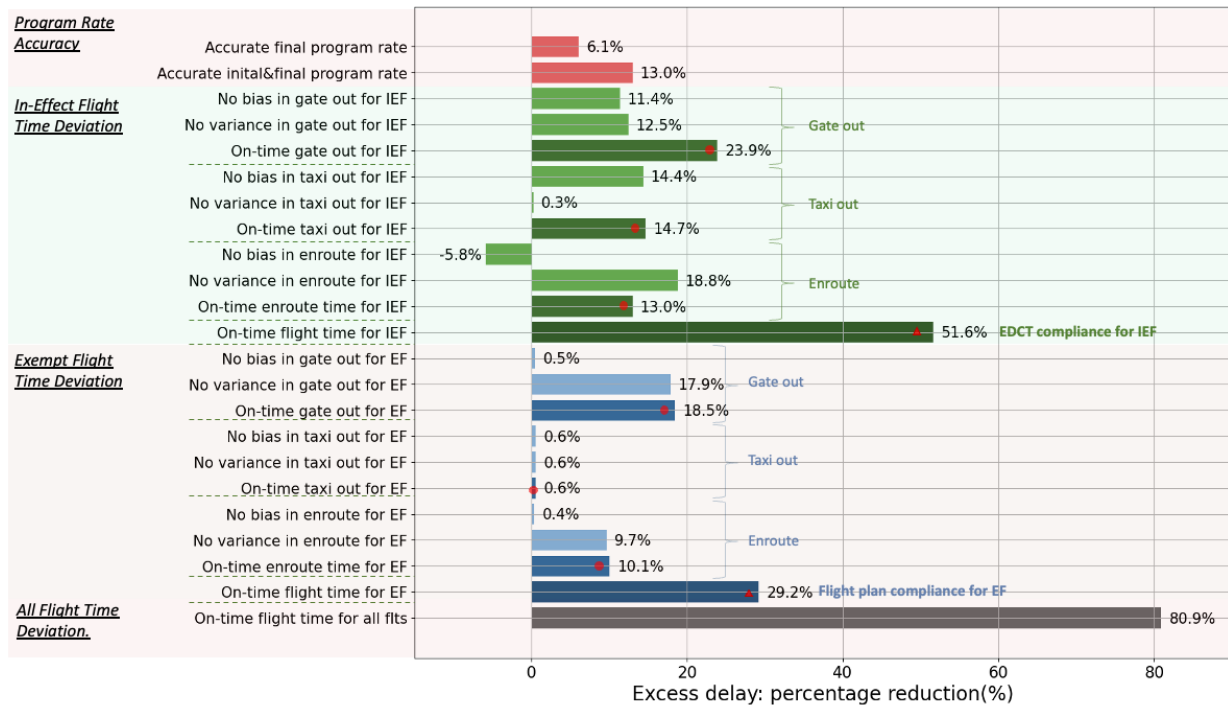


Figure 52. Relative factor contribution in reduction of average GDP excess delay

5 Conclusion

In this study, we proposed and implemented a method to assess efficiency of Ground Delay Programs (GDPs). Our method focuses on two closely related forms of inefficiency—excess delay and wastes slots. The performance metrics were assessed under the counterfactual scenario with the GDPs removed from the system by utilizing a deterministic queuing model to simulate the arrival performance under varying capacity settings. The generated excess delay was further explored through statistical models to identify influential factors related to GDP planning and execution stages and their factor contribution. Our methodology was applied to 1204 viable GDPs implemented by 33 airports in the NAS during 2019. Under the "Case 2" arrival capacity scenario, which adjusts for over-delivery due to aircraft mix or miscalculated rate estimates, the removal of GDPs culminated in an average of 29.03-min excess delay per restricted flight with a standard deviation of 31.07 min/RF. The excess delay ratio indicates that 70% of GDP delay would be eliminated, as opposed to being converted to airborne delay, if the GDPs were removed. Removing a GDP would allow 0.64 flights per quarter hour to land but also result in 5.6 additional holding flights on average. For the statistical model of observations of excess delay of individual GDPs, Ridge regression yields the best model as compared to OLS and LASSO regression. The model insights highlight a delicate balance between control power, which is dictated by factors such as scope and gap time, and the potential for increased unnecessary delay. The program rate setting has a significant impact on GDP excess delay, conservative program rates leading to excessive ground delays. However, uncertainties in gate out and taxi out time for in-scope flights have a much stronger impact on excess delay. A counterfactual analysis based on the Ridge model shows that uncertainties in program rate setting contribute to 13% of the excess delay, while flight time deviations account for 80.9% of excess delay.

In this study, we have enriched the understanding of GDP efficiency at the level of the National Airspace System (NAS), with an emphasis on arrival delay and airport capacity utilization. It can improve our understanding of the relationship between GDP excess delay and operational conditions, particularly those related to GDP planning and execution stages. Furthermore, our research aids in identifying areas of priority for improvements, such as more accurate flight time estimation and program rate planning. Importantly, our findings underline the importance of carefully balancing the benefits and costs associated with different GDP operational strategies. While the benefits of mitigating excess delay and fully utilizing airport capacity are evident, these advantages must be weighed against the potential costs, such as the potential increase in airborne queueing delay. Our findings reveal that the costs of converting airborne delay to ground delay are far from zero, but this does not mean that the costs outweigh the benefits. Airborne delay is more expensive than ground delay in terms of fuel, maintenance, and other factors, while also increasing controller workload and reducing safety. It is therefore reasonable to accept a substantial increase in total delay in order to reduce airborne delay. Our results document the terms of trade of this transaction for present-day GDPs.

Equally important, our analysis identifies causes of excess delay in GDPs and thus suggests strategies for alleviating it. Our counterfactual scenarios highlight the relative importance of precision in program rate prediction and execution to reduce excess delay. As expected, discrepancies between planned and actual program rates lead to excess delays but the contribution is relatively small. On the other hand, our results show the crucial role of accurate flight time estimation and strict compliance with planned times in reducing GDP excess delay. Enhancements in the accuracy of flight plans and predictions, along with strict adherence to time schedules, can profoundly impact flight delivery by minimizing unnecessary delays. This, in turn, could lead to a more efficient and effective implementation of GDPs.

Based on our current work, future studies can explore several avenues. First, the flight cancellation, flight substitution, compression and pop-up behavior resulting from GDPs warrant further analysis, particularly with respect to airline value functions and flight delivery accuracy. One virtue of GDPs that is not captured in our analysis is the flexibility they allow in re-allocating slots to reduce overall delay costs. Secondly, the deterministic queueing model can be augmented with a simulation model that includes GDP design parameters, multiple runways as servers, system uncertainties, and actual airport acceptance rates. This model could facilitate a more robust analysis of individual GDP performance under different scenarios when better information or higher levels of system uncertainties are present. Furthermore, it could account for impacts of more precise flight time estimation from TBO and better GDP decision-making. Lastly, additional feature engineering and use of more advanced statistical models or deep learning modes with larger dataset could result in more reliable analysis of GDP efficiency and accurate estimates of the factors that drive performance.

Bibliography

1. Air China LLC. (2016). 2016 Annual Report. Retrieved from: http://www.airchina.com.cn/en/investor_relations/images/financial_info_and_roadshow/2017/04/25/F554ED4E11C792B421F419738143B8E2.pdf
2. Ball, M., Barnhart, C., Dresner, M., Hansen, M., Neels, K., Odoni, A. R., Peterson, E., Sherry, L., Trani, A, Zou, B. (2010). Total delay impact study: a comprehensive assessment of the costs and impacts of flight delay in the United States. Research Report prepared for the U.S. Federal Aviation Administration.
3. Ball, M., Hoffman, R., & Mukherjee, A. (2010). Ground Delay Program Planning Under Uncertainty Based on the Ration-by-Distance Principle. *Transportation Science* Vol. 44, No. 1, 1-14.
4. Ball, M., Vossen, T., & Hoffman, R. (2001). Analysis of demand uncertainty effects in ground delay programs. 4th USA/Europe air traffic management R&D seminar, 51-60.
5. Ballin, M., & Erzberger, H. (1996). An analysis of landing rates and separations at the Dallas/Fort Worth International Airport. *NASA Technical Memorandum 110397*.
6. Baum, C. F., Schaffer, M. E. and Stillman, S. (2003). Instrumental variables and GMM: Estimation and testing. *The Stata Journal*, 3(1), 1-31.
7. Belkoura, S., Peñalva, J. M., & Zanin, M. (2016). Generation and recovery of airborne delays in air transport. *Transportation Research Part C: Emerging Technologies*, 69, 436-450.
8. Bilimoria, K. (2016). Analysis of Additional Delays Experienced by Flights Subject to Ground Holding. 16th AIAA Aviation Technology, Integration, and Operations Conference .
9. Bloem, M., & Bambos, N. (2014). Ground Delay Program Analytics with Behavioral Cloning and Inverse Reinforcement Learning. *J. Aerosp. Inf. Syst.*, 12, 299-313.
10. Bloem, M., Hattaway, D., & Bambos, N. (2012). Evaluation of Algorithms for a Miles-in-Trail Decision Support Tool. Fifth International Conference on Research in Air Transportation (ICRAT 2012).
11. Britto, R., Dresner, M. and Voltes, A. (2012). The impact of flight delays on passenger demand and societal welfare. *Transportation Research Part E: Logistics and Transportation Review*, 48(2), pp.460-469.
12. Brueckner, J. K. (2002a). Airport congestion when carriers have market power. *American Economic Review*, 92(5), 1357-1375.
13. Brueckner, J. K. (2002b). Internalization of airport congestion. *Journal of Air Transport Management*, 8(3), 141-147.
14. Brueckner, J. K. (2005). Internalization of airport congestion: A network analysis. *International Journal of Industrial Organization*, 23(7-8), 599-614.
15. Brueckner, J. K., & Van Dender, K. (2008). Atomistic congestion tolls at concentrated airports? Seeking a unified view in the internalization debate. *Journal of Urban Economics*, 64(2), 288-295.
16. Bureau of Transportation Statistics. (2020). Airport rankings 2019. Retrieved from <https://www.bts.dot.gov/airport-rankings-2019>
17. CAAC. (2012). Notice to Print and Distribute the On-Time Performance Statistical Methods for Civil Aviation Flights
18. CAAC. (2016) Weekly observation: stringent management to improve on-time flight performance. Retrieved from: http://www.caac.gov.cn/ZTZL/RDZT/LHDBRYMH/201603/t20160310_29826.html
19. CAAC. (2017). Statistical Data on Civil Aviation of China.
20. CAAC. (2020). Standard airborne time of airports. Retrieved from:
21. Callaham, M., & Dearmon, J.S. (2001). - 1-- Assessing NAS Performance : Normalizing for the Effects of Weather.
22. Campbell-Hill. (2017). The Economic Impact of Emirates in the United States. Campbell-Hill Aviation Group.

23. China Eastern LLC. (2016). Annual Report. Retrieved from http://www.airchina.com.cn/en/investor_relations/images/financial_info_and_roadshow/2017/04/25/F554ED4E11C792B421F419738143B8E2.pdf
24. Christien, R., Hoffman, E., & Zeghal, K. (2019). Enroute traffic overflows versus arrival management delays. Thirteenth USA/Europe Air Traffic Management Research and Development Seminar.
25. Churchill, A.M., Lovell, D.J., Ball, M.O. (2010). Flight Delay Propagation Impact on Strategic Air Traffic Flow Management.
26. Clausen, J., Larsen, A. Larsen, J., Rezanova, N.J. (2010). Disruption management in the airline industry-Concepts, models and methods. *Comput. Oper. Res.* 37(5):809-821.
27. Cox, J., & Kochenderfer, M. (2016). Ground Delay Program Planning Using Markov Decision Processes. *J. Aerosp. Inf. Syst.*, 13, 134-142.
28. Coy, S. (2006). A global model for estimating the block time of commercial passenger aircraft. *Journal of Air Transport Management*, pp. 300-305.
29. Delgado, L.; Prats, X.; Sridhar, B. (2013). Cruise speed reduction for ground delay programs: A case study for San Francisco International Airport arrivals. *Transportation Research Part C*. pp. 83-96.
30. Deshpande, V., & Mazhar, A. (2012). The Impact of Airline Flight Schedules on Flight Delays. *Manufacturing & Service Operations Management*, 423-440.
31. Ding, W., Zhang, Y., & Hansen, M. (2018). Downstream impact of flight rerouting. *Transportation Research Part C*, 176-186.
32. Durbin, M., Noonan, K., Knorr, D., Bonn, J., & Lloyd, L. (2021). Trajectory Based Operations: Challenges Managing Trajectory Uncertainty. 021 ICNS. IEEE.
33. EPA (Environmental Protection Agency). (2006) Greenhouse Gas Inventory Guidance: Direct Emissions from Stationary Combustion Sources
34. Estes, A., & Ball, M. (2017). Data-Driven Planning for Ground Delay Programs. *Transportation Research Record*, 2603, 13-20.
35. Evans, S., Evans, A., Idris, H., & Kozarsky, D. (2004). Modeling Distance Based Metering of Arrival Operations into Major Airports. AIAA 4th Aviation Technology, Integration and Operations (ATIO) Forum.
36. FAA (2015). Simultaneous Dependent and Independent Parallel Approach Operations. N JO 7110.693. Air Traffic Organization Policy.
37. FAA IPH. (2022) Instrument Procedures Handbook (IPH): Chapter: 3. Arrivals.
38. FAA Office of Aviation Policy and Plans. (2016). Economic Values for Evaluation of FAA Investment and Regulatory Decisions: Section 4 Aircraft operating costs. U.S. Federal Aviation Administration Washington.
39. FAA. (2008). Density Altitude. FAA-P-8740-2 o AFS-8 (2008).
40. FAA. (2009). Traffic Flow Management in the National Airspace System. FAA National Headquarters.
41. FAA. (2011). ORDER JO 7110.65T: Section 5. Radar Separation. Retrieved from <https://tfmlearning.faa.gov/publications/atpubs/ATC/atc0505.html#PiK190JACK>
42. FAA. (2011). Section 6. Traffic Management Initiatives. From Subject: Facility Operation and Administration: <https://tfmlearning.faa.gov/publications/atpubs/fac/1706.html>
43. FAA. (2014). Traffic Management Initiative Interaction. From <https://nbaa.org/wp-content/uploads/aircraft-operations/airspace/tfm/FAA-TMI-Interaction.pdf>
44. FAA. (2017). Instrument Procedures Handbook (IPH): Chapter 3: Arrivals.
45. FAA. (2017). NEXTGEN PRIORITIES: Joint Implementation Plan Update Including the Northeast Corridor (October 2017).
46. FAA. (2017). NEXTGEN PRIORITIES: Joint Implementation Plan Update Including the Northeast Corridor.
47. FAA. (2018). Time Based Flow Management (TBFM) Study Report (2018 April). FAA Report.
48. FAA. (2019). FAA Aerospace Forecast: Fiscal Year 2019-2039 .

49. FAA. (2020). Chapter 18: Traffic Management National, Center, and Terminal, Order JO 7210.3BB - Facility Operation and Administration.
50. FAA. (2020). Order JO 7210.3BB - Facility Operation and Administration. From Chapter 18. Traffic Management National, Center, and Terminal:
https://www.faa.gov/air_traffic/publications/atpubs/foa_html/part_5.html
51. FAA. (2021). Section 10. Ground Delay Programs. Retrieved from Chapter 18: Traffic Management National, Center, and Terminal: Chapter 18: Traffic Management National, Center, and Terminal
52. FAA. (2022). Air Traffic by the Numbers: May 2022.
53. FAA. (2022). FAA Order JO 7110.65Z - Air Traffic Control: Section 10. Arrival Procedures and Separation. FAA National Headquarters (FOB-10B) Publications & Administration (AJV-P12).
54. FAA. (2022). PILOT/CONTROLLER GLOSSARY Chapter 7. Safety of Flight - Section 4. Wake Turbulence.
55. Fan, T. P. (2019). Schedule creep-In search of an uncongested baseline block time by examining scheduled flight block times worldwide 1986-2016. *Transportation Research Part A: Policy and Practice*, pp. 192-217.
56. Federal Aviation Administration (2019). Air Traffic By the Numbers.
57. FlightAware. (2023). IFR Plates: KJFK Procedures. Retrieved from <https://flightaware.com/resources/airport/KJFK/APD/AIRPORT+DIAGRAM>
58. Ghalebsaz, J. B., Donohue, G. L., & Shortle, J. F. (2009). A statistical analysis of the aircraft landing process. *Journal of Industrial and Systems Engineering*, 3(3), 152-169.
59. Glaab, P. C., Tamburro, R., & Lee, P. U. (2016). Analysis of the capacity potential of current day and novel runway configurations for New York's John F. Kennedy Airport. *16th AIAA Aviation Technology, Integration, and Operations Conference*.
60. Glover, C., & Ball, M. (2013). Stochastic optimization models for ground delay program planning with equity-efficiency tradeoffs. *Transportation Research Part C-emerging Technologies*, 33, 196-202.
61. Hainan Airlines LLC. (2016). 2016 Annual Report. Retrieved from https://www.hainanairlines.com/go/Annual_Report/cn/2016AnnualReport.pdf
62. Hampton, M. E. (2015). FAA Has Not Effectively Deployed Controller Automation Tools That Optimize Benefits of Performance-Based Navigation, Assistant Inspector General for Aviation Audits(August 2015).
63. Hansen, M. (2002). Micro-level analysis of airport delay externalities using deterministic queuing models: A case study. *Journal of Air Transport Management*, 8, 73-87.
64. Hansen, Mark. (2004). Post-Deployment Analysis of Capacity and Delay Impacts of an Airport Enhancement: Case of a New Runway at Detroit. *Air Traffic Control Quarterly*. 12. 339-365. 10.2514/atcq.12.4.339.
65. Hao, L., Hansen, M. (2014). Block time reliability and scheduled block time setting. *Transportation Research Part B: Methodological*, pp. 98-111.
66. Hsu, K. (2014). Air defense identification zone intended to provide China greater flexibility to enforce East China Sea claims. US-China Economic and Security Review Commission.
67. ICAO. (2005). Annex 2: Rules of the Air: 3.9-3.9 VMC visibility and distance from cloud minima.
68. International Airport Review. (2018). Port authority board approves \$355 million runway revamp at JFK Airport. Retrieved from <https://www.internationalairportreview.com/news/76160/runway-revamp-jfk-airport/>
69. Juneyao Airlines LLC. (2016). 2016 Annual Report. Retrieved from http://www.juneyaoair.com/pages/investor/pdf/H2_AN2016_1.pdf
70. Kafle, N., Zou, B. (2016). Modeling flight delay propagation: A new analytical-econometric approach. *Transportation Research Part B: Methodological*, pp. 520-542.
71. Kang, L., Hansen, M. (2017). Behavioral analysis of airline scheduled block time adjustment. *Transportation Research Part E: Logistics and Transportation Review*, pp. 56-68.
72. Kang, L., Hansen, M. (2018). Assessing the impact of tactical airport surface operations on airline schedule block time setting. *Transportation Research Part C: Emerging Technologies*, pp. 133-147.

73. King, G. (1998). *Unifying Political Methodology: The Likelihood Theory of Statistical Inference*. New York: Cambridge University Press.
74. Kuhn, K. (2013). Ground delay program planning: delay, equity, and computational complexity. *Transp. Res. Part C: Emerg. Technol.* 2013 (35), 193-203. *Transportation Research Part C Emerging Technologies* 35, 193-203.
75. Liu, Y., & Hansen, M. (2014). Evaluation of the Performance of Ground Delay Programs. *Transportation Research Record*. *Transportation Research Record*, 2400(1), 54-64.
76. Liu, Y., Hansen, M., Ball, M. and Lovell, J. (2021). Causal analysis of flight en route inefficiency, *Transportation Research Part B: Methodological*. Pages 91-115.
77. Liu, Y., Liu, Y., Hansen, M., Pozdnukhov, A., & Zhang, D. (2019). Using machine learning to analyze air traffic management actions: Ground delay program case study. *Transportation Research Part E: Logistics and Transportation Review*.
78. Lorentson, M. (2011). The Use of Aviation System Performance Metrics in Qualitative Forecast Evaluation. *General Topics in Aviation, Range, and Aerospace Meteorology*, Part 1.
79. Mangortey, E., Pinon-Fischer, O., Puranik, T., & Mavris, D. (2019). Predicting the occurrence of weather and volume related ground delay programs. *AIAA Aviation 2019 Forum*, 3188.
80. Manley, B., & Sherry, L. (2010). Analysis of performance and equity in ground delay programs. *Transportation Research Part C*, 910-920.
81. Marla, L., Vaaben, B. and Barnhart, C., 2016. Integrated disruption management and flight planning to trade off delays and fuel burn. *Transportation Science*, 51(1), pp.88-111.
82. Mayer, C., and Sinai, T. (2003). "Network Effects, Congestion Externalities, and Air Traffic Delays: Or Why Not All Delays Are Evil." *American Economic Review*, 93 (4): 1194-1215.
83. More, D., Sharma, R. (2014). The turnaround time of an aircraft: a competitive weapon for an airline company. *DECISION: Official Journal of Indian Institute of Management Calcutta*, 489-497.
84. Mukherjee, A., & Hansen, M. (2007). A Dynamic Stochastic Model for the Single Airport Ground Holding Problem. *Transportation Science* Vol. 41, No. 4, 444-456.
85. Mukherjee, A., Grabbe, S., & Sridhar, B. (2014). Predicting Ground Delay Program at an Airport Based on Meteorological Conditions. 14th AIAA Aviation Technology, Integration, and Operations Conference.
86. Myers, T., Klopfenstein, M., Mintzer, J., & Wilmoth, G. (2005). A Preliminary Analysis Of The Impact Of Miles-In-Trail Restrictions On Nas Flight Operations.
87. NBAA. (2020). Miles-in-Trail (MIT)/Minutes-in-Trail (MINIT). From National Business Aviation Association: <https://nbaa.org/aircraft-operations/airspace/tfm/tools-used-for-traffic-flow-management/miles-in-trail-mitminutes-in-trail-minit/>
88. OpenFlight. (2019). Airport, airline and route data. Retrieved from <https://openflights.org/data.html>
89. Ostwald, P., & Topiwala, T. (2006). The Miles-in-Trail Impact Assessment Capability. American Institute of Aeronautics and Astronautics.
90. Pawitan, Y. (2013). *In All Likelihood: Statistical Modelling And Inference Using Likelihood*. Oxford University Press.
91. Rakas, J., & Yin, H. (2005). Statistical modeling and analysis of landing time intervals: Case study of Los Angeles International Airport, California. *Transportation Research Record*, 69-78.
92. Ren, K., Kim, A., & Kuhn, K. (2018). Exploration of the Evolution of Airport Ground Delay Programs. *Transportation Research Record*, 2672(23), 71-81.
93. Ren, P., Li, L. (2018). Characterizing air traffic networks via large-scale aircraft tracking data: A comparison between China and the US networks. *Journal of Air Transport Management*, 181-196.
94. Rios, J. (2010). Aggregate Statistics of National Traffic Management Initiatives. 10th AIAA Aviation Technology, Integration, and Operations (ATIO) Conference, AIAA 2010-9382.
95. Ryerson, M.S., Hansen, M. and Bonn, J., 2014. Time to burn: Flight delay, terminal efficiency, and fuel consumption in the National Airspace System. *Transportation Research Part A*, 69, pp.286-298.
96. Schummer, J., & Vohra, R. V. (2013). Assignment of Arrival Slots. *American Economic Journal: Microeconomics*, 164-185.

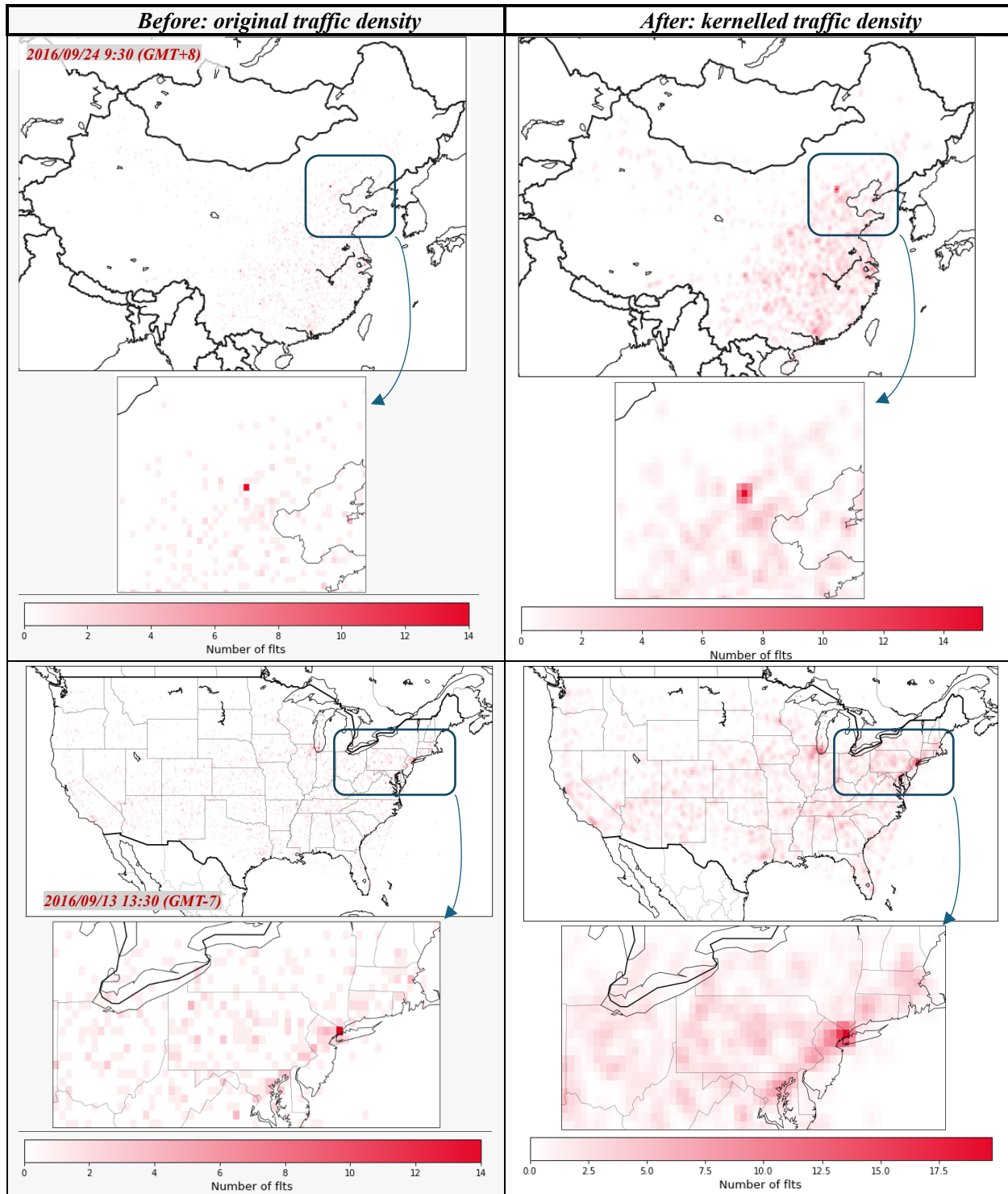
97. SciPy User Guide. (2023). Optimization and root finding. Retrieved from <https://docs.scipy.org/doc/scipy/reference/generated/scipy.optimize.minimize.html>
98. Sgorcea, R., Weitz, L., Huleatt, R., Levitt, I., & Mount, R. (2019). Integrated Time-Based Management and Performance-Based Navigation Design for Trajectory-Based Operations.
99. Sheth, K., Gutierrez-Nolasco, S., & Petersen, J. (2013). Analysis and Modeling of Miles-in-Trail Restrictions in the National Airspace System. Aviation Technology, Integration, and Operations Conference.
100. Sheth, K., Gutierrez-Nolasco, S., & Petersen, J. (2014). Development of Miles-in-Trail Passback Restrictions for Air Traffic. IEEE/AIAA 33rd Digital Avionics Systems Conference (DASC).
101. Shortle, J., & Sherry, L. (2013). A model for investigating the interaction between go-arounds and runway throughput. *2013 Aviation Technology, Integration, and Operations Conference*, American Institute of Aeronautics and Astronautics.
102. Sölveling, G., & Clarke, J. P. (2014). Scheduling of airport runway operations using stochastic branch and bound methods. *Transportation Research Part C: Emerging Technologies*, 45, 119–137.
103. Spring Airlines LLC. (2016). 2016 Annual Report. Retrieved from <https://www.ch.com/Content/images/IntoSpring/investor/%E6%98%A5%E7%A7%8B%E8%88%AA%E7%A9%BA%202016%E5%B9%B4%E5%B9%B4%E5%BA%A6%E6%8A%A5%E5%91%8A.pdf>
104. Sternberg, A., Soares, J., Carvalho, D., & Ogasawara, E. (2017). A Review on Flight Delay Prediction. arXiv:1703.06118v2.
105. Vandevenne, H. F., & Lippert, M. A. (2000). Using maximum likelihood estimation to determine statistical model parameters for landing time separations. 92PM-AATT-0006. *Advanced Air Traffic Technologies, FAA*.
106. Vandevenne, H., & Lippert, M. A. (1994). Evaluation of runway assignment and aircraft sequencing algorithms in terminal area automation. *The Lincoln Laboratory Journal*.
107. Wang, Y., & Kulkarni, D. (2011). Modeling Weather Impact on Ground Delay Programs. SAE International Journal of Aerospace, 4,, 1207-1215.
108. Wong, J.-T., Tsai, S.-C. (2012). A survival model for flight delay propagation. *Journal of Air Transport Management*, pp. 5-11.
109. World Bank (2018). Retrieved from <http://data.worldbank.org/indicator/IS.AIR>. PSGR
110. Yin, J., Hu, M., Ma, Y., Han, K., & Chen, D. (2019). Airport taxi situation awareness with a macroscopic distribution network analysis. *Network and Spatial Economics*, 19(3), 669–695.
111. Zou, B., Hansen, M. (2012). Impact of operational performance on air carrier cost structure: evidence from US airlines. *Transportation Research Part E: Logistics and Transportation Review*, pp. 1032-1048.
112. Zou, B., Hansen, M. (2014). Flight delay impact on airfare and flight frequency: A comprehensive assessment. *Transportation research part E: logistics and transportation*, 54-74, pp. 54-74.

Appendix 1: Top 30 airports in China and the U.S. with commercial flights

Top 30 airports selected in China		
IATA	Airport	Flights (in 1000)
PEK	Beijing Capital International	289.11
CAN	Guangzhou Baiyun International	215.13
SHA	Shanghai Hongqiao International	179.22
SZX	Shenzhen Bao'an International	179.18
CTU	Chengdu Shuangliu International	170.54
XIY	Xi'an Xianyang International	159.97
KMG	Kunming Changshui International	157.86
CKG	Chongqing Jiangbei International	157.69
PVG	Shanghai Pudong International	154.69
HGH	Hangzhou Xiaoshan International	151.73
NKG	Nanjing Lukou International	123.11
CGO	Zhengzhou Xinzheng International	114.87
WUH	Wuhan Tianhe International	111.19
XMN	Xiamen Gaoqi International	107.60
TAO	Qingdao Liuting International	103.86
CSX	Changsha Huanghua International	103.04
HAK	Haikou Meilan International	85.03
TSN	Tianjin Binhai International	81.93
SYX	Sanya Phoenix International	79.66
KWE	Guiyang Longdongbao International	78.15
URC	Urumqi Diwopu International	76.52
DLC	Dalian Zhoushuizi International	70.31
HRB	Harbin Taiping International	65.07
LHW	Lanzhou Zhongchuan International	64.89
SHE	Shenyang Taoxian International	64.70
NNG	Nanning Wuxu International	62.63
FOC	Fuzhou Changle International	61.29
TYN	Taiyuan Wusu International	59.10
CGQ	Changchun Longjia International	45.29
HET	Hohhot Baita International	35.13

Top 30 airports selected in the U.S.		
IATA	Airport	Flights (in 1000)
ATL	Hartsfield-Jackson Atlanta International	870.46
ORD	Chicago O'Hare International	818.82
LAX	Los Angeles International	628.30
DFW	Dallas/Fort Worth International	565.45
DEN	Denver International	522.24
SFO	San Francisco International	422.36
JFK	John F. Kennedy International	398.32
IAH	George Bush Intercontinental/Houston	390.60
LAS	McCarran International	373.08
EWK	Newark Liberty International	369.71
BOS	Logan International	352.57
PHX	Phoenix Sky Harbor International	346.79
MSP	Minneapolis-St Paul International	345.03
MIA	Miami International	335.38
SEA	Seattle/Tacoma International	318.02
DTW	Detroit Metro Wayne County	315.25
CLT	Charlotte Douglas International	308.17
LGA	LaGuardia	304.02
MCO	Orlando International	296.78
PHL	Philadelphia International	266.67
SLC	Salt Lake City International	245.20
FLL	Fort Lauderdale-Hollywood International	236.84
BWI	Baltimore/Washington International Thurgood Ma...	215.28
DCA	Ronald Reagan Washington National	215.07
MDW	Chicago Midway International	199.85
SAN	San Diego International	175.88
TPA	Tampa International	157.76
PDX	Portland International	135.36
IAD	Washington Dulles International	134.21
MEM	Memphis International	50.06

Appendix 2: Example of traffic density at a specific time in China and the U.S.



Appendix 3: Major hub airports in China and the U.S.

	Airlines	Hub airports
China	Air China	Beijing Capital International Airport (PEK) Chengdu Shuangliu International Airport (CTU) Shanghai Pudong International Airport (PVG) Shanghai Hongqiao International Airport (SHA)
	Eastern Airlines	Shanghai Pudong International Airport (PVG) Shanghai Hongqiao International Airport (SHA) Beijing Capital International Airport (PEK) Kunming Changshui International Airport Xi'an Xianyang International Airport (XIY)
	Southern Airlines	Guangzhou Baiyun International Airport (CAN) Beijing Capital International Airport (PEK) Ürümqi Diwopu International Airport (URC) Chongqing Jiangbei International Airport (CKG) Shanghai Pudong International Airport (PVG)
	Hainan Airlines	Haikou Meilan International Airport (HAK) Beijing Capital International Airport (PEK) Xi'an Xianyang International Airport (XIY)
	Xiamen Airlines	Xiamen Gaoqi International Airport (XIA) Fuzhou Changle International Airport (FOC) Hangzhou Xiaoshan International Airport (HGH) Changsha Huanghua International Airport (CSX)
	Shanghai Airlines	Shanghai Pudong International Airport (PVG) Shanghai Hongqiao International Airport (SHA)
	Shandong Airlines	Jinan Yaoqiang International Airport (TNA) Qingdao Liuting International Airport (TAO) Xiamen Gaoqi International Airport (XIA)
U.S.	Southwest Airlines	Denver International Airport (DEN) McCarran International Airport (LAS) Phoenix Sky Harbor International Airport (PHX) Hartsfield-Jackson Atlanta International Airport (ATL) Orlando International Airport (MCO) Chicago Midway Airport (MDW) Baltimore/Washington International Thurgood Marshall Airport (BWI)
	United Airlines	San Francisco International Airport (SFO) Los Angeles International Airport (LAX) Denver International Airport (DEN) George Bush Intercontinental Airport (IAH) Newark Liberty International Airport (EWR) Chicago O'Hare International Airport (ORD) Washington Dulles International Airport (IAD)
	Delta Airlines	Hartsfield-Jackson Atlanta International Airport (ATL) Detroit Metropolitan Wayne County Airport (DTW) Minneapolis-Saint Paul International Airport (MSP) Los Angeles International Airport (LAX) Salt Lake City International Airport (SLC) John F. Kennedy International Airport (JFK)
	American Airlines	Philadelphia International Airport (PHL) Phoenix Sky Harbor International Airport (PHX) Dallas-Fort Worth International Airport (DFW) John F. Kennedy International Airport (JFK) Los Angeles International Airport (LAX) Chicago O'Hare International Airport (ORD) Charlotte Douglas International Airport (CLT)

Appendix 4: Flight share of airline and aircraft types of core 30 airports in China & the U.S.

China: aircraft type				
Aircraft	Flight count	Cumulative flight count	Flight share(%)	Cumulative flight share(%)
B738	683473	683473	39.94	39.94
A320	483667	1167140	28.26	68.20
A321	220046	1387186	12.86	81.06
A319	84531	1471717	4.94	86.00
B737	71687	1543404	4.19	90.19
A333	45701	1589105	2.67	92.86
A332	30823	1619928	1.80	94.66
E190	19581	1639509	1.14	95.80
OTHERS	71811	1711320	4.20	100.00

U.S.: aircraft type				
Aircraft	Flight count	Cumulative flight count	Flight share(%)	Cumulative flight share(%)
B738	440854	440854	17.41	17.41
A320	367396	808250	14.51	31.92
A321	250609	1058859	9.90	41.82
B737	242043	1300902	9.56	51.38
B739	212124	1513026	8.38	59.76
A319	185238	1698264	7.32	67.08
E170	118067	1816331	4.66	71.74
B752	114320	1930651	4.52	76.26
MD88	72860	2003511	2.88	79.14
B712	67466	2070977	2.66	81.80
E190	66312	2137289	2.62	84.42
MD90	51300	2188589	2.03	86.45
CRJ7	49708	2238297	1.96	88.41
B733	41389	2279686	1.63	90.04
E75L	32465	2312151	1.28	91.33
B753	30432	2342583	1.20	92.53
CRJ9	27983	2370566	1.11	93.63
B763	26932	2397498	1.06	94.70
OTHERS	134228	2531726	5.30	100.00

China: airline				
Airline	Flight count	Cumulative flight count	Flight share(%)	Cumulative flight share(%)
CSN	354227	354227	20.70	20.70
CES	255918	610145	14.95	35.65
CCA	193880	804025	11.33	46.98
CHH	141908	945933	8.29	55.28
CXA	127134	1073067	7.43	62.70
CSZ	104336	1177403	6.10	68.80
CSC	79613	1257016	4.65	73.45
CDG	62141	1319157	3.63	77.08
CSH	53153	1372310	3.11	80.19
DKH	44878	1417188	2.62	82.81
GCR	35836	1453024	2.09	84.91
CBJ	32653	1485677	1.91	86.81
CQH	30212	1515889	1.77	88.58
CHB	21056	1536945	1.23	89.81
LKE	20324	1557269	1.19	91.00
UEA	18499	1575768	1.08	92.08
OTHERS	135552	1711320	7.92	100.00

U.S.: airline				
Airline	Flight count	Cumulative flight count	Flight share(%)	Cumulative flight share(%)
AAL	669945	669945	26.46	26.46
DAL	554483	1224428	21.90	48.36
UAL	415921	1640349	16.43	64.79
SWA	330870	1971219	13.07	77.86
JBU	127536	2098755	5.04	82.90
NKS	96564	2195319	3.81	86.71
SKW	82061	2277380	3.24	89.95
ASA	78799	2356179	3.11	93.07
VRD	54567	2410746	2.16	95.22
FFT	53897	2464643	2.13	97.35
OTHERS	67083	2531726	2.65	100.00

Appendix 5: Components of direct operating cost for major and minor Chinese airlines

2016 (billion RMB)	Fuel	Depreciation	Crew salaries	Aircraft Maintenance	Rentals	Total	Ratio
Major Airlines							
Air China	21.98	13.47	20.08	4.65	6.25	66.43	0.33
China Eastern	19.63	12.15	18.15	4.96	2.86	57.75	0.34
China Southern	23.80	12.62	92.15	11.32	7.33	64.28	0.37
Hainan Airline	7.86	3.66	2.88	3.99	4.35	22.74	0.35
Minor Airlines							
Juneyao Airlines	1.92	1.53	1.57	0.44	0.14	5.61	0.34
Spring Airlines	1.88	1.33	1.39	0.41	/	5.01	0.38