Data-Driven Modeling of Air Traffic Flows for Advanced Air Traffic Management

by

Mayara Condé Rocha Murça

B.E., Aeronautics Institute of Technology (2011)
M.S., Aeronautics Institute of Technology (2013)

Submitted to the Department of Aeronautics and Astronautics
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Aeronautics and Astronautics

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

September 2018

Author .....
Department of Aeronautics and Astronautics
August 23, 2018

Certified by .....
R. John Hansman
T. Wilson Professor of Aeronautics and Astronautics
Thesis Supervisor

Certified by .....
Hamsa Balakrishnan
Associate Professor of Aeronautics and Astronautics
Thesis Committee Member

Certified by .....
Tom G. Reynolds
Associate Group Leader, MIT Lincoln Laboratory
Thesis Committee Member

Accepted by .....
Hamsa Balakrishnan
Associate Professor of Aeronautics and Astronautics
Chair, Graduate Program Committee

Data-Driven Modeling of Air Traffic Flows for Advanced Air Traffic Management

by

Mayara Condé Rocha Murça

Submitted to the Department of Aeronautics and Astronautics
on August 23, 2018, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy in Aeronautics and Astronautics

Abstract

The Air Traffic Management (ATM) system enables air transportation by ensuring a safe and orderly air traffic flow. As the air transport demand has grown, ATM has become increasingly challenging, resulting in high levels of congestion, flight delays and environmental impacts. To sustain the industry growth foreseen and enable more efficient air travel, it is important to develop mechanisms for better understanding and predicting the air traffic flow behavior and performance in order to assist human decision-makers to deliver improved airspace design and traffic management solutions. This thesis presents a data-driven approach to modeling air traffic flows and analyzes its contribution to supporting system level ATM decision-making.

A data analytics framework is proposed for high-fidelity characterization of air traffic flows from large-scale flight tracking data. The framework incorporates a multi-layer clustering analysis to extract spatiotemporal patterns in aircraft movement towards the identification of trajectory patterns and traffic flow patterns. The outcomes and potential impacts of this framework are demonstrated with a detailed characterization of terminal area traffic flows in three representative multi-airport (metroplex) systems of the global air transportation system: New York, Hong Kong and Sao Paulo.

As a descriptive tool for systematic analysis of the flow behavior, the framework allows for cross-metroplex comparisons of terminal airspace design, utilization and traffic performance. Novel quantitative metrics are created to summarize metroplex efficiency, capacity and predictability. The results reveal several structural, operational and performance differences between the metroplexes analyzed and highlight varied action areas to improve air traffic operations at these systems.

Finally, the knowledge derived from flight trajectory data analytics is leveraged to develop predictive and prescriptive models for metroplex configuration and capacity planning decision support. Supervised learning methods are used to create prediction models capable of translating weather forecasts into probabilistic forecasts of the metroplex traffic flow structure and airport capacity for strategic time horizons. To process these capacity forecasts and assist the design of traffic flow management strategies, a new optimization model for capacity allocation is developed. The proposed models are found to outperform currently used methods in predicting throughput performance at the New York airports. Moreover,

when used to prescribe optimal Airport Acceptance Rates in Ground Delay Programs, an overall delay reduction of up to 9.7% is achieved.

Thesis Supervisor: R. John Hansman
Title: T. Wilson Professor of Aeronautics and Astronautics

Acknowledgments

The author would like to thank the support from the Aeronautics Institute of Technology (ITA), the Brazilian Air Force (FAB), the Brazilian Federal Agency for Support and Evaluation of Graduate Education (CAPES) and the Latin American Scholarship Program of American Universities (LASPAU). In addition, the author would like to thank the members of the doctoral committee, Prof. John Hansman, Prof. Hamsa Balakrishnan and Dr. Tom Reynolds, for all the guidance and feedback throughout the development of the research, as well as the thesis readers, Dr. James Jones and Prof. Carlos Müller, for their technical and editorial comments. The author also thanks all the collaborators, especially Richard DeLaura and Dr. Lishuai Li, for providing data and valuable insights.

THIS PAGE INTENTIONALLY LEFT BLANK

Contents

THIS PAGE INTENTIONALLY LEFT BLANK

List of Figures

THIS PAGE INTENTIONALLY LEFT BLANK

List of Tables

List of Acronyms and Abbreviations

AAR	Airport Acceptance Rate
ADS-B	Automatic Dependent Surveillance - Broadcast
ANSP	Air Navigation Service Provider
ARSI	Arrival Route Status and Impact
ARTCC	Air Route Traffic Control Center
ASPM	Aviation System Performance Metrics
ATC	Air Traffic Control
ATCSCC	Air Traffic Control System Command Center
ATCT	Air Traffic Control Tower
ATFM	Air Traffic Flow Management
ATM	Air Traffic Management
CDM	Collaborative Decision-Making
CGH	Sao Paulo/Congonhas Airport
CNS	Communication, Navigation and Surveillance
CTOP	Collaborative Trajectory Options Program
EDCT	Expected Departure Clearance Time
EWR	Newark International Airport
FAA	Federal Aviation Administration
FCA	Flow Constrained Area
GDP	Ground Delay Program
GHP	Ground Holding Problem
GPS	Global Positioning System

GRU	Sao Paulo/Guarulhos International Airport
HKG	Hong Kong International Airport
IATA	International Air Transport Association
ICAO	International Civil Aviation Organization
ILS	Instrument Landing System
IMC	Instrument Meteorological Conditions
JFK	John F. Kennedy International Airport
LGA	LaGuardia Airport
METAR	Meteorological Terminal Aviation Routine Weather Report
MFM	Macau International Airport
NAS	National Airspace System
NASA	National Aeronautics and Space Administration
NextGen	Next Generation Air Transportation System
NEXTOR	National Center of Excellence for Aviation Operations Research
RAPT	Route Availability Planning Tool
RBS	Ration-By-Schedule
SESAR	Single European Sky ATM Research
STAR	Standard Terminal Arrival Route
SZX	Shenzhen Bao'an International Airport
TAF	Terminal Aerodrome Forecast
TBO	Trajectory-Based Operations
TFMS	Traffic Flow Management System
TMI	Traffic Management Initiative
TRACON	Terminal Radar Approach Control
VCP	Viracopos International Airport
VMC	Visual Meteorological Conditions

Chapter 1

Introduction

1.1 Motivation

Air transportation is critical to promoting a nation's economic growth and social development. As the fastest means for worldwide transportation, it is essential for global business and tourism. In 2017, aviation transported 4.1 billion passengers and 59.9 million tonnes of freight [4]. According to the latest study on the economic and social benefits of air transport, the industry supported 62.7 million jobs worldwide and its global economic impact (direct, indirect, induced and enabling) was estimated at US$ 2.7 trillion, equivalent to 3.5% of the world's Gross Domestic Product, in 2014 [5]. The importance of the industry is expected to continue to increase over the next decades. The International Air Transport Association (IATA) expects the air traffic demand to double by 2035, reaching 7.2 billion passengers globally [6]. In the U.S., the Federal Aviation Administration (FAA) forecasts an average increase in system enplanements of 2.2% per year over the next 20 years [7].

Enabling air transportation over the years, Air Traffic Management (ATM) is the system of systems responsible for promoting a safe and orderly flow of aircraft through the airspace. Though serving the aviation needs since its beginning, this system has reached its operational limits. In 1983, the International Civil Aviation Organization (ICAO) first recognized the need for new air navigation systems and procedures in order to facilitate the civil aviation growth at a global scale [8]. Yet, present ATM systems are still limited by technologies and operational procedures from the past century, which constrain capacity and lead to ineffi-

ciencies such as unnecessary delays and emissions. Demand-capacity mismatches currently impose significant delays for passengers and costs for the airline industry and the economy as a whole. In the U.S., for example, 16.5% of all domestic flights were delayed by more than 15 minutes in 2016 [9]. In Europe, the percentage of delayed flights was 18.2% in 2015 [10]. According to a study developed by the FAA National Center of Excellence for Aviation Operations Research (NEXTOR), the annual cost of flight delays to the U.S. economy was estimated to be US$ 31.2 billion in 2007, including direct costs for airlines and passengers and indirect costs associated with demand and Gross Domestic Product losses [11].

In order to enable the foreseen air traffic growth sustainably and address future air transportation challenges, a global effort is currently underway towards the modernization and harmonization of ATM systems. In the U.S., the Next Generation Air Transportation System (NextGen) initiatives aim to achieve an in-depth transformation of the National Airspace System (NAS) that will increase safety and efficiency and reduce the environmental impacts of aviation [12]. Likewise, the Single European Sky ATM Research (SESAR) initiative coordinates and concentrates all EU research and development activities to define, develop and deploy the new generation of ATM system in Europe [13]. Similar programs have been created in Japan (CARATS [14]), Brazil (SIRIUS [15]), China (CAAMS [16]) etc.

Overall, these initiatives envision the deployment of new technologies for Communication, Navigation and Surveillance (CNS) and ATM automation as well as new operational procedures that best leverage the on-going technological advancements. For example, digital data-link systems are expected to replace voice-based systems for faster and more accurate air traffic controller-pilot communication. Satellite-based navigation will enable more direct and efficient routes and provide more stringent navigational accuracy than ground-based navaids. Satellite-based surveillance will complement radar-based systems, allowing for greater coverage and increased precision in the monitoring and control of aircraft. New decision support systems will increase automation and enhance strategic and tactical traffic flow and airspace management. Taking advantage of all these capabilities, Trajectory-Based Operations (TBO) is envisioned to be the new concept of operations for future ATM, bringing together modern procedures to allow aircraft to follow precise 4D trajectories pre-negotiated between flight operators and Air Navigation Service Providers (ANSPs) towards increasing

the efficiency, predictability and flexibility in trajectory planning and execution [8, 17, 18].

Besides the implementation of new technologies and operational procedures, efficient use and management of air transport system data and information is key to achieve the desired transformation of ATM systems. Every day, large amounts of data are generated as air traffic operations occur, both at the planning and execution stages. Data types include environmental conditions (e.g., weather reports), system state (e.g., air traffic control facility reports), flight demand (e.g., airline schedules, flight plans, surveillance tracks), etc. Furthermore, they are expected to become increasingly available and accessible as new technologies are deployed. For instance, flight trajectory data has become increasingly accessible with the advent of new surveillance technologies. Historically, there has been limited sharing of this type of data by ANSPs for national security reasons. With Automatic Dependent Surveillance-Broadcast (ADS-B), open-source flight tracking data has become publicly available, making it possible for the first time to analyze aircraft movement at a global scale. The development of innovative ways to leverage the system raw data can play a key role in the ATM transformation by providing means for better assessing and understanding operational performance, increasing the system predictive power and creating new decision support tools to assist the planning at strategic and tactical levels. It can help moving from a system that often works reactively and still much relies on experience, intuition and "golden rules" towards a system that is more intelligent, proactive and adaptive.

This research seeks to contribute to this goal by exploring the development of high-fidelity air traffic flow models from large-scale aircraft tracking data and analyzing their potential towards assisting ATM decision-making. Air traffic flows can be far more diverse and complex than what published routes and procedures alone suggest. For example, Figure 1-1 shows the trajectories for arrival flights into John F. Kennedy International airport (JFK) during two different time periods of one same day overlaid with the published Standard Terminal Arrival Routes (STAR). It is observed that the spatial distribution of the traffic does not fully match the STARs and it can vary reasonably over time. Knowledge about these spatiotemporal characteristics of the traffic is key for accomplishing ATM by informing various airspace management and traffic flow management related decisions, as exemplified below:

- *Strategic airspace design:* It supports the assessment of the actual utilization of the

airspace for better design of sectors and routes.

- *System capacity estimation:* It supports the estimation of how much traffic can be accommodated in airspace resources; airspace capacity is a function of air traffic control complexity, which is a function of the flow structure and associated difficulty of control.
- *Demand estimation:* It supports estimating the demand on airspace fixes, sectors or larger airspace volumes.
- *Design of traffic flow management strategies:* It supports the design of flow management strategies to balance demand with capacity and reduce delays; critical regions of the airspace (more prone to see demand-capacity imbalances) can be more easily identified and become a target for flow adjustment programs.
- *Tactical airspace management:* It supports the allocation of staff and/or the tactical reconfiguration of the airspace to balance workload, increase capacity and throughput; periods of higher air traffic control complexity, with complex flow structure, can be more easily anticipated, indicating that changes in staffing or in the airspace structure may be necessary to have an expeditious flow.

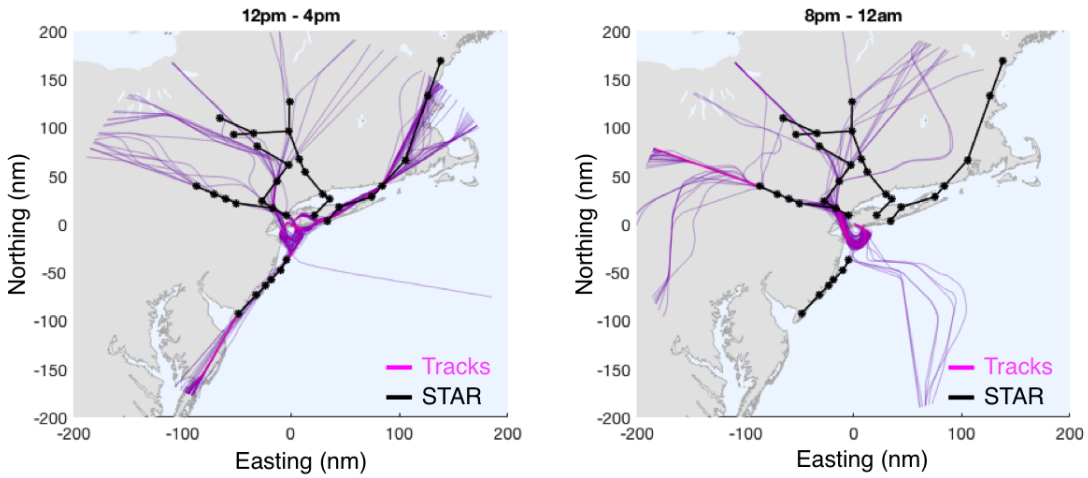


Figure 1-1: Trajectories of arrival flights into JFK during two different time periods on June 23, 2015 overlaid with the Standard Terminal Arrival Routes.

Currently, this knowledge about the traffic flow behavior is primarily obtained through human experience. Based on the continuous observation of flight trajectories through air

traffic situation displays from surveillance systems, air traffic controllers and managers are able to form mental models and create abstractions of the traffic flow in a region. Consequently, they create an intuition about the location of major flows and about the time they tend to occur, which inform their decisions. This reliance on human experience/intuition sometimes yields inefficiencies as decisions tend to be taken based on local aspects rather than on a systems perspective. Moreover, it limits the capabilities of automated decision support systems for air traffic management and hinders the implementation of more advanced operational concepts that require high-fidelity flow information.

Hence, there is a need to develop automated tools that can best characterize and predict the behavior of the traffic and be integrated into decision-support tools for ATM. It is worth mentioning that this need is even more patent in the envisioned TBO environment, where the less structured airspace resulting from more flexible trajectory planning will require more efforts to understand traffic flows.

1.2 Thesis Objectives

This thesis has two main objectives:

1. Investigate how operational data in the ATM system can be mined to automatically identify and predict air traffic flows;

2. Analyze the potential contribution of such knowledge towards obtaining a better understanding of the traffic flow behavior and performance and supporting system-level decision-making in ATM.

1.3 Methodological Approach

This thesis leverages data analytics to achieve the objectives aforementioned. Hu et al. [19] define data analytics as the process of using algorithms to uncover information concealed in data, such as hidden patterns or unknown correlations, with a variety of purposes, such as to give advice, to diagnose and infer reasons for fault or to predict what will occur in

the future. Generally, data analytics can be classified in three levels according to the ways by which it can affect decision-making: descriptive, predictive and prescriptive [20]. These paths are illustrated in Figure 1-2. The first path is the most straightforward; descriptive analytics exploits historical data to diagnose, identify relevant explanatory factors for an event outcome and create insights for guiding future decisions. The second path leverages data to predict the most likely outcome of an event, which can be the state of a variable affecting the decision or even the decision itself. Finally, the third path leverages insights from data to develop realistic models that can prescribe actions to control the outcome of the decision for a given objective. Ultimately, data analytics can be described as the science of using data to build models, improve decisions, and add value to organizations and individuals [21].

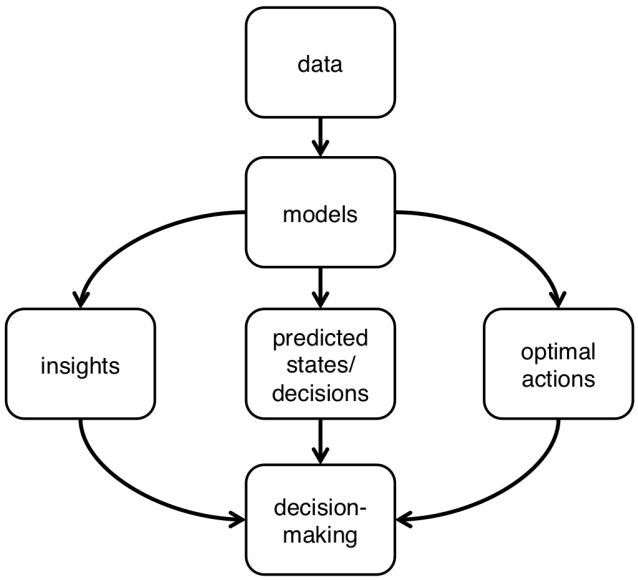


Figure 1-2: General ways by which data analytics affects decision-making.

In this thesis, a flight trajectory data analytics framework is developed based on the application of machine learning methods to exploit flight tracking data and discover spatial and temporal patterns in the movement of aircraft through the airspace in order to provide a high-fidelity characterization of air traffic flows. This characterization provides the foundation for obtaining new insights about the traffic behavior and performance as well as for the development of predictive models to support airspace management and traffic flow

management decisions. The approach is illustrated in Figure 1-3.

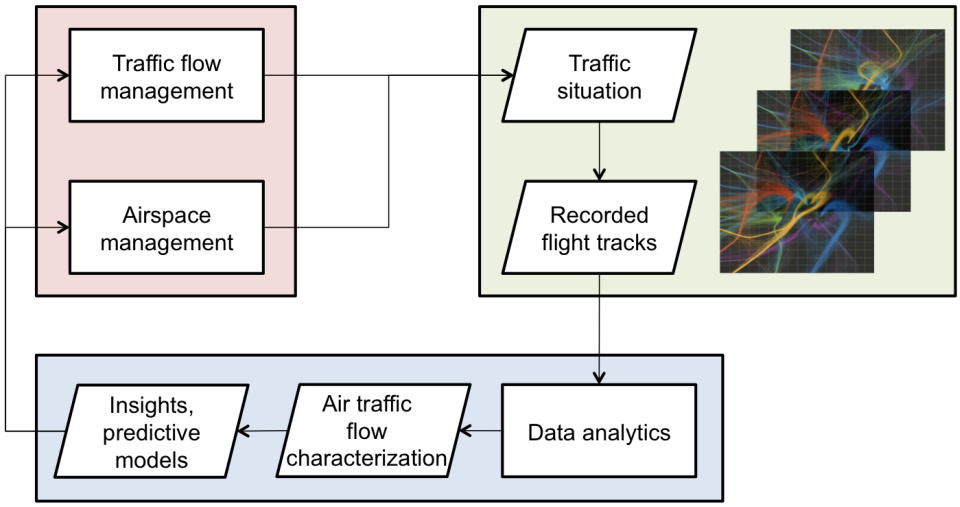


Figure 1-3: Ways by which the flight trajectory data analytics framework can impact ATM.

In order to demonstrate this data analytics framework and discuss its potential impacts on ATM decision-making, this thesis focus on terminal area operations, particularly of multi-airport (metroplex) systems. Large metropolitan regions are typically served by two or more significant airports, which are often closely located and share most of their surrounding airspace (terminal area). As a result of the high traffic density and the interdependent utilization of both airport and airspace resources, the metroplex terminal airspace tends to feature complex traffic dynamics. Besides, the terminal area phase has significant contribution to both individual flight and system level efficiency. The high density of aircraft converging in the terminal area along with the more constrained airspace structure often result in sub-optimal flight paths. These inefficiencies are even more pronounced in multi-airport systems because of the even higher traffic volume and the existence of inter-airport flow interactions. Therefore, better understanding the dynamics and the performance of the traffic flows at the terminal area scale, especially for super dense and complex terminal airspace that characterize metroplex systems, has one of the greatest potential towards supporting airspace and traffic flow management decisions at the systems and enhancing system level ATM performance.

1.4 Background and Related Literature

1.4.1 An Overview of Air Traffic Management

The primary function of Air Traffic Management (ATM) is to ensure a safe and orderly flow of aircraft through the airspace. ATM can be generally subdivided in three distinct functions [22]:

- Air Traffic Control;
- Airspace Management;
- Air Traffic Flow Management.

The main purpose of Air Traffic Control (ATC) is to guarantee safe separation between aircraft and between aircraft and obstacles and avoid collisions. ATC is also responsible for providing flight information services to aircraft, such as weather reports and updates on facility conditions, as well as alerting about aircraft in need of search and rescue. In order to enable the ATC activity, some elementary systems are required: communication, navigation, surveillance and flight and weather information. A detailed description of these systems can be found in [23].

Airspace Management concerns the allocation of the airspace to its various users and its design to provide ATC services. In general, the airspace is subdivided into regions that are assigned to different control facilities. The standard airspace and control structure [24] is composed of Control Tower, which is responsible for the traffic in the nearby airspace around airports and on their surface, Approach Control (called Terminal Radar Approach Control - TRACON - in the U.S.), which controls the terminal airspace (or terminal area) containing the arrival and departure procedures of one or more airports, and Area Control Center (called Air Route Traffic Control Center - ARTCC - in the U.S.), which is responsible for volumes of the airspace containing airways. These different regions of the airspace are further subdivided into sectors, which are individually controlled by an air traffic controller. In the U.S., the NAS contains 22 ARTCCs that are subordinated to a central control unit called Air Traffic Control System Command Center (ATCSCC). This airspace structure is

quite rigid and only shows some level of flexibility in terms of internal sectorization as sectors can be merged or split depending on traffic conditions.

Finally, Air Traffic Flow Management (ATFM) has the goal of matching demand with available capacity by adjusting the flows on a national or regional basis when imbalances are predicted to occur because of weather conditions, traffic volume or special events. In the NAS, ATFM is accomplished through different types of Traffic Management Initiatives (TMIs) issued at strategic (i.e., planning horizons of 2 to 8 hours) and tactical (i.e., real time to planning horizons of 2 hours) time frames. Common strategic TMIs include delaying aircraft on the ground at the origin airport via a Ground Delay Program (GDP) when capacity at the destination airport is reduced or via an Airspace Flow Program when the aircraft is planned to traverse a Flow Constrained Area (FCA), i.e., an airspace region that is impacted (primarily because of weather). Still at strategic time frame, flights can also be rerouted pre-departure to avoid congested or weather impacted areas. The National Playbook contains a set of alternative routes that have been pre-validated and coordinated with ARTCCs and that can be used in support of strategic rerouting. While airborne, flights are subject to tactical TMIs implemented in a more reactive way. For example, an aircraft may be tactically rerouted around weather or may be slowed down with a Miles-in-Trail restriction, a strategy commonly used to control arrival rates into an area through additional spacing at airspace fixes. Recently, efforts have been made to develop new ATFM strategies that integrate existing ones and provide a more flexible and efficient solution for airspace users. The Collaborative Trajectory Options Program (CTOP) is a new strategic TMI developed by the FAA that simultaneously assigns ground delays and/or reroutes around one or more FCAs after explicitly taking into account the priorities of flight operators under a Collaborative Decision-Making (CDM) approach [25]. The National Aeronautics and Space Administration (NASA) is also currently working towards the development of the Integrated Demand Management concept, which attempts to coordinate strategic and tactical flow scheduling by using a CTOP TMI to pre-condition the flows into the Time-Based Flow Management system so that an arrival schedule with lower delays can be built [26].

The planning and execution of the three functions described above is a complex task accomplished by the ATM system. The difficulty of ATM can be generally attributed to

three major factors:

High volume and heterogenous demand

In the U.S., an average of more than 42,000 flights operate per day, and more than 5,000 aircraft can be present in the sky at any given time [27]. Large hub airports serving dense metropolitan regions concentrate a significant part of the traffic; in 2016, 30 airports handled more than 70% of the revenue passenger enplanements [28]. Besides, the system has to deal with a variety of aircraft and mission types, including large commercial jets, small general aviation aircraft, rotorcraft, etc. The rise of unmanned and urban air mobility markets is expected to further increase the volume and heterogeneity of demand, accentuating the pressure over the system.

Dynamic system capacity

The difficulty of ATM can also be attributed to the dynamic behavior of system capacity. Knowledge of capacity is key for allocating demand to ATM resources efficiently, but the large number of capacity-determining factors and their dynamics makes it difficult to estimate flow rates precisely, especially for long time horizons. Figure 1-4 provides a decomposition of the factors that influence airport/airspace capacity and their interrelations. The uncertainty in system capacity can be largely attributed to the dynamic weather behavior. It can be observed that weather conditions plays a major role in determining achievable capacity, both indirectly by influencing the selection of an ATM configuration, and directly by impacting the throughput performance of a given ATM configuration.

Fragmented, multi-stakeholder and human-based decision-making

Additional complexity arises because ATM decisions are taken by multiple human decision-makers in multiple sub-systems dealing with different phases of a flight at different time horizons, requiring lots of coordination. Moreover, they have to accommodate interests from multiple stakeholders, including air traffic service providers, flight operators and passengers.

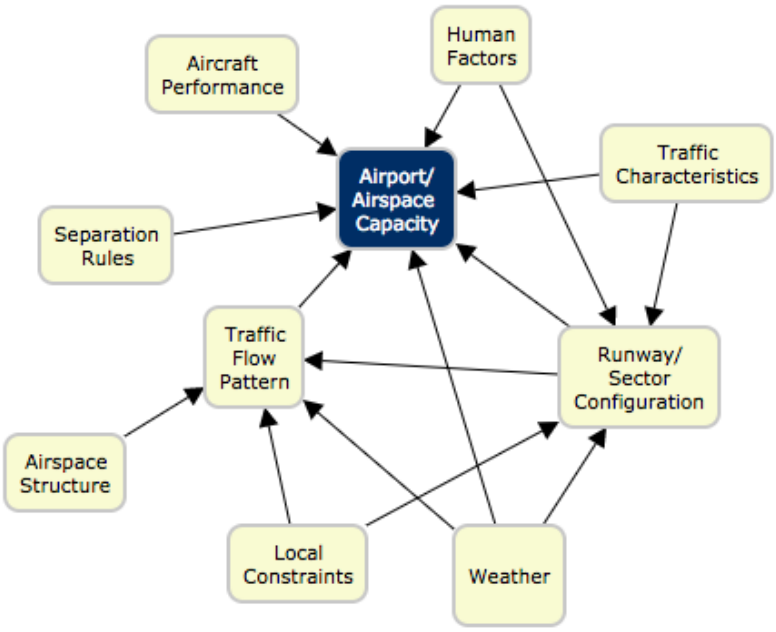


Figure 1-4: Airport/airspace capacity influencing factors.

The ATM functional structure is summarized in the simplified diagram shown in Figure 1-5. The processes and information flows that occur throughout the life-cycle of a flight are indicated, as well as the time horizons associated with each particular process. At the long-term planning horizon ranging from months to years in advance, airline scheduling processes generates the flight schedules to meet the airline's business goals, while strategic Airspace Management is accomplished by the ANSP to design the airspace and the operational procedures. Closer to the flight departure date, which can range from hours to days in advance, airlines elaborate their flight plans describing in detail the proposed aircraft flight to realize the schedule. Flight plans are submitted to the ANSP central ATFM unit, which will approve or require adjustments, for instance, to mitigate an expected demand-capacity imbalance at a particular resource. Expected system capacity is a function of expected ATM configurations and forecast weather conditions at that time. At the tactical time horizon ranging from real time to hours in advance, tactical ATFM is accomplished at the facility level to adjust the flow rates in response to dynamic changes in weather conditions, ATM configuration and capacity. Finally, ATC delivers the final clearances and advisories to as-

sure the safe separations, while aircraft guidance systems enable the execution of the agreed trajectory plan.

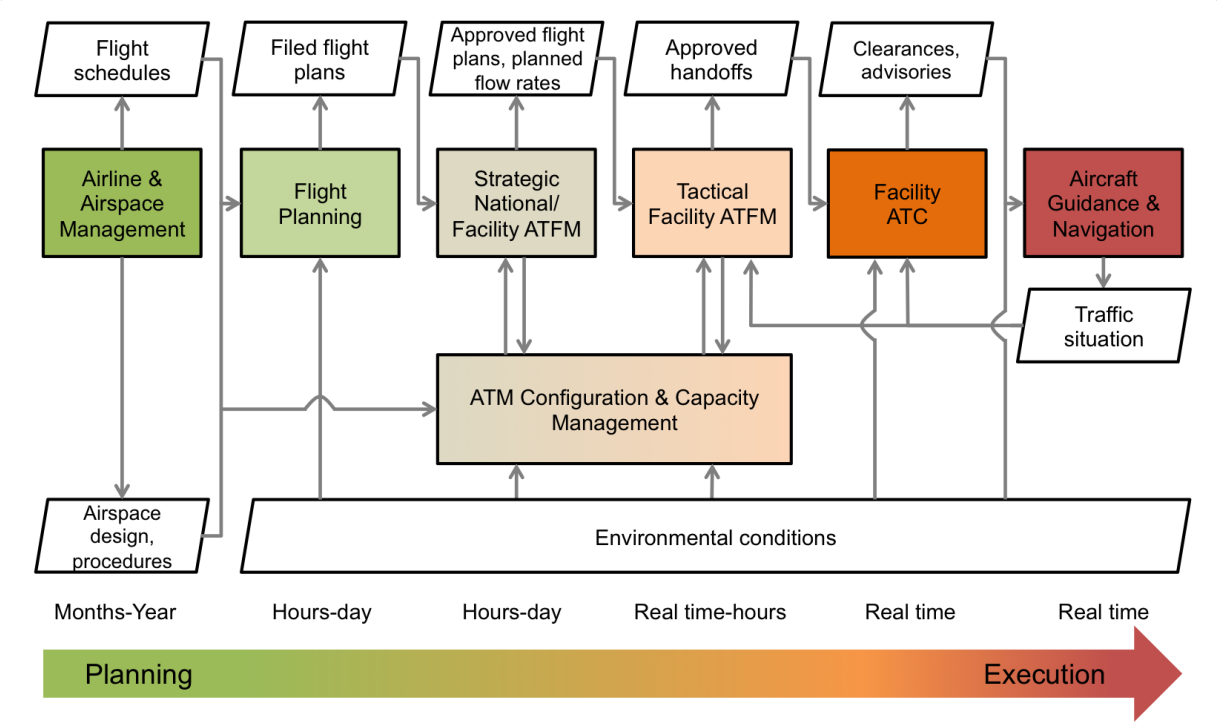


Figure 1-5: ATM functional structure.

1.4.2 Challenges in Metroplex Air Traffic Management

A multi-airport system (also referred to as metroplex) can be defined as a set of two or more significant airports that serve commercial traffic in a metropolitan region, without regard to ownership or political control of individual airports [29]. Multi-airport systems have emerged worldwide as a response to congestion problems, allowing the air transportation system to scale and meet the increasing demand [30]. Bonnefoy (2008) identified two fundamental mechanisms governing the emergence of a multi-airport system historically: construction of a new airport in a region served by a single airport system, with partial or total transfer of traffic; or use of existing airports that were previously sub-utilized in the metropolitan region [30]. The second mechanism has been especially observed after the emergence of low-cost carriers. Through this scaling process, multi-airport systems have accommodated a significant portion of the global traffic and have become key nodes of the air transportation

system. Based on data of 2006, Bonnefoy (2008) identified the existence of 59 multi-airport systems worldwide, which handled 50% of the global passenger traffic.

With the emergence of multi-airport systems, the ATM system has to deal with increasing levels of complexity in the management of the traffic surrounding these systems. The sharing of the terminal airspace results in operational interdependencies and interactions between dense arrival and departure routes from the multiple airports in the same geographical area, creating a complex environment for traffic management. An extensive study developed by Ren et al. [31] and Clarke et al. [3] identified twelve major general metroplex dependency issues that impact their operations based on site visits, domain expert evaluation and qualitative analysis of four major U.S. multi-airport systems (Atlanta, Los Angeles, New York and Miami). These issues are listed in Table 1.1. For example, the use of arrival and departure routes by an airport tends to be dictated not only by its individual runway configuration, but also by the neighboring airport configuration. These interdependencies increase the impact and the workload associated with runway configuration changes (issues 3, 5). Besides, the sharing of airspace resources (arrival and departure fixes, path segments) creates flow interactions that affect the overall performance of the metroplex (issues 1, 2, 6, 7, 8, 11).

As mentioned previously, the difficulty of ATM can be generally attributed to three major factors: high volume; dynamic system capacity; and fragmented decision-making. At multi-airport systems, all of these factors are intensively present: metroplex airspace are the densest of all; metroplex interdependencies result in high variability in ATM configuration and performance; metroplex decisions have to be coordinated between facilities managing the individual airports and airspace. As a result, these systems are characterized by a complex operating environment for ATM. For instance, at the same time they are more reliant on ATFM to adjust demand with the more dynamic system capacity, their characteristics also make the planning of ATFM very challenging. Because there is higher uncertainty in the ATM configuration and its performance, estimating future capacity precisely is more difficult and the allocation of airport and airspace resources under uncertainty becomes more subject to inefficiencies.

The metroplex airspace design and the operating strategies governing the use of the

Table 1.1: Major airport/airspace dependency issues that affect metroplex operations [3].

	Metroplex issue	Definition	Impact
1	Multi-airport departure merge over common departure fix	Occurs when flights from at least two separate airports are procedurally merged over at least one common departure fix.	Very high
2	Major volume-based traffic flow management (ATFM) restrictions	Occurs when a significant level of ATFM restrictions due to demand-to-capacity overloads exist at airspace fixes or at airports.	Very high
3	Proximate-airport configuration conflicts	Occurs when an airport configuration change of one of at least two proximate airports puts restrictions on flights flying to/from other proximate airport(s). This involves flows from one impacting another airport's flows, causing significant rerouting or delays.	High
4	Slow inter-airport ground connectivity	Occurs when inadequate surface transportation of passengers between airports cause significant delays and consequently limits the efficient use of airports by passengers.	High
5	Inefficient/high workload airport configuration changes	Occurs when any major airport configuration change requires significant workload due to reasons, such as coordination of a large number of personnel, FAA facilities, and airports or sector reconfigurations.	High
6	Inefficient multi-airport departure sequencing	Occurs when departure sequencing of flights from multiple airports requires conservative flight restrictions.	High
7	Major secondary airport flow constraints	Occurs when conflicts between a primary airport and a secondary airport lead to constraints on secondary airport flows. Typically, secondary airport traffic will be held below primary airport traffic flows or will be routed around the primary airport traffic patterns resulting in longer flight paths.	High
8	Inefficient flushing of airport flows	Occurs when ATC uses a flushing technique that constrains other airport traffic flows in order to expedite one airport's arrival or departure flights as a way to solve a particular congestion problem (e.g., airport arrival gridlock).	Medium
9	External special use airspace (SUA) causes flow dependencies	Occurs when SUA external to the TRACON constricts TRACON flows into narrow corridors and forces inter-airport traffic flow dependencies.	Medium
10	Terrain causes flow dependencies	Occurs when terrain internal to the TRACON constricts TRACON flows into narrow corridors and forces inter-airport traffic flow dependencies.	Medium
11	Severe limitations on instrument procedures due to proximate airport	Occurs when the use of instrument procedures is severely constrained due to the existence of a proximate airport.	Low
12	Insufficient regional airport capacity	Occurs when there is generally not enough TRACON runway capacity to efficiently serve the air traffic demand.	Low

airspace are recognized to be key factors affecting the coupling of operations, the traffic dynamics and the overall performance of a metroplex. Yet, studies that investigate the interplay between these factors and their impacts on the traffic dynamics and performance are limited. Visual inspection of flight trajectory data has been used to qualitatively evaluate airspace design and identify potentially constraining flow interactions [31]. Delay and fuel burn impacts of decoupled metroplex design and new traffic scheduling strategies have been assessed through simulation [3]. Donaldson and Hansman presented an empirical study aimed at quantifying capacity impacts associated with metroplex interactions [32]. Throughput performance at the airport level and at the system level for particular combinations of runway configurations in the New York metroplex were evaluated using operational data. Capacity discrepancies as great as 60 operations per hour were observed, emphasizing that terminal airspace flow interactions tend to be an important constraining factor driving the capacity of metroplex airports. In this thesis, we demonstrate the potential of the flight trajectory data analytics framework to provide a deeper understanding of the metroplex traffic behavior and performance and discuss how this understanding can be useful for ATM decision support at these systems.

1.4.3 Optimization Models for Air Traffic Management

There is a vast stream of literature dedicated to developing analytical models for optimization of ATM decisions towards assisting human decision makers. The area of ATFM has received particular attention given the complex decision-making nature of this ATM function and its impacts on system-level efficiency. The problem of adjusting air traffic flows in real time in order to balance demand with capacity was first modeled by Odoni [33] using optimization techniques. Since then, several models have been conceptualized for the flow management problem at different scales and considering various types of capacitated resources and control mechanisms.

At the national and strategic level, ATFM is typically concerned about regulating flights destined to a capacity-constrained airport or planned to traverse a capacity-constrained airspace region. The problem of regulation of traffic towards a capacity-constrained airport is the most discussed in the literature and it is commonly referred to as the Ground Holding

Problem (GHP). When demand exceeds capacity at an airport, the arrival traffic will be subject to delays. There are two primary mechanisms by which flight delays can be incurred: delays can be absorbed in the air, for example, when the aircraft is slowed down or put on a holding pattern; or delays can be absorbed on the ground if the departure time is postponed. Because airborne delays are typically more expensive than ground delays and there is an upper bound on the amount of time that a flight can be held in the air, ground delays are generally preferred from both an operating cost and a safety perspective. Nevertheless, as ground delays are strategic in nature and capacity cannot be predicted exactly a few hours in advance, a trade-off between ground and airborne delays will exist when the uncertain nature of capacity is accounted for. Several mathematical programming formulations for the GHP have been developed in order to determine the optimal allocation of ground delays that minimizes the expected overall delay costs. The first model formulations for the GHP were deterministic and static, i.e., capacity was assumed to be known with certainty and decisions were made once at the beginning of the planning horizon [34]. Stochastic and static models were proposed by Richetta and Odoni [35] and by Ball et al. [36] in order to account for uncertainty in the airport capacity profiles. In order to also incorporate the ability to revise decisions as updated capacity information becomes available during the planning horizon, dynamic models were proposed by Richetta and Odoni [37] and Mukherjee and Hansen [38].

Most of the GHP models consider that the planning horizon for capacity allocation is known a priori and determined by the ANSP in a centralized manner. Liu and Hansen [39] took an alternative approach and proposed stochastic GHP models based on deterministic queueing theory in order to optimize the duration of the ATFM plan. Instead of incorporating uncertainty in the arrival rates, they assumed the existence of deterministically known normal and low capacity levels and introduced uncertainty on the time when the airport capacity moves from the low level to the normal level. In their models, the decision on the planning horizon was made to minimize the expected delay costs, which included unpredictability premiums associated with revisions due to early or late capacity recovery. A body of research has also been dedicated to developing decentralized (or "airline-driven") GHP models that incorporate individual flight operator's preferences with respect to the design of the ATFM plan using voting mechanisms [40, 41, 42, 43]. The underlying motivation

for these studies is the fact that airlines have different recovery capabilities due to their operational characteristics and, therefore, they can have different risk tolerances regarding the design of the ATFM plan. For example, an airline that operates under higher flight frequency, lower load factors and more homogeneous fleet mix may have a higher recovery potential from disruption since it can more easily perform aircraft swaps or re-accommodate passengers. Such an airline would tend to be less risk tolerant and prefer a more predictable ATFM plan with longer duration in order to take advantage of its recovery capability [43]. By contrast, airlines with lower recovery potential might be willing to take more risks in order to take advantage of any extra capacity that may be realized (such as an early capacity recovery). These airlines would tend to prefer more aggressive ATFM plans with shorter duration.

The Air Traffic Flow Management Problem (TFMP) extended the GHP by considering that not only the airport but also the airspace can be capacity constrained and by introducing additional control mechanisms such as airborne delays, rerouting and cancellations. Two distinct modeling approaches have been pursued for the TFMP: *Lagrangian* approaches are aircraft-based and seeks to control departure times and traversing times throughout the airspace from origin to destination for each individual flight [44, 45, 46, 47, 48, 49]; whereas *Eulerian* approaches are flow-based and seeks to control aggregate flow rates at nodes and links of the air traffic route network [50, 51, 52, 53, 54].

At the regional and tactical level, ATFM also plays an important role in the regulation of the traffic at individual airports or terminal areas. Many analytical models have been derived for more efficient coordination of airport arrival and departure flows, with the objectives of minimizing delays and maximizing resource throughput while maintaining fairness among airspace users. Runway sequencing and scheduling models have been developed to determine the optimal sequence and schedule of runway usage by arrival and departure aircraft based on the safety requirements between specific aircraft types and their wake turbulence categories [55, 56, 57, 58]. Runway configuration selection models have been developed to optimally schedule airport runway configuration changes based on expected demand and meteorological conditions in order to best balance capacity with the demand of arrivals and departures [59]. Taxiway scheduling models have been created to schedule taxi operations and determine

the optimal routing in the taxiway system in order to prevent "stop-and-go" situations and minimize taxi times [60]. Finally, departure metering models have been developed to manage the demand of departures by holding aircraft at the gate, with engines off, until the right time for its release (i.e., the optimal time to leave the gate and reach the runway at its assigned slot for takeoff) towards minimizing taxi and runway delays and mitigating airport surface congestion [61, 62].

Besides traffic flow management, there has also been substantial research focused on developing analytical models for airspace management decision support. Several approaches have been derived for the problem of determining an efficient airspace sectorization given a particular air traffic situation. Some of them leverage the existing airspace structure to determine sector merging or splitting as a function of demand [63, 64, 65], whereas others focus on redesigning airspace boundaries from scratch [66, 67, 68, 69, 70, 71, 72]. In order to do so, a variety of methods have been used such as mathematical programming [63, 66, 67], computational geometry [68, 72], heuristics [69] and graph theory [71]. Common design objectives for an efficient sectorization were the balancing of monitoring workload among sectors (i.e., the workload associated with monitoring and conflict-avoidance tasks) and the minimization of coordination workload (i.e., the workload associated with hand-offs between sectors). Other design goals included the maximization of sector dwell times and the maximization of sector capacity.

Research in ATM has produced useful models to optimize the traffic flows and airspace configurations, showing efficient computational times on practical size instances and significant potential towards improving the efficiency of air traffic operations. Yet, it is still observed a large gap between theory and practical implementation. One reason for the existing gap between theory and practice is the fact that many analytical models make assumptions that do not hold under the high levels of dynamism and uncertainty that characterize the ATM operating environment, or require information that is not currently automatically available in the ATM system. For instance, models for airspace configuration management typically require detailed traffic information (e.g., dominant flows, critical points) that is not automatically generated by the current systems that support ATM. Models for large-scale air traffic flow optimization make simplistic assumptions about the air traffic route network

that typically do not match the actual spatial patterns of air traffic flows. Besides, they often assume airport and airspace capacities are deterministic and can be known in advance with certainty. Even more realistic models that account for the uncertain behavior of system capacity face barriers towards implementation. For example, most stochastic GHP models assume the uncertain airport capacity profiles can be represented in the form of scenario trees, in which capacity changes between few discrete states assumed to be known to the decision-maker [37, 38]. Yet, generating such scenarios in practice is not straightforward. Besides, actual capacity may be better represented as a continuous variable rather than with few discrete states. Other models assume that nominal and off-nominal capacity values can be treated deterministically and the time of capacity recovery can be modeled with probability distributions [39, 43], but they lack of empirical validation.

1.4.4 Data Analytics for Air Traffic Management

Given the increasing amounts of operational data generated and recorded every day by the ATM system and the advancements in computational science, a field of research has emerged to investigate how data analytics can be leveraged to create useful information from the system raw data for advanced ATM. A broad range of studies have exploited the various types of data available in the ATM system for multiple different tasks, with the overarching goals of better measuring and predicting ATM performance and facilitating decision-support tool development. For example, a non-exhaustive list of topics addressed by these studies is presented below.

Data-driven modeling of air traffic flows

Work has been done to model air traffic flows from flight tracking data for performance assessment and monitoring, airspace design and traffic flow management purposes. For example, Gariel et al. [73] developed a framework for terminal airspace monitoring through the use of clustering techniques to learn nominal spatial trajectory patterns and to assess the conformance of flight trajectories. Sabhnani et al. [74] also applied clustering methods on flight trajectory data in order to identify en route sectors with highly structured traffic

patterns for airspace redesign purposes.

Data-driven modeling of airport and airspace capacity

Another stream of literature has focused on modeling airport and airspace capacity from weather forecast data for traffic flow management applications. For example, Provan et al. [75] used supervised learning to translate raw terminal aerodrome forecasts into probabilistic airport capacity predictions for strategic planning horizons. Pfeil and Balakrishnan [76] also applied machine learning methods on raw convective weather forecasts in order to generate probabilistic predictions of terminal area route blockage because of convective weather.

Data-driven modeling of air traffic delays

Work has also been done to model the dynamics of delay from on-time performance records for prediction and control of air traffic delays. For example, Rebollo and Balakrishnan [77] applied clustering methods to identify characteristic delay states for the NAS and used supervised learning to develop prediction models that leverages this network state information to forecast departure delays for specific origin-destination pairs up to 24 h in the future.

Data-driven modeling of aircraft performance

Another group of studies has been dedicated to model aircraft performance from flight recorder data for performance monitoring applications. For example, Li et al. [78] applied clustering techniques on flight recorder data to model nominal aircraft behavior and identify anomalous flights towards improved safety monitoring. Chati [79] developed data-driven models of aircraft engine performance also using operational data from flight recorders in order to predict a flight's fuel burn directly from its trajectory for enhanced fuel efficiency assessment and monitoring.

A more detailed review of the literature on the topics of data-driven air traffic flow modeling and airport capacity modeling, which are relevant in the context of this thesis, are provided in the next two sections.

1.4.5 Data-Driven Modeling of Air Traffic Flows

The characterization of air traffic flows from aircraft tracking data can be framed as a problem of identification of spatial and temporal trends in the collective movement of aircraft through the airspace. Such problem appears in a variety of other domains in which the movement of objects tends to exhibit correlations in both spatial and temporal dimensions (e.g., vehicles in a road network). Therefore, this section will provide a broader review of works (not restricted to the air traffic domain) that aimed to investigate spatial and temporal patterns in the collective behavior of moving targets. Before doing so, we will first make important distinctions about types of spatial and temporal trends commonly investigated in trajectory data.

Figure 1-6 exemplifies spatiotemporal patterns in a general trajectory dataset [1]. Using the same terminology provided in [1], the following patterns are highlighted: a *flock* of three entities with similar spatial pattern over five time-steps; a *periodic pattern* for an individual entity, i.e., one entity repeats a specific spatial pattern with some periodicity; a *meeting place* where three entities remain for some period; and a *frequent location* where a single entity remains for some period. In this figure, the *flock pattern* provides a good illustration of the spatiotemporal patterns that we seek to mine for flow identification: a group of entities with similar spatial behavior during a period of time. A small distinction is that the entities do not have to be at the same location at exactly the same time. By contrast, they traverse the same sequence of locations within the same time interval. In summary, this thesis considers the following definition of flow:

Definition 1: A *flow* is a pattern in which entities move along the same paths/routes in a spatial dimension, and the movements start/end within the same time interval.

Spatiotemporal Pattern Recognition in General Trajectory Data

The problem of discovering spatiotemporal patterns in actual trajectory datasets has received growing attention in the recent literature across a variety of domains. The increased

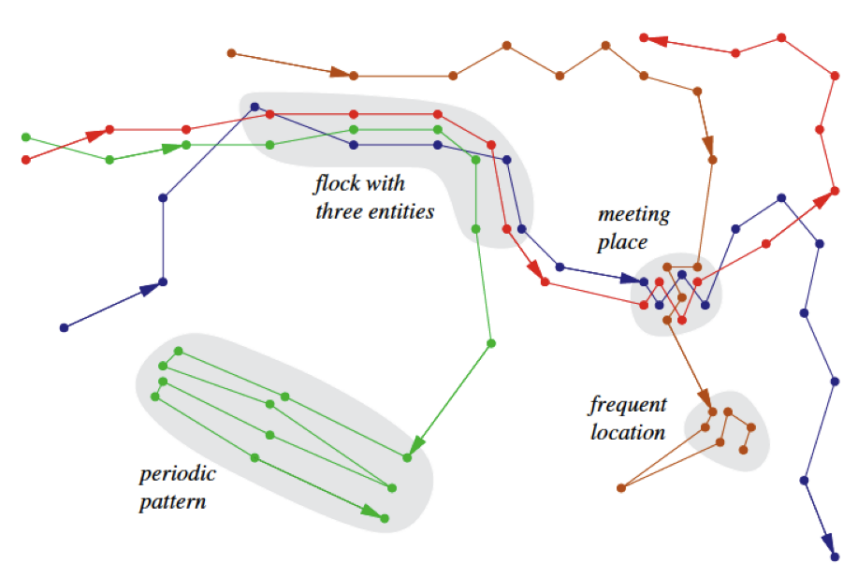


Figure 1-6: Example spatiotemporal patterns in trajectory datasets [1].

use of position sensing technologies (e.g., Global Positioning System (GPS)) has produced large amounts of data and created an opportunity to gain insights through trajectory data mining. Examples include tracking datasets of vehicles, people, animals, weather phenomena (e.g., hurricanes) etc, which are exploited for a variety of purposes. Analysis of vehicle trajectory patterns is used for traffic management in order to discover hot spots in a transportation network and support route planning. Analysis of pedestrian flows can help identifying suspicious behavior in a monitored environment; it can also support urban planning and guide infrastructure investments. Scientists investigate animal movement behavior in order to identify regions frequently visited, investigate social structures within a group of animals and better understand migration patterns. Investigation of hurricane/cyclone movement patterns is performed to understand their behavior and increase the predictability of severe-weather events for improved disaster relief management.

Early approaches to spatial pattern data mining involved indexing trajectory databases and performing basic analysis such as nearest neighbor queries [1]. More recently, trajectory clustering has been extensively used for the identification of common trajectory patterns and outlier detection in such datasets. Antonini and Thiran [80] used hierarchical clustering to group trajectories of redundant targets associated with a person for automatic counting of pedestrians in video sequences. Fu et al. [81] also analyzed video surveillance data of

real-time traffic to identify vehicle motion patterns. With a first-layer spectral clustering, similar trajectories were grouped together, and with a subsequent hierarchical clustering, dominant paths and lanes were identified. Gaffney et al. [82, 83] proposed a probabilistic modeling of trajectories as individual sequences of points generated from regression mixture models (the spatial position was modeled with a polynomial regression model in which time is the independent variable) and used the expectation-maximization algorithm to estimate the parameters of the mixture. They applied the model-based clustering to identify spatial patterns in extra-tropical cyclone tracks over the North Atlantic using meteorological data. Lee et al. [84] developed a partition-and-group framework to discover common sub-trajectories as an alternative to clustering trajectories as a whole. They argue that in some applications (especially when there are regions of special interest for the analysis) it may be useful to find portions of trajectories that have similar behavior even if they are dissimilar when compared as a whole. Their proposed algorithm TRACLUS first partitions a trajectory into a set of line segments at characteristic points where the behavior of the trajectory changes rapidly. Then it groups similar line segments into clusters using a density-based clustering scheme. They used the framework to identify movement patterns in very noisy datasets from hurricane and animal tracking.

The works described so far primarily focused on the identification of similar trajectory patterns in the spatial dimension. A much smaller body of work has incorporated the temporal dimension in the analysis to investigate trends in the occurrence of spatial clusters over time. However, such analysis is key for understanding the dynamics of flow behavior and generating the knowledge necessary for prediction. Kim and Mahmassani [85] identified spatial travel patterns in a road network through density-based clustering of actual vehicle trajectory data and then used k-means to cluster daily time series of traffic for each identified spatial travel pattern and discover daily trends in their use. In a similar way, Wen et al. [86] developed an algorithm to extract shipping route from vessel's historical position point cloud using local polynomial regression and then clustered time series of vessel's frequency of occurrence in multiple zones defined along the frequent routes to find daily traffic patterns.

Spatiotemporal Pattern Recognition in Flight Trajectory Data

In the aviation domain, clustering techniques have also been extensively used to identify spatial traffic patterns from flight track data for performance assessment, airspace monitoring, airspace design and traffic flow management purposes [74, 87, 73, 88, 89, 53, 90, 54, 91]. Eckstein [87] developed a flight track taxonomy for monitoring aircraft behavior based on filtering, segment identification, track decomposition and clustering in order to evaluate how well individual flight trajectories are performing their procedures in the terminal airspace. Gariel et al. [73] also developed a framework for terminal airspace monitoring using a density-based clustering algorithm to learn typical patterns of operation and to assess the conformance of flight trajectories. Similarly, Rehm [88] and Enriquez [89] relied on hierarchical and spectral clustering respectively to identify nominal and abnormal spatial traffic patterns to/from a specific airport. Trajectory clustering has also been used to identify spatial traffic patterns at the en route phase. Sabhnani et al. [74] developed a greedy grid-based trajectory clustering algorithm in order to learn standard flows and critical points and identify en route sectors with highly structured traffic patterns for the goal of airspace redesign. Arneson et al. [90] used a density-based clustering algorithm to identify dominant routing structures between the Forth Worth Center and New York Center and assess the impacts of convective weather on the flow rate capacity along the commonly used routes. Similarly, Marzuoli et al. [53] and Bombelli et al. [54, 91] clustered trajectories between various origin and destination pairs in order to identify common routing structures and model the air traffic route network towards developing higher-fidelity models for traffic flow management optimization.

Studies that aimed to investigate patterns in the spatial organization of the traffic flows over time are also limited in the aviation domain. Song et al. [92] manually extracted flow features from actual flight trajectories in a 15-minute basis, which were aggregated into a vector-based representation and clustered to identify traffic flow patterns in airspace sectors. Flights entering and exiting an airspace sector through the same neighboring sectors were considered to be part of the same flow regardless the shape of their trajectories. With the methodology, they idealized a framework for sector capacity prediction for longer look-ahead times, although little discussion was provided on the predictability of the flow features.

Sidiropoulos et al. [93] proposed a framework to identify spatiotemporal patterns in traffic flows crossing the terminal area boundary in metroplex systems using flight plan data. The approach first identified temporal changes in the local demand detected for discretized zones of the terminal area boundary using a threshold for the difference in the number of flights between consecutive time periods. Subsequently, temporal clusters in the aggregate metroplex demand were identified using a threshold on the number of zones with detected local demand change. These thresholds were optimized to account for uncertainty in demand data. Finally, within each temporal cluster, flights were spatially clustered based on the direction in which they enter or exit the terminal area boundary using k-means algorithm. With the framework, their goal was to provide enhanced demand estimates over shared arrival and departure fixes for supporting metroplex traffic flow management. However, as the spatial analysis was restricted to the location of crossings at the terminal area boundary, their framework did not provide complete information about terminal area traffic flow patterns.

1.4.6 Data-Driven Modeling of Airport Capacity

As discussed in Section 1.4.1, the ATM system capacity is determined by a number of structural, operational and environmental factors. The capacity of an airport is usually determined by the capacity of its most constraining airside element. Typically, this element is the runway system. A common definition of capacity is given by the expected number of movements (landings and takeoffs) that can be performed per unit of time (usually, 1 hour), in the presence of continuous demand and complying with all ATC separation requirements [23]. The capacity of a single runway is primarily determined by the minimum safety separation requirements between consecutive operations, which vary according to the wake turbulence categories of the leading and trailing aircraft. As a consequence, the mix of aircraft types using the runway can significantly affect its capacity. Other important factors that impact the capacity of a single runway include its geometric characteristics (e.g., the existence of high-speed exits reduces runway occupancy times), aircraft performance (e.g., faster aircraft can clear the runway more quickly) and the mix of operations (i.e., whether the runway is dedicated exclusively for arrivals or departures or for both types of operation simultaneously).

Most airports have multiple runways that can be arranged in various configurations for

serving the demand. The layout of these configurations will determine the degree of dependence between operations conducted in different runways. Therefore, a major driver of airport capacity is the active runway configuration [94], i.e., the set of runways and active runway thresholds selected by ATC personnel at any given time to serve the arrival and departure demand. For example, Figure 1-7 shows two different runway configuration commonly used at JFK. In the first one, two runways are used, with arrival operations on 31L/31R (blue arrows) and departure operations on 31L (green arrow). In the second one, three runways are used, with arrivals on 22L/22R and departures on 22R and 31L. The selection of the runway configuration in which an airport will operate is a subjective human-based decision-making process, which is per se affected by various factors such as meteorological conditions (wind speed and direction, ceiling and visibility), demand, noise and workload related restrictions and terminal airspace constraints.

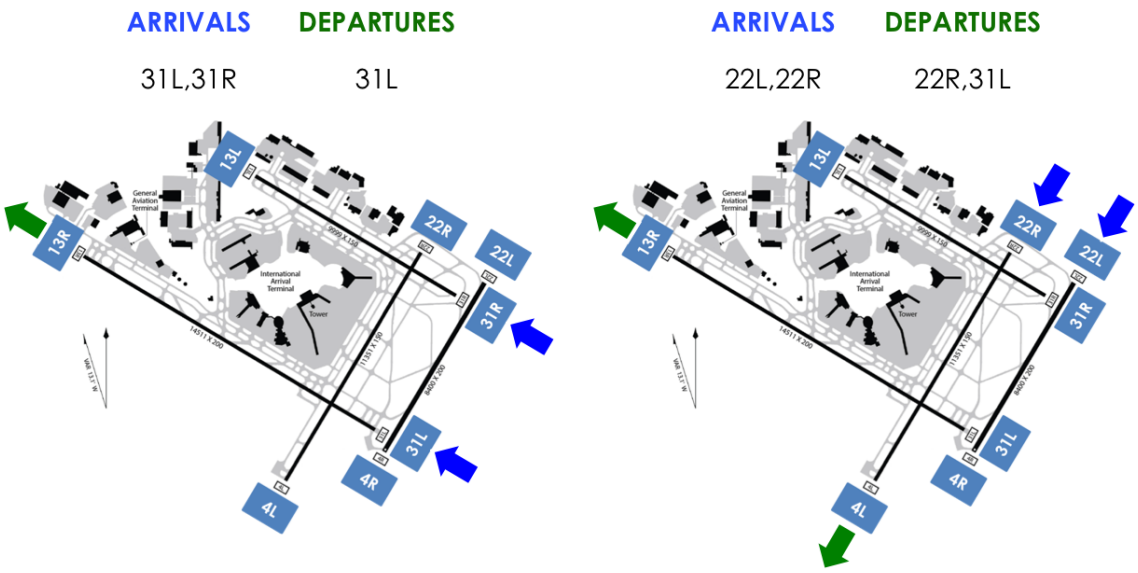


Figure 1-7: Example of typical runway configurations at JFK.

Finally, another key driver of airport capacity is weather. Not only does weather influence the selection of runway configurations, but also directly impacts their throughput performance. Adverse weather conditions such as low ceiling and visibility or high surface winds can drastically reduce capacity and even cause the closure of the airport at extreme cases.

The problem of airport capacity estimation has been widely addressed using analytical [95], empirical [94] and simulation [96] models. For example, Gilbo [94] introduced the notion of capacity curves (also called Pareto frontier) to characterize the trade-offs between arrival and departure operations given a particular runway configuration. Figure 1-8 shows example curves derived empirically from actual observed throughput for both runway configurations shown in Figure 1-7 under Visual Meteorological Conditions (VMC). The curves envelope the peak operating points and can provide an estimate of airport capacity under the assumption that the peak data reflects the airport performance near capacity for that particular condition. However, during real-time ATFM planning, predicting airport capacity is a more difficult problem because runway configurations, weather conditions and demand are not deterministically known a priori, especially for long forecast horizons.

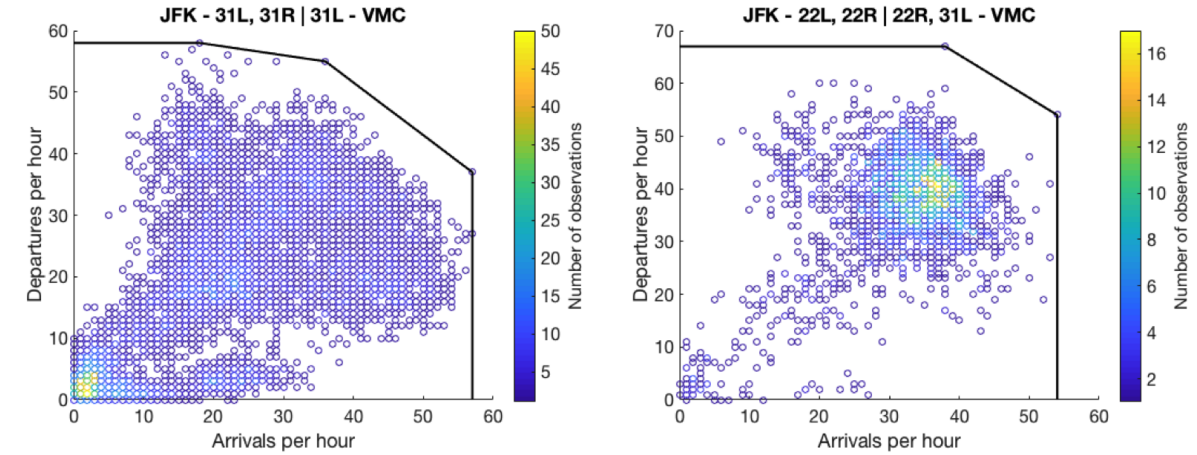


Figure 1-8: Capacity curves for two different runway configurations at JFK (data from 2013-2015).

Efficient planning of airport capacity (typically referred to as Airport Acceptance Rates (AARs)) is key for the overall efficiency of TMIs such as GDP. An overestimation of capacity can amplify airborne delays and require reactive undesirable actions (e.g., diversions, excessive holding), whereas an underestimation of capacity can lead to resource sub-utilization and unnecessary ground delays. Yet, the planning of runway/airspace configurations and capacity is typically done on the basis of experience and through the use of rules-of-thumb. Despite the availability of many weather forecast sources, few tools are available to directly assist traffic managers in the translation of weather forecasts into operational impact.

A body of literature has emerged to investigate whether data analytics techniques can be leveraged to provide capacity-related information for improved ATFM decision support [97, 98, 75, 99, 100, 101, 102, 103, 104]. Different methods for machine learning have been used to translate weather forecasts and/or historical arrival rates into probabilistic airport capacity profiles for strategic planning horizons [97, 98, 75, 99]. Liu et al. [97] used clustering techniques to create airport capacity profiles from historical AAR data. These profiles were represented by the centroids of the obtained clusters from daily AAR time series. Buxi and Hansen [98] adopted a similar approach, but also considered weather forecasts (Terminal Aerodrome Forecast - TAF) to generate the profiles. Provan et al. [75] developed the Weather Translation Model for GDP Planning (WTMG), which uses an ensemble of decision trees to predict the AAR up to 12 hours in the future based on weather forecasts (TAF) and the current state of the airport. Cox and Kochenderfer [99] considered the same variables to build probability distributions for the AAR up to 6 hours in advance using Bayesian networks. Jones et al. [100, 101] included additional features for capturing compression phenomena using the High-Resolution Rapid Refresh (HRRR) forecast and developed a boosted regression tree model to issue capacity predictions in the form of quantiles, which were then applied to chance-constrained integer programming models in order to prescribe AARs. Given the significant impact of runway configuration selection on capacity, another group of studies has used historical airport operational data to model the airport runway configuration selection process and develop tools for runway configuration prediction [102, 103, 104]. For example, Ramanujam and Balakrishnan [103] and Avery and Balakrishnan [104] used discrete-choice models to generate probabilistic forecasts of the runway configuration for look-ahead times of up to 6 hours.

One caveat of the existing data-driven approaches for airport capacity prediction concerns the fact that airports are usually treated as isolated nodes and system level aspects that arise in the most complex metroplex operational environment and affect throughput are not completely incorporated, reducing their performance and applicability in such cases. For instance, Avery [105] found that predicting the runway configurations individually for the New York airports and then combining the results to obtain the aggregate metroplex configuration rather than directly predicting the combined multi-airport runway configuration

led to much lower accuracy in the configuration forecasts. Donaldson and Hansman [32] also ratified the existence of metroplex effects after identifying capacity discrepancies for specific combinations of runway configurations in the New York metroplex through the comparison between capacity envelopes for the individual airports and for the multi-airport system.

1.5 Contributions of the Thesis

Review of the literature in ATM has shown that many analytical models for optimization of ATM decisions have been developed in order to assist human decision-makers, but they either make simplistic assumptions about the ATM operating environment or require information that is not currently automatically generated by the ATM system, hindering their practical implementation. The increasing availability of data and the advancements in computational science have created new opportunities for improving ATM decision-making through the development of data-driven models that can better assess and predict system performance and facilitate the design and implementation of decision-support tools. Several studies have already tapped into this opportunity. For example, trajectory data analytics has been particularly useful to unravel spatial patterns in air traffic for a wide range of ATM applications. Yet, existing approaches are still limited in their capabilities.

The main contributions of this thesis are:

1. *Development of a flight trajectory data analytics framework for high-fidelity characterization of air traffic flows*

   This thesis develops a data analytics framework to exploit large-scale aircraft tracking data in order to discover both spatial and temporal patterns in aircraft movement and enable a high-fidelity traffic flow characterization. The framework has three main capabilities: (1) identification of spatial trajectory patterns; (2) identification of air traffic flows; (3) identification of temporal patterns in air traffic flows. In order to provide these capabilities, the framework is composed of three modules that perform a sequential application of machine learning methods. In the first module, a trajectory clustering scheme enables the identification of spatial trajectory patterns, which define

the as-flown route structure in the airspace of interest. In the second module, a trajectory classification scheme enables the assessment of flight trajectory conformance against the learned airspace structure and the identification of air traffic flows. Finally, in the third module, a clustering analysis at temporal scale enables the identification of traffic flow patterns. With these capabilities, the framework allows to automatically learn the airspace structure, assess the use of the airspace and identify patterns of usage of the airspace.

2. *Development of systematic descriptive approach for metroplex operational behavior and performance assessment*

   Through the use of the flight trajectory data analytics framework to characterize air traffic flows in the terminal area of multi-airport systems, we develop a systematic descriptive approach to analyze metroplex airspace design and utilization and to assess operational performance. Novel quantitative metrics are created to summarize metroplex performance in the areas of efficiency, capacity and predictability. The approach is used to perform a comparative analysis of terminal area operations in three relevant multi-airport systems of different regions of the world: New York, Hong Kong and Sao Paulo. As a result of the analysis, we gain better insights about structural and operational factors driving metroplex performance, and highlight different areas of action to be considered at each multi-airport system towards improving their air traffic operations.

3. *Development of predictive and prescriptive models for metroplex configuration and capacity management*

   Through the use of the flight trajectory data analytics framework to exploit patterns in the metroplex traffic flows, we identify recurrent utilization patterns of runways and airspace and learn the major configuration modes in which the metroplex collectively operates as a system as well as key intervening factors. We leverage this knowledge to develop a data-driven end-to-end approach for metroplex configuration and airport capacity planning. First, predictive models are developed to deliver probabilistic forecasts of the metroplex configuration and airport capacity for strategic planning horizons. As

they incorporate operational interdependency aspects that arise in metroplexes and that are not captured by existing approaches, our predictive models are better suited for this more complex operational environment. Second, we demonstrate how these predictive models can be used to support the design of strategic traffic flow management plans. We develop a novel capacity allocation model that directly incorporates the probabilistic capacity predictions to prescribe optimal AARs for strategic traffic flow management based on an user-defined robustness level.

The contributions of the thesis in the context of the ATM decision-making process previously introduced are illustrated in Figure 1-9. We take a "reverse engineering" approach by developing a data analytics framework to exploit flight trajectory data and provide a high-fidelity characterization of air traffic flows from actual operations. This characterization provides the foundation for obtaining new insights about the traffic behavior and performance as well as for the development of predictive models to support airspace management and traffic flow management decisions.

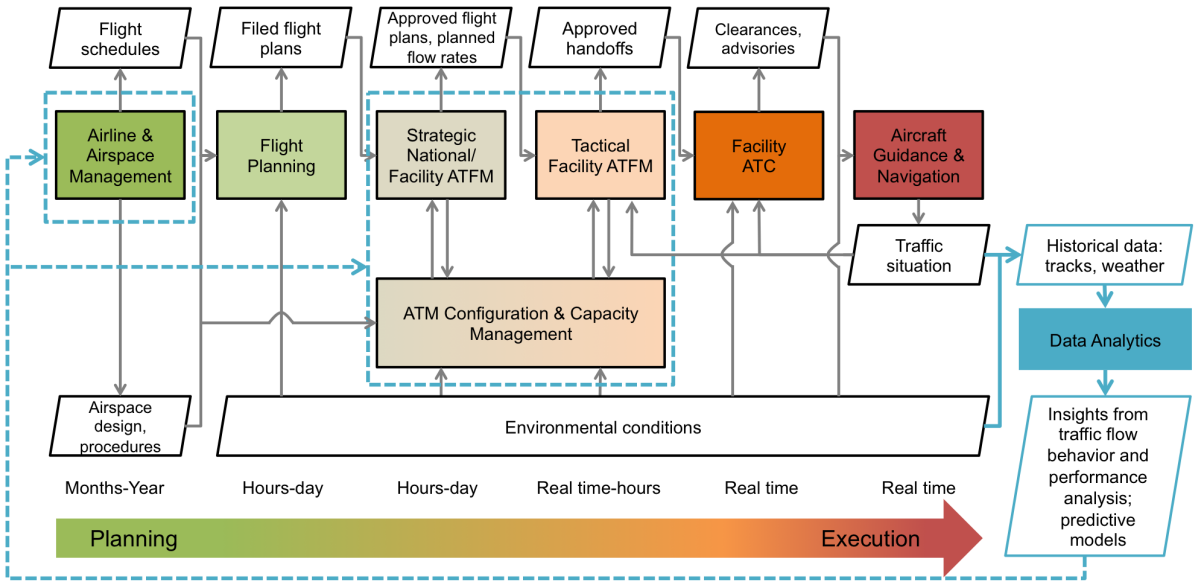


Figure 1-9: View of thesis contributions in the context of the ATM decision-making process.

1.6 Organization of the Thesis

The thesis is organized as follows. Chapter 2 describes the flight trajectory data analytics framework for characterization of air traffic flows. We present the data processing techniques and machine learning methods used in each module of the framework. The methodology is demonstrated with flight tracking data for arrival flights at LaGuardia airport (LGA). Chapter 3 describes the application of the framework for the characterization of air traffic flows in the terminal area of multi-airport systems and demonstrates its use as a descriptive tool to analyze metroplex airspace design and utilization and to assess operational performance. The results of the detailed comparative analysis between New York, Hong Kong and Sao Paulo metroplex operations are discussed. Chapter 4 describes the development of the data-driven approach for metroplex configuration and airport capacity planning based on the knowledge derived from trajectory data analytics. First, we describe the predictive models for traffic flow pattern and capacity prediction, presenting the datasets used, machine learning algorithms tested and predictive performance analysis results. Subsequently, we present the optimization model for arrival rate prescription and discuss its impacts in the planning of GDPs. Finally, Chapter 5 summarizes the major achievements and discusses potential future research directions.

Chapter 2

Flight Trajectory Data Analytics Framework for Characterization of Air Traffic Flows

In this chapter, we describe the flight trajectory data analytics framework for characterization of air traffic flows. The framework has three main capabilities in order to provide a high-fidelity traffic flow characterization: (1) it enables the identification of spatial patterns of aircraft movement; (2) it enables the identification of air traffic flows; (3) it enables the identification of temporal patterns in air traffic flows. With these capabilities, the framework can be used to automatically learn the airspace structure, assess the use of the airspace and identify patterns of usage of the airspace.

2.1 Overview of Methodological Approach

In order to provide these capabilities, the framework is composed of three modules that use machine learning techniques to mine spatial and temporal patterns in flight tracking data. The framework is illustrated in Figure 2-1. In the first module, clustering at spatial scale is performed with a trajectory clustering scheme to identify spatial patterns of aircraft movement, which are referred to as *trajectory patterns* and define the as-flown route structure in the airspace of interest. Based on this knowledge, the second module uses a trajectory

classification scheme to match new flight trajectories with the learned airspace structure and classify them as conforming (to one of the learned routes) or non-conforming. With trajectories classified, flows are identified as temporally associated flight trajectories conforming to the same standard route. The last module of the framework seeks to discover temporal patterns in air traffic flows. For this, a clustering analysis at temporal scale is performed to identify similar structures in the set of flows observed during a period of time, which are referred to as *traffic flow patterns*. A detailed description of each module of the framework is provided next.

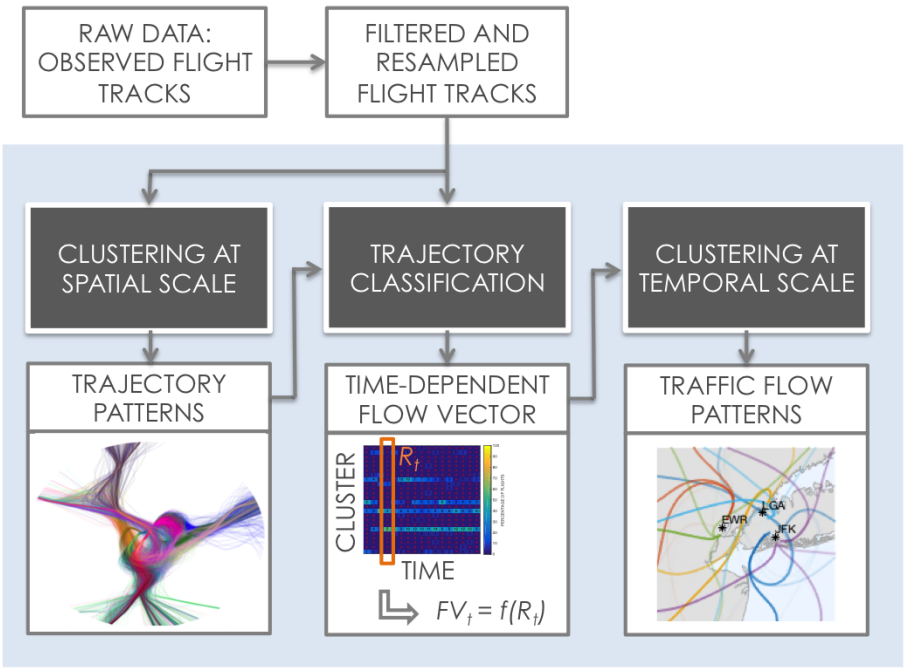


Figure 2-1: Schematic overview of flight trajectory data analytics framework for characterization of air traffic flows.

2.2 Dataset Description

The framework aims at providing a high-fidelity traffic flow characterization by exploiting flight tracking data. This type of data is continuously generated by surveillance systems responsible for tracking aircraft through the airspace and enabling real-time monitoring and control of the traffic by ANSPs around the world. In the U.S., the FAA uses the Traffic Flow

Management System (TFMS) for real-time traffic monitoring. The TFMS displays and logs fused flight track records from the network of domestic radars. TFMS records are logged every minute and provide historical flight trajectory data in the domestic airspace. The data fields include flight ID, latitude, longitude, altitude, speed, origin airport, destination airport and aircraft type.

With the advent of Automatic Dependent Surveillance - Broadcast (ADS-B), flight trajectory data can now be more easily obtained through public repositories. ADS-B is a new type of surveillance technology that relies on aircraft avionics, a constellation of GPS satellites, and a network of ground stations. Briefly explained, aircraft position determined by on-board satellite navigation systems can be broadcasted by the ADS-B transponder on aircraft and then picked up by ADS-B receivers on the ground. Platforms such as *FlightAware* and *FlightRadar24* have a huge network of crowdsourced ADS-B receivers around the world that receive flight information from ADS-B equipped aircraft and send this information to their servers in order to provide open-source live flight tracking.

In order to demonstrate each module of the data analytics framework described in this chapter, we used a sample of trajectory data for arrival flights at LaGuardia airport (LGA). The sample consists of a set of 16 days of flight tracks from the TFMS.

2.3 Clustering at Spatial Scale: Trajectory Clustering

In the first module of the framework, a trajectory clustering scheme is developed to identify spatial patterns of aircraft movement. Clustering is an unsupervised learning method that aims at identifying groups of similar observations in a dataset without prior knowledge about the existence of these groups or about how the observations are distributed among them. In the trajectory clustering problem, the goal is to find groups of similar trajectories in the spatial dimension. We define a group of spatially similar trajectories as a trajectory pattern.

In a general perspective, any clustering methodology requires:

- a data representation;
- a similarity/distance function;

- a clustering method.

The dataset representation will ensure the alignment with the clustering goals and the use of an appropriate format for the subsequent step of similarity assessment. The choice of similarity/distance function is many times driven by the data representation, but it can also be determined by the choice of clustering method, as some algorithms require a particular function to be used. In the latter case, the data representation should satisfy the similarity/distance function requirements. Let us consider the general trajectory clustering problem to exemplify the interaction between data representation and choice of similarity/distance function. The trajectory of a moving object is usually given by a time-series of positions in a two or three-dimensional space. The majority of trajectory datasets has the following characteristics:

- *Different lengths:* trajectories may have different number of data points depending on the duration of the object movement;
- *Different sampling rates:* the time interval between consecutive points may vary between trajectories (e.g., the collection of data may be performed every 1 s for one object and every 2 s for another object) or within the same trajectory (e.g., the sensor may fail to collect the data for some time period).

Because of these characteristics, the assessment of similarity between trajectories is the first challenge that arises in the trajectory clustering process. As most of the distance functions deal with vectors of the same length, one approach is to convert the variable-length trajectory data into a vector of fixed dimension by constraining the trajectory length of and using resampling mechanisms. One con of this approach is that we may lose the smoothness in the original trajectory data if trajectories vary too much in length and the choice of vector size is significantly smaller than the length of the longer trajectories. Another approach is to use more complex similarity/distance functions that can handle different lengths. For example, Vlachos et al. [106] created similarity functions based on the Longest Common Subsequence (LCS) to discover similar multidimensional trajectories. The LCS first appeared in the area of string matching and its basic idea is to find the longest common subsequence

of two sequences of different lengths by allowing them to stretch without rearranging the sequence of the elements but allowing some elements to be unmatched [106]. Zhu et al. [107] used the Fréchet distance (commonly used to measure the dissimilarity between curves) to assess the dissimilarity between trajectories of different lengths. One major drawback of using more complex distance functions is the increase in computational time.

The final requirement in the clustering methodology is the method for grouping similar observations. A number of clustering algorithms exist to perform this task and their selection is primarily determined by the characteristics of the data in analysis. In general, they can be categorized in four classes: hierarchical, partitioning, density-based and grid-based. Hierarchical algorithms create a hierarchical decomposition of the data space based on similarity/dissimilarity measures between observations/groups. This decomposition is represented by a tree, also called dendrogram, which progressively splits the database into smaller subsets from the root to the leaves (divisive approach) or merges individual subsets into groups from the leaves to the root (agglomerative approach) [108]. The clusters are then extracted using a distance threshold to horizontally cut the dendrogram. Partitioning algorithms creates partitions of the database based on the desired number of clusters. They start with an initial random partition and then iteratively optimize an objective function such as minimizing the intra-cluster distance. K-means is the most popular algorithm in the partitioning category [109]. Density-based clustering algorithms are based on the fact that clusters can be determined by high-density regions separated by low-density regions in the data space. They are highly used in spatial databases and have the advantage of handling clusters of arbitrary shape. Finally, grid-based methods divide the data space into a finite number of cells and perform the clustering based on the population of each cell.

Most of the clustering algorithms produce as output a partition of the given dataset. However, in many cases, the dataset is not entirely structured. There may be a significant portion that is unstructured and that can affect the clustering of the structured portion. Usually, this unstructured part is labeled noise or outlier, although it is not necessarily composed of observations that deviate too much from the rest of the dataset as in the common outlier definition. Therefore, the clustering analysis should take into account the existence of this unstructured part in order to produce reasonable clusters that represent

the underlying patterns of the dataset. This can be done either by applying noise detection algorithms prior the clustering or using clustering algorithms that automatically handles noise.

The data representation, distance function and clustering method used in the flight trajectory clustering problem are presented below.

2.3.1 Trajectory Representation

A data resampling approach is used for representing the flight trajectories with feature vectors of fixed dimension. The raw flight trajectory data are time-series of aircraft position sampled every minute from the origin to the destination airport. First, a filtering procedure is implemented to extract the portion of the trajectory associated with the scale of operation under analysis. For the terminal area scale, we consider the trajectory information between the airport runway threshold and the terminal area boundary, which is modeled as a circle of 60-mile radius with its center at the airport. The filtered flight trajectories are characterized by time-series of different lengths, depending on the time spent in the airspace regions under consideration. A data resampling is performed to transform each time-series into a high-dimensional feature vector of fixed dimension. The resampling approach normalizes the time stamps for each trajectory into the interval $\tau = [0, 1]$, divides τ into a fixed number of equally sized time intervals and linearly interpolates the spatial position for the fixed number of normalized time stamps in τ. The result is a feature vector of 2D spatial position evenly spaced in time:

$$F_i = (x_{i1}, y_{i1}, x_{i2}, y_{i2}, ..., x_{in}, y_{in})$$

It is worth mentioning that the flight trajectory dataset is not characterized by very large differences in length and a fair representation of the trajectory can be achieved with resampling. Figure 2-2 exemplifies the post-processed trajectory resulting from this procedure.

2.3.2 Trajectory Similarity Assessment

To allow each feature of the trajectory vector to equally contribute to the distance measure, the trajectory vectors are standardized so that each individual dimension is centered to have

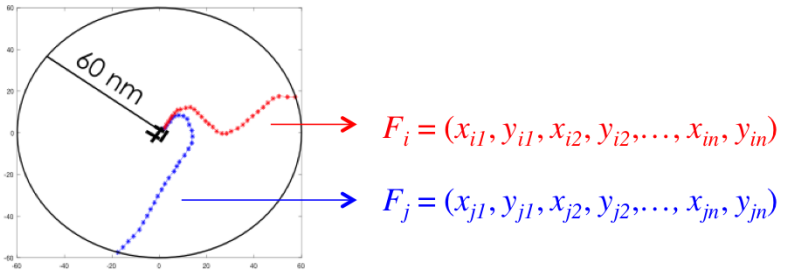


Figure 2-2: Resampled flight trajectory data.

mean 0 and scaled to have standard deviation 1, as follows:

$$\hat{x}_{in} = \frac{x_{in} - \bar{x}_n}{\sigma_n} \tag{2.1}$$

where x_{in} is the value of feature n for trajectory i, $\hat{x}_{in}$ is the value of feature n for trajectory i after standardization, $\bar{x}_n$ and σ_n are the sample mean and standard deviation of feature n.

With trajectory vectors of equal dimension and standardized, Euclidean distance is used to determine the similarity between flight trajectories:

$$d(F_i, F_j) = \|F_i - F_j\|_2 = \sqrt{\sum_{k=1}^{n}(x_{ik} - x_{jk})^2} \tag{2.2}$$

2.3.3 Trajectory Clustering Method

A density-based clustering algorithm - Density-Based Spatial Clustering of Applications with Noise (DBSCAN) [110] - is used for flight trajectory clustering. As the name of the algorithm suggests, this method is suitable for datasets with background noise. In flight trajectory datasets, the standard routes and adaptions produce the core underlying patterns, yet abnormal trajectories can also occur for a variety of reasons and can be considered noise. DBSCAN enables the identification of the core trajectory patterns in the presence of abnormal trajectory profiles. Other advantages of DBSCAN include the ability to discover non-convex clusters and no need to set the number of clusters a priori.

DBSCAN relies on two input parameters in order to cluster the data space:

- $MinPts$: a minimum number of points (observations);

- ε: a distance threshold.

The algorithm is built on three fundamental concepts (illustrated in Figure 2-3):

1. ε-neighborhood

   The ε-neighborhood of an observation F_i contains all observations that are within a distance ε and it is defined as:

$$N_\varepsilon(F_i) = \{F_j \in D / d(F_i, F_j) \leq \varepsilon, d(F_i, F_j) = \|F_i - F_j\|_2\} \tag{2.3}$$

2. Density-reachability

   An observation F_j is directly density-reachable from an observation F_i if:

   - $F_j \in N_\varepsilon(F_i)$
   - $|N_\varepsilon(F_i)| \geq MinPts$ (core point condition)

   An observation F_j is density-reachable from an observation F_i if there exists a chain $F_i, ..., F_j$ such that each observation is directly density-reachable from the predecessor.

3. Density-connectivity

   An observation F_j is density-connected to an observation F_i if there exists another observation F_k such that both F_j and F_i are density-reachable from F_k.

Based on these concepts, the clustering process works as follows. The algorithm starts with an arbitrary instance of the database and determines its ε-neighborhood. If it contains at least $MinPts$ (core point condition), the observation and its neighbors start a cluster. The ε-neighborhood of the neighbors is iteratively retrieved and the same procedure is applied until all the border points of the cluster are achieved. Otherwise, the observation is labeled as noise and the algorithm visits the next instance of the database. It is worth mentioning that an instance labeled noise can later become part of a cluster if it is reachable from a new discovered core point. A cluster is then defined as a set of density-connected points.

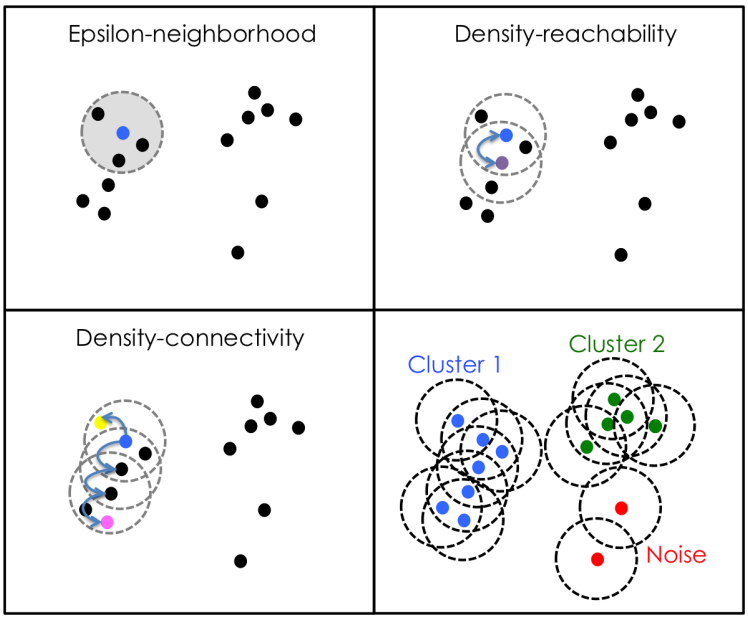


Figure 2-3: Illustration of DBSCAN concepts.

The clustering solution provided by DBSCAN is dependent on the input parameters $MinPts$ and ε. DBSCAN is not particularly sensitive to the $MinPts$ choice, which will determine the smallest number of observations to define a cluster. Once $MinPts$ is chosen, the choice of ε will involve a trade-off between undesirable creation of clusters from minor variations in a major cluster and undesirable merger of distinct clusters.

Given the absence of a correct answer for establishing a baseline for comparison, the clustering process is inherently subjective and its success is primarily determined by the meaningfulness and interpretability of structural patterns extracted from the data. As an attempt to provide a quantitative measure to evaluate the clustering solution quality, several clustering validity indices have been developed and can be found in the literature. A detailed description and comparison of some of these indices is provided in [111]. In general, these different metrics evaluate similar properties such as the compactness and separability between the clusters. A good clustering solution is usually associated with high levels of compactness and separability between the clusters. However, in a practical perspective, these indices should be primarily used as a validation mechanism rather than a tool to blindly select the number of clusters.

In this work, analysis of Silhouette plots [112] was used as a decision aid in the cluster-

ing process on a case-by-case basis. The Silhouette Index provides a quantitative measure about the correctness of assignment of an observation to a cluster. Let $a(i)$ be the average dissimilarity of observation i to the observations in the same cluster, $d(i, C)$ be the average dissimilarity of observation i to the observations in a different cluster C and $b(i)$ be the minimum $d(i, C)$, the Silhouette Index $s(i)$ of observation i is then given by Eq. 2.4. From the Silhouette Index formulation, it is possible to note that it varies between -1 and 1: values close to 1 indicate the observation is well clustered, values close to -1 indicate the observation is probably assigned to the wrong cluster.

$$s(i) = \begin{cases} 1 - a(i)/b(i) & \text{if } a(i) < b(i) \\ 0 & \text{if } a(i) = b(i) \\ b(i)/a(i) - 1 & \text{if } a(i) > b(i) \end{cases} \quad \Rightarrow \quad s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \tag{2.4}$$

2.3.4 Example Application

This section shows the results of the application of the trajectory clustering methodology to the sample data of LGA arrival trajectories.

The sensitivity of the DBSCAN solution with respect to the parameters $MinPts$ and ε is shown in Figure 2-4. It is noticed the percentage of noise observations is not particularly sensitive to $MinPts$, but very sensitive to ε. As ε increases, more observations are captured in the clusters.

The average Silhouette Index for the clustering output obtained with different parameter settings is shown in Figure 2-5. For smaller values of ε, the value of the metric is closer to zero, indicating a lower clustering solution quality. Figure 2-6 suggests this is driven by the creation of clusters from minor variations in the data, as revealed by the high number of clusters when ε is small. As ε increases, the Silhouette Index increases and starts to stabilize. Between $\varepsilon = 1$ and $\varepsilon = 1.2$, the metric stabilizes at its highest level. After that point, increasing ε generates a small degradation in the clustering solution quality. Figure 2-6 reveals this degradation is caused by merger of distinct clusters.

Based on these observations, we used $MinPts = 6$ and $\varepsilon = 1.2$ to cluster the sample data. The clustering output for the chosen parameters as well as the detailed Silhouette

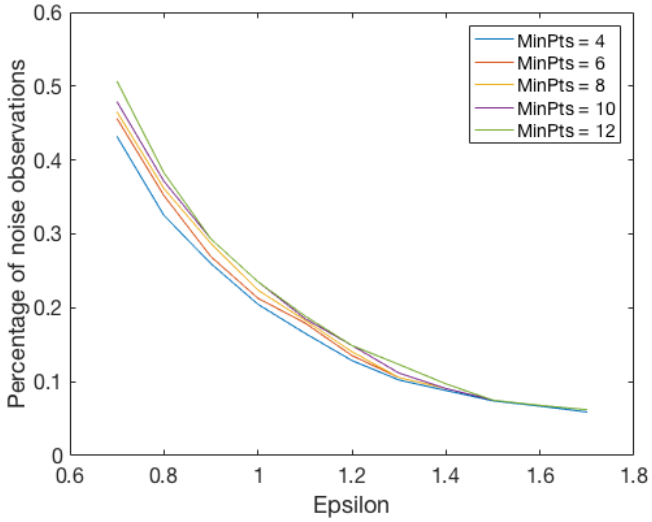


Figure 2-4: Percentage of noise associated with the clustering output obtained with different parameter settings ($MinPts$, ε).

values for each observation in each cluster are shown in Figure 2-7.

2.4 Trajectory Classification

In the second module, a trajectory classification scheme is developed to match new flight trajectories with the airspace structure identified in the first module, including the ability to identify non-conforming behaviors (those that do not conform with one of the identified trajectory patterns). The purpose of the trajectory classification module is two-fold: 1) given the computational effort associated with the trajectory clustering process, the classification procedure provides a more efficient way for processing large sets of trajectory data for spatial pattern identification during offline applications; 2) it is key for online applications that depend on monitoring of airspace use, as it enables consistent processing of new batches of trajectory data continuously generated by the surveillance system.

In the trajectory classification problem, the classes are defined by the trajectory patterns learned with the spatial clustering analysis (training dataset). Given a new flight trajectory defined by its vector of spatial position, the goal is to predict whether it conforms to one of the learned trajectory patterns. Obviously, not all flight trajectories will match the identified airspace structure and the classification scheme should also have the ability to detect these

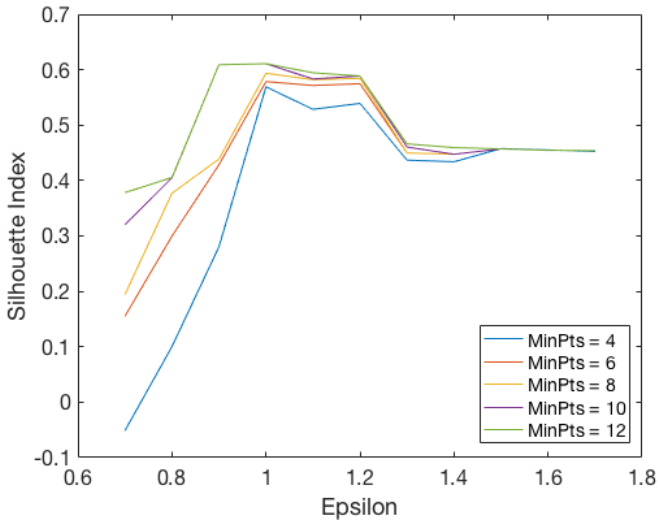


Figure 2-5: Silhouette Index associated with the clustering output obtained with different parameter settings ($MinPts$, ε).

non-conforming behaviors.

Differently than clustering, classification is a supervised learning method that aims at predicting the class/label of an observation given a set of predictors/features based on prior knowledge extracted from a training dataset [113]. Figure 2-8 provides a schematic representation of the supervised learning classification process. Given a training set $S_n = \{(x_i, y_i), i = 1, ..., n\}$, where $x_i \in \mathbb{R}^d$ is a feature (input) vector and $y_i \in \{1, 2, 3, ..., C\}$ is the corresponding label (output) in a multi-way classification problem with C labels, a classifier h is defined as a mapping $h : \mathbb{R}^d \to \{1, 2, 3, ..., C\}$. The classifier outputs a decision boundary (linear or not) that partitions the input space and can therefore classify new instances in a test set.

A fair classifier is expected to replicate its performance in the training set to the test set and therefore shows both low training (Eq. 2.5) and generalization errors (Eq. 2.6). In order to achieve these objectives and avoid over-fitting (when a low training error is obtained but it does not translate into a low generalization error), cross-validation should be performed during the training phase in order to adjust the settings of the classifier. In this work, we used the k-fold cross-validation method, with $k = 5$, for this purpose. It consists of partitioning the training dataset into k different folds and using $(k - 1)$ folds to train the classifier and one fold to evaluate its performance in k different runs (each run uses a different fold for validation).

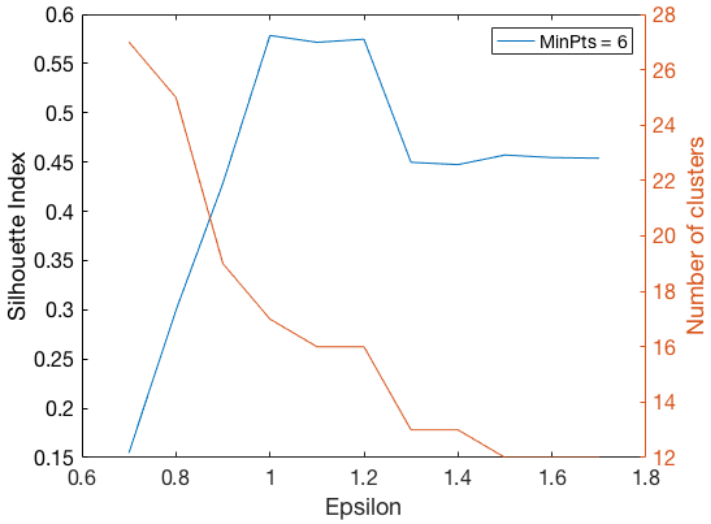


Figure 2-6: Silhouette Index versus number of clusters identified.

$$\varepsilon(h) = \frac{1}{n}\sum_{i=1}^{n}[\![h(x_i) \neq y_i]\!] \tag{2.5}$$

$$\varepsilon'(h) = \frac{1}{n'}\sum_{i=n+1}^{n+n'}[\![h(x_i) \neq y_i]\!], n' \to \infty \tag{2.6}$$

2.4.1 Matching Flight Trajectories with Learned Trajectory Patterns

The trajectory classification process starts with the development of a classifier that can successfully discriminate the learned trajectory patterns. The dataset of learned trajectory patterns was partitioned into 80%/20% subsets for training and testing the classifier.

Multi-way Trajectory Classification Model

A multi-way classifier using the Random Forests algorithm was created. Random Forests was selected because it is widely recognized as one of the highest performing methods in terms of classification accuracy and because it runs efficiently on large databases.

Random Forests is a non-parametric method for both classification and regression consisting of an ensemble of decision tree learners [114]. A decision tree is defined by a hierarchical structure of decision nodes that progressively partitions the input space using a sequence of

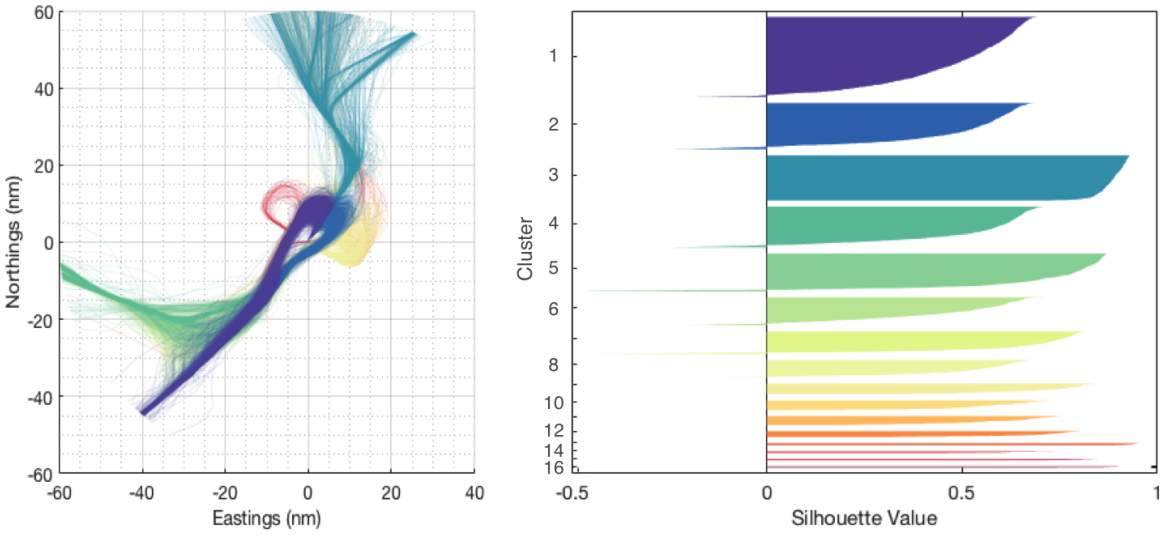


Figure 2-7: (a) Clusters of arrival trajectories. (b) Silhouette plot.

binary splits from the root node to the leaves (terminal nodes). Each decision node evaluates the value of a feature from the input set to decide which branch will be taken from it. The goodness of a split is quantified with an impurity measure. In a classification setting, the Gini index is typically used. Given a training set with n observations and n_k observations in a decision node k after a sequence of splits, with $n_{k,c}$ observations belonging to class c such that $\sum_c n_{k,c} = n_k$, the Gini index measures the impurity of node k as follows:

$$G_k = 1 - \sum_c p_{k,c}^2 \tag{2.7}$$

where $p_{k,c} = n_{k,c}/n_k$ is an estimate for the probability of belonging to class c.

It follows that the minimum value of G_k is zero and it is achieved when the node is pure, whereas the maximum value of G_k is $1 - 1/C$ and it is achieved when the classes $c = 1, ..., C$ have equal probabilities. At each decision node, the method searches for the partition that will minimize the impurity among all input variables and their potential splits. The splitting process is repeated until a stopping criterion is met (e.g., node is pure, minimum number of observations is achieved). The result is a rectangular partition of the input space. Every leaf node l then corresponds to a rectangular subspace R_l. For every training observation x_i, there is only one leaf node such that $x_i \in R_l$. The prediction for a new observation x' is

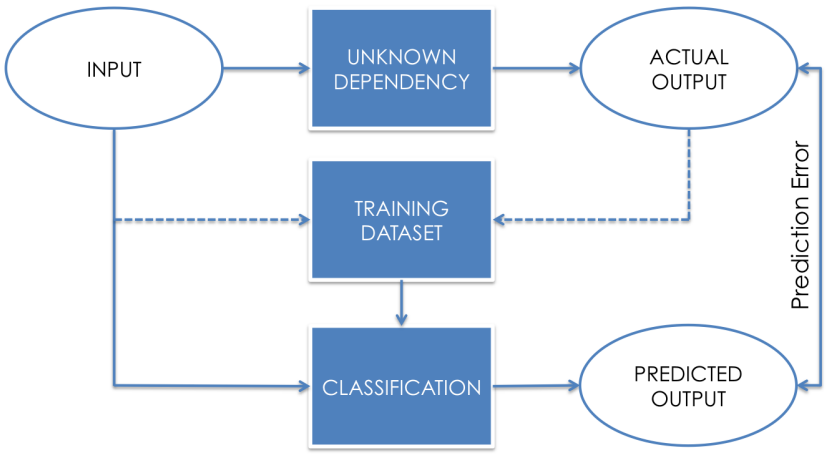


Figure 2-8: Supervised learning classification process.

obtained by applying a majority rule for the leaf $l(x')$ that this observation falls into when it is passed through the tree:

$$\hat{y}' = \arg\max_{c} \sum_{i:x_i \in l(x')} I(c = y_i) \tag{2.8}$$

Figure 2-9 illustrates a classification tree, based on a toy dataset characterized by a two-dimensional input space and two labels. In this fairly simple example, the resulting decision boundary (solid black line) perfectly partitions the input space into local regions, and can therefore classify new instances.

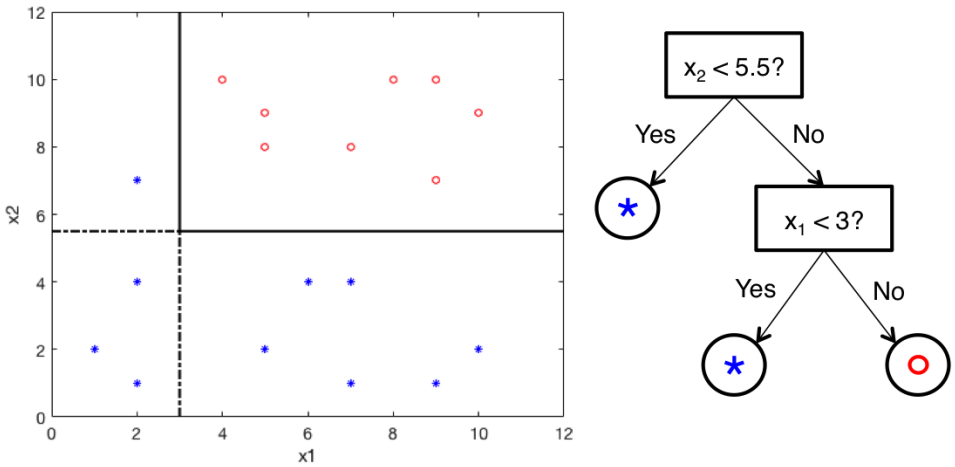


Figure 2-9: Example of dataset and resulting decision tree for classification.

As any ensemble method, Random Forests creates multiple decision trees with bootstrap

samples of the training data, but it has the unique characteristic of selecting only a random subset of the features to determine each split. The final response of the model is computed by aggregating the results of the individual trees:

$$\hat{y}'_{bag} = \arg\max_{c} \sum_{t=1}^{T} I(c = \hat{y}'_t) \tag{2.9}$$

where $\hat{y}'_t$ is the prediction from tree t in the ensemble.

The algorithm has few parameters to tune: T is the number of trees in the ensemble, M is the number of features to sample at each split and N is the minimum number of observations in a leaf node. Based on cross-validation performance, the Random Forests classifier was created with $T = 50$, $M = 10$ and $N = 1$.

Multi-way Trajectory Classification Performance

The Random Forests trajectory classification accuracy achieved on the test data of LGA arrival trajectories was 99%, revealing that the classifier is able to precisely distinguish the spatial patterns learned with trajectory clustering.

2.4.2 Identifying Non-Conforming Trajectories

The Random Forests classifier successfully discriminates the spatial patterns between the conforming trajectories. However, given a new flight trajectory, it is not known a priori whether it conforms to the standard routes. In order to enable the identification of non-conforming behaviors, we developed a Conformal Prediction semi-supervised anomaly detection framework.

Anomaly detection is an important problem that has been researched within various areas and application domains. Chandola et al. [115] defines an anomaly as "a pattern that does not conform to expected normal behavior". Anomaly detection approaches typically operate in three settings: supervised, semi-supervised and unsupervised. In supervised mode, a training dataset with representative instances of normal and anomalous behavior is used to create a predictive model that can discriminate between them. In semi-supervised mode, representative instances of normal behavior are used for creating a model for the normal

data, which is then used for identifying anomalies. Finally, in unsupervised mode, a training dataset is not required; techniques in this category make the implicit assumption that normal instances are much more frequent than anomalies.

We assume that the trajectory patterns identified with the spatial clustering analysis are representative of the normal trajectory behavior in the airspace region analyzed. The trajectory classifier developed in the previous section models the normal behavior. Therefore, it can also be used to identify non-conforming behaviors, if a measure of confidence for its predictions can be obtained. Conformal Prediction provides such measure rigorously, as described below.

Non-conforming Trajectory Detection Model

Conformal Prediction is a technique that intends to generate confidence intervals for predictions made by machine learning algorithms [116]. Therefore, it is built upon existing classification and regression methods. The concept involves the estimation of a p-value for each candidate label that can be attributed to a test example based on a non-conformity metric, which measures how different this observation is relative to a set of examples that have the candidate label. The labels having a p-value less than the specified significance level ε are then excluded from the prediction set.

Formally, given a training dataset $Z : (x_1, y_1), ..., (x_{n-1}, y_{n-1})$, where $x_i \in X$ is a vector of attributes and $y_i \in Y$ is the class label for observation i, non-conformity scores α_i can be calculated for each observation based on non-conformity functions. These non-conformity functions are typically tailored to the machine learning algorithm used to create the mapping between the input space and the output. Since Random Forests was the underlying classification method for conformal prediction, a non-conformity score based on the proportion of tree votes in the ensemble was used [117, 118, 119]:

$$\alpha_i = 1 - \frac{|tree_j \in ensemble : y_i = \hat{y}_i^j|}{|ensemble|} \tag{2.10}$$

For a new observation $z_n = (x_n, \hat{y}_n)$ for which we want to predict the class label, a non-conformity score $\alpha_n^{\hat{y}_n}$ can be calculated for each tentative class label $\hat{y}_n$ and compared with

the non-conformity scores of the training dataset in order to generate an associated p-value:

$$p(\hat{y}_n) = \frac{|z_i \in Z : \alpha_i \geq \alpha_n^{\hat{y}_n}| + 1}{|Z| + 1} \tag{2.11}$$

The p-value is a measure of confidence in the prediction. The higher the p-value, the higher the confidence. For a significance level ε, if $p \geq \varepsilon$, $\hat{y}_n$ becomes part of the prediction set with confidence $1 - \varepsilon$.

Non-conforming flight trajectories can be identified when the p-value is lower than the significance level for all possible labels, i.e., when the prediction set is empty. In contrast to previous methods, a key property of conformal anomaly detection is that it provides a well-founded approach for tuning the confidence thresholds for anomaly detection that can be directly related to the expected or desired false alarm rate [120].

Non-conforming Trajectory Detection Performance

An anomaly detection framework should seek to achieve a low rate of false anomaly discovery (false alarm rate) and a high rate of anomaly detection (true positive rate). Ideally, we would like to have a dataset of labeled non-conforming arrival trajectories to evaluate the performance of our model not only in terms of false alarm rate but also true positive rate. In the absence of this labeled data, we used the noise trajectories identified by DBSCAN during the spatial clustering process, as the noise typically includes the non-conforming behaviors.

We compared the performance of the Conformal Prediction (CP) non-conforming trajectory detection approach with two other approaches commonly used for anomaly detection. The first one is a nearest neighbor based approach. A classifier is created to discriminate the trajectory patterns using the K-Nearest Neighbors (KNN) method. For a new trajectory i, the KNN based model retrieves the k-closest neighbors on the training dataset of trajectory patterns using Euclidean distance and computes the posterior probability of belonging to trajectory pattern j as follows:

$$p(j/i) = \frac{\sum_{m \in M(i)} \pi_j [\![y_m = j]\!]}{\sum_{j=1}^{K} \sum_{m \in M(i)} \pi_j [\![y_m = j]\!]} \tag{2.12}$$

where $M(i)$ is the set of k nearest neighbors of trajectory i, π_j is the prior probability of

trajectory pattern j and y_m is the known trajectory pattern of neighbor observation m.

A non-conforming trajectory is then identified when the maximum posterior probability is smaller than an user-specified threshold.

The second is a statistical approach. We model the learned trajectory patterns using a Gaussian Mixture Model (GMM), where each component describes one trajectory pattern. GMM is a parametric model that assumes the probability density function of the data is given by a weighted sum of Gaussian component densities:

$$P(x;\theta) = \sum_{j=1}^{K} \pi_j \mathcal{N}(x;\mu_j,\Sigma_j) \tag{2.13}$$

where x is the set of feature vectors in the training dataset, K is the number of components in the mixture, π_j are the component weights and $\mathcal{N}(x;\mu_j,\Sigma_j)$ are the component Gaussian densities, with mean μ_j and covariance matrix Σ_j.

The posterior probability that a new observation (trajectory) i has been generated from component (trajectory pattern) j is given by:

$$p(j/i) = \frac{\pi_j \mathcal{N}(x^{(i)};\mu_j,\Sigma_j)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x^{(i)};\mu_j,\Sigma_j)} \tag{2.14}$$

As in the KNN based approach, a non-conforming trajectory is identified when the maximum posterior probability is smaller than an user-specified threshold.

We evaluated the performance of the three non-conforming trajectory detection models on the test data of LGA arrival trajectories. We used $k = 30$ in the KNN model, based on cross-validation performance. The GMM model was created with 16 components (number of learned trajectory patterns). In order to detect non-conforming trajectories, a significance level $\varepsilon = 0.05$ was used in the CP model. Since the significance level is directly related to the expected false alarm rate in CP, we calibrated the thresholds for non-conforming trajectory detection in the KNN and GMM models in order to achieve false alarm rates close to 5%. The following performance metrics were calculated and compared:

True positive rate - TPR (or sensitivity or recall)

$$TPR = \frac{tp}{tp + fn} \tag{2.15}$$

where tp is the number of true positives (actual non-conforming trajectories predicted as non-conforming) and fn is the number of false negatives (actual non-conforming trajectories predicted as conforming).

False positive rate - FPR

$$FPR = \frac{fp}{fp + tn} \tag{2.16}$$

where fp is the number of false positives (actual conforming trajectories predicted as non-conforming) and tn is the number of true negatives (actual conforming trajectories predicted as conforming).

Positive predictive value - PPV (or precision)

$$PPV = \frac{tp}{tp + fp} \tag{2.17}$$

Accuracy - ACC

$$ACC = \frac{tp + tn}{tp + tn + fp + fn} \tag{2.18}$$

F1-score - F1

$$F1 = 2 * \frac{PPV * TPR}{PPV + TPR} \tag{2.19}$$

The performance results are shown in Table 2.1. First, it is observed the CP model with significance level $\varepsilon = 0.05$ indeed led to FPR $< 5\%$, as expected. For similar levels of FPR, the CP model outperformed the others in detecting non-conforming trajectories, as it showed the highest TPR. As a result, the CP model also showed the highest precision and accuracy. In summary, the CP model achieved the best balance between precision and sensitivity, as measured by the F1 score, and therefore was selected for non-conforming trajectory detection.

Table 2.1: Non-conforming trajectory detection performance.

Model	**CP**	**KNN**	**GMM**
TPR(%)	69.5	49.4	17.3
FPR(%)	4.3	5.0	5.0
PPV(%)	73.0	61.4	35.6
ACC(%)	92.1	88.6	84.1
F1(%)	71.1	54.7	23.2

2.4.3 Flow Identification

The trajectory classification scheme enables the processing of new sets of data for spatial pattern identification. Once trajectories are classified, flows can be identified as temporally associated flight trajectories conforming to the same standard route. For this, the output of the trajectory classification is organized and stored in daily flow matrices $W \in \mathbb{R}^{n \times p}$, with n rows (number of trajectory patterns) and p columns (number of time periods during the day). Each matrix element w_{ij} indicates the number of trajectories conforming to trajectory pattern i during time period j. A flow is identified whenever $w_{ij} > 1$. The matrix columns are vectors $\{R_t \in \mathbb{R}^n, t = 1, ..., p\}$ that indicate the time-dependent traffic flows. An example of daily flow matrix is shown in Figure 2-10. For this particular date, two major changes were observed in the arrival traffic flow structure at LGA.

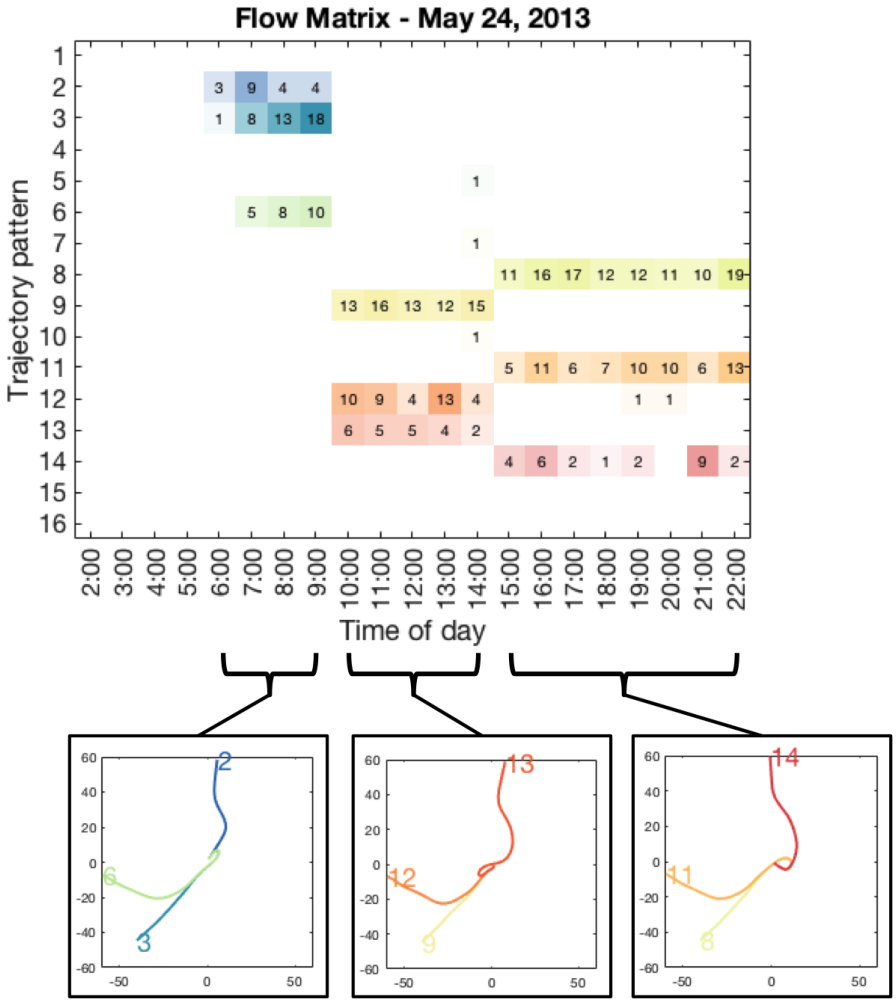


Figure 2-10: Flow matrix for LGA arrivals on May 24, 2013.

2.5 Clustering at Temporal Scale: Time-Dependent Flow Vector Clustering

The third module in the data-driven framework has the goal of identifying patterns in the traffic flow structure (group of traffic flows in a given time period). For this, a second-layer clustering analysis is performed on the set of time-dependent traffic flows R_t. The vector of time-dependent traffic flows is transformed to a time-dependent characteristic flow vector FV_t to represent the traffic flow structure:

$$FV_t = f(R_t)$$

The flow vectors are clustered to identify traffic flow patterns. Since the flow vector representation is tailored to the application, the details about the clustering process are presented in the next chapter.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 3

Comparative Analysis of Terminal Area Operations in Multi-Airport Systems

In this chapter, we demonstrate how the flight trajectory data analytics framework for characterization of air traffic flows developed in this thesis can be used to obtain a deeper understanding of metroplex operational behavior and performance based on a systematic approach backed by operational data. For this, we performed a comparative analysis of terminal area operations in three representative multi-airport systems of the global air transportation system: New York, Hong Kong and Sao Paulo. Through the characterization of terminal area air traffic flows, we identified structural and operational differences between these systems and evaluated the impacts of such differences in the key ATM performance areas of efficiency, capacity and predictability.

3.1 Case Studies

For the purpose of this research, a *multi-airport system* was defined as a set of two or more *significant* airports that serve commercial passenger traffic in a metropolitan region. An airport was defined as *significant* if it serves at least 5,000,000 passenger movements per year and if it is located within 50 miles from the center of the major city in the metropolitan region.

Three relevant multi-airport systems of different regions of the world were selected for

the comparative analysis: New York, Hong Kong and Sao Paulo. All three metropolitan regions are served by three significant commercial airports: New York is served by John F. Kennedy International (JFK), Newark International (EWR) and LaGuardia (LGA); Hong Kong is served by Hong Kong International (HKG), Shenzhen Bao'an International (SZX) and Macau International (MFM); Sao Paulo is served by Sao Paulo/Guarulhos International (GRU), Sao Paulo/Congonhas (CGH) and Viracopos International (VCP). Figure 3-1 shows the location of the significant airports in each metroplex and the distances between them.

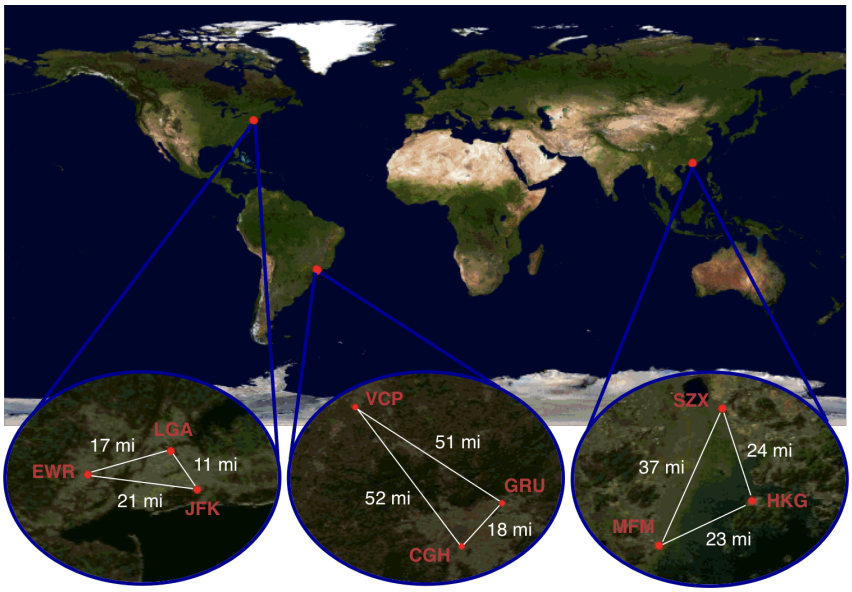


Figure 3-1: Location of airports in the New York, Hong Kong and Sao Paulo metroplexes.

The detailed passenger and aircraft movement statistics for the year 2016 [121, 122, 123, 124, 125, 126, 127] at these three multi-airport systems is presented in Figures 3-2 and 3-3. The New York system is the busiest, followed by Hong Kong and Sao Paulo. The Hong Kong system has the highest ratio between passenger movements and aircraft movements, indicating higher participation of larger aircraft in its mix of operations.

Figure 3-4 shows the lateral trajectories of arrival and departure flights during one day of terminal area operations in these regions and provides a good illustration of the air traffic density and complexity that characterize metroplex systems.

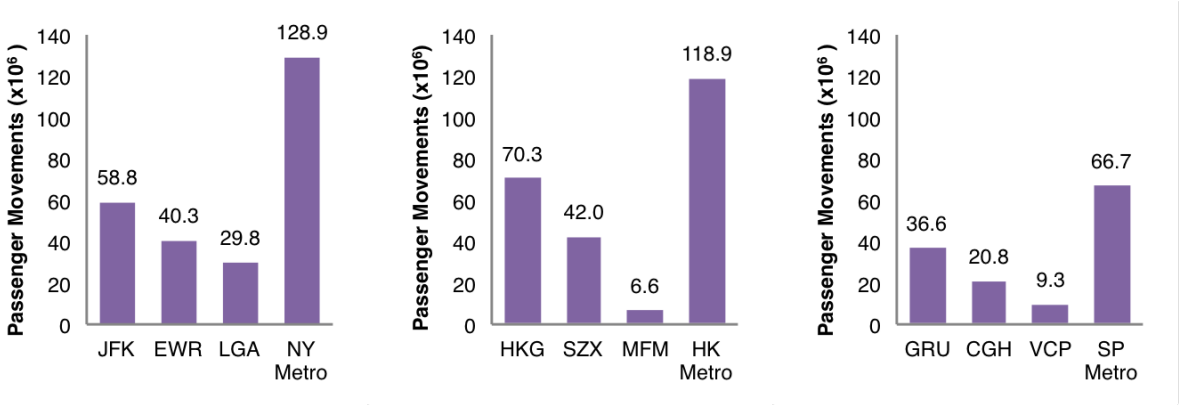


Figure 3-2: Passenger movement by airport at the New York, Hong Kong and Sao Paulo metroplexes in 2016.

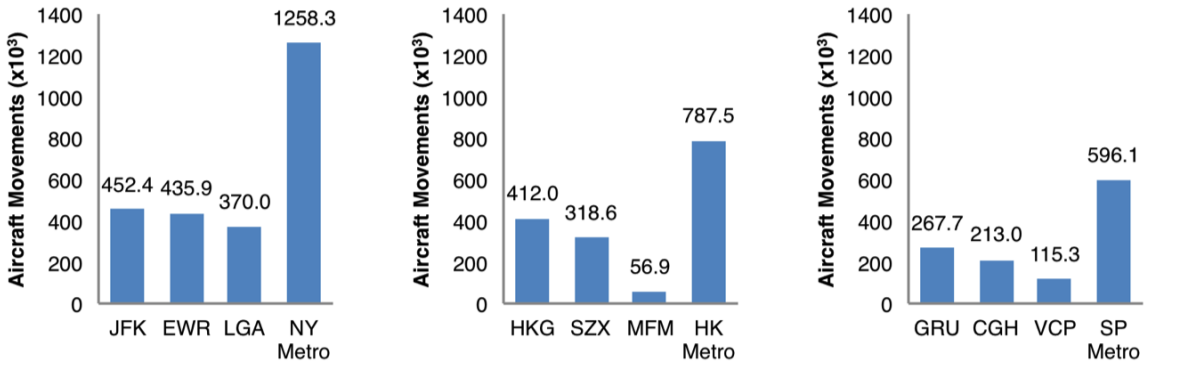


Figure 3-3: Aircraft movement by airport at the New York, Hong Kong and Sao Paulo metroplexes in 2016.

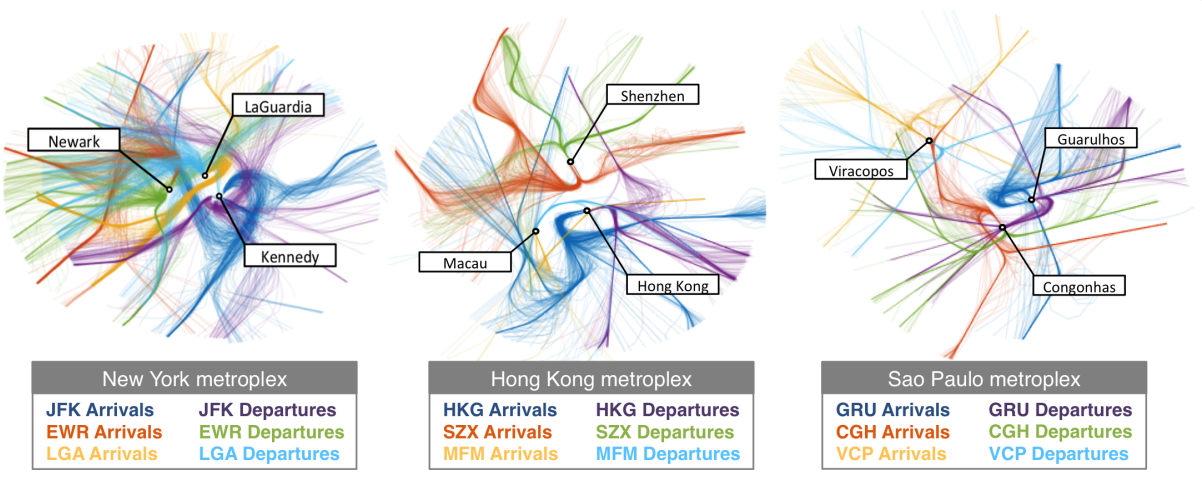


Figure 3-4: Arrival and departure trajectories for one day of terminal area operations in the New York, Hong Kong and Sao Paulo multi-airport systems.

3.1.1 New York Multi-Airport System

The multi-airport system that serves the New York metropolitan region is composed of three primary [1] commercial airports: JFK, EWR and LGA. Together, they served 128.9 million passengers in 2016, being the world's second busiest multi-airport system, after London [128]. Given the high traffic volume (which is accentuated by the presence of other smaller airports serving the demand for business and charter flights such as Teterboro airport) and the proximity between the airports, the New York metroplex is often recognized as one of the most operationally complex multi-airport systems in the world.

ATC services are provided to the major commercial airports by the individual Air Traffic Control Towers (ATCT) and one consolidated TRACON (N90), which also provides services to other smaller airports in the New York metropolitan region. ATCT and TRACON rely on "Letters of Agreement" to use specific operating procedures and delegate airspace regions that need to be shared [2]. As an example, Figure 3-5 shows some airspace regions that are shared between JFK and LGA.

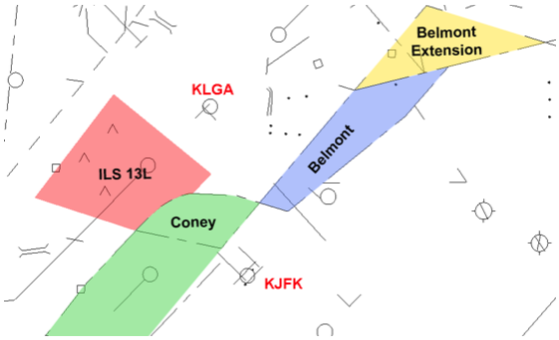


Figure 3-5: Shared airspace regions between JFK and LGA [2].

In terms of runway infrastructure, the New York metroplex has the highest installed capacity among the three systems, with a total of nine runways. JFK has two pairs of parallel runways - 4L-22R, 4R-22L, 13L-31R and 13R-31L - aligned at right angles (Figure 3-6). Parallel runways 13-31 are 6,700 ft apart and parallel runways 4-22 are 3,000 ft apart. While runways 13-31 meet the geometry criteria for independent approaches [2], they are often

[1] A primary airport was defined as a significant airport serving at least 20% of the total passenger traffic in the multi-airport system, whereas a secondary airport was defined as a significant airport with less than 20% of traffic share.

[2] In order to operate simultaneous independent instrument parallel approaches, FAA Order JO 7210.3Z

operated with a 2-mile stagger in order to avoid the need of additional final monitor staff [130]. Runways 4-22 are often operated with a 1-mile stagger.

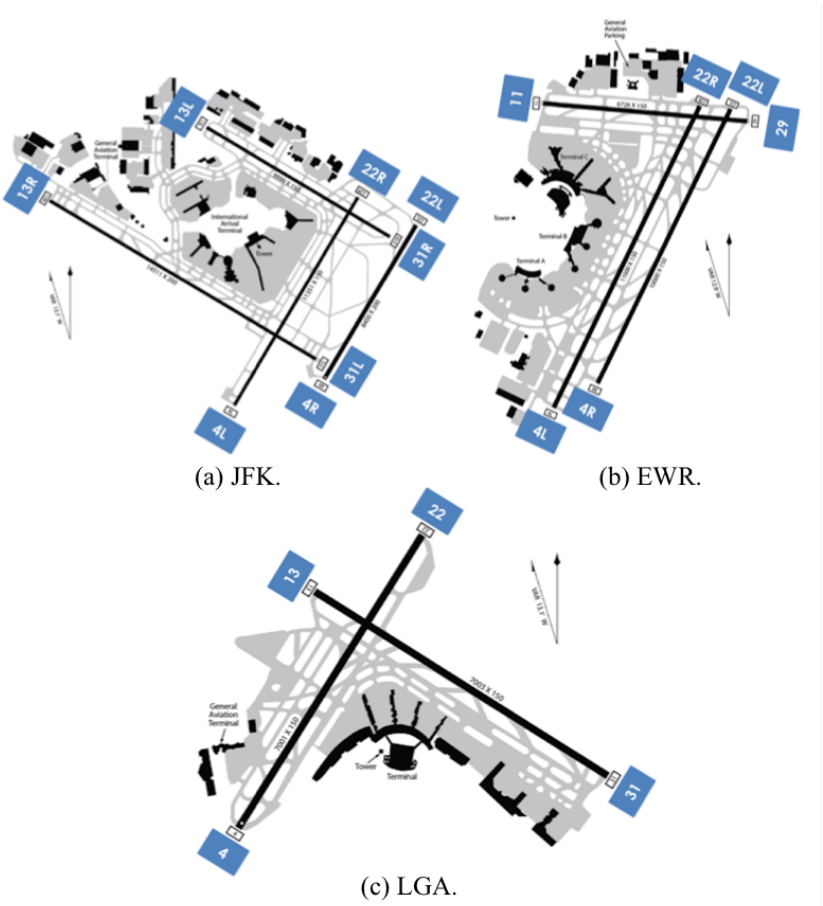


Figure 3-6: Runway infrastructure in the New York metroplex.

EWR has two parallel runways, 4R-22L and 4L-22R, and a third crosswind runway, 11-29. The parallel runways are separated by 950 ft. Because of the small separation, they are often operated in segregated mode, primarily using runway 4R-22L for landings and 4L-22R for departures. Runway 11-29 is often used as an overflow arrival runway during high demand

requires a minimum separation of 3,600 ft between runway centerlines (if the airport field elevation is 2,000 ft MSL or less) and the use of a final monitor controller for each runway operating independently [129]. Simultaneous independent approaches may be conducted without final monitors if the runway centerlines are separated by at least 9,000 ft. If parallel approaches cannot be executed independently, simultaneous dependent operations may take place. FAA Order JO 7110.65W [129]) requires the use of diagonal separations between each aircraft and the trailing aircraft in the adjacent runway in order to ensure aircraft remain staggered on parallel approaches and reduce the risk of collision from deviation from the final approach path. The diagonal separations are defined based on the distance between the runway centerlines as follows: 1.0 NM, if 2500 ft <= distance <= 3600 ft; 1.5 NM, if 3600 ft < distance <= 4300 ft ; 2.0 NM, if 4300 ft < distance <= 9000 ft.

periods, but it also provides flexibility during adverse high wind conditions.

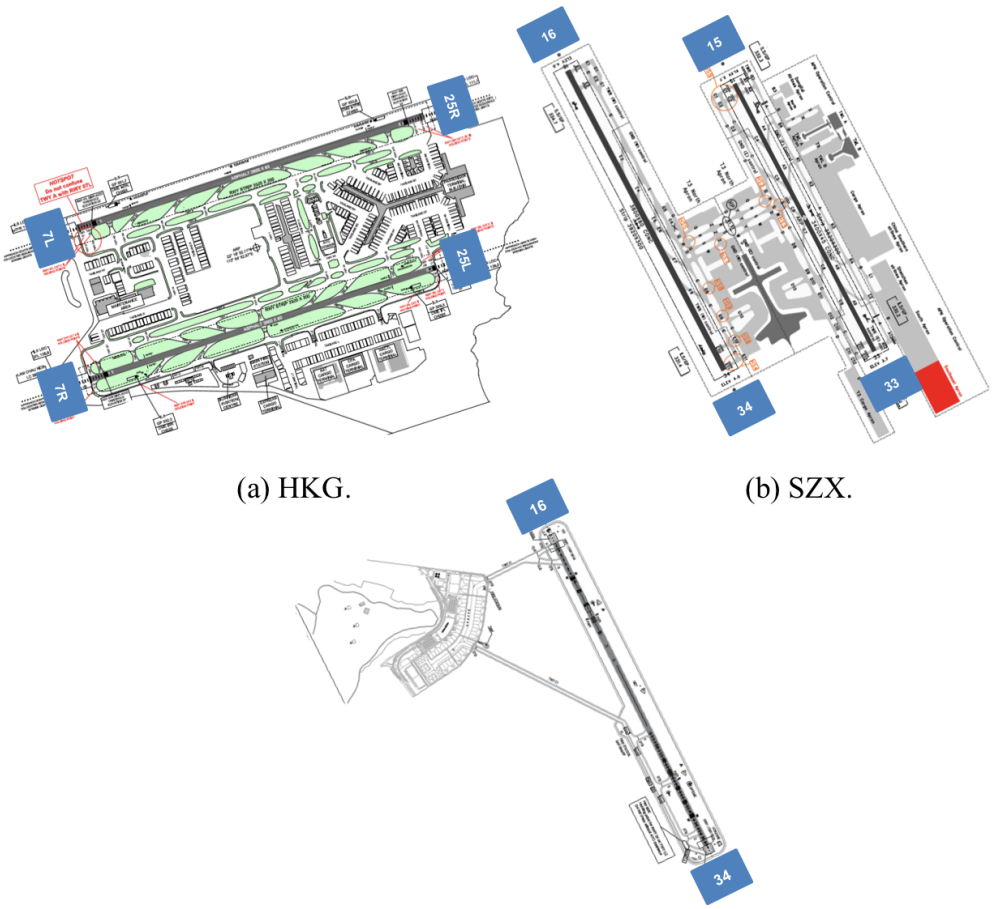
Finally, LGA has two crossing runways, 4-22 and 13-31, and typically dedicates one runway for landings and one runway for takeoffs. Standard operating procedures dictate that a departing aircraft can be cleared for takeoff after a preceding arriving aircraft crosses the intersection after landing [129]. For that reason, configuration 22-13 is optimal for LGA since the intersection is closer to the active runway thresholds, leading to the minimum spacing to get departures out.

3.1.2 Hong Kong Multi-Airport System

Hong Kong is part of the Pearl River Delta metropolitan region and is served by two primary commercial airports, HKG and SZX, and a secondary airport, MFM. As a major gateway to China and Asia, HKG is one of the top-ten international airports in terms of passenger movement and it is the world's largest air cargo hub [128]. Together with SZX and MFM, the Hong Kong multi-airport system served more than 100 million passengers in 2016 and it is one of the busiest systems in the world.

Approach control services for HKG and MFM are provided by the Hong Kong Air Traffic Control Centre [131], which is part of the Civil Aviation Department of the Hong Kong Special Administrative Region. For SZX, approach control is provided by the Guangzhou Air Traffic Control Centre, under authority of the Civil Aviation Administration of China.

The airport runway systems are shown in Figure 3-7 and totalize five runways for the metroplex. HKG has two parallel runways, 07L-25R and 07R-25L, separated by 5,000 ft. SZX also has two parallel runways, 15-33 and 16-34, separated by 5,000 ft. The large distance between the parallel runway centerlines at HKG and SZX enable them to be operated independently. MFM is served by only one runway, 16-34.



(a) HKG. (b) SZX.

(c) MFM.

Figure 3-7: Runway infrastructure in the Hong Kong metroplex.

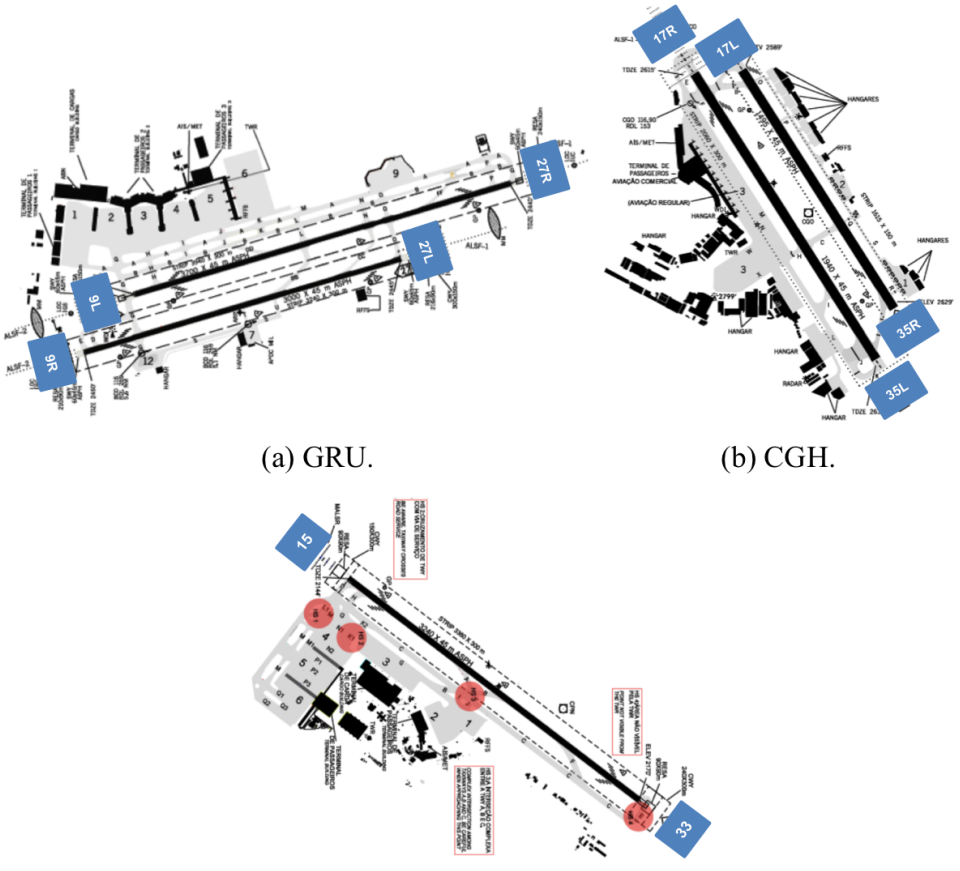
3.1.3 Sao Paulo Multi-Airport System

The multi-airport system that serves the Sao Paulo metropolitan region is composed of two primary commercial airports, GRU and CGH, and a secondary airport, VCP. Although VCP is located 50 miles from the Sao Paulo city center, today it serves an important part of the metropolitan region air travel demand given the physical limitations for capacity expansion at the primary airports. The Sao Paulo metroplex is the busiest multi-airport system in Latin America, with GRU and CGH being the busiest airports in Brazil in terms of passenger movement.

Approach control services for GRU, CGH and VCP is provided by the Regional Service of Flight Protection of Sao Paulo (SRPV-SP) through a consolidated Terminal Maneuvering

Area (TMA-SP).

Similarly to the Hong Kong metroplex, the airport runway systems totalize five runways and are shown in Figure 3-8. GRU has two parallel runways, 09L-27R and 09R-27L, which are separated by 1230 ft. The runways are typically operated dependently in segregated mode, with runway 09R-27L dedicated for arrivals and runway 09L-27R dedicated for departures. CGH has two parallel runways, 17L-35R and 17R-35L, separated by 700 ft. It typically operates in a single runway configuration using the main runway 17R-35L, given the small separation and length constraints of the auxiliary runway 17L-35R. VCP is served by only one runway, 15-33.



(a) GRU. (b) CGH.

(c) VCP.

Figure 3-8: Runway infrastructure in the Sao Paulo metroplex.

3.2 Data Description

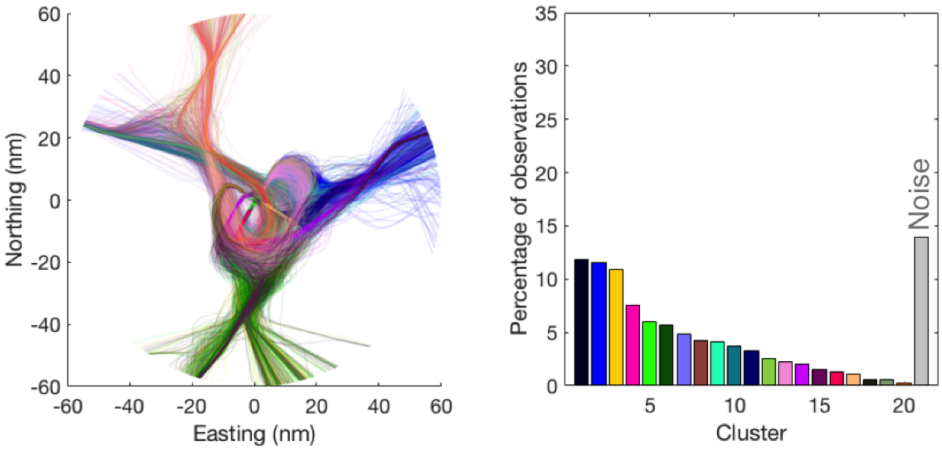
Flight tracking data was used for the identification and characterization of air traffic flows. For the New York metroplex region, we used a set of 69 days of radar tracks, within the period 2013-2015, from the TFMS. For both the Hong Kong and Sao Paulo metroplex regions, we used a set of 60 days of flight tracks, within the period 2016-2017, from the *FlightRadar24* global flight tracking service [132]. The datasets report one-minute updates of aircraft state, including flight ID, latitude, longitude, altitude, speed, origin airport, destination airport and aircraft type. Dates were empirically selected to ensure coverage of a broad set of operational conditions; 55% are fair weather condition days and 45% are weather-impacted days (further categorized as convective weather impacts - 30% - and non-convective weather impacts, such as adverse wind conditions or low ceiling/visibility - 15%). Historical Meteorological Terminal Aviation Routine Weather Report (METAR) data was used to categorize the weather conditions for each day of operations.

3.3 Characterization of Structural Differences through the Analysis of Terminal Area Route Structures

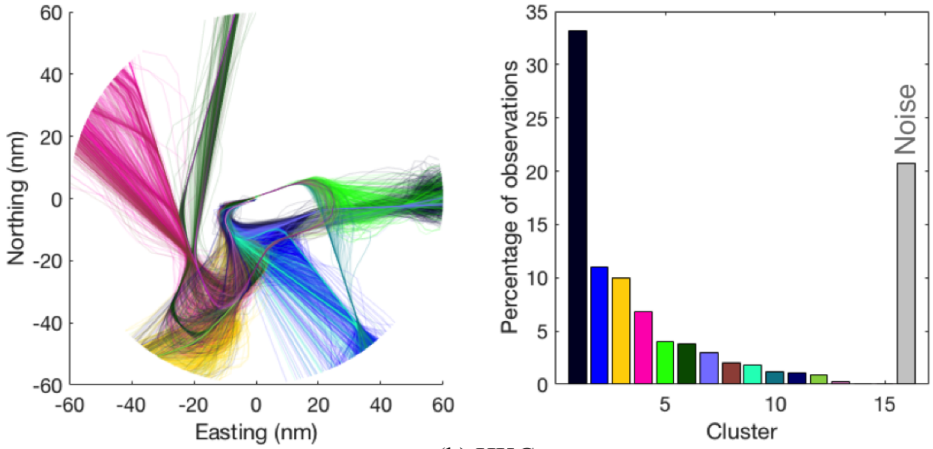
3.3.1 Identification of Trajectory Patterns

The trajectory clustering methodology described in Section 2.3 was used to identify spatial trajectory patterns in the terminal area. A subset of the data was used to reduce the computational effort. For the New York metroplex airports, a set of sixteen days was used. For the Hong Kong and Sao Paulo metroplex airports, given their lower number of operations, a set of thirty days was used. As an example, Figure 3-9 shows the identified arrival clusters for the major airport in each metroplex. The distribution of observations in the clusters is also presented. For instance, it is observed that both HKG and GRU have one dominant trajectory pattern that concentrates more than 30% of the observations, while JFK has a lower concentration of trajectories in one particular pattern. It is also observed that JFK has a lower percentage of noise observations than HKG and GRU. The clustering results for

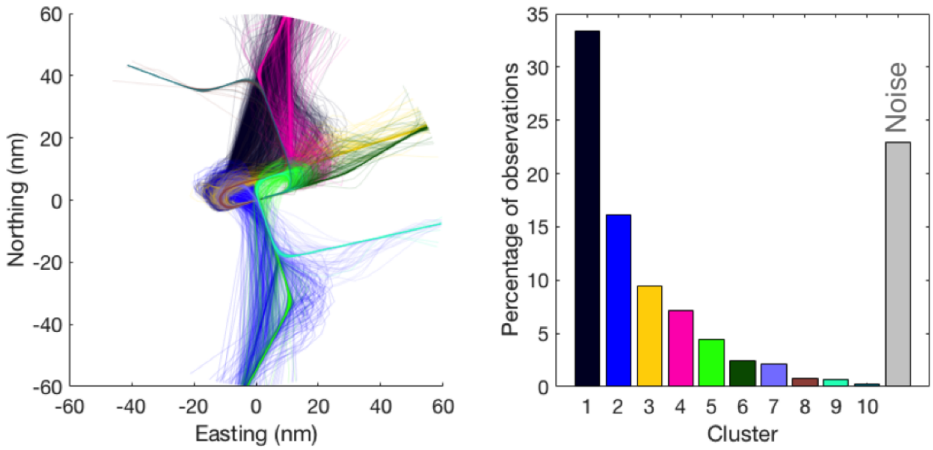
the other airports are presented in Appendix A, B and C.



(a) JFK.



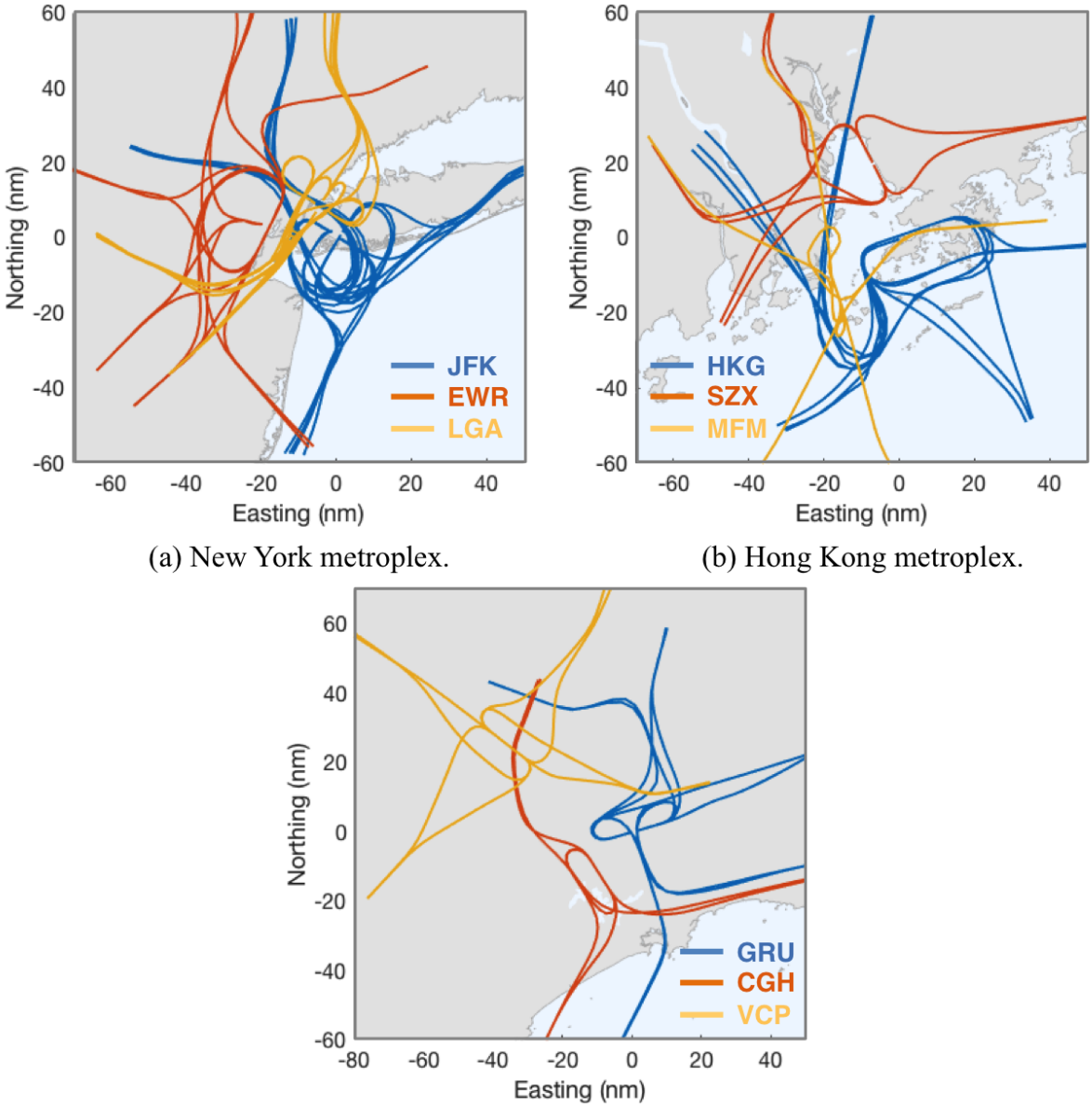
(b) HKG.



(c) GRU.

Figure 3-9: Clusters of arrival trajectories; each color represents one cluster; grey bar represents the percentage of noise observations.

A total of 50, 29 and 24 arrival trajectory clusters and 55, 28 and 36 departure trajectory clusters were identified for the New York, Hong Kong and Sao Paulo multi-airport systems, respectively. The centroids of the arrival and departure clusters are shown in Figures 3-10 and 3-11.

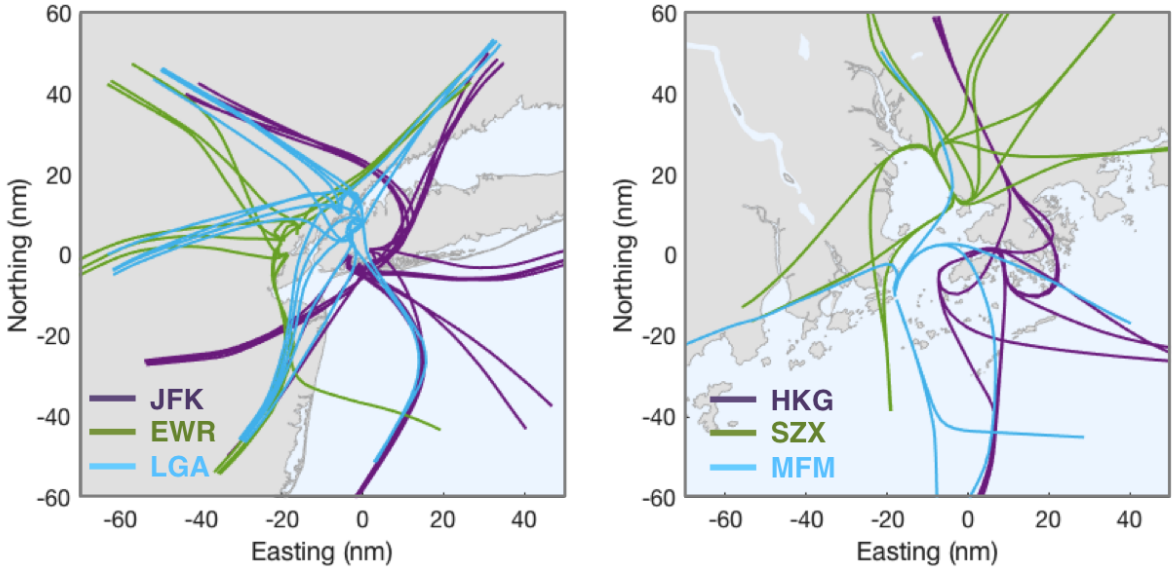


(a) New York metroplex.

(b) Hong Kong metroplex.

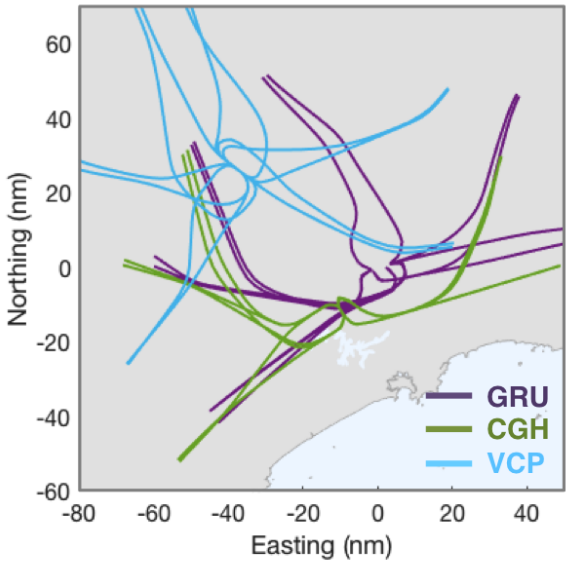
(c) Sao Paulo metroplex.

Figure 3-10: Centroids of metroplex arrival trajectory clusters.



(a) New York metroplex.

(b) Hong Kong metroplex.



(c) Sao Paulo metroplex.

Figure 3-11: Centroids of metroplex departure trajectory clusters.

At a first glance, it appears the New York metroplex has the most complex airspace structure. First, there are more trajectory patterns, which can be explained not only because of the higher number of runways, but also because of the presence of more than one trajectory pattern for some combinations of arrival/departure gate and runway threshold. An example is shown in Figure 3-12 for LGA arrival trajectories. Two trajectory patterns were identified

for flights entering the terminal area through the west gate and landing at runway 22, which are driven by different maneuvers to intercept the final approach leg. Second, the trajectory patterns are more closely located, with a higher number of lateral crossings.

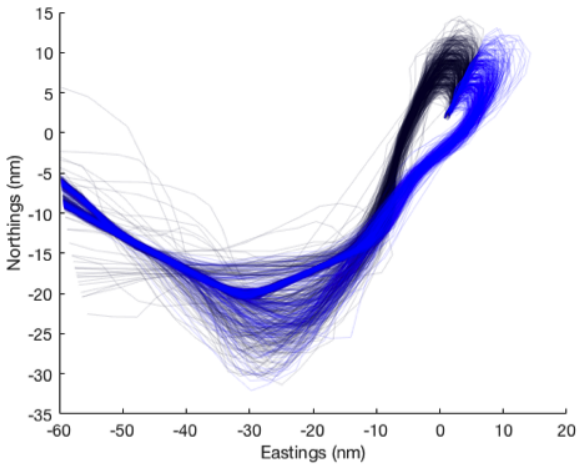
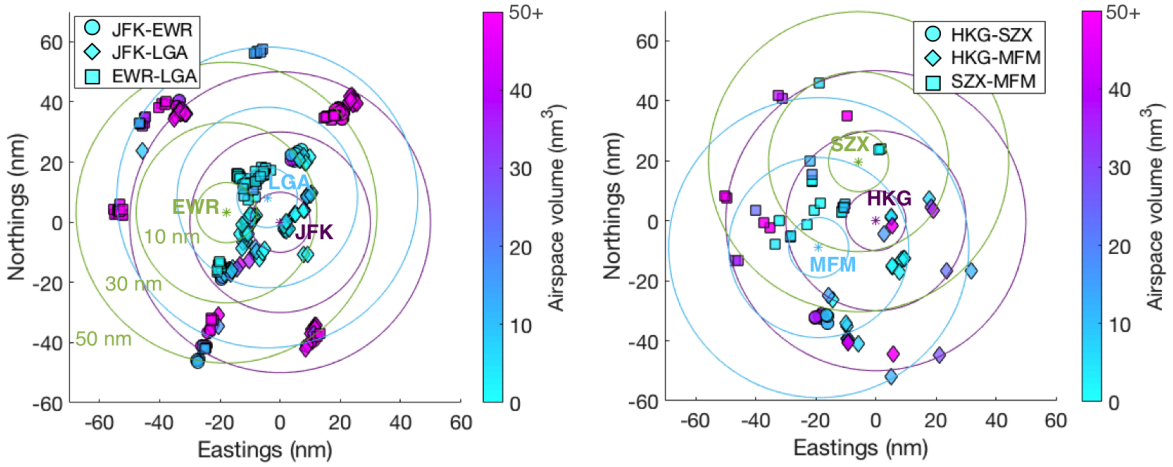


Figure 3-12: Example of different trajectory patterns from arrival fix to runway threshold 22 at LGA.

3.3.2 Identification of Route Intersections

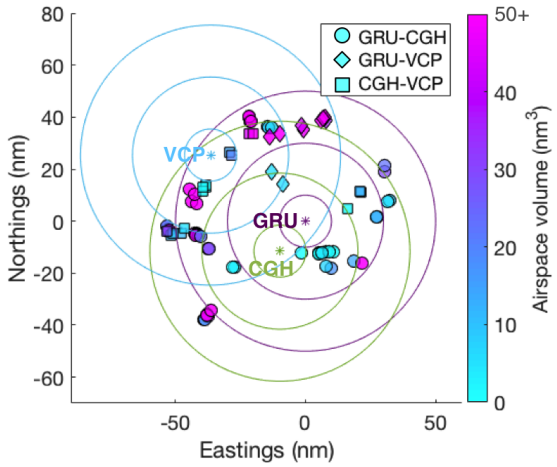
The observed lateral crossings of trajectory patterns in Figures 3-10 and 3-11 do not necessarily indicate they are closely located at the vertical dimension to potentially create flow interactions. In order to investigate the complexity of the airspace structure in terms of the potential to create flow interactions between the airports, we introduce the concept of trajectory tubes. A trajectory tube is defined for each spatial cluster based on the dispersion of their trajectory members. Since all trajectories are described with the same number of resampling points $i = 1, ..., n$, the tubes are discretized in $n - 1$ parts. Each discrete part $i = 1, ..., n - 1$ is determined by resampling points i and $i + 1$. For each part i, the width of the tube is determined by the 95^{th} percentile of the lateral spread of all i^{th} and $(i + 1)^{th}$ resampled points around the cluster centroid. The height of the tube is determined by the difference between the 95^{th} and the 5^{th} percentile of all i^{th} and $(i+1)^{th}$ vertical positions. We calculated the intersection volume between all pairs of trajectory tubes from different airports. They indicate the terminal airspace regions where the metroplex trajectory patterns intersect laterally and vertically and flows can potentially interact.

Table 3.1 shows the number of trajectory tube intersections identified by pair of metroplex airports and Figure 3-13 shows their location and size. First, it is noticeable the New York metroplex has the most conflicted airspace structure. JFK and LGA are the pair of airports with the highest number of intersections. A significant number of the New York metroplex route intersections are located less than 10 nautical miles from the airports (especially for JFK and LGA). An example is shown in Figure 3-14. It displays the intersection of a LGA arrival trajectory tube and a JFK arrival trajectory tube, as the airspace volume colored in red. It is also observed some clusters of high volume trajectory tube intersections located close to the terminal area boundary, revealing the sharing of arrival/departure gates.



(a) New York metroplex.

(b) Hong Kong metroplex.



(c) Sao Paulo metroplex.

Figure 3-13: Characterization of trajectory tube intersections.

Table 3.1: Number of trajectory tube intersections identified by pair of metroplex airports.

Multi-airport system	**Airport pair**	**Number**	**Detailed airport pair by type of operation**	**Number**
New York	JFK - EWR	42	JFK Arr - EWR Arr	0
			JFK Dep - EWR Dep	38
			JFK Arr - EWR Dep	4
			JFK Dep - EWR Arr	0
	JFK - LGA	130	JFK Arr - LGA Arr	34
			JFK Dep - LGA Dep	61
			JFK Arr - LGA Dep	29
			JFK Dep - LGA Arr	6
	EWR - LGA	91	EWR Arr - LGA Arr	17
			EWR Dep - LGA Dep	65
			EWR Arr - LGA Dep	0
			EWR Dep - LGA Arr	9
Hong Kong	HKG - SZX	12	HKG Arr - SZX Arr	0
			HKG Dep - SZX Dep	0
			HKG Arr - SZX Dep	12
			HKG Dep - SZX Arr	0
	HKG - MFM	28	HKG Arr - MFM Arr	2
			HKG Dep - MFM Dep	10
			HKG Arr - MFM Dep	14
			HKG Dep - MFM Arr	2
	SZX - MFM	30	SZX Arr - MFM Arr	10
			SZX Dep - MFM Dep	13
			SZX Arr - MFM Dep	2
			SZX Dep - MFM Arr	5
Sao Paulo	GRU - CGH	53	GRU Arr - CGH Arr	7
			GRU Dep - CGH Dep	33
			GRU Arr - CGH Dep	9
			GRU Dep - CGH Arr	4
	GRU - VCP	10	GRU Arr - VCP Arr	0
			GRU Dep - VCP Dep	2
			GRU Arr - VCP Dep	6
			GRU Dep - VCP Arr	2
	CGH - VCP	16	CGH Arr - VCP Arr	0
			CGH Dep - VCP Dep	6
			CGH Arr - VCP Dep	8
			CGH Dep - VCP Arr	2

The Hong Kong metroplex shows the less conflicted airspace structure. The airspace structures of the two primary airports are highly de-conflicted, with the exception of a small cluster of intersections between HKG arrival trajectory patterns and SZX departure trajectory patterns located in the south departure area of SZX. Most of the intersections are located more than 10 nautical miles away from the airports. A few high volume intersections are observed between 10 and 30 nautical miles from the airports, revealing the sharing of route segments. An example of a shared route segment between a SZX arrival trajectory tube and a MFM arrival trajectory tube is shown in Figure 3-15. It is also observed some clusters of high volume intersections located close to the terminal area boundary, revealing the sharing of arrival/departure gates, especially between SZX and MFM.

Finally, in the Sao Paulo metroplex, GRU and CGH have the highest number of intersections, especially between departure trajectory patterns. Most of these intersections are located close to the terminal area boundary, revealing the sharing of arrival/departure gates. Figure 3-16 shows an example of departure trajectory patterns from GRU and CGH sharing the same departure gate. A few high volume intersections are also observed between 10 and 30 nautical miles from the airports, revealing the sharing of route segments.

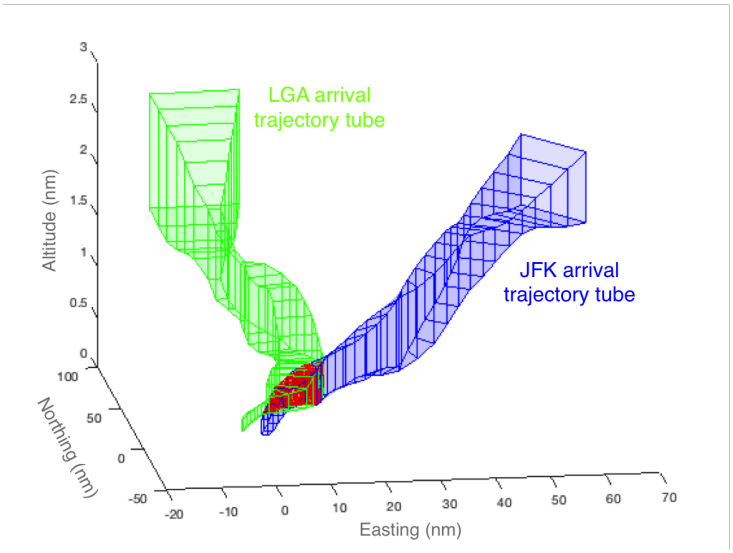


Figure 3-14: Example of intersection between JFK and LGA arrival trajectory tubes.

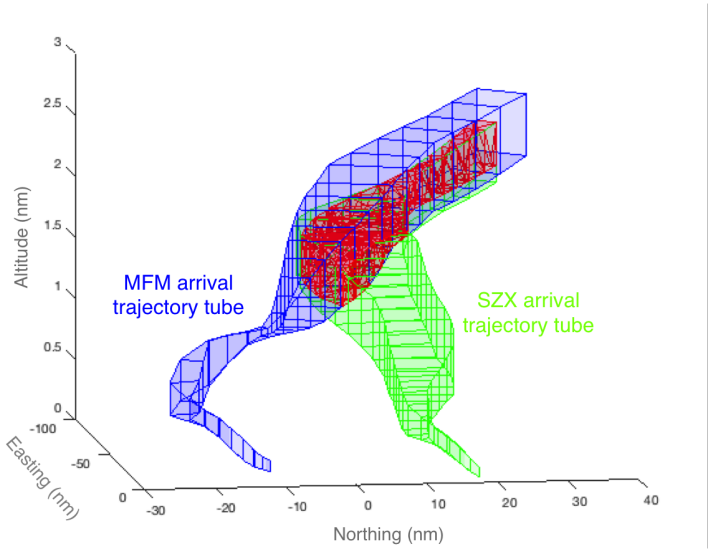


Figure 3-15: Example of intersection between SZX and MFM arrival trajectory tubes.

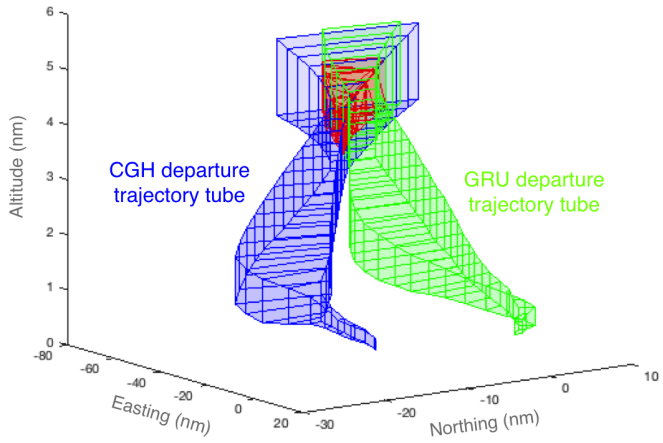


Figure 3-16: Example of intersection between GRU and CGH departure trajectory tubes.

3.4 Characterization of Operational Differences through the Analysis of Traffic Flow Dynamics

3.4.1 Identification of Metroplex Flow Patterns

Once the metroplex airspace structures were identified, Random Forests classifiers were created to discriminate the trajectory patterns for each type of operation (arrival/departure) at

each metroplex airport. All trajectory classifiers showed a multi-way classification accuracy higher than 98%. The classifiers were used to assign trajectories in the remaining dataset (not used for the spatial clustering) to the learned routes and the Conformal Prediction framework was applied with a significance level $\varepsilon = 0.05$ to identify non-conforming behaviors. Flow matrices were then created for each day of operations. As described in Chapter 2, the matrices were designed so that each element w_{ij} indicates the number of flights using a particular route i during a given hour j of the day. A flow is identified whenever $w_{ij} > 1$. The matrix columns are vectors $\{R_t \in \mathbb{R}^n, t = 1, ..., p\}$ that indicate the time-dependent traffic flows.

In order to investigate the presence of patterns in the metroplex terminal area flow behavior, we used the third module of the trajectory data analytics framework and performed a second-layer clustering analysis on the set of time-dependent traffic flows R_t. For this, a data representation was established for the time-dependent metroplex flow structure. The columns $\{(R^{m,Arr})_t \in \mathbb{R}^{|K^{m,Arr}|}, t = 1, ..., p\}$ and $\{(R^{m,Dep})_t \in \mathbb{R}^{|K^{m,Dep}|}, t = 1, ..., p\}$ of the daily arrival and departure flow matrices created for each metroplex airport m were aggregated into flow vectors to represent the hourly terminal area flow structure in the metroplex:

$$\left\{FV_t = \left((C_i^{m,Arr})_t, (C_j^{m,Dep})_t\right), m \in M, i \in A, j \in D, t = 1, ..., p\right\}$$

where:

$$(C_i^{m,Arr})_t = \begin{cases} \underset{k \in K_i^{m,Arr}}{\arg\max}\{(R_k^{m,Arr})_t\} & \text{if } \underset{k \in K_i^{m,Arr}}{\max}\{(R_k^{m,Arr})_t\} > 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.1}$$

$$(C_j^{m,Dep})_t = \begin{cases} \underset{l \in K_j^{m,Dep}}{\arg\max}\{(R_k^{m,Dep})_t\} & \text{if } \underset{k \in K_j^{m,Dep}}{\max}\{(R_k^{m,Dep})_t\} > 1 \\ 0 & \text{otherwise} \end{cases} \tag{3.2}$$

where M is the set of metroplex airports, A is the set of arrival gates in the terminal area boundary, D is the set of departure gates, $K_i^{m,Arr} \subset K^{m,Arr}$ is the set of arrival trajectory patterns between arrival gate $i \in A$ and airport $m \in M$, and $K_j^{m,Dep} \subset K^{m,Dep}$ is the

set of departure trajectory patterns between airport $m \in M$ and departure gate $j \in D$. Arrival/departure gates are defined as zones in the terminal area boundary intersected by a group of trajectory patterns.

It follows that the hourly flow vector FV_t is a categorical vector of dimension $|M| \times |A| + |M| \times |D|$ and indicates the dominant arrival flows from each arrival gate to each airport and the dominant departure flows from each airport to each departure gate.

Hierarchical clustering was applied using the Hamming distance and complete linkage for clustering the set of hourly flow vectors. This clustering algorithm creates a hierarchical decomposition of the data space (usually represented by a dendrogram) based on a dissimilarity matrix that progressively splits the database into smaller subsets using either a divisive (from the root to the leaves) or an agglomerative (from the leaves to the nodes) approach [108]. Under the complete linkage agglomerative approach, distances between groups of observations are calculated considering the maximum distance between elements in these groups. A distance threshold is then used to horizontally cut the dendrogram and determine the clusters. The Hamming distance was used to assess the similarity between the categorical flow vectors and it corresponds to the number of elements in which they differ. In this case, zero entries indicate absence of flow between the arrival/departure gate and the airport and were disregarded during the computation of the Hamming distance.

The clustering was performed for the hourly flow vectors observed between 6:00 am and 11:00 pm (local time). A distance threshold equal to two was used, establishing a maximum tolerable dissimilarity of two between flow vectors in the same cluster. The following rationale was used for selecting the distance threshold. Consider a particular flow structure at a given time represented by its flow vector. If there is a meaningful arrival (departure) flow change at one airport because of a runway configuration switch such that different arrival (departure) trajectory patterns are observed between the gates at the terminal area boundary and the airport, a new flow structure will be observed and the flow vector will change. The dissimilarity between the new flow vector and the previous one will correspond to the number of arrival (departure) gates at the terminal area boundary for that particular airport. Since the minimum number of arrival or departure gates by airport is three across all multi-airport systems, the distance threshold should be at most two in order to prevent meaningfully

dissimilar flow structures to be included in the same cluster.

Figure 3-17 shows the number of clusters identified at each multi-airport system and their cumulative frequency of occurrence. It is noticeable that a few clusters dominate and capture the majority of the observations, revealing the primary operational modes at each metroplex. The remaining clusters are composed of very few members and are characterized by less frequently used metroplex configurations and abnormal flow structures (primarily associated with periods of runway configuration changes). We defined a cluster of flow structures as a *Metroplex Flow Pattern* (MFP). The clustering results are summarized in Table 3.2. The New York metroplex stands out with the highest number of operational patterns and the highest operational variability. For instance, the most frequent New York MFP accounts for only 15.8% of the observations. At the Hong Kong and Sao Paulo metroplexes, the most frequent MFP accounts for 31.2% and 53.5% of the observations, respectively. The results suggest the New York metroplex is the most dynamic system. Indeed, it is observed the New York terminal area has a higher number of flow structure changes during the course of a day and the MFP dwell times are lower.

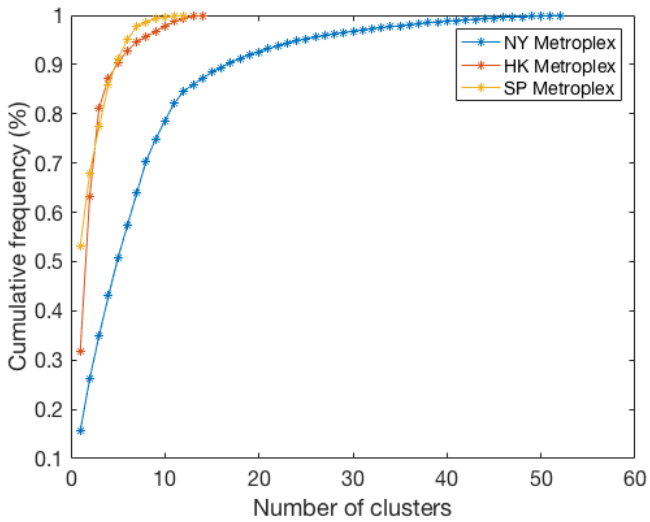


Figure 3-17: Cumulative percentage of observations for the clusters of hourly flow vectors identified.

A detailed characterization of the dominant patterns for each metroplex was performed in order to obtain insights about some of the factors that drive the behavior of the metroplex flows. For this, we considered the most frequent MFPs that account for 85% of the

Table 3.2: Results of the metroplex hourly flow vector clustering analysis.

Multi-airport system	**New York**	**Hong Kong**	**Sao Paulo**
Frequency of occurrence of the most frequent MFP (%)	15.8	31.2	53.5
Median number of flow structure changes per day	5	2	2.5
Median dwell time of the most frequent MFP (h)	3.5	8	5

observations.

New York Metroplex

In the New York metroplex, twelve dominant MFPs were identified. Figure 3-18 shows the different uses of the airspace structure associated with each MFP; specifically, it shows the average number of aircraft following each spatial trajectory pattern. MFPs are numbered in descending order of frequency of observation.

The detailed characterization of the New York MFPs is shown in Table 3.3. The most evident factor driving the behavior of the metroplex flows is clearly the runway configuration in use at each airport. Five MFPs were characterized as south flow (which means that all airports operate in a south landing flow configuration), five MFPs were characterized as north flow and two MFPs were characterized as mixed of north and south flows (Figure 3-19 illustrates the convention about landing flow direction using JFK as an example). It is noticeable the tendency of alignment between runway configurations across the airports. Table 3.4 lists the most frequently observed runway configuration for each MFP. They are represented in the form "A | B", where "A" indicates the arrival runways and "B" indicates the departure runways.

Besides the runway configuration, some differences are also driven by the airspace configuration in terms of the delegation of airspace regions that are shared. For instance, MFPs 2 and 4 are characterized by the same runway configuration at LGA. However, restricted climb procedures from runway 13 are observed for MFP 2. Analysis of current operating procedures reveals that whenever JFK uses runway 31L for departures, it remains with control of the shared Coney airspace, restricting the use of multiple climbs by LGA. This fact

also explains the differences between MFPs 6 and 12.

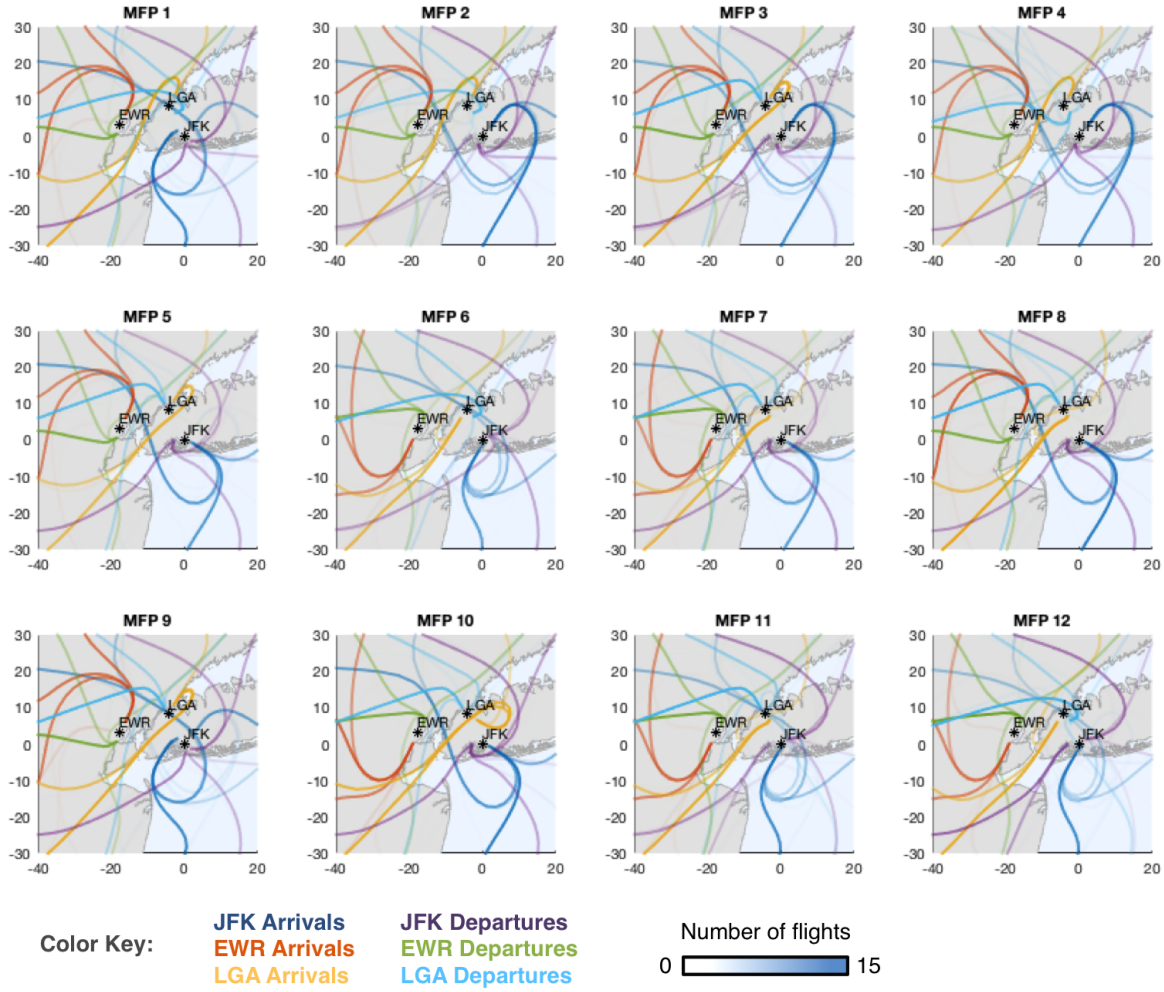


Figure 3-18: Most frequently observed New York MFPs.

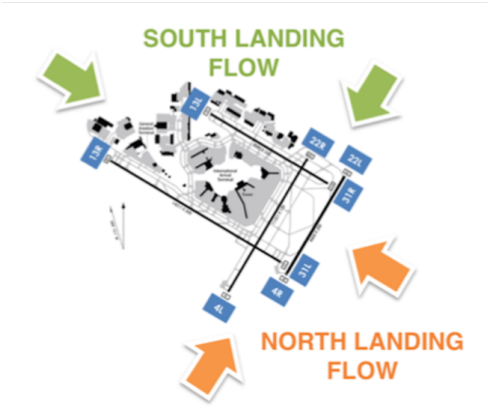


Figure 3-19: Convention on flow direction.

Table 3.3: Description of the New York MFPs.

MFP		Description
South Flow Arrival Priority (South-AP)	1	South landing flow configuration at all airports; Favors arrival operations; Right-hand arrival traffic pattern at LGA determined by departure configuration (runway 13); Primarily observed during VMC.
	9	South landing flow configuration at all airports; Favors arrival operations; Left-hand arrival traffic pattern at LGA determined by departure configuration (runway 31); Primarily observed during VMC.
	4	South landing flow configuration at all airports; Favors arrival operations; Right-hand arrival traffic pattern at LGA determined by departure configuration (runway 13); Primarily observed during IMC; Multiple climbs from runway 13 at LGA with Coney Airspace release.
South Flow Departure Priority (South-DP)	2	South landing flow configuration at all airports; Favors departure operations; Right-hand arrival traffic pattern at LGA determined by departure configuration (runway 13); Primarily observed during VMC; Restricted climbs from runway 13 at LGA.
	3	South landing flow configuration at all airports; Favors departure operations; Left-hand arrival traffic pattern at LGA determined by departure configuration (runway 31); Primarily observed during VMC.
North Flow Arrival Priority (North-AP)	6	North landing flow configuration at all airports; Favors arrival operations; Primarily observed during IMC; Multiple climbs from runway 13 at LGA with Coney Airspace release.
	7	North landing flow configuration at all airports; Favors arrival operations; Visual approach at LGA runway 31; Primarily observed during VMC.
	10	North landing flow configuration at all airports; Favors arrival operations; Instrument approach at LGA runway 31; Primarily observed during IMC.
North Flow Departure Priority (North-DP)	11	North landing flow configuration at all airports; Favors departure operations; Visual approach at LGA runway 31; Primarily observed during VMC.
	12	North landing flow configuration at all airports; Favors departure operations; Primarily observed during VMC; Restricted climbs from runway 13 at LGA.
Mixed Flow Arrival Priority (Mixed-AP)	5	Mixed flow configuration with south landing flow dominance (EWR and LGA); Favors arrival operations; Primarily observed during VMC.
	8	Mixed flow configuration with north landing flow dominance (JFK and LGA); Favors arrival operations; Primarily observed during VMC.

Table 3.4: Airport runway configurations associated with each New York MFP.

MFP	Runway configuration		
	JFK	**EWR**	**LGA**
1	13L, 22L | 13R	22L | 22R	22 | 13
2	22L, 22R | 22R, 31L	22L | 22R	22 | 13
3	22L, 22R | 22R, 31L	22L | 22R	22 | 31
4	22L, 22R | 22R	22L | 22R	22 | 13
5	31L, 31R | 31L	22L | 22R	22 | 31
6	4L, 4R | 4L	4R | 4L	4 | 13
7	31L, 31R | 31L	4R | 4L	31 | 4
8	31L, 31R | 31L	22L | 22R	31 | 4
9	13L, 22L | 13R	22L | 22R	22 | 31
10	31L, 31R | 31L	4R | 4L	31 | 4
11	4L, 4R | 4L, 31L	4R | 4L	31 | 4
12	4L, 4R | 4L, 31L	4R | 4L	4 | 13

Figure 3-20 shows that the time of the day is an important variable determining the occurrence of MFPs. As an example, MFPs 1 and 4 are more likely to be observed in the afternoon period, whereas MFPs 2 and 3 have a higher frequency of observations in the morning and evening periods. This was observed to be correlated with the daily demand profile in the New York metroplex. Analysis of the daily demand patterns reveals the existence of banks of arrivals in the afternoon, primarily driven by JFK and EWR demand profiles, and departures in the morning, as shown in Figure 3-21. Indeed, this demand profile, with unbalanced mix, is one important factor that drives the selection of runway configuration at these airports in order to favor arrival or departure operations. Figure 3-22 shows the distribution of arrival and departure throughput with a boxplot for each MFP. A clear imbalance is observed. MFPs 1, 4, 5, 6, 7, 8, 9 and 10 tend to favor arrival operations (third quartile was used for comparison). Indeed, they are more frequently observed in the afternoon period. By contrast, MFPs 2, 3, 11 and 12 tend to favor departure operations, and they are more frequently observed in the morning and evening periods.

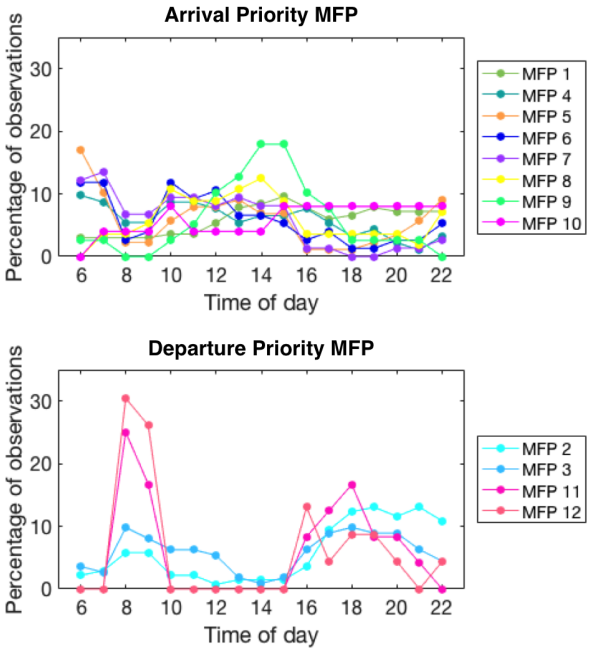
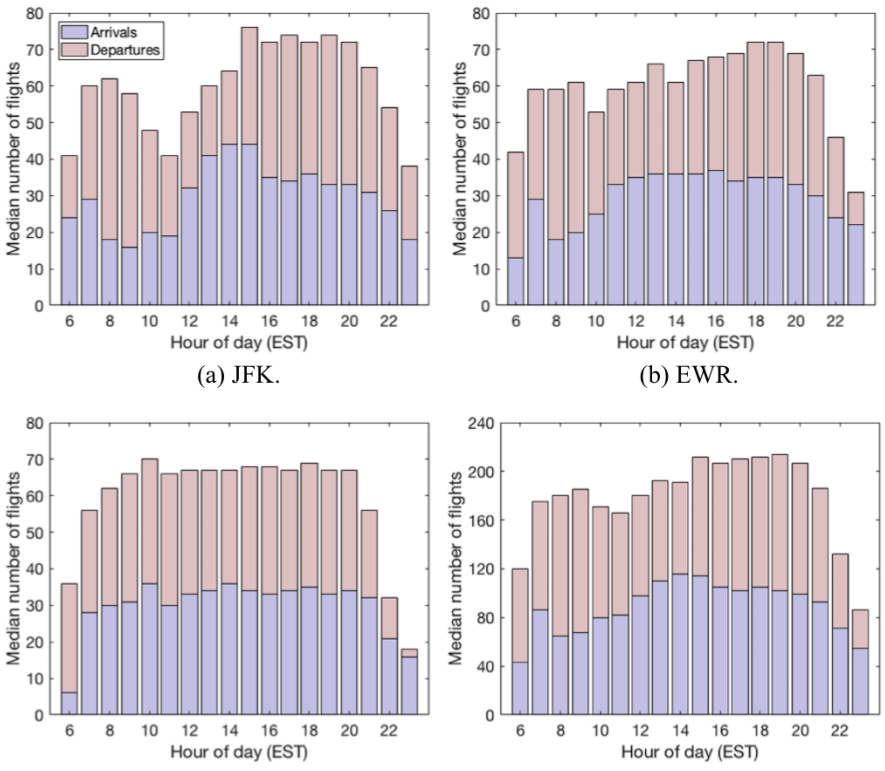


Figure 3-20: New York MFP frequency of occurrence by time of day.



(a) JFK. (b) EWR.

(c) LGA. (d) New York metroplex.

Figure 3-21: Daily New York demand patterns, based on data from years 2013-2015.

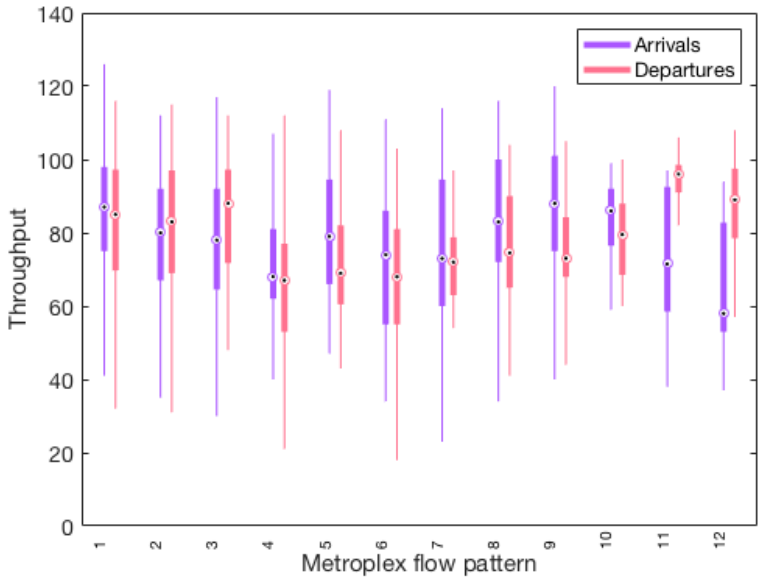


Figure 3-22: Distribution of metroplex arrival and departure throughput for the New York metroplex MFPs.

Finally, a particular influence of meteorological conditions in the operational mode in the New York metroplex is also observed. Figure 3-23 shows, for each MFP, the percentage of observations associated with periods of Visual Meteorological Conditions (VMC) and Instrument Meteorological Conditions (IMC). For instance, MFPs 4, 6 and 10 are more likely to be observed during IMC. MFP 10 has 80% of its observations associated with periods of IMC. If we contrast MFP 10 and MFP 7, which was only observed during VMC, it is noted that the major difference between these two MFPs is determined by the LGA arrival flows: in MFP 7, they are tailored to a visual approach to runway 31, whereas a long and straight Instrument Landing System (ILS) approach to the same runway is noticeable in MFP 10. In other words, part of the New York metroplex behavior is driven by the existence of multiple routes tailored to the approach procedure, with use governed by meteorological conditions.

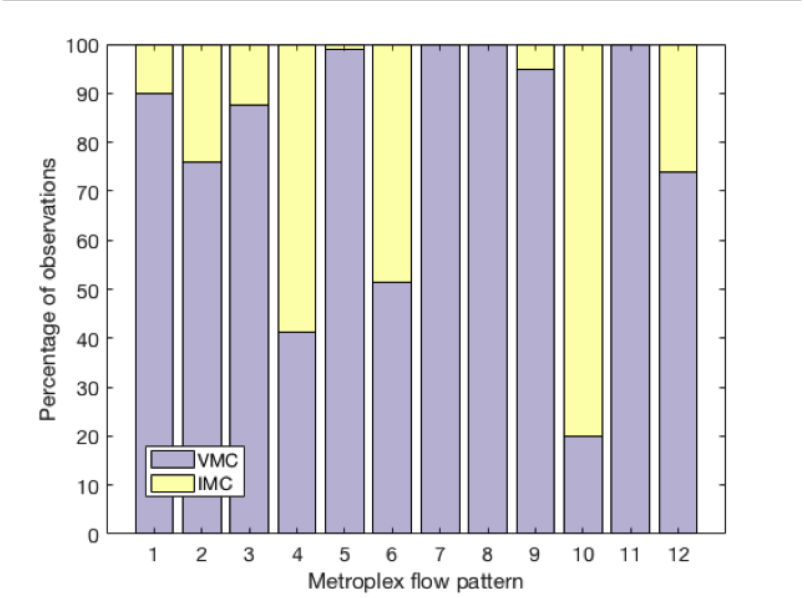


Figure 3-23: New York MFP frequency of occurrence under VMC and IMC.

Hong Kong Metroplex

In the Hong Kong metroplex, four dominant MFPs were identified, which are shown in Figure 3-24 and described in Table 3.5. In the most frequent MFP, all airports operate in a north landing flow configuration. The other MFPs are characterized by a mixed flow configuration. Table 3.6 lists the airport runway configurations inferred for each MFP.

Time of the day also appears to play a role in the occurrence of MFPs in the Hong Kong metroplex. Figure 3-25 shows the frequency of occurrence of each MFP during the day. While the most frequent MFP 1 is equally likely to be observed throughout the day, MFP 2 is more frequently observed in the morning and night periods and MFPs 3 and 4 are more likely to be observed in the afternoon period, revealing a preference for use of runways 25L/R during the afternoon at HKG.

Unlike the New York metroplex, meteorological conditions did not appear to have a significant influence on the operational mode in the Hong Kong metroplex. Figure 3-26 shows that none of the MFPs showed a higher likelihood of being observed during IMC.

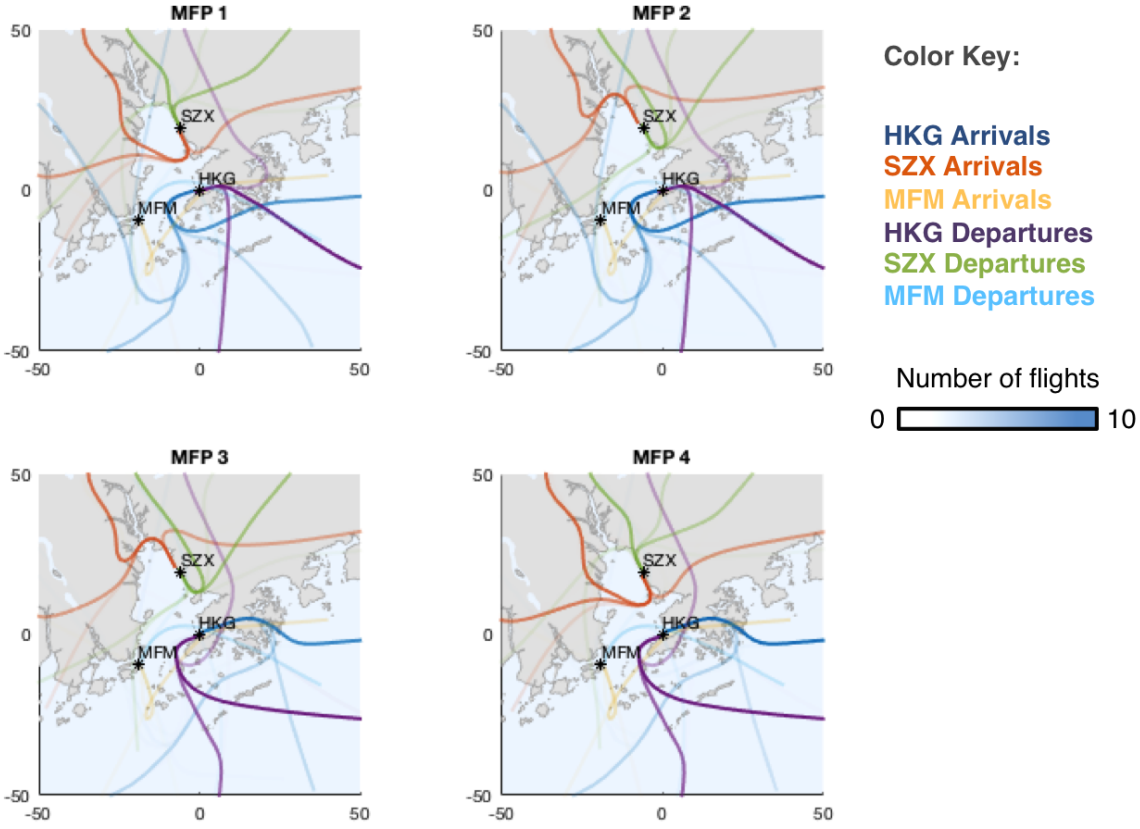


Figure 3-24: Most frequently observed Hong Kong MFPs.

Table 3.5: Description of the Hong Kong MFPs.

MFP		Description
North Flow	1	North landing flow configuration at all airports.
Mixed Flow - N,S,N	2	North landing flow configuration at HKG and MFM; South landing flow configuration at SZX.
Mixed Flow - S,S,N	3	South landing flow configuration at HKG and SZX; North landing flow configuration at MFM.
Mixed Flow - S,N,N	4	South landing flow configuration at HKG; North landing flow configuration at SZX and MFM.

Table 3.6: Airport runway configurations associated with each Hong Kong MFP.

MFP	Runway configuration		
	HKG	SZX	MFM
1	07L | 07R	34 | 33	34 | 34
2	07L | 07R	16 | 15	34 | 34
3	25R | 25L	16 | 15	34 | 34
4	25R | 25L	34 | 33	34 | 34

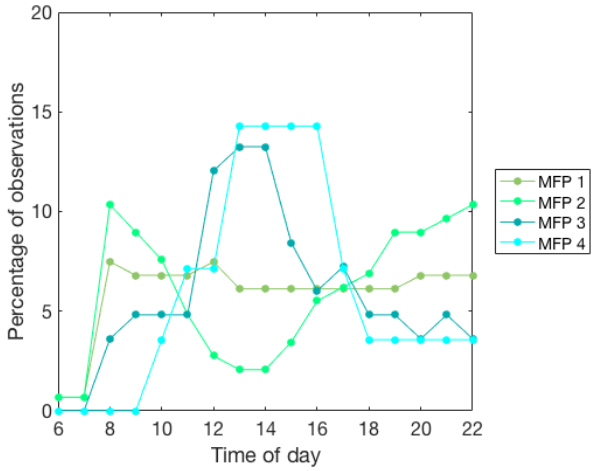


Figure 3-25: Hong Kong MFP frequency of occurrence by time of day.

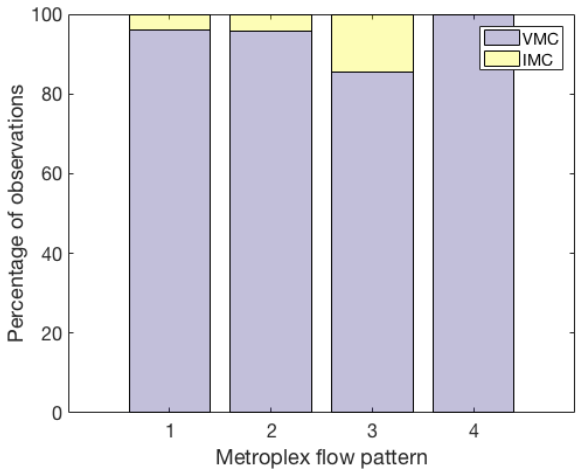


Figure 3-26: Hong Kong MFP frequency of occurrence under VMC and IMC.

Sao Paulo Metroplex

In the Sao Paulo metroplex, four dominant MFPs were also identified, which are shown in Figure 3-27 and described in Table 3.7. The most frequent MFP is characterized by a mixed flow configuration, with arrivals landing north at GRU and south at CGH and VCP. As in the Hong Kong metroplex, it was not observed a clear tendency of alignment between the airport runway configurations. Table 3.8 lists the airport runway configurations inferred for each MFP.

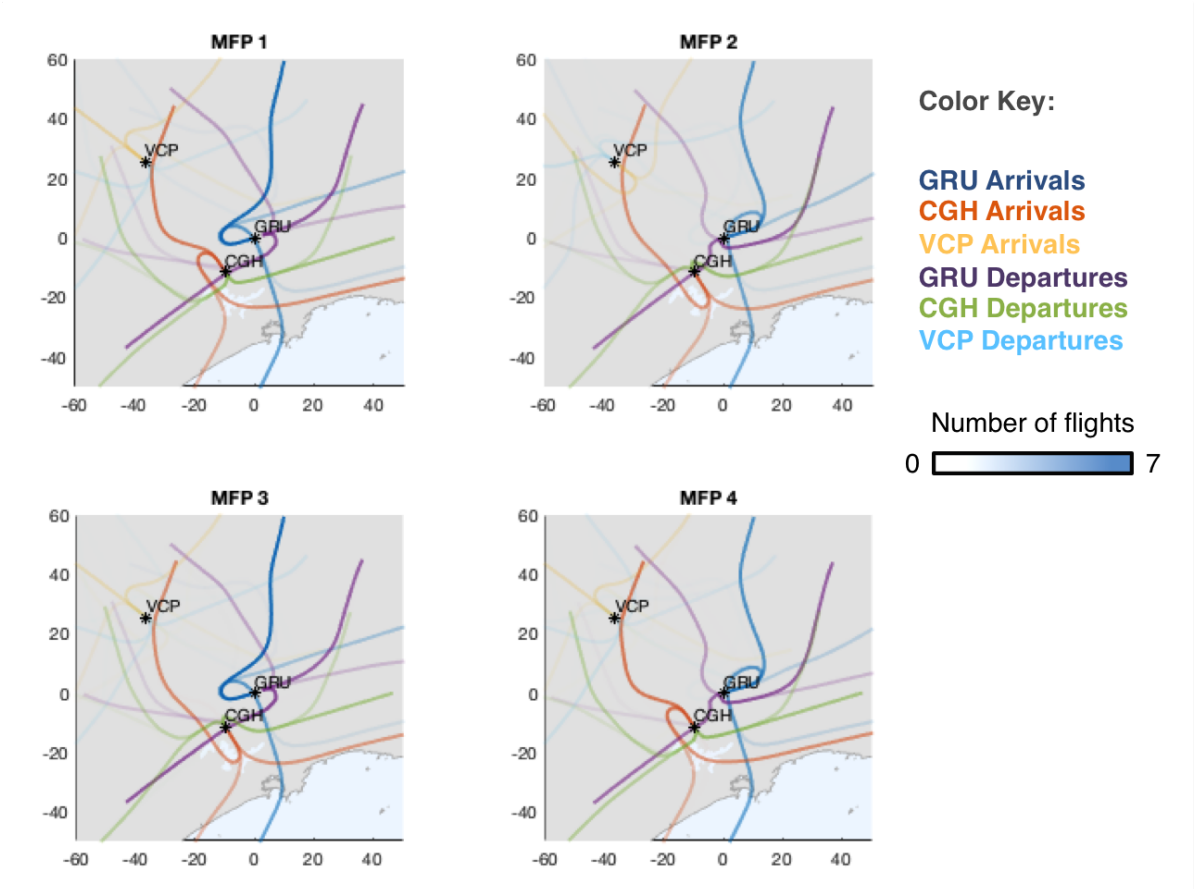


Figure 3-27: Most frequently observed Sao Paulo MFPs.

Figure 3-28 shows the frequency of occurrence of each MFP during the day. MFPs 1 and 3 are more likely to be observed in the morning and night periods, whereas MFPs 2 and 4 are more likely to be observed in the afternoon period. This seems to be correlated with the daily demand patterns at GRU, as shown in Figure 3-29. There is a preference for using runways 09L/R during the busiest hours of the day in the morning and night periods, as

Table 3.7: Description of the Sao Paulo MFPs.

MFP		Description
Mixed Flow - N,S,S	1	North landing flow configuration at GRU; South landing flow configuration at CGH and VCP.
Mixed Flow - S,N,N	2	South landing flow configuration at GRU; North landing flow configuration at CGH and VCP.
Mixed Flow - N,N,S	3	North landing flow configuration at GRU and CGH; South landing flow configuration at VCP.
South Flow	4	South landing flow configuration at all airports.

Table 3.8: Airport runway configurations associated with each Sao Paulo MFP.

MFP	Runway configuration		
	GRU	CGH	VCP
1	09R | 09L	17R | 17R	15 | 15
2	27L | 27R	35L | 35L	33 | 33
3	09R | 09L	35L | 35L	15 | 15
4	27L | 27R	17R | 17R	15 | 15

configuration 09R | 09L is optimal for GRU [3].

Finally, the operational mode in the Sao Paulo metroplex was not observed to be particularly influenced by meteorological conditions. Figure 3-30 shows that none of the MFPs showed a higher likelihood of being observed during IMC.

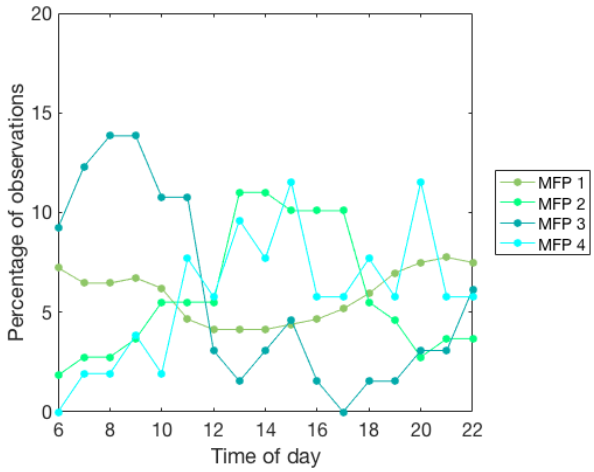


Figure 3-28: Sao Paulo MFP frequency of occurrence by time of day.

[3]The offset of 590 m between the runway thresholds 09L and 09R has a positive impact on capacity when arrivals use 09R and departures use 09L, whereas the offset of 1140 m between the runway thresholds 27L and 27R has a negative impact on capacity when arrivals use 27L and departures use 27R

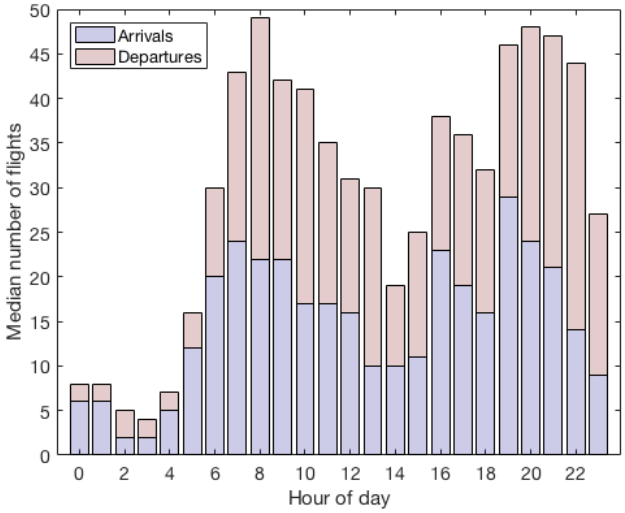


Figure 3-29: Daily demand pattern at GRU, based on data of 2017.

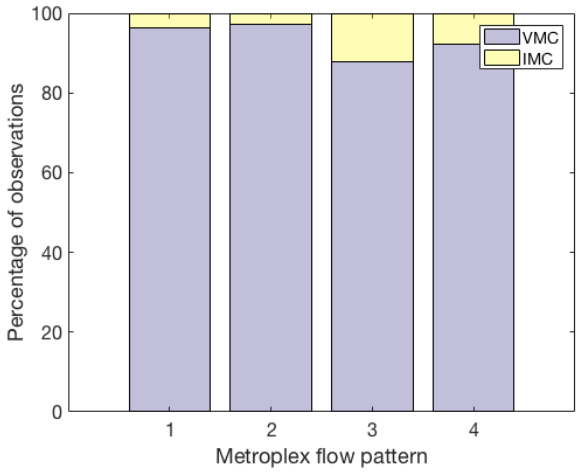


Figure 3-30: Sao Paulo MFP frequency of occurrence under VMC and IMC.

3.4.2 Identification of Metroplex Flow Interactions

We leveraged the knowledge about the flow patterns to identify relevant inter-airport flow interactions driven by the metroplex airspace design and use. The trajectory tube intersections identified in Section 3.3.2 can be generally translated into two types of flow interactions, depending on the use of the trajectory patterns involved in the intersection:

- *Flow dependency:* It occurs when the use of one trajectory pattern inhibits the use of the other one; therefore, flows are dependent and not observed simultaneously. From

an operational perspective, a flow dependency is expected to make the management of runway/airspace configuration more challenging, as coordination is required between the airports.

- *Flow crossing:* It occurs when the use of one trajectory pattern does not inhibit the use of the other one; therefore, both flows can be observed simultaneously and they overlap at their shared airspace region. From an operational perspective, a flow crossing is expected to increase air traffic control complexity, as it tends to increase workload associated with conflict detection and avoidance in the shared airspace region.

We used the knowledge about the metroplex flow patterns to identify relevant flow crossings and flow dependencies as follows:

A trajectory tube intersection is translated into a relevant *flow crossing* if there exists at least one MFP for which the trajectory patterns are used simultaneously and meaningfully. We define a meaningful use of a trajectory pattern if the median number of aircraft using the trajectory pattern is higher than one.

A trajectory tube intersection is translated into a relevant *flow dependency* if, for all MFPs, the trajectory patterns are not used simultaneously. For this, we consider that whenever one trajectory pattern is used meaningfully (median number of aircraft is higher than one), the median number of aircraft using the other trajectory pattern should be equal to zero.

For the New York Metroplex, 60 flow dependencies were identified. For the Hong Kong and Sao Paulo metroplexes, flow dependencies were not observed. Figure 3-31 shows the location of the trajectory tube intersections characterized as flow dependencies in the New York metroplex. JFK and LGA arrival flows have the highest number of dependencies. Most of the flow dependencies are associated with shared airspace regions located within 30 nm from the airports. Clusters of trajectory tube intersections are observed close to the airports, revealing the existence of infeasible combinations of runway configurations. The results reveal the New York metroplex is the most interdependent system, suggesting the management of runway and airspace configuration tends to be more challenging at this system.

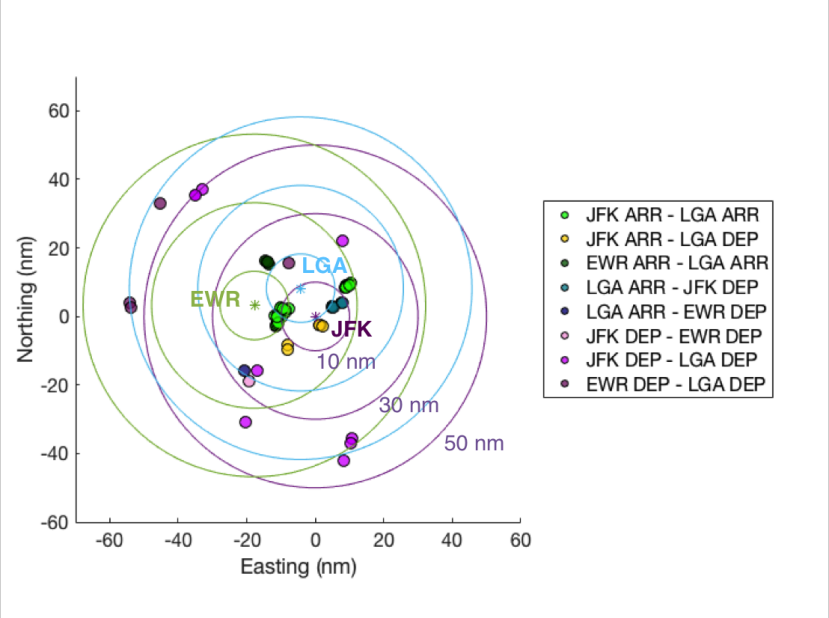


Figure 3-31: Trajectory tube intersections associated with flow dependencies in the New York metroplex.

Table 3.9 shows the number of relevant flow crossings identified by MFP. The New York metroplex also has the highest number of flow crossings by MFP, revealing the typical terminal area flow structures are potentially characterized by higher levels of air traffic control complexity.

Table 3.9: Number of flow crossings by MFP.

MFP	New York	Hong Kong	Sao Paulo
1	15	2	7
2	10	7	6
3	17	6	7
4	19	4	7
5	14	-	-
6	14	-	-
7	12	-	-
8	15	-	-
9	13	-	-
10	13	-	-
11	12	-	-
12	15	-	-

3.5 Characterization of Performance Differences

The performance analysis was focused on arrival operations, given they are typically subject to higher levels of inefficiency as most of the arrival sequencing and scheduling is performed within the terminal area. We focused on the three performance areas of Efficiency, Capacity and Predictability, which are defined by ICAO as key areas for ATM performance measurement [133].

3.5.1 Efficiency

In this performance area, we were primarily concerned with flight trajectory efficiency under the notion of how actual trajectories compare to reference ideal trajectories. With this notion, we analyzed the efficiency of metroplex airspace design and use and the efficiency of traffic flows.

Metroplex Airspace Design and Use Efficiency

We first compared the efficiency of the learned airspace structure for the three multi-airport systems in terms of trajectory lateral efficiency. Based on the trajectory patterns identified for each metroplex airport, we established a structural path stretch metric S, which is defined as the weighted average of the path stretch associated with the representative trajectory of each learned route:

$$S = \sum_{i=1}^{N} p_i s_i \tag{3.3}$$

Where S is the structural path stretch of an airspace with N trajectory patterns, p_i is the weight of trajectory pattern i and s_i is the path stretch associated with the centroid of trajectory pattern i. If the weights are equal for all trajectory patterns, the metric assesses the efficiency of metroplex airspace design regardless airspace use. If the weights are defined as the frequency of occurrence of trajectory patterns, the metric evaluates the combined efficiency of metroplex airspace design and use. The path stretch is defined as the difference between the actual trajectory total length and the length of the shortest path that connects

the initial and the end points of the trajectory, as illustrated in Figure 3-32.

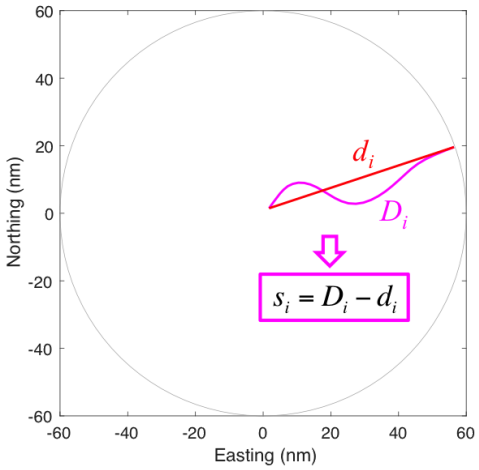


Figure 3-32: Example of path stretch calculation.

From the perspective of lateral performance, an airspace designed with lower structural path stretch can be considered more efficient as aircraft fly smaller distances between the terminal area boundary and the runway threshold. The calculated structural path stretch for arrival operations for each metroplex and each of their airports is shown in Figure 3-33. The metric was calculated considering both equal weights and use-based weights.

The Sao Paulo metroplex presents the most efficient airspace design, with the lowest structural path stretch. Interestingly, when airspace use is taken into account, the New York metroplex outperforms the others and shows the most efficient combination of airspace design and use. The metric is significantly higher for the Hong Kong metroplex. HKG stands out as the airport with the lowest efficiency, which is potentially due to the high level of deconfliction with the neighboring airports. At the other end, VCP stands out as the airport with the highest efficiency, ratifying the advantages of four-corner post airspace design.

Figure 3-34 shows the structural path stretch associated with different metroplex configurations. The metric was calculated with use-based weights, based on the average number of aircraft following each spatial trajectory pattern observed for each MFP. The results reveal that the south flow configurations in the New York metroplex (MFPs 1, 2, 3, 4, 5 and 9) are less efficient as they consistently show higher levels of structural path stretch. In the Hong Kong metroplex, the most frequently used configurations are also the least efficient. This is

particularly driven by the north flow configuration at HKG; trajectory patterns associated with runway 07L showed high levels of path stretch. Finally, Sao Paulo showed the lowest variability in MFP structural path stretch.

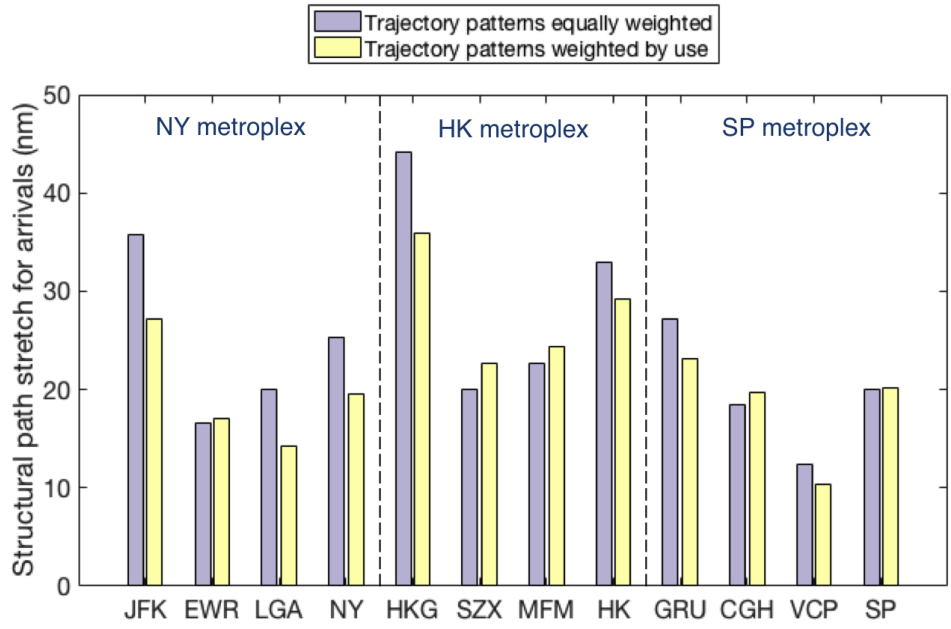


Figure 3-33: Structural path stretch for arrivals.

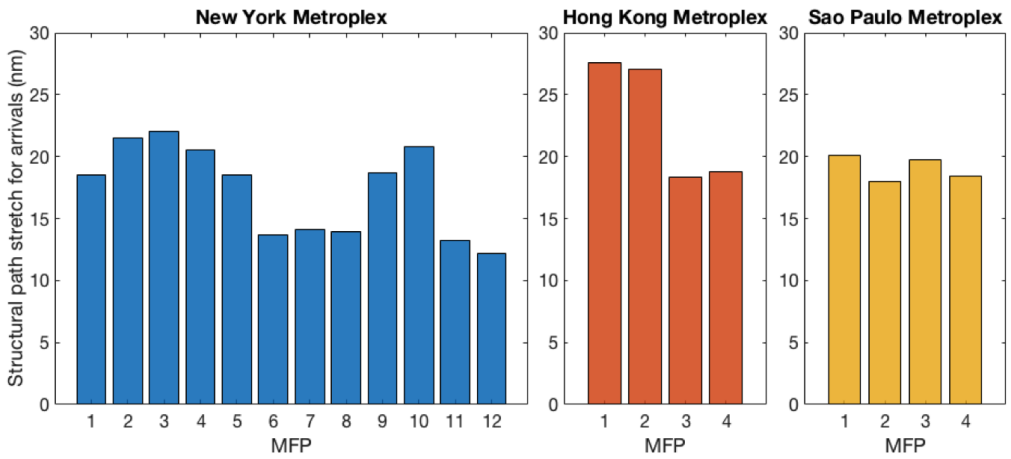


Figure 3-34: Structural path stretch for arrivals by MFP.

Traffic Flow Efficiency

We compared the overall traffic flow efficiency by analyzing the efficiency of actual trajectories, both at spatial and temporal dimensions. Two metrics were defined. The lateral traffic flow efficiency is a distance-based efficiency metric. For a set of N flight trajectories associated with the traffic flow structure observed during a given time period, it is defined as:

$$Eff_{lateral} = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} D_i} \tag{3.4}$$

where d_i is the length of the shortest path connecting the initial and end points of trajectory i and D_i is the actual length of trajectory i.

The temporal traffic flow efficiency is a time-based efficiency metric and it is defined in a similar fashion:

$$Eff_{temporal} = \frac{\sum_{i=1}^{N} t_i}{\sum_{i=1}^{N} T_i} \tag{3.5}$$

where t_i is the unimpeded flight time associated with trajectory i and T_i is the actual flight time associated with trajectory i. For each trajectory, its unimpeded flight time is estimated as the minimum between its actual flight time and the 10^{th} percentile of the distribution of flight times for trajectories in the same spatial pattern.

Based on their definitions, both metrics range between 0 and 1. Values closer to one indicate higher efficiency. The lateral and temporal efficiencies were computed on an hourly basis for the metroplex flows. The median efficiency values for each metroplex are shown in Figure 3-35. Overall, the New York metroplex presents the highest traffic flow efficiency, both spatially and temporally. The Hong Kong metroplex shows a significantly lower lateral efficiency, which is in part driven by its least efficient airspace design. It also presents the lowest temporal efficiency. Under inclement weather conditions, however, the New York metroplex shows significant drops in efficiency, becoming the lowest performing system temporally.

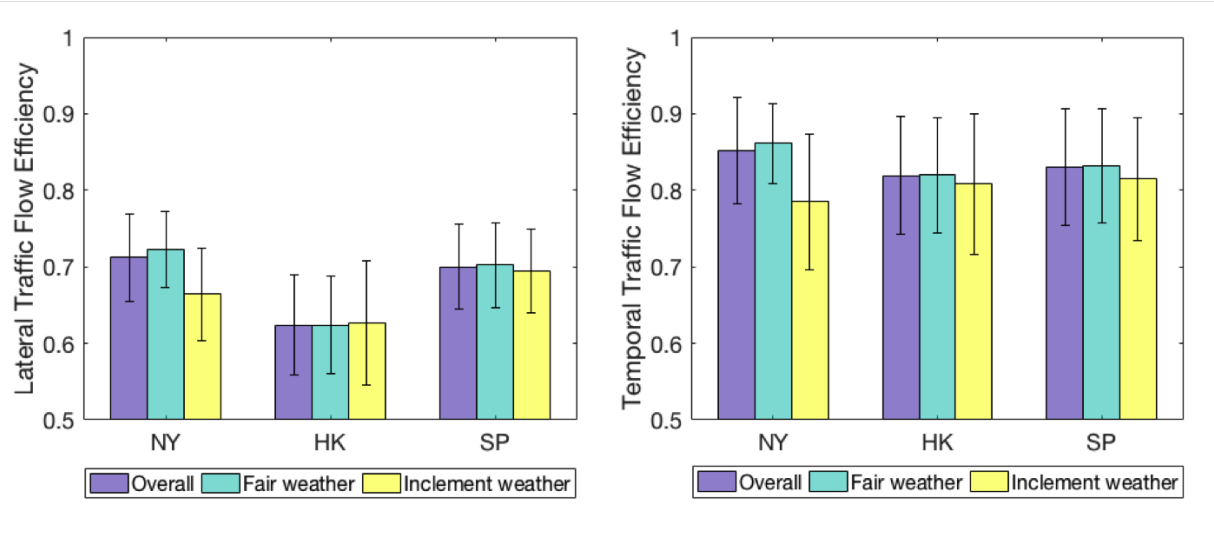


Figure 3-35: Lateral and temporal traffic flow efficiency.

We also calculated the median traffic flow efficiency for each of the dominant operational modes observed in each metroplex. Figure 3-36 shows the distribution of the MFP efficiency values in a boxplot. While the New York metroplex on average outperforms the others both in terms of lateral and temporal efficiency, it exhibits the highest variability in MFP efficiency.

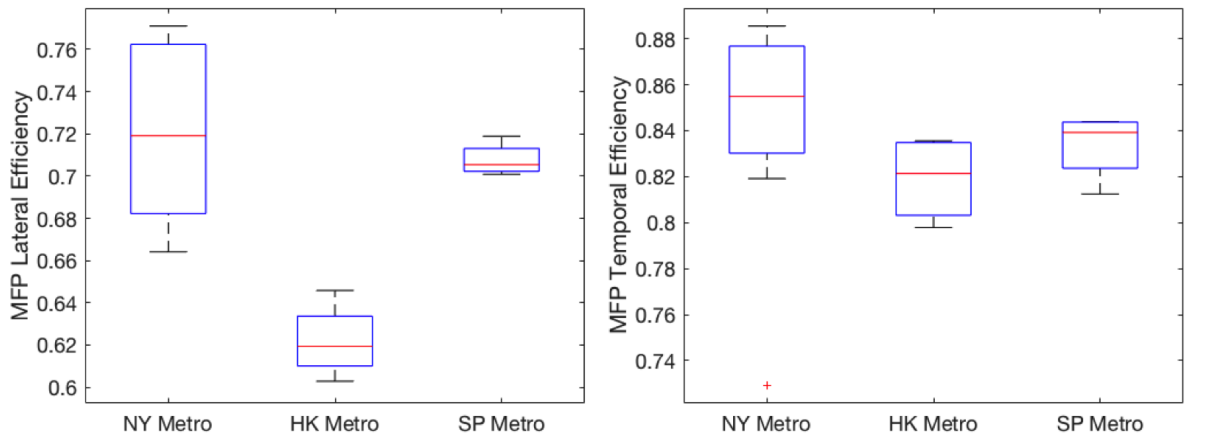


Figure 3-36: Distribution of traffic flow efficiency by MFP.

Trajectory Conformance Efficiency

We also evaluated the lateral conformance of flight trajectories against the learned airspace structure. While the lateral performance evaluated by the lateral traffic flow efficiency metric

is a result of both airspace design and tactical terminal area traffic flow management, the analysis of trajectory conformance provides insights about the sole contribution of tactical traffic flow management. Figure 3-37 shows the average daily percentage of non-conforming arrival trajectories at each metroplex airport. The results are displayed separately for days with fair weather conditions and for days with weather impacts. Overall, it is observed the New York metroplex exhibits the best level of trajectory conformance, both for nominal and off-nominal conditions. This is consistent with its high lateral traffic flow efficiency. HKG stands out as the airport with the highest percentage of non-conforming behaviors, well above the other airports, even for nominal conditions. This reveals that the observed low lateral traffic flow efficiency for the Hong Kong metroplex is not only a result of its least efficient airspace design, but also of terminal area traffic flow management. At the Sao Paulo metroplex, the primary airport GRU also shows the lowest level of trajectory conformance.

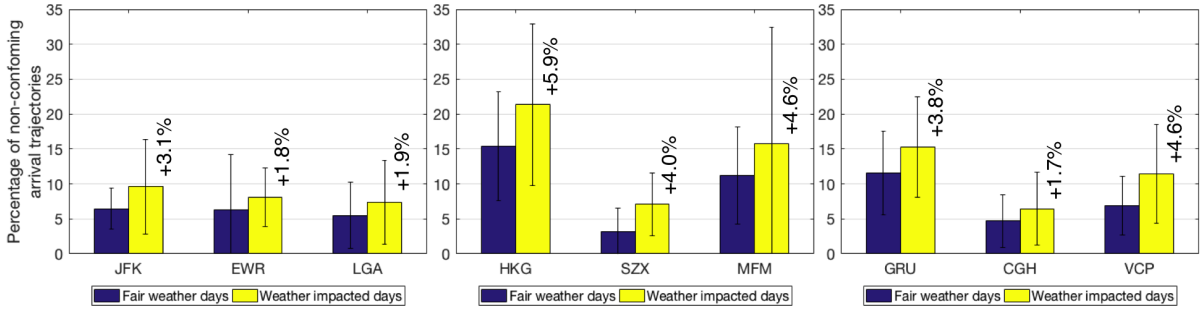


Figure 3-37: Average daily percentage of non-conforming arrival trajectories.

It is also observed that, for all airports, the percentage of non-conforming trajectories increases for days impacted by adverse meteorological conditions such as convection, low ceiling/visibility or strong winds. This suggests that weather is an important factor affecting the efficiency of terminal area operations at all three multi-airport systems.

As an example, the non-conforming trajectories identified for a convective weather day at JFK, HKG and GRU are shown in Figure 3-38. It is observed that most of the trajectory deviations are caused by airborne holding, excessive vectoring or rerouting, in other words, tactical ATC instructions that tend to increase the length of the trajectory.

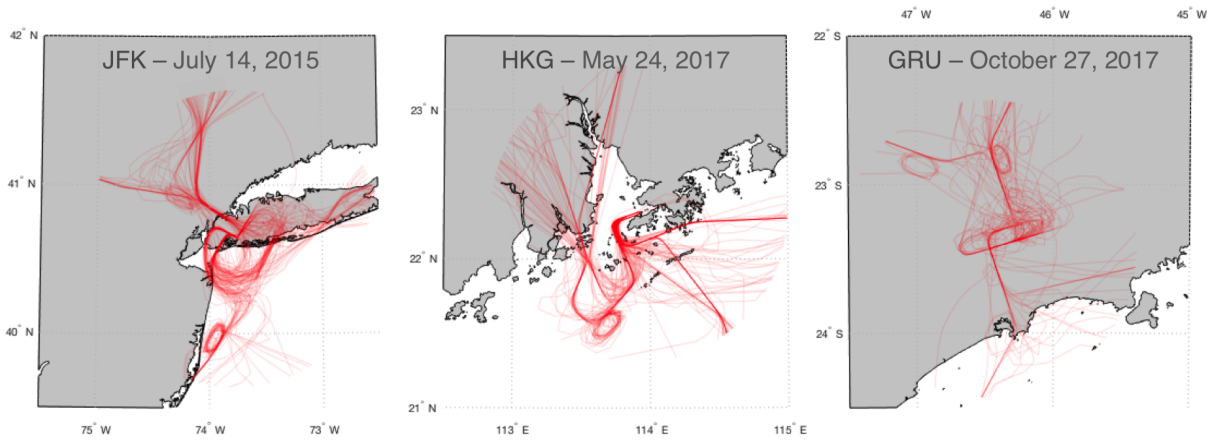


Figure 3-38: Example non-conforming arrival trajectories identified for a convective weather day at JFK, HKG and GRU.

3.5.2 Capacity

Capacity generally refers to an upper bound on the allowable throughput of a facility. Two common definitions of capacity are the *maximum capacity* (or *saturation capacity*) and the *practical capacity* [29]. The *maximum capacity* is defined as the expected number of movements under persistent demand. The *practical capacity* extends this definition by including the notion of level of service and specifying a threshold on the expected level of delay experienced by the users of the facility. It recognizes that a certain level of delay should be acceptable in order to ensure a steady stream of demand at the facility, but it should also be reasonable to ensure a sustainable operation of the facility through time. Indeed, this balance between throughput and delay is key for the capacity declaration process.

We developed an empirical approach for assessing metroplex capacity (maximum and practical) from historical terminal area throughput and delay performance. Rather than treating each airport individually, metroplex airports are looked from a systems perspective. Figures 3-39 and 3-40 shows how the arrival throughput and the excess transit times in the New York terminal area change as a function of the arrival demand, under VMC. Specifically, for every minute, the arrival throughput in the next 15-minute period and the excess transit time for each flight that landed in the same 15-minute period is plotted as a function of the number of arriving aircraft simultaneously present in the terminal area. Clearly, as the

arrival demand increases, the arrival throughput increases in a disproportionate fashion until reaching a saturation level, i.e., point at which delivering more aircraft to the terminal area does not significantly increase throughput because the system has reached its capacity. This behavior is reflected in the excess transit time curve. When the arrival demand increases, more aircraft are competing for the same resources and will encounter a higher probability of a delay assignment during the runway sequencing and scheduling process. When the saturation level is reached, the excess transit time increases much more rapidly. The plots also reveal significant variability in throughput and delay for the same level of demand. Our assumption is that part of this variability can be explained by differences in performance across metroplex configurations.

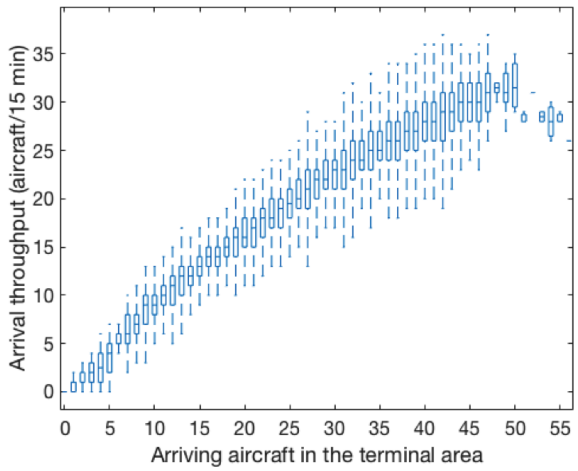


Figure 3-39: New York metroplex arrival throughput as a function of the arrival demand in the terminal area.

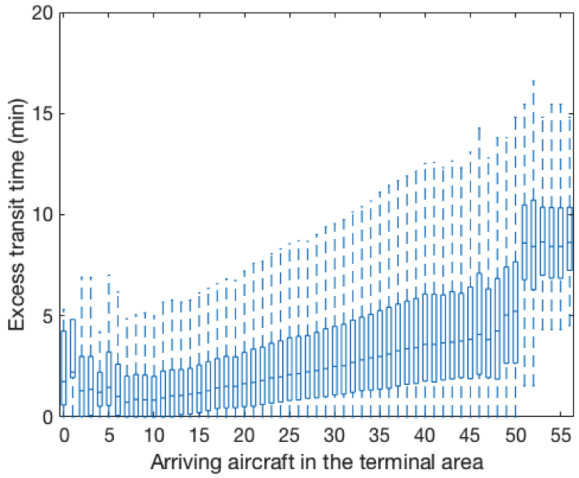


Figure 3-40: Excess terminal area transit time as a function of the arrival demand.

Our goal was to estimate the arrival throughput and the excess terminal area transit time as a function of the arrival demand in order to obtain performance curves that correlate throughput and level of delay in the terminal area for each MFP. In that way, the curves would allow to quantify the expected level of delay associated with a specific throughput level and could be used to determine arrival rates tailored to a given level of service.

We formalize the throughput estimation problem as a regression problem. Considering the observed throughput and delay profiles shown in Figures 3-39 and 3-40, we modeled the arrival throughput and the excess transit time in the terminal area with a piecewise

linear function of the arrival demand (number of arriving aircraft in the terminal area), with additional constraints arising from expected operational behavior: the arrival throughput was modeled as a non-decreasing and concave function of the arrival demand and the excess transit time was modeled as a non-decreasing function of the arrival demand. The parameters of the curves were estimated using quantile regression [134].

The quantile regression problem was formulated as a linear programming model and solved with Gurobi 6.5.0. Given N observations of arrival demand a, arrival throughput t and excess transit time d, LP-I determines the p-quantile regression fit of arrival throughput as a function of arrival demand ($\tilde{t}_n = \alpha_i + \beta_i a_n$) and LP-II determines p-quantile regression fit of excess transit time (delay) as a function of arrival demand ($\tilde{d}_n = \alpha_i + \beta_i a_n$). In LP-I, Equations 3.6, 3.7 and 3.8 frame the quantile regression problem as a linear programming model, as described in [134]. Equations 3.9 and 3.10 model the arrival throughput as a concave and non-decreasing function of the arrival demand, while Equation 3.11 ensures the continuity of the piecewise function. Similarly, in LP-II, Equations 3.12, 3.13 and 3.14 frame the quantile regression problem as a linear programming model, Equation 3.15 models the excess transit time as a non-decreasing function of the arrival demand and Equation 3.16 ensures the continuity of the piecewise function.

LP-I:

$$\min_{\alpha,\beta,y} \sum_{n=1}^{N} y_n \tag{3.6}$$

S.T.:

$$y_n \geq p\big(t_n - (\alpha_i + \beta_i a_n)\big), \quad i-1 \leq a_n \leq i, \quad \forall n \tag{3.7}$$

$$y_n \geq (1-p)\big((\alpha_i + \beta_i a_n) - t_n\big), \quad i-1 \leq a_n \leq i, \quad \forall n \tag{3.8}$$

$$\beta_{k+1} - \beta_k \leq 0, \quad \forall k = 1, ..., a_{max} - 1 \tag{3.9}$$

$$\alpha_k + \beta_k k \leq \alpha_{k+1} + \beta_{k+1}(k+1), \quad \forall k = 1, ..., a_{max} - 1 \tag{3.10}$$

$$\alpha_k + \beta_k k = \alpha_{k+1} + \beta_{k+1} k, \quad \forall k = 1, ..., a_{max} - 1 \tag{3.11}$$

LP-II:

$$\min_{\alpha,\beta,y} \sum_{n=1}^{N} y_n \tag{3.12}$$

S.T.:

$$y_n \geq p\big(d_n - (\alpha_i + \beta_i a_n)\big), \quad i-1 \leq a_n \leq i, \quad \forall n \tag{3.13}$$

$$y_n \geq (1-p)\big((\alpha_i + \beta_i a_n) - d_n\big), \quad i-1 \leq a_n \leq i, \quad \forall n \tag{3.14}$$

$$\alpha_k + \beta_k k \leq \alpha_{k+1} + \beta_{k+1}(k+1), \quad \forall k = 1, ..., a_{max} - 1 \tag{3.15}$$

$$\alpha_k + \beta_k k = \alpha_{k+1} + \beta_{k+1} k, \quad \forall k = 1, ..., a_{max} - 1 \tag{3.16}$$

The performance curves obtained with a median regression fit ($p = 0.5$) for each MFP in the New York, Hong Kong and Sao Paulo metroplexes are presented in Figures 3-41, 3-42, 3-43 , 3-44, 3-45 and 3-46. The curves were estimated using VMC observations in order to marginalize out weather related throughput impacts and enable a cleaner assessment of metroplex configuration performance.

The plots reveal that New York is the highest throughput system, followed by Hong Kong and Sao Paulo. The median arrival throughput of the dominant MFP saturates at 34 aircraft/15min (or 136 aircraft/hour) in the New York metroplex, 14 aircraft/15min (or 56 aircraft/hour) in the Hong Kong metroplex and 13 aircraft/15min (or 52 aircraft/hour) in the Sao Paulo metroplex.

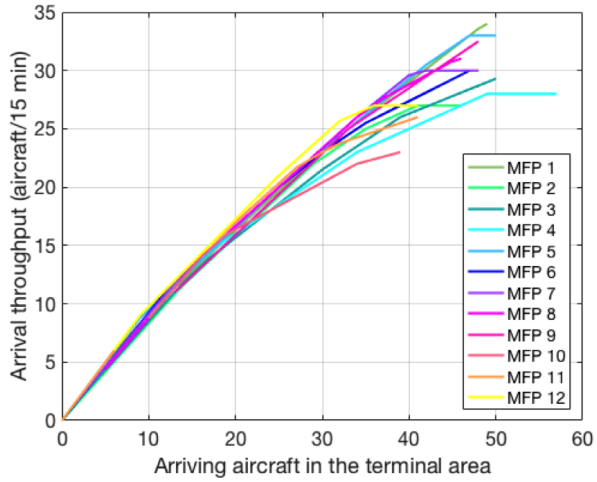


Figure 3-41: Throughput-demand curves for the New York MFPs.

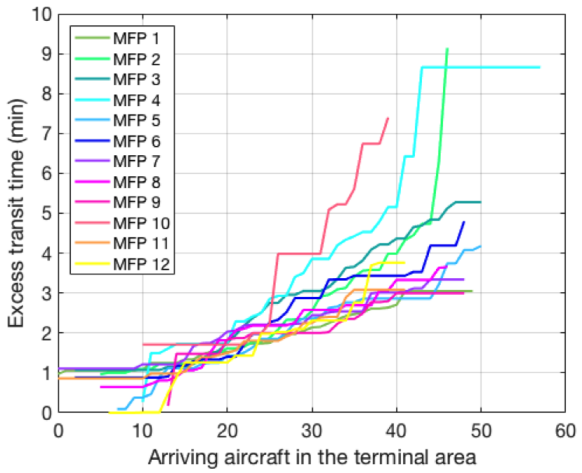


Figure 3-42: Delay-demand curves for the New York MFPs.

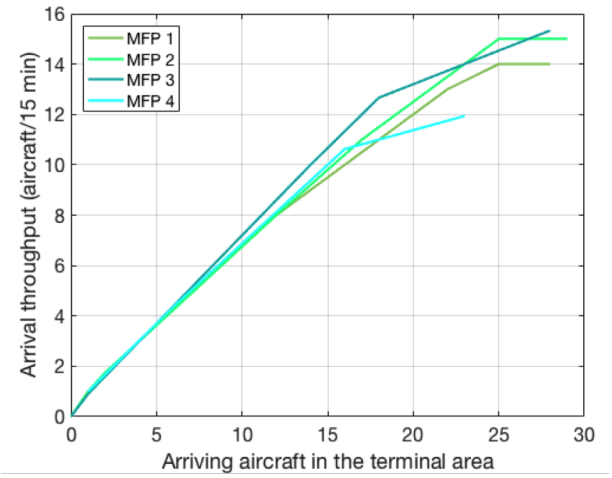


Figure 3-43: Throughput-demand curves for the Hong Kong MFPs.

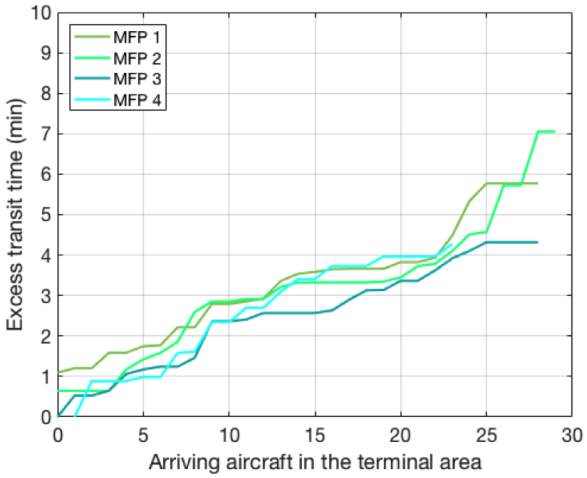


Figure 3-44: Delay-demand curves for the Hong Kong MFPs.

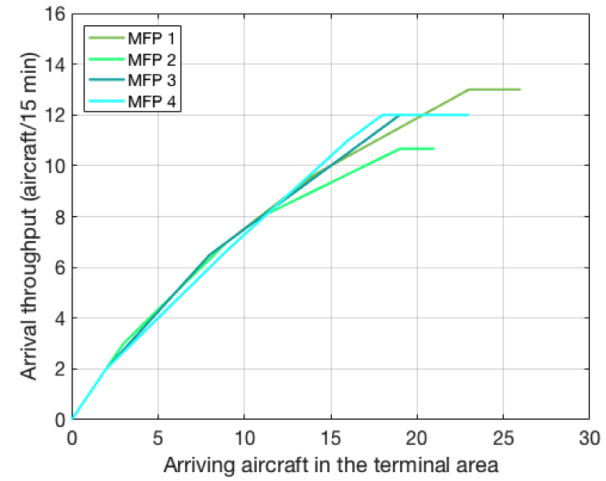


Figure 3-45: Throughput-demand curves for the Sao Paulo MFPs.

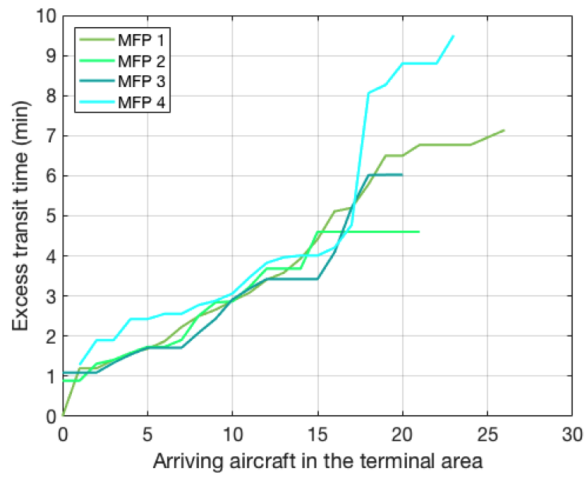


Figure 3-46: Delay-demand curves for the Sao Paulo MFPs.

The maximum throughput as well as the median level of delay associated with it are presented in Table 3.10 for all MFPs. The maximum throughput normalized by the number of active arrival runways associated with each particular MFP is also presented in Figure 3-47 for facilitating the comparison between metroplexes. The New York metroplex stands out with the largest standard deviation in throughput performance, emphasizing its operational variability, as also observed in the previous section. MFP 1 shows the highest arrival rate (136 aircraft/hour), while MFP 10 shows the lowest arrival rate (92 aircraft/hour). The median level of delay at saturation is 3.7 minutes, but the standard deviation is high. MFPs

4 and 10 show the highest delay levels of 8.7 minutes and 7.4 minutes, respectively. In the Hong Kong and Sao Paulo metroplexes, there is smaller variation in throughput performance. For Hong Kong, the median level of delay at saturation is 4.4 minutes. Sao Paulo stands out with a relatively higher median delay level of 6.4 minutes.

Table 3.10: Metroplex arrival rates under persistent demand.

MFP	New York		Hong Kong		Sao Paulo	
	Rate (aircraft/h)	Delay (min)	Rate (aircraft/h)	Delay (min)	Rate (aircraft/h)	Delay (min)
1	136	3.1	56	5.8	52	6.8
2	108	4.3	60	4.6	44	4.6
3	116	5.3	60	4.3	48	6.0
4	112	8.7	48	4.3	48	8.1
5	132	3.7	-	-	-	-
6	120	4.2	-	-	-	-
7	120	3.1	-	-	-	-
8	124	3.7	-	-	-	-
9	130	3.0	-	-	-	-
10	92	7.4	-	-	-	-
11	104	3.1	-	-	-	-
12	108	2.8	-	-	-	-
Median	**118**	**3.7**	**58**	**4.4**	**48**	**6.4**
Standard deviation	**12.8**	**1.9**	**5.7**	**0.7**	**3.3**	**1.4**

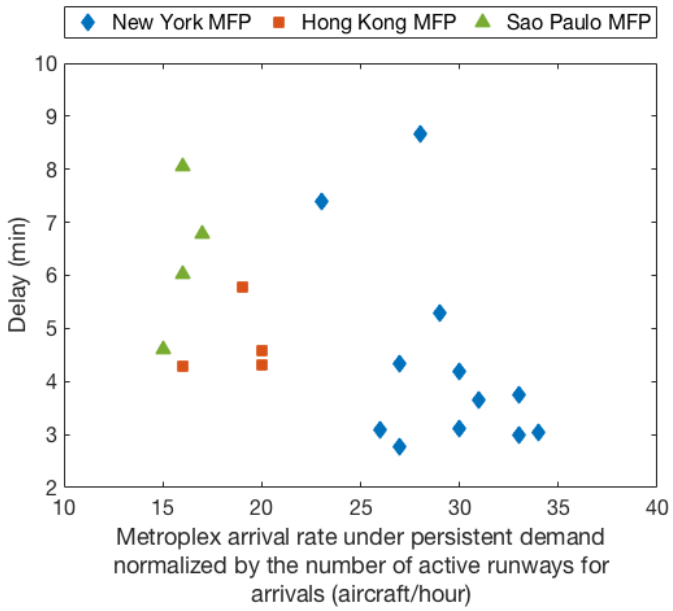


Figure 3-47: Metroplex arrival rates under persistent demand and associated level of delay.

In order to exemplify the use of the performance curves for also assessing the practical capacity, we obtained the metroplex arrival rates corresponding to a specific level of service of 4-minute of terminal area delay for each New York MFP. They are presented in Table 3.11 and were compared with baseline metroplex arrival rates obtained by summing the individual airport capacity declared by the FAA [135] for the runway configurations associated with each MFP. The comparison reveals that, for some MFPs, the differences between the empirical and the baseline rates are large, indicating that higher levels of delay should be expected when the New York system is operating at the declared capacity under these operational modes. It is also observed that MFPs with the same theoretical runway system capacity can show very different throughput performance. For instance, MFPs 2, 3 and 4 have the same declared runway system capacity, but the differences in metroplex practical capacity can be as great as 20 aircraft per hour. The results highlight the impacts of the terminal area flow structure on the multi-airport system capacity and emphasize the importance of taking a systems perspective for metroplex airport capacity and flow management.

Table 3.11: Comparison between empirical and baseline metroplex arrival rates for each New York MFP.

MFP	**Empirical rates**	**Baseline rates**	**Difference (baseline - empirical)**
1	136	138	2
2	108	126	18
3	100	126	26
4	88	126	38
5	132	139	7
6	120	123	3
7	120	137	17
8	124	137	13
9	132	138	6
10	84	111	27
11	104	123	19
12	108	123	15

3.5.3 Predictability

Predictability is defined by ICAO [133] as the ability of airspace users to provide consistent and dependable levels of performance. It is typically assessed by measuring the variation in

system performance experienced by the users [136]. In this section, we analyze the variability in performance across MFPs and within a MFP and discuss interrelations between ATM configuration predictability and ATM performance predictability.

ATM performance variability due to ATM configuration variability

As indicated in Table 3.2, the New York metroplex shows the lowest MFP dwell times and the highest number of changes in the flow structure per day, suggesting the ATM configuration is less predictable in this system. We translated this variation in ATM configuration into daily capacity variation and structural path stretch variation in order to analyze how the ATM configuration predictability affects ATM performance predictability.

Table 3.12 shows the average accumulated change in capacity and terminal area structural path stretch per day for each metroplex. For this, we considered the observed MFPs for each day and their associated maximum achievable throughput (normalized by the number of runways) and structural path stretch, as shown in Table 3.10 and 3-34. It is observed that the New York system presents the highest capacity variation per day (11 aircraft). For Hong Kong and Sao Paulo, there is a small capacity variation of less than 2 aircraft. From an operational perspective, this means the New York system is more reliant on traffic flow management in order to adjust demand with the more dynamic system capacity. In terms of structural path stretch variation, Hong Kong stands out with the highest value, which suggests lower trajectory lateral performance predictability. It is important to mention that, in the context of TBO, predictability of the ATM configuration and associated distances to be flown is important for effective trajectory optimization.

Table 3.12: Daily capacity and structural path stretch variation.

	New York	**Hong Kong**	**Sao Paulo**
Capacity variation	11.2	1.8	1.4
Structural path stretch variation (nm)	8.5	9.8	1.4

ATM performance variability within an ATM configuration

We investigated the variability in ATM configuration performance by analyzing the predictability power of the capacity curves empirically derived from the data. For this, we calculated the Normalized Root Mean Squared Error (NRMSE) associated with the median regression fit for each MFP throughput-demand curve:

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{sd(\mathbf{y})} \tag{3.17}$$

where $\hat{y}$ is the predicted throughput based on the capacity curve, y is the actual observed throughput, n is the total number of observations and $sd(\mathbf{y})$ is the standard deviation of the observed throughput.

Table 3.13 shows the NRMSE for the MFP throughput-demand curves for each metroplex. It is observed that the New York metroplex has both the most predictable MFP (lowest NRMSE) and the least predictable MFP (highest NRMSE) from a capacity standpoint. In all metroplexes, it is observed that the least predictable MFPs were those with the lowest throughput performance.

Table 3.13: NRMSE for the MFP throughput-demand curves.

MFP	**New York**	**Hong Kong**	**Sao Paulo**
1	0.375	0.453	0.434
2	0.412	0.440	0.517
3	0.419	0.534	0.386
4	0.549	0.575	0.481
5	0.414	-	-
6	0.356	-	-
7	0.348	-	-
8	0.423	-	-
9	0.527	-	-
10	0.692	-	-
11	0.435	-	-
12	0.394	-	-

3.6 Discussion

The characterization of terminal area traffic flows in the New York, Hong Kong and Sao Paulo metroplexes revealed structural and operational differences between these multi-airport systems. First, the New York metroplex is observed to have a more complex route design, with a higher number of routes and a higher number of interactions between them. A significant number of these interactions are located close to the airport and were characterized as flow dependencies, revealing the existence of impracticable combinations of runway configuration and routes. On the other hand, for the Hong Kong and Sao Paulo systems, most of the inter-airport route interactions are located further from the airports and were characterized by larger airspace volumes resulting from the sharing of both airspace fixes and path segments.

The analysis also showed the New York system is the one with more dynamic changes in the terminal area flow behavior during daily operations. While the metroplex traffic flows can be categorized in few patterns in the Hong Kong and Sao Paulo systems, they show much higher variability in the New York system. This variability can be generally attributed to a couple of factors: higher variability in the set of runway configurations by airport; existence of inter-airport flow dependencies that create operationally infeasible combinations of runway configuration and routes; higher airspace design complexity that enables ATM to tailor operations for VMC and IMC. The New York metroplex was also found to be the most interdependent system, with a significant number of flow dependencies, and the typical terminal area flow structures were found to have a higher number of flow crossings, suggesting higher levels of air traffic control complexity.

By analyzing the impacts of such structural and operational differences on performance, we found the more complex New York airspace presents the most efficient design and use in terms of trajectory lateral performance. The Hong Kong airspace design showed a high level of structural path stretch, which is potentially the cost of its more decoupled airspace, with fewer flow interactions with the neighboring airports. On average, the New York metroplex showed the highest traffic flow efficiency, both spatially and temporally. Yet, it exhibited the highest variability in traffic flow efficiency as well as more pronounced efficiency drops during inclement weather conditions. The Hong Kong metroplex showed a significantly lower lateral

traffic flow efficiency, which was observed to be a result not only of its least inefficient airspace design, but also of tactical traffic flow management, as revealed by its lower temporal traffic flow efficiency and trajectory conformance. In terms of capacity, the New York metroplex was found to be the highest throughput system, but stood out with the largest variability in throughput performance. Differences in metroplex arrival capacity across MFPs were as great as 44 aircraft per hour. Such differences were associated not only with its more diverse set of runway configurations, but also with differences in the terminal area flow structure.

The more dynamic New York metroplex flow behavior, particularly influenced by meteorological conditions, suggests the coupled ATM configuration in this multi-airport system as well as its performance might be less predictable. Indeed, we observed the New York system exhibited the highest variability in traffic flow efficiency and capacity by MFP. From an operational perspective, this means the New York system is more reliant on traffic flow management in order to adjust demand with the more dynamic system capacity. At the same time, it also means traffic flow management tends to become more challenging, since it has to deal with increased levels of uncertainty in the ATM configuration and associated constraints.

Overall, the results highlight different action areas to be considered at each system towards improving their terminal area operations. The high level of structural path stretch in the Hong Kong metroplex, particularly for HKG, reveals opportunities for improvement in the airspace structure in order to increase trajectory lateral efficiency. Enhancements in airspace design in the Hong Kong and Sao Paulo metroplexes may also be performed to decouple routes from different airports that share the same airspace region in order to achieve a more expeditious flow. The lower level of trajectory conformance and temporal efficiency in the Hong Kong and Sao Paulo metroplexes also reveals opportunities for improvement in tactical traffic flow management towards better sequencing and scheduling of arriving flights. Finally, the higher level of airport interdependency, dynamism, weather dependency and performance variability in the New York metroplex reveal opportunities for (1) improving the airspace design in order to de-conflict flows close to the airports and (2) integrating weather forecasting with decision-making in order to better anticipate weather conditions, translate them into operational impact and plan traffic flow management efficiently.

THIS PAGE INTENTIONALLY LEFT BLANK

Chapter 4

Data-Driven Approach for Metroplex Configuration and Capacity Planning

In Chapter 3, we demonstrated the use of the flight trajectory data analytics framework to identify and characterize traffic flow patterns in the terminal area of multi-airport systems. As the observed patterns in the terminal area traffic flow revealed recurrent utilization patterns of runways and airspace, we were able to identify the major configurations in which a metroplex system collectively operates as well as key intervening factors. The results also emphasized that metroplex systems can have very dynamic airspace use and show high variability in throughput performance across different configurations. From an operational perspective, this has two major implications. First, they are more reliant on ATFM to adjust demand with the more dynamic system capacity. Second, the planning of ATFM is more challenging. Because there is higher uncertainty in the ATM configuration and its performance, estimating future capacity precisely is more difficult and the allocation of airport and airspace resources under uncertainty becomes more subject to inefficiencies. In this chapter, we show how the knowledge generated by the characterization of metroplex flow patterns, provided by the flight trajectory data analytics framework, can be leveraged to develop descriptive and prescriptive models for metroplex configuration and capacity planning towards improved traffic flow management decision support. The analysis is focused on the New York metroplex.

4.1 Metroplex Configuration and Airport Capacity Planning Framework

The data-driven approach for metroplex configuration and airport capacity planning developed in this chapter has two components: a *capacity prediction* module and a *capacity allocation* module. It is illustrated in Figure 4-1.

The first contribution is the development of predictive models for capacity estimation that specifically accounts for operational interdependency aspects existing in a metroplex environment. The capacity prediction module uses machine learning methods to translate weather forecasts into probabilistic forecasts of the metroplex configuration and airport capacity for strategic time horizons. A traffic flow pattern prediction model first delivers a forecast of the most likely flow structure to be seen in the terminal airspace on an hourly basis. At any given time, the terminal area flow pattern reflects coupled runway and airspace configuration decisions and interactions between arrival and departure procedures, providing a complete picture of the multi-airport system state. An airport capacity prediction model then takes as input the metroplex flow pattern forecast and weather features representing the forecast environmental conditions in the terminal area to deliver arrival capacity forecasts.

The second contribution is the development of an AAR planning model for airport capacity allocation that is intended to support the design of strategic TMIs such as GDPs. The capacity allocation module uses optimization methods in order to prescribe optimal AARs for strategic traffic flow management. The AAR planning model builds upon existing GHP optimization models, but instead of relying on theoretical capacity modeling or capacity scenario trees, it directly incorporates the output generated by the capacity prediction module. Besides, it is designed to incorporate robustness goals. In summary, the model has the following key features: (1) it is stochastic, i.e., capacity uncertainty is accounted for in the form of capacity profiles drawn from actual predictive distributions; (2) it enables revision of the plan in order to take advantage of updated information and more accurate predictions throughout the planning horizon; (3) it incorporates robustness goals, i.e., the decision-maker has the ability to select a desired level of robustness for the ATFM plan based on its relative valuation between efficiency and predictability.

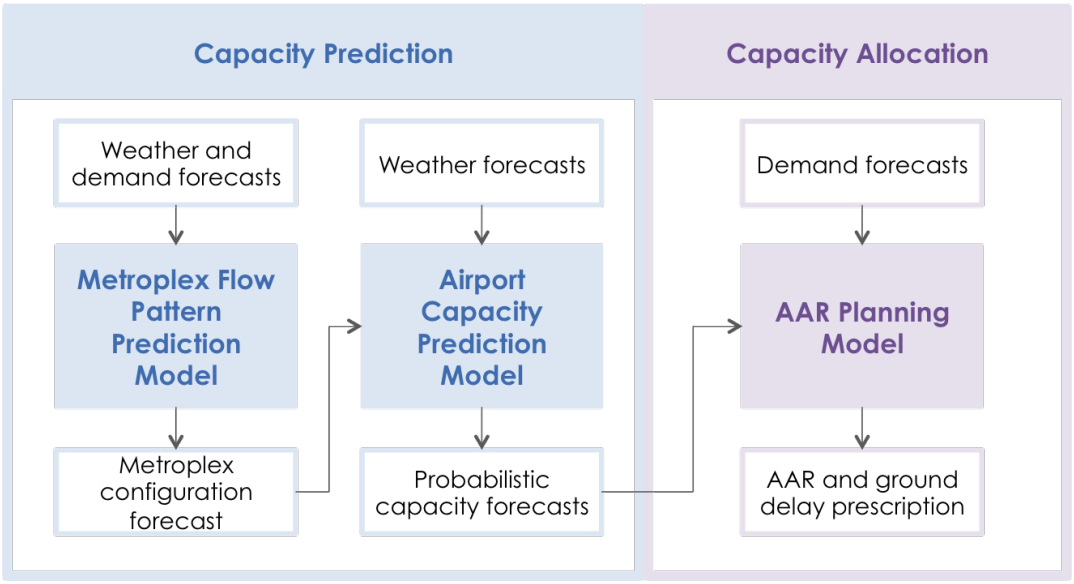


Figure 4-1: Metroplex configuration and airport capacity planning framework.

4.2 Data Description

The datasets used for this study covers the same period of 69 days from 2013-2015 (discretized in hourly periods from 6:00 am to 11:00 pm (local time)) considered for the analysis of traffic flows in the New York metroplex in Chapter 3; 37 are fair weather days and 32 are weather-impacted days, which include convective weather impacts and non-convective impacts, such as adverse wind conditions and low ceiling/visibility. The data was randomly partitioned into training and test datasets, composed of 70% of the days and 30% of the days, respectively. The partitioning was performed separately for fair weather days and weather-impacted days in order to ensure the training and test datasets had a high variability of weather conditions as in the original dataset. The data sources are described below.

4.2.1 Terminal Aerodrome Forecast (TAF)

TAF is a concise weather report that states the expected meteorological conditions at an airport during a specified period of time (usually 24 hours). It is routinely issued every 6 hours (at 0, 6, 18 and 24 UTC), but an amended TAF can also be issued at any time if the expected meteorological conditions change significantly. The following weather features were

extracted from TAF: wind direction, wind speed, visibility and ceiling. Although ceiling is not directly reported in TAF, it can be extracted as the lowest height of broken or overcast cloud layer, as defined by the FAA [137]. For each time t marking the beginning of a forecast period, the most recent TAF (issued at $t' \leq t$) was considered and the weather features corresponding to each hourly time period in the forecast horizon were extracted for prediction. For instance, if $t' = t - 1$, a 1-hour look-ahead capacity prediction for $[t, t + 1]$ will use a 2-hour look-ahead weather forecast in TAF.

4.2.2 Arrival Route Status and Impact (ARSI) Forecast

ARSI is the arrival version of the Route Availability Planning Tool (RAPT), a decision support tool that provides information about departure routes affected by convective weather [138]. Starting at the current time and extending out 30 minutes into the future in 5-minute intervals, RAPT assigns a departure route status that indicates the expected convective weather impact along the route (0 for clear, 1 for insignificant weather encountered, 2 for partial or uncertain blockage and 3 for blocked). Similarly, ARSI assigns an arrival route status in 15-minute intervals up to 60 minutes into the future. As the ARSI forecast is made with respect to the time the flight is expected to reach the transition fix in the arrival route, its conditions will be reflected at the airport after some lead time (corresponding to the length of the arrival routes embedded in the tool, which is approximately 1 hour). Therefore, ARSI forecasts were used as predictive features for up to 2-hour arrival capacity forecasts. For a given time t, the mean of the ARSI forecast $\{(t - 60min) + 0min, (t - 45min) + 0min, (t - 30min) + 0min, (t - 15min) + 0min\}$ was used as a predictive feature for a 1-hour prediction, whereas the mean of the ARSI forecast $\{t + 0min, t + 15min, t + 30min, t + 45min\}$ was used as a predictive feature for a 2-hour prediction.

4.2.3 Hourly Airport Reports

Hourly airport reports ("Airport Information" and "Airport Efficiency Information") from the FAA Aviation System Performance Metrics (ASPM) database were used to obtain data about flight schedules and arrival/departure rates.

4.2.4 Flight Tracks

One of the challenges for capacity prediction using supervised learning concerns the derivation of the knowledge about actual capacity (output "truth"). In this thesis, the airport arrival capacity information (actual output) required for building the machine learning models was derived from historical arrival throughput after filtering out periods of low demand in which the airport is not running at capacity. For this, we leveraged flight track data and the empirical approach for capacity estimation described in Section 3.6.2 of Chapter 3. Specifically, we empirically estimated a threshold N^* for the number of arriving aircraft present in the terminal area (arrival demand) for which the airport system can be considered to be under pressure and only used the arrival throughput values for which the demand was equal or above the threshold ($N \geq N^*$). Figure 4-2 shows the histogram of hourly arrival rates resulting from the filtering process.

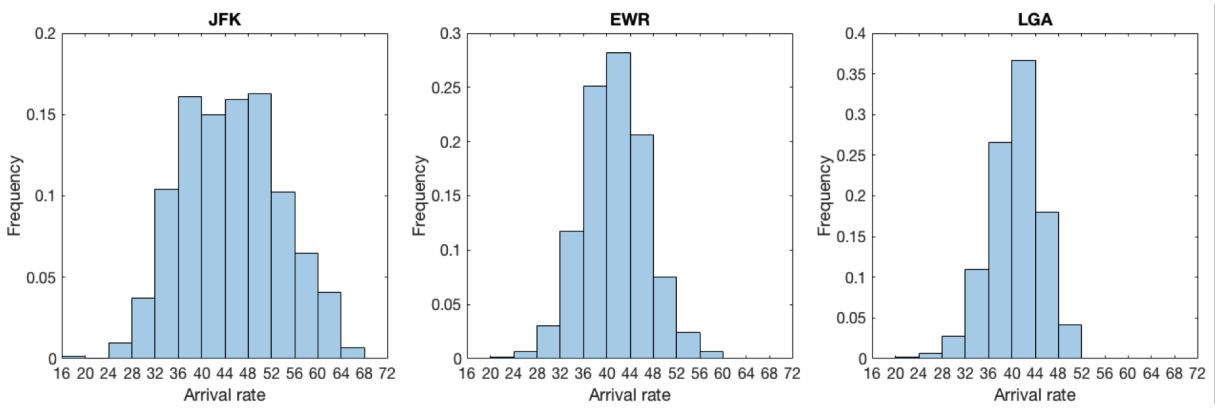


Figure 4-2: Histogram of hourly airport arrival rates.

4.3 Prediction Models and Features

The capacity prediction module relies on two models: a traffic flow pattern prediction model and an airport arrival capacity prediction model. The prediction problems are framed as supervised learning classification/regression problems. In supervised learning, the goal is to learn a mapping between a set of input variables and a given output variable based on knowledge extracted from a training dataset.

A multi-way classification model is first developed to predict the most likely pattern

in which the metroplex flows might be for strategic time horizons. For this, we leverage the knowledge about the metroplex flow patterns obtained with the flight trajectory data analytics framework. As they reflect utilization patterns of runway configurations and arrival/departure route structure, the traffic flow pattern prediction model enables the forecasting of the global metroplex configuration state. The multi-way classification model is represented as

$$z_t = g(u_t, z_{t-1}) \tag{4.1}$$

where z_t is the metroplex flow pattern at time t and u_t is the input vector with the relevant factors driving the metroplex flow dynamics, as discussed in Chapter 3:

$$u_t = \Big(t, (wspeed_{k,t})_{k\in M}, (wdir_{k,t})_{k\in M}, (imc_{k,t})_{k\in M}, arr_t, dep_t\Big) \tag{4.2}$$

where t is the time of the day, $(wspeed_{k,t})_{k\in M}$ is the vector of forecast wind speed at time t for each airport k in the set M of metroplex airports, $(wdir_{k,t})_{k\in M}$ is the vector of forecast wind direction, $(imc_{k,t})_{k\in M}$ is the vector of forecast meteorological conditions (0, if VMC, and 1, if IMC) and arr_t and dep_t are the scheduled aggregate arrival and departure demands for the metroplex, respectively, at time t.

In a subsequent step, a regression model is developed to map forecast weather conditions and metroplex configuration state into flow rates for each individual metroplex airport. It is represented as

$$y_{k,t} = f(t, w_{k,t}, y_{k,t-1}, z_t) \tag{4.3}$$

where $y_{k,t}$ is the arrival rate output at time t for airport k, z_t is the expected metroplex configuration state at time t and $w_{k,t}$ is a feature vector encompassing all weather forecast features at time t for airport k, as follows:

$$w_{k,t} = \Big(ceil_{k,t}, vis_{k,t}, head_{k,t}, tail_{k,t}, cross_{k,t}, convec^N_{k,t}, convec^S_{k,t}, convec^W_{k,t}\Big) \tag{4.4}$$

where $ceil_{k,t}$ is the ceiling forecast at time t for airport k, $vis_{k,t}$ is the visibility forecast, $head_{k,t}$, $tail_{k,t}$ and $cross_{k,t}$ are the headwind, tailwind and crosswind forecasts, respectively, at the dominant arrival runway at airport k for metroplex configuration z_t, and, finally, $convec^N_{k,t}$, $convec^S_{k,t}$ and $convec^W_{k,t}$ are the forecast convective blockage status for the dominant north, south and west arrival routes for airport k.

Given the ATFM planning needs for capacity forecasts with proper quantification of uncertainty, a key requirement of the prediction framework is to provide the complete predictive distribution of the output. Besides, it has to be able to generate forecasts for each time period throughout the ATFM planning horizon. This is accomplished through an iterative procedure in which the regression model is sequentially used to make predictions for each time period based on the output from the previous time period, as illustrated in Figure 4-3. In this process, a Monte Carlo sampling approach is used to propagate the uncertainty about the metroplex configuration state and about the arrival rate from the beginning to the end of the planning horizon. For instance, given the predictive distribution of arrival rates for the first forecast hour, a random sample is taken from this distribution and used to predict the arrival rate for the second hour, resulting in a conditioned predictive distribution for the second hour arrival rate. The process continues for the rest of the planning horizon and, once completed, is repeated n times. As a result, we also obtain n potential capacity profiles drawn from the predictive distributions that can be directly used for ATFM planning. The following pseudo code describes the algorithmic process for arrival rate forecasting throughout the planning horizon and generation of capacity profiles. The subscript k indicating each airport was suppressed to not overload the notation.

for test instance $i = 1, ..., p$

  for iteration $n = 1, ..., 1000$

    Draw z^i_{t+1} from $Z_{t+1} \sim g_{Z_{t+1}}\left(u^i_{t+1}, z^i_t\right)$

    Draw y^i_{t+1} from $Y_{t+1} \sim f_{Y_{t+1}}\left(w^i_{t+1}, y^i_t, z^i_{t+1}\right)$

    for forecast horizon $h = 2, ..., H$

      Draw z^i_{t+h} from $Z_{t+h} \sim g_{Z_{t+h}}\left(u^i_{t+h}, z^i_{t+h-1}\right)$

      Draw y^i_{t+h} from $Y_{t+h} \sim f_{Y_{t+h}}\left(w^i_{t+h}, y^i_{t+h-1}, z^i_{t+h}\right)$

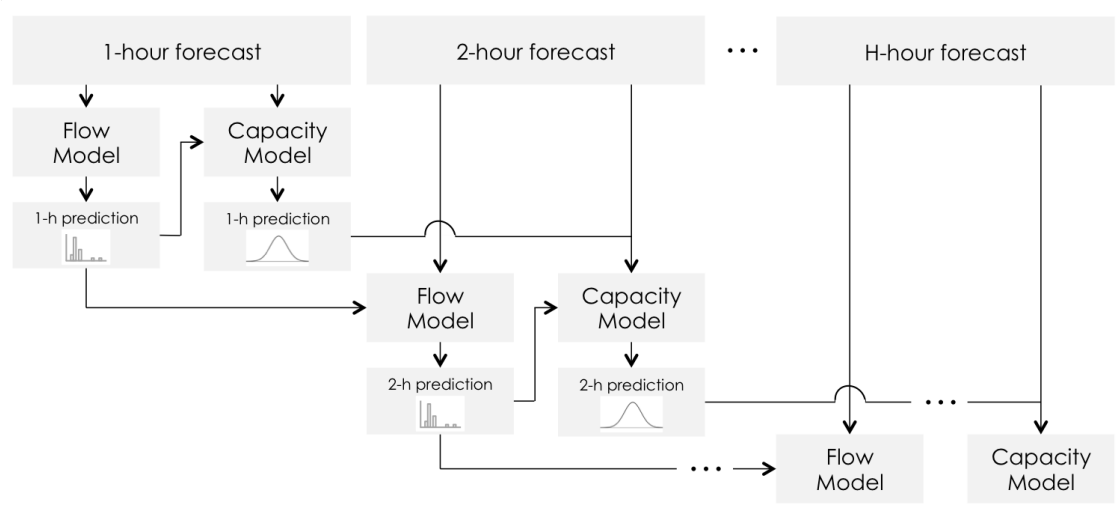


Figure 4-3: Forecasting procedure throughout the planning horizon.

4.4 Prediction of Traffic Flow Patterns

The prediction of traffic flow patterns is framed as a supervised multi-way classification problem. We used state-of-the-art methods for machine learning (Neural Network, Random Forests and Support Vector Machine) to build a set of candidate classification models for this purpose, which were then evaluated with respect to predictive performance. The following sub-sections provide the details about the models created. To simplify the notation, the traffic flow pattern mapping $z_t = g(u_t, z_{t-1})$, defined in Equation 4.1, is represented as $z = g(x)$. All models were developed with MATLAB R2017a.

4.4.1 Classification Models

Neural Network Classification Model (NN)

An artificial neural network is a non-parametric method that can be used for both classification and regression. It consists of a sequence of layers containing artificial nodes (neurons) that compute some function of the previous layer in a feed-forward fashion from the input layer to the output layer [139].

Different network architectures were tested, with varied number of layers and neurons. Based on 5-fold cross validation performance (Table 4.1), the neural network model designed

for the multi-way classification problem was composed of 1 input layer, 1 hidden layer with 30 neurons (m = 30) and 1 output layer. With this architecture, the output z is mapped from the input space $x \in \mathbb{R}^d$ as follows:

$$a_j = \sum_{i=1}^{d} x_i A_{ij} + A_{0j} \tag{4.5}$$

$$b_k = \sum_{j=1}^{m} h(a_j) B_{jk} + B_{0k} \tag{4.6}$$

$$z_k = h(b_k) \tag{4.7}$$

where a_j is the output of neuron j in the hidden layer, b_k is the output of neuron k in the output layer, $\theta = A_{ij}, A_{0j} \cup B_{jk}, B_{0k}$ are the layer parameters (weights and bias), tuned with back-propagation, and $h(a_j) = \frac{1}{1+e^{-a_j}}$ is the logistic activation function used in the hidden layer. Equation 4.5 maps the input layer to the hidden layer, Equation 4.6 maps the hidden layer to the output layer and Equation 4.7 applies a softmax function $h(b_k) = \frac{e^{b_k}}{\sum_{k=1}^{K} e^{b_k}}$ to the output layer in order to transform the neural network scores into posterior probabilities.

Table 4.1: Classification accuracy of NN model for various network architectures based on 5-fold cross validation performance.

		Number of nodes							
		5	10	15	20	25	30	35	40
Number of hidden layers	1	79.1	80.9	81.1	79.3	82.4	82.4	82	79.7
	2	76.6	81.3	81.4	80.7	81.3	82	80.5	80.9

Random Forests Classification Model (RF)

Random Forests is a non-parametric method for both classification and regression consisting of an ensemble of decision tree learners [114]. As the method has already been described in Chapter 2, the reader is referred to Section 2.4.1.

The algorithm has few parameters to tune: T is the number of trees in the ensemble, M is the number of features to sample at each split and N is the minimum number of observations

in a leaf node. Based on cross-validation performance (Figure 4-4), the RF model for traffic flow pattern prediction was created with $T = 100$, $M = 10$ and $N = 1$.

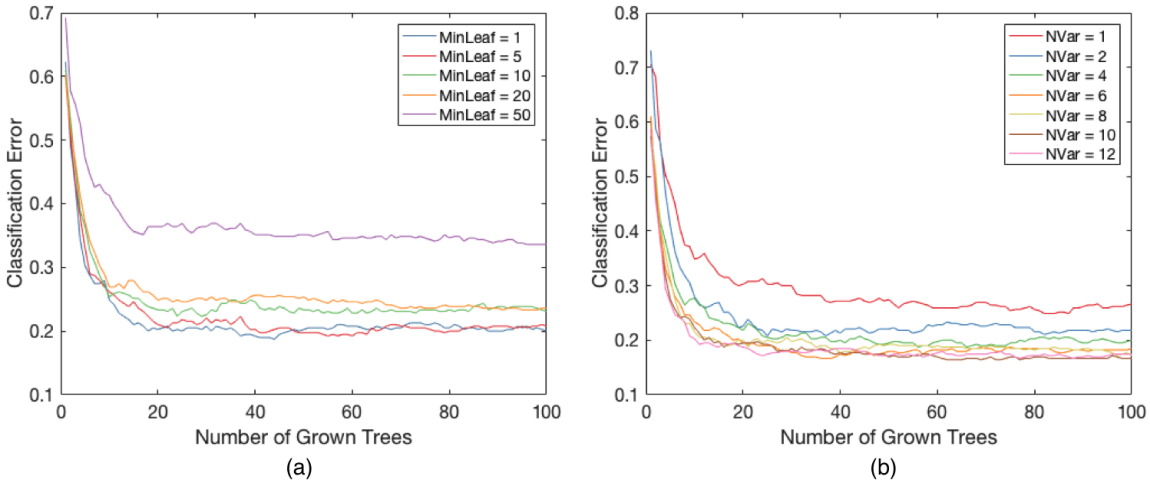


Figure 4-4: Parameter sensitivity of RF model based on 5-fold cross validation performance. (a) Impacts of the minimum size of leaf node on the classification error. (b) Impacts of the number of predictors to sample at each split on the classification error.

Support Vector Machine Classification Model (SVM)

Support Vector Machine outputs a decision boundary for binary classification by finding an optimal separating hyperplane $g(x) = a^T x + b$ such that observations mostly fall on the right side of the hyperplane ($g(x) \geq +1$ for $z = +1$, and $g(x) \leq -1$ for $z = -1$) and the distance from the hyperplane to the observations closest to it on either side (margin) is maximized for best generalization [113].

This problem can be formulated as an optimization problem:

$$\min \quad \frac{1}{2}\|a\|^2 + \lambda \sum \xi \tag{4.8}$$

Subject to:

$$z(a^T x + b) \geq 1 + \xi \tag{4.9}$$

$$\xi \geq 0 \tag{4.10}$$

In Equation 4.8, the first term of the objective function seeks to maximize the margin and the second term introduces a penalty factor for misclassifications. Equation 4.9 ensures that observations fall on the right side of the hyperplane, but with allowance for misclassifications based on the introduction of non-negative slack variables (Equation 4.10).

For the multi-way classification problem, SVM was applied with an ensemble of binary learners using a one-versus-one encoding. A binary SVM learner $l = 1, ..., L$ was created for each pairwise combination of classes. The final class $\hat{z}$ of a new observation was determined by maximizing the posterior probability over the binary learners:

$$\hat{z} = \arg\max_{k} \left\{ \frac{\sum_{l=1}^{L} |m_{kl}| p_l(z = m_{kl})}{\sum_{l=1}^{L} |m_{kl}|} \right\} \tag{4.11}$$

where $m_{kl} = 1$ if class k is coded as the positive class in binary learner l, $m_{kl} = -1$ if class k is coded as the negative class in binary learner l, and $m_{kl} = 0$ if class k is not considered in binary learner l.

For each binary learner, posterior probabilities were estimated by fitting a sigmoid function to the scores $g(x)$ based on cross-validation [140]:

$$p(z = 1/g(x)) = \frac{1}{1 + \exp(Ag(x) + B)} \tag{4.12}$$

In order to compute the final posterior probabilities from the ensemble of binary learners, we followed the approach of Wu and Lin [141]:

$$\min \quad \sum_{l=1}^{L} [-p_l(z = 1) \sum_{k=1}^{K} \hat{p}_k I(m_{kl} = -1) + (1 - p_l(z = 1)) \sum_{k=1}^{K} \hat{p}_k I(m_{kl} = +1)]^2 \tag{4.13}$$

Subject to:

$$0 \leq \hat{p}_k \leq 1 \tag{4.14}$$

$$\sum_{k} \hat{p}_k = 1 \tag{4.15}$$

We tested both a linear SVM model with a linear kernel $K(x, x') = x^T x'$ as well as a non-linear SVM model with the radial basis function kernel $K(x, x') = \exp(-\frac{\|x-x'\|^2}{2\sigma^2})$. Based on 5-fold cross-validation performance, the non-linear SVM model was chosen, as it showed classification accuracy of 83.0%, whereas the classification accuracy was 78.7% for the linear SVM model.

4.4.2 Performance Evaluation

The predictive performance of the classification models for traffic flow prediction was evaluated for up to six hours of look-ahead time, which is a typical time horizon for strategic traffic flow management. In order to have useful predictions for decision-making under uncertainty, probabilistic forecasts were generated for each time period throughout the planning horizon. For this, Bayes' rule was sequentially applied: for the first hour, the prediction model was run and a probability for each MFP was obtained; for the subsequent hours in the forecast horizon, the prediction model was run assuming a given outcome for the previous hour and the final probability was computed conditioned on the probability of the assumed outcome for the previous hour. For each hour t, the predicted flow pattern z_t^* then corresponds to the feasible one with maximum probability p_t among all C flow patterns:

$$z_t^* = \arg\max_x \{p_{t,x}\} \tag{4.16}$$

where:

$$p_{t,x} = \sum_{c=1}^{C} p(z_t = x / z_{t-1} = c).p_{t-1,c} \tag{4.17}$$

The feasibility of each MFP was determined based on crosswind and tailwind thresholds derived statistically from the data. Specifically, these thresholds represent the 99^{th} percentile of the observed crosswind and tailwind values on the major arrival/departure runways at each airport (Figures 4-5 and 4-6). For any given time, a MFP is considered to be feasible if the crosswind and tailwind values on the corresponding runways do not exceed the statistically derived thresholds.

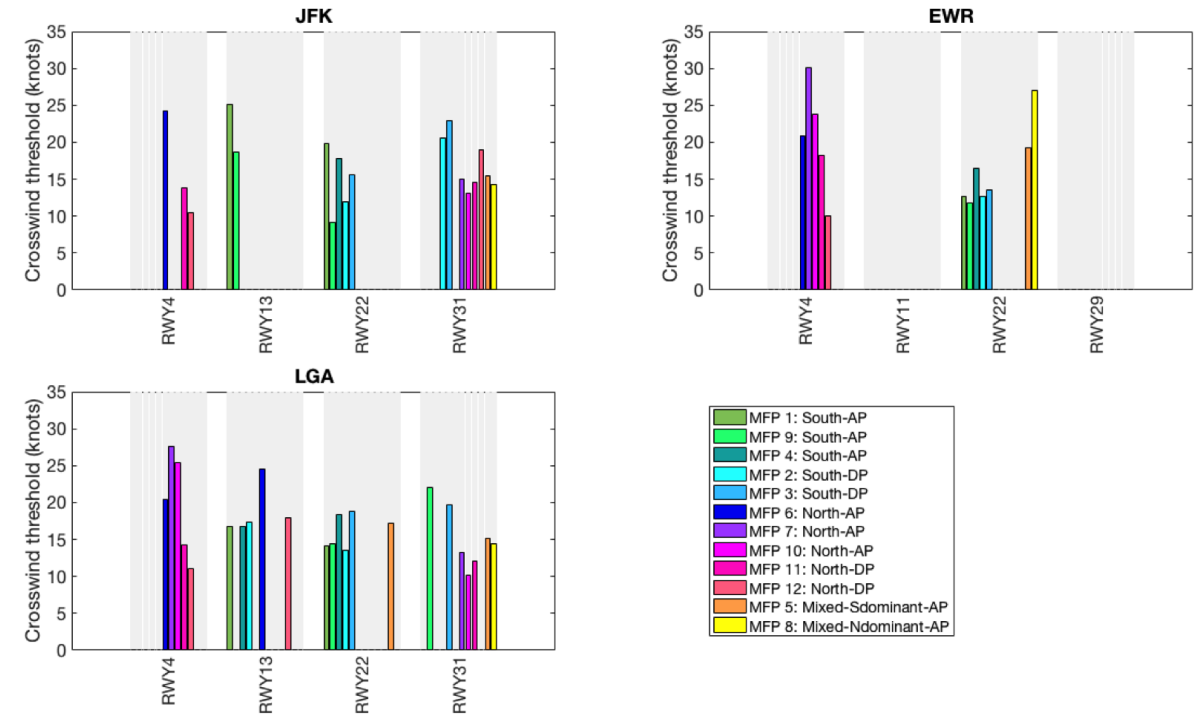


Figure 4-5: Statistically derived crosswind thresholds by MFP.

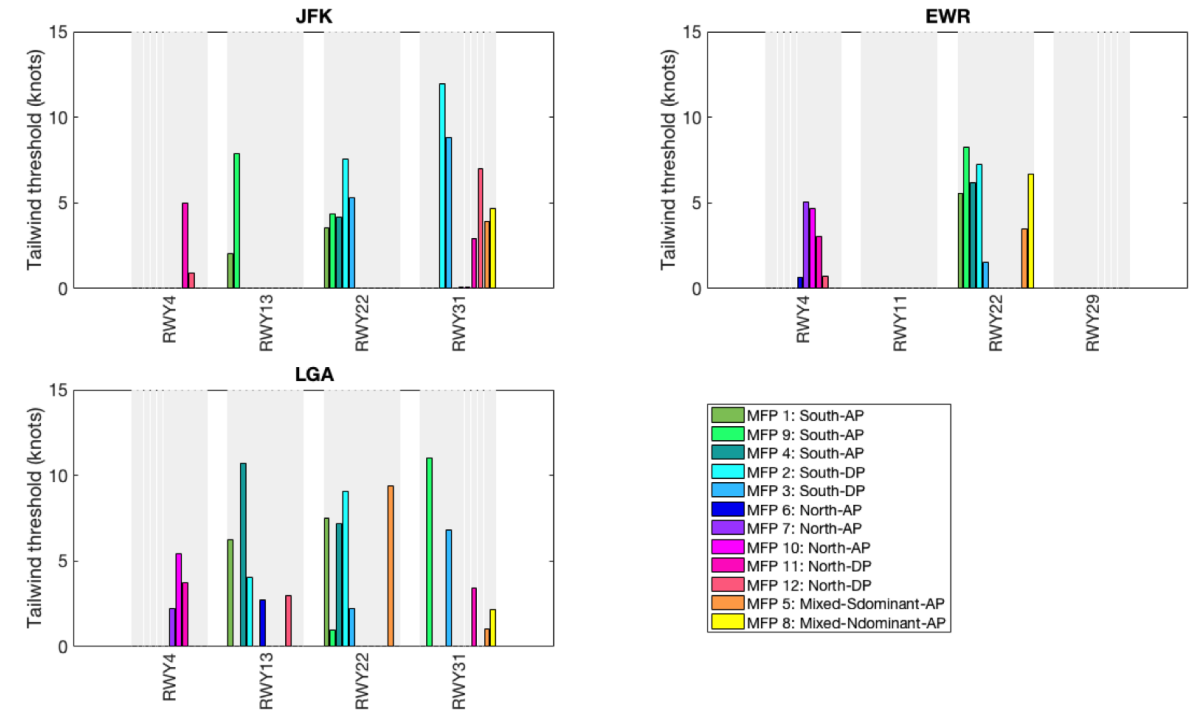


Figure 4-6: Statistically derived tailwind thresholds by MFP.

First, the accuracy of point predictions was evaluated. Figure 4-7 shows the multi-way classification accuracy obtained for each model as a function of the look-ahead time of the prediction. It is observed that the SVM model slightly outperformed the others for shorter time horizons, whereas the RF model showed higher classification accuracy for longer look-ahead times. On average, the accuracy was approximately 83% for a 1-hour prediction, 64% for a 3-hour prediction, and 50% for a 6-hour prediction.

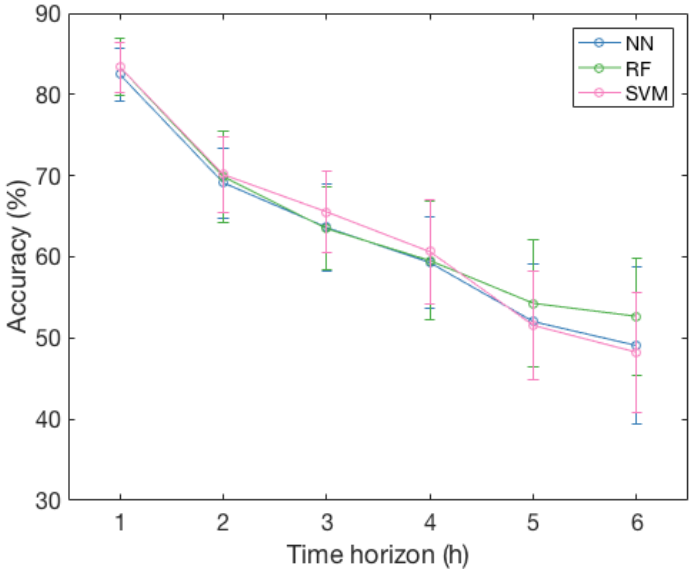


Figure 4-7: Multi-way classification accuracy of NN, RF and SVM traffic flow pattern prediction models for forecast horizons of 1 h to 6 h.

In order to evaluate the accuracy of the probabilistic forecasts, the Brier score was used:

$$BS = \frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C}(p_{ic} - I(z_i = c)) \tag{4.18}$$

where N is the total number of observations in the test set, C is the number of classes, p_{ic} is the posterior probability of observation i belonging to class c and $I(z_i = c)$ is an indicator function that equals to 1, if the actual class of observation i is c, and 0, otherwise.

Based on the formulation of Brier score, the lower the value of this metric, the higher the accuracy of the probabilistic forecast. Figure 4-8 shows the Brier score obtained for each model throughout the forecast horizon. First, it is observed that the score increases throughout the forecast horizon, which is consistent with the decrease in classification accuracy for

longer look-ahead times. Overall, the RF model showed the lowest Brier score, followed by SVM and NN. Based on the accuracy of point predictions and posterior probability estimation, the RF model was chosen for prediction of traffic flow patterns.

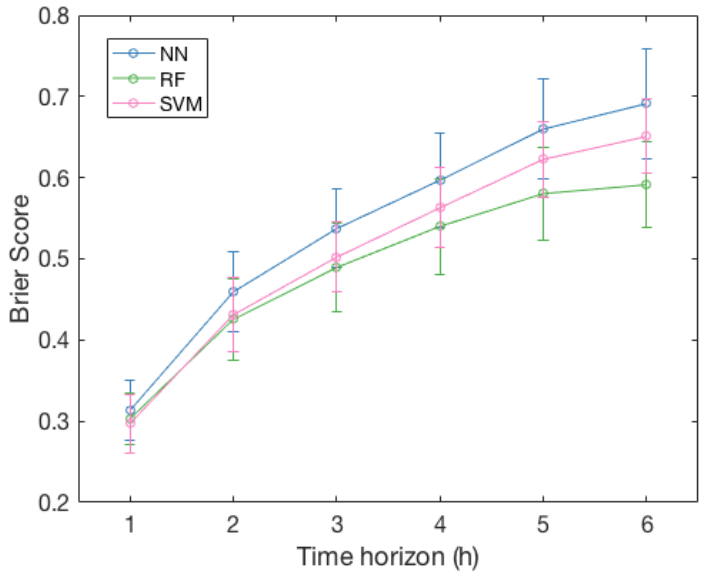


Figure 4-8: Brier score of NN, RF and SVM traffic flow pattern prediction models for forecast horizons of 1 h to 6 h.

We also compared the traffic flow pattern prediction model developed in this thesis with a static model that assumes the flow pattern observed at the beginning of the planning horizon remains throughout the entire planning horizon. The results are shown in Figure 4-9. It is observed that, for short look-ahead times, the RF model outperforms the static model only by a small amount, revealing that the current flow pattern carries most of the predictive power for short-term predictions. Indeed, there is a lot of inertia in the runway/airspace configuration selection process since changes can generate a significant increase in workload. Therefore, a given flow pattern is expected to be seen until conditions turn it operationally infeasible. As the prediction horizon increases, the likelihood that current conditions will remain decreases and inertia starts to have a smaller impact (changes become necessary). The RF model better predicts these changes as the gap between the models increases with the prediction horizon. For a six-hour forecast horizon, the performance gap was 13%.

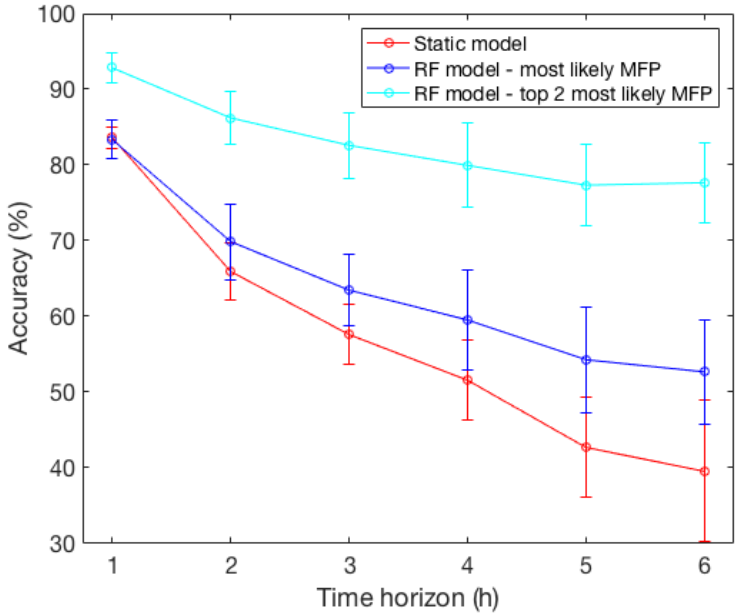


Figure 4-9: Accuracy of traffic flow pattern forecasts for RF model and static model as a function of the look-ahead time of the prediction.

Figure 4-9 also shows the prediction accuracy obtained if the top two most likely flow patterns in the probabilistic forecast are considered instead of just the most likely one. For long-term predictions, the accuracy stabilizes at about 77%. The result indicates that, for many instances in which the model incorrectly predicted the MFP, it was right on the second best option, emphasizing the value of the probabilistic forecast. Indeed, the classification confusion matrix for a 3-h forecast horizon (Figure 4-10) reveals the majority of the incorrect predictions occurred because of confusion between very similar flow patterns, especially for the south flow patterns and because of incorrect prediction of JFK flows.

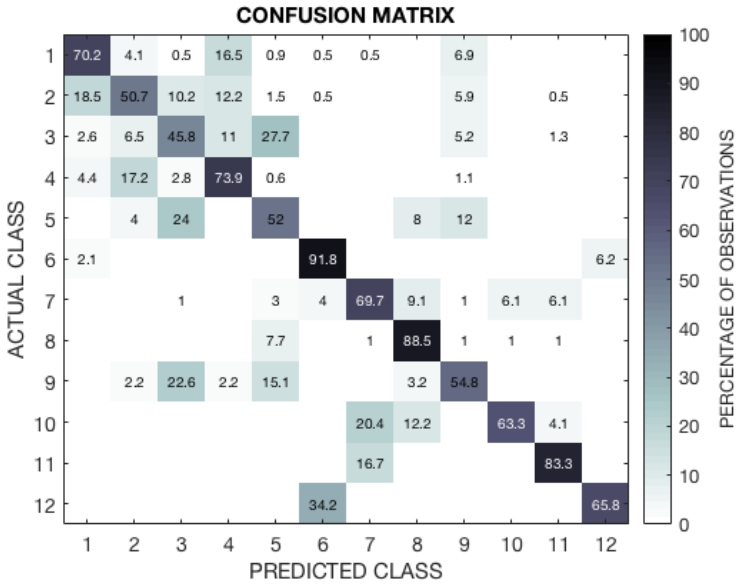


Figure 4-10: Classification confusion matrix for traffic flow pattern predictions for a forecast horizon of 3 h.

4.5 Prediction of Airport Arrival Capacity

4.5.1 Regression Models

The prediction of airport arrival capacity is framed as a regression problem, and the complete predictive distributions are desired. We used different methods for machine learning to build a set of three candidate regression models for this purpose, which were then evaluated with respect to predictive performance. We started with the simplest regression model under a parametric approach by assuming the output is a linear function of the input variables. The linear regression model was constructed in a Bayesian setting, as we are interested in the predictive distribution. The other two models were constructed using non-parametric methods - Random Forests and Gaussian Process - that can learn non-linear mappings directly from the data. While Random Forests is not probabilistic in nature, it can provide an empirical predictive distribution based on the Quantile Regression Forests framework [142]. Besides, it has shown reasonable performance in previous studies in the same domain [75]. The following sub-sections provide the details about the models created. To simplify the notation, the arrival capacity mapping $y_{k,t} = f(t, w_{k,t}, y_{k,t-1}, z_t)$, defined in Equation 4.3,

is represented as $y = f(x)$. All models were developed with MATLAB R2017a.

Bayesian linear Regression Model (BR)

In a linear regression model, the output is assumed to be a linear function of the input variables plus a random disturbance, which is assumed to be independent and identically distributed Gaussian white noise:

$$y = w^T x + \varepsilon \tag{4.19}$$

$$\varepsilon \sim \mathcal{N}(0, \frac{1}{\beta}) \tag{4.20}$$

Under a Bayesian approach, a prior probability over the parameters of the model $p(\theta)$ is defined, where $\theta = [w, \beta]$, and Bayes' rule is used to calculate the posterior probability distribution given the training data (x, y):

$$p(\theta/x, y) = \frac{p(\theta)p(x, y/\theta)}{p(x, y)} = \frac{p(\theta)p(x, y/\theta)}{\int p(x, y/\theta)p(\theta)d\theta} \tag{4.21}$$

The posterior distribution is then used to derive a predictive distribution for y' given new data x':

$$p(y'/x, y) = \int p(y'/\theta)p(\theta/x, y)d\theta \tag{4.22}$$

In this study, the Bayesian linear regression model was developed using the normal-gamma prior over the parameters. In the absence of specific prior knowledge, a vague prior was given, with $p(w/\beta) \sim \mathcal{N}(0, 100^2)$ and $p(\beta) \sim \mathcal{G}(1, 100)$. If we define a gamma prior over the noise precision, $p(\beta) \sim \mathcal{G}(a_0, b_0)$, and a normal prior over the input weights conditioned on the noise precision, $p(w/\beta) \sim \mathcal{N}(\mu_0, \beta\Sigma_0)$, such that the joint prior probability is normal-gamma, $p(w, \beta) = p(\beta)p(w/\beta) \sim \mathcal{NG}(\mu_0, \Sigma_0, a_0, b_0)$, it can be shown that the posterior is analytically tractable and it is also normal-gamma [113]:

$$p(w, \beta/x, y) \sim \mathcal{NG}(\mu_N, \Sigma_N, a_N, b_N) \tag{4.23}$$

where:

$$\Sigma_N = (x^T x + \Sigma_0)^{-1} \tag{4.24}$$

$$\mu_N = \Sigma_N(x^T y + \Sigma_0 \mu_0) \tag{4.25}$$

$$a_N = a_0 + \frac{N}{2} \tag{4.26}$$

$$b_N = b_0 + \frac{1}{2}(y^T y + \mu_0^T \Sigma_0 \mu_0 - \mu_N^T \Sigma_N \mu_N) \tag{4.27}$$

The predictive distribution for y' given new data x' can then be easily obtained by sampling from the posterior $(\tilde{w}, \tilde{\beta}) \sim \mathcal{NG}(\mu_N, \Sigma_N, a_N, b_N)$ and subsequently from the likelihood $\tilde{y} \sim \mathcal{N}(\tilde{w}^T x', \frac{1}{\tilde{\beta}})$.

Random Forests Regression Model (RF)

As described in Section 4.4.1, Random Forests can be used for both classification and regression. In a regression context, the prediction of a decision tree for a new observation x' is obtained by averaging over the observed output values in the leaf $l(x')$ that this observation falls into when it is passed through the tree:

$$\hat{y}' = \frac{1}{|i : x_i \in l(x')|} \sum_{\{i:x_i \in l(x')\}} y_i \tag{4.28}$$

The final response of the ensemble is computed by aggregating the results of the individual trees:

$$\hat{y}'_{bag} = \frac{1}{\sum_{t=1}^{T} \alpha_t} \sum_{t=1}^{T} \alpha_t \hat{y}'_t \tag{4.29}$$

where $\hat{y}'_t$ is the prediction from tree t in the ensemble and α_t is the weight of tree t.

While predicting for a new observation x', if we fully record the observed output values in the leaf $l(x')$ that this observation falls into instead of averaging over them, an empirical

predictive distribution can be built [142].

The parameters of algorithm (T, M and N, as described in Section 4.4.1) were tuned with 5-fold cross validation. Figure 4-11 shows the parameter sensitivity results for arrival capacity prediction at JFK. Based on these results, the RF model was created with $T = 100$, $M = 3$ and $N = 10$.

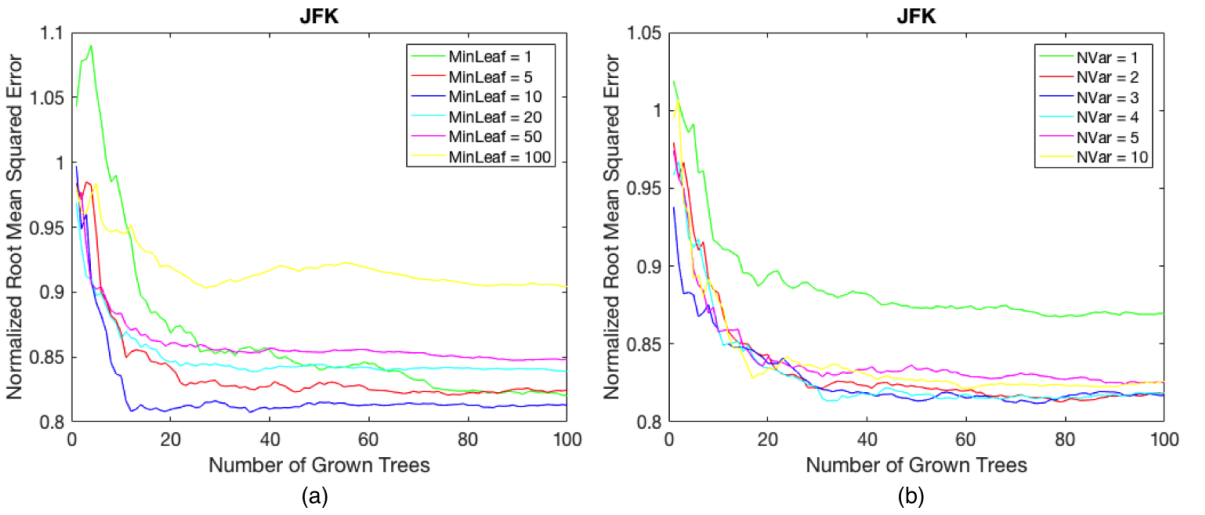


Figure 4-11: Parameter sensitivity of RF regression model based on 5-fold cross validation performance. (a) Impacts of the minimum size of leaf node on the prediction error. (b) Impacts of the number of predictors to sample at each split on the prediction error.

Gaussian Process Regression Model (GP)

The final model was created using a Gaussian Process probabilistic framework, which is also a non-parametric method. In a Gaussian Process regression model, the output variable y is assumed to be a function f of the input variables x following a Gaussian Process. A Gaussian Process is defined as a collection of random variables, any finite number of which have a joint Gaussian distribution [143]. Alternatively, it can be described as a distribution over functions and be written as:

$$f(x) \sim \mathcal{GP}(m(x), k(x, x')) \tag{4.30}$$

where $m(x)$ is the mean function (typically assumed to be zero):

$$m(x) = E[f(x)] \tag{4.31}$$

and $k(x, x')$ is the covariance/kernel function:

$$k(x, x') = E[(f(x) - m(x))(f(x') - m(x'))] \tag{4.32}$$

The kernel function governs the covariance between functions values and it is the main feature that enables the modeling of different types of functions flexibly. Different kernel functions can be used to generate a valid covariance matrix (i.e., symmetric and positive semi-definite) [143]. In this work, we used the canonical squared exponential kernel function, with a separate length scale for each predictor:

$$k(x_p, x_q) = \sigma_f^2 \exp\left(-\frac{1}{2} \sum_{i=1}^{d} \frac{(x_{p,i} - x_{q,i})^2}{l_i^2} \right) \tag{4.33}$$

where σ_f^2 is the signal variance and l is d-dimensional vector of length scales. These kernel parameters are normally referred to as hyperparameters.

As in a typical regression framework, we assumed the observations to be noisy, with independent and identically distributed Gaussian white noise with zero mean and variance σ_n^2:

$$y = f(x) + \varepsilon \tag{4.34}$$

$$\varepsilon \sim \mathcal{N}(0, \sigma_n^2) \tag{4.35}$$

Under the assumption of a zero mean function and Gaussian white noise, the observations y also follow a Gaussian Process:

$$y \sim \mathcal{GP}(0, k'(x_p, x_q)) \tag{4.36}$$

$$k'(x_p, x_q) = k(x_p, x_q) + \sigma_n^2 \delta_{pq} \tag{4.37}$$

where δ_{pq} is a Kronecker delta.

Finally, the joint distribution of the observed target values y and the function values f_* at the test locations is also Gaussian and can be written as

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N}\left(0, \begin{bmatrix} K(X,X) + \sigma_n^2 I & K(X,X_*) \\ K(X_*,X) & K(X_*,X_*) \end{bmatrix}\right) \tag{4.38}$$

where $K(X, X_*)$ denotes the matrix of covariances evaluated at all pairs of n training points and n_* test points (similarly for $K(X,X)$, $K(X_*,X)$, $K(X_*,X_*)$).

By conditioning the joint Gaussian prior distribution on the observations, we can derive the joint posterior distribution over function values and obtain the predictive equations:

$$f_*/X, y, X_* \sim \mathcal{N}\Big(\bar{f}_*, cov(f_*)\Big) \tag{4.39}$$

where:

$$\bar{f}_* = K(X_*,X)[K(X,X) + \sigma_n^2 I]^{-1} y \tag{4.40}$$

$$cov(f_*) = K(X_*,X_*) - K(X_*,X)[K(X,X) + \sigma_n^2 I]^{-1} K(X,X_*) \tag{4.41}$$

The parameters of the GP model can be estimated using Bayesian inference. If θ is the vector of model parameters (kernel hyperparameters σ_f^2 and l; and noise variance σ_n^2), it can be estimated by maximizing its log posterior probability:

$$\hat{\theta} = \arg\max_{\theta}\{\log p(\theta/X,y)\} = \arg\max_{\theta}\Big\{\log p(\theta) - \frac{1}{2} y^T K_y^{-1} y - \frac{1}{2}\log|K_y| - \frac{n}{2}\log 2\pi\Big\} \tag{4.42}$$

where $p(\theta)$ is the prior distribution over the hyperparameters, n is the number of training observations, y is the vector of observed output values, and K_y is the covariance matrix obtained by applying the kernel function over pairs of input vectors in X.

GPstuff [144] was used for hyperparameter inference in this study. Due to the absence of specific prior knowledge, the hyperparameters were given a broad gamma prior $\mathcal{G}(1, 100)$.

Gradient-based optimization was used for maximizing the log posterior probability.

4.5.2 Performance Evaluation

Evaluating the accuracy of point capacity predictions

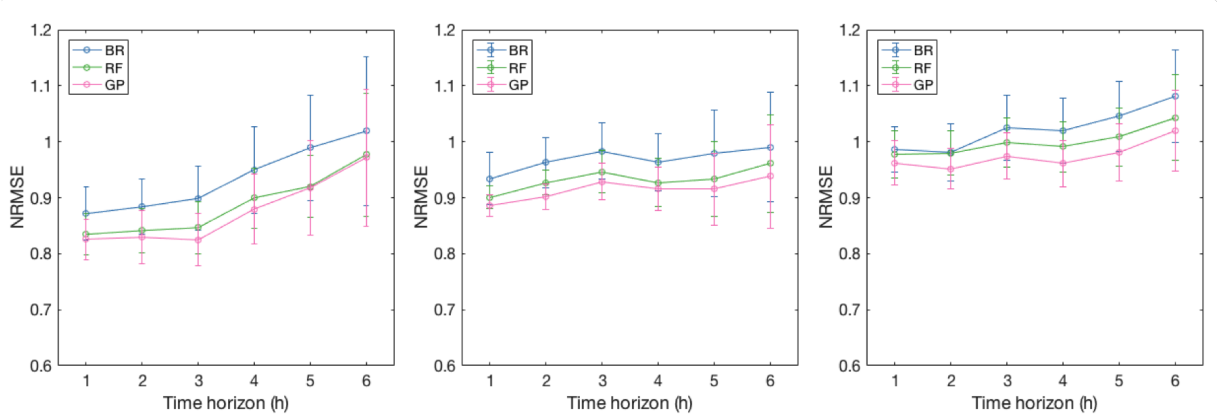
The predictive performance of the regression models was also evaluated for up to six hours of look-ahead time. Two metrics were used to evaluate the accuracy of point capacity predictions: the Normalized Root Mean Squared Error (NRMSE), as defined in Equation 4.43, and the Mean Absolute Prediction Error (MAPE), as defined in Equation 4.44.

$$NRMSE = \frac{\sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2}}{sd(y)} \tag{4.43}$$

where $\hat{y}_i$ is the predicted arrival rate for observation i, y_i is the observed arrival rate for observation i, n is the total number of observations in the test set and $sd(y)$ is the standard deviation of the vector of observed arrival rates.

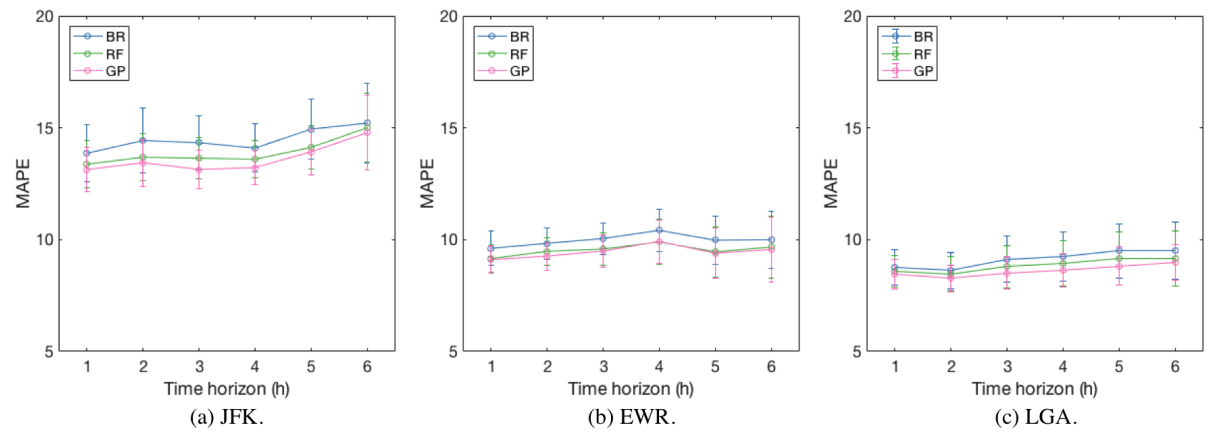
$$MAPE = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{\hat{y}_i - y_i}{y_i}\right| \tag{4.44}$$

Figures 4-12 and 4-13 shows the NRMSE and MAPE values for each regression model throughout different forecast horizons. Overall, it is observed that the GP model outperformed the RF and BR models for all three airports. The GP and RF models exhibited more similar performance, while the BR model systematically showed higher error rates. It is also observed that the error tends to increase slightly with the prediction horizon. This can be explained by the decreasing contribution of the current arrival rate feature and by larger uncertainties in weather conditions and metroplex configuration state. On average, the MAPE obtained with the GP model was 13.6% for JFK, 9.5% for EWR, and 8.6% for LGA.



(a) JFK. (b) EWR. (c) LGA.

Figure 4-12: NRMSE of BR, RF and GP capacity prediction models for forecast horizons of 1 h to 6 h.



(a) JFK. (b) EWR. (c) LGA.

Figure 4-13: MAPE of BR, RF and GP capacity prediction models for forecast horizons of 1 h to 6 h.

Figure 4-14 shows the predictor importance values generated by the RF model. For any feature, this measure is the increase in the prediction error if the values of that feature are permuted across the out-of-bag observations. It is computed for every tree, averaged over the entire ensemble and divided by the standard deviation over the ensemble. The results show that the previous rate is the dominant predictive feature for the three airports, followed by the metroplex configuration feature. This result emphasizes the importance of anticipating the metroplex configuration state for predicting airport capacity. Among the weather variables, ceiling seems to be the most important feature for the three airports, followed by visibility. Headwind showed a relatively high importance value for JFK and

EWR. In high wind conditions, the changes in wind direction and speed along the arrival route can cause a loss of horizontal separation between aircraft (known as compression) and affect throughput. This compression phenomena is more of a factor during strong headwind conditions near the ground, as the spacing between consecutive aircraft will tend to decrease during the approach if they encounter significant wind changes during the descent towards the airport. Finally, convective weather features are also observed to be relevant for the three airports and convective impacts seem to be more important when they block the west flows.

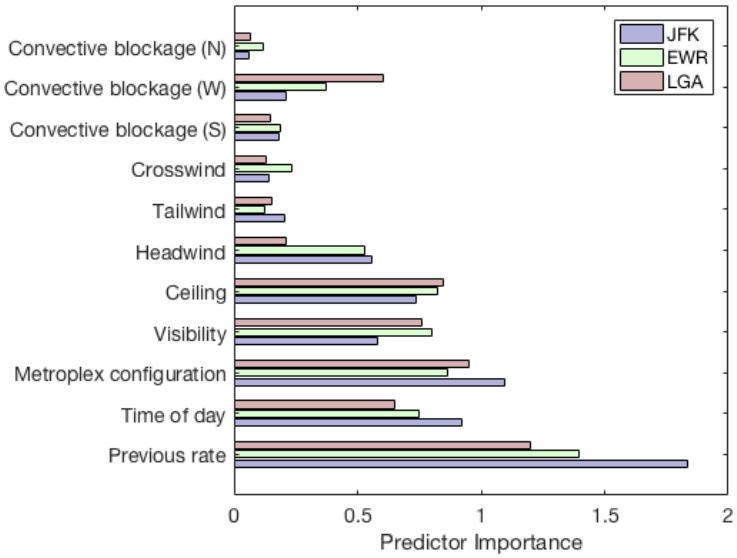


Figure 4-14: Predictor importance estimates from RF model.

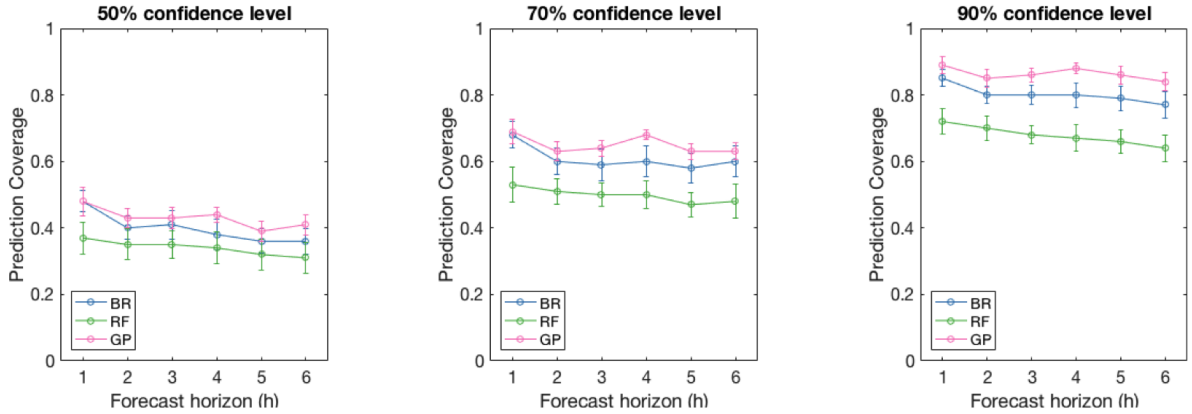
Evaluating the accuracy of probabilistic capacity predictions

The Prediction Coverage (PC), as defined in Equation 4.45, was the quantitative metric used to assess the quality of the predictive distributions generated by the models.

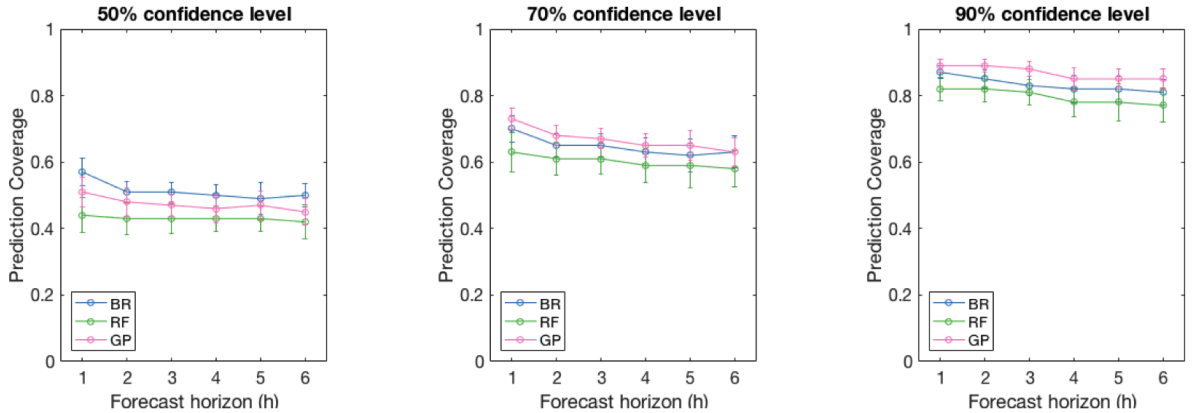
$$PC_\alpha = \frac{\left(\sum_{i=1}^{n} I(l_\alpha^{\hat{y}_i} \leq y_i \leq u_\alpha^{\hat{y}_i}) \right)}{n} \tag{4.45}$$

where $[l_\alpha^{\hat{y}_i}, u_\alpha^{\hat{y}_i}]$ is the $\alpha\%$ confidence interval for the predicted arrival rate $\hat{y}_i$ and I is an indicator function that equals to 1 if the argument expression is true (i.e., the actual arrival rate falls within the confidence interval) and 0, otherwise.

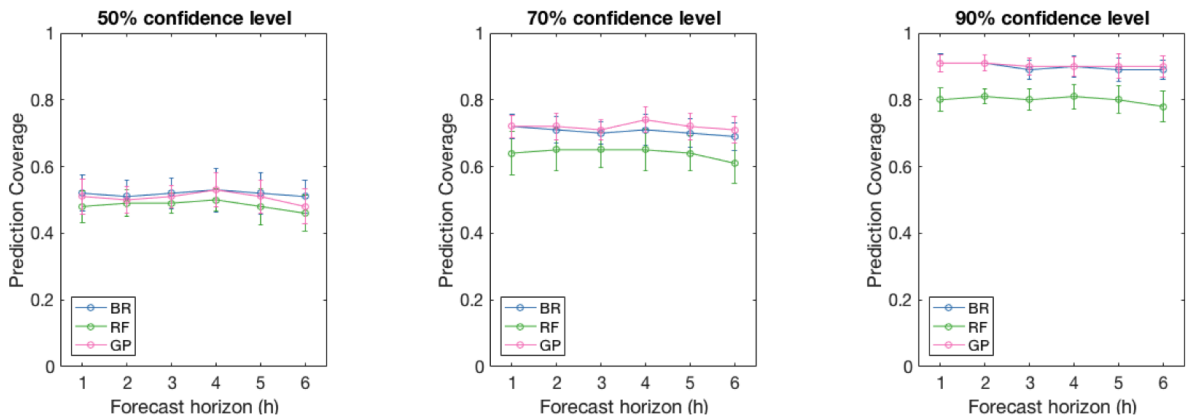
For a given confidence level α%, the PC indicates the percentage of observations in the test set for which the actual arrival rates fall within the α% confidence interval of the predicted arrival rates. Figure 4-15 shows the PC obtained for each regression model, considering three confidence levels: 50%, 70% and 90%.



(a) JFK.



(b) EWR.



(c) LGA.

Figure 4-15: PC of BR, RF and GP capacity prediction models for 50%, 70% and 90% confidence levels and forecast horizons of 1 h to 6 h.

Ideally, one would like to obtain a strong match between the PC and the target confidence level, as it would indicate a good matching between the actual and the estimated distributions. For JFK and EWR, it is observed that the GP model outperformed the others in estimating the uncertainty of the predictions since its PC values were closer to the target confidence levels. The RF model systematically showed PC values that are lower than the target confidence levels, revealing the predicted distributions tend to be narrower than the actual. For LGA, the GP and BR model exhibited similar performance. Based on the accuracy results in terms of point prediction and uncertainty quantification, the Gaussian Process model was chosen for arrival capacity prediction.

Analysis of metroplex system state predictive value

The prediction framework developed in this thesis uniquely accounts for metroplex conditions in the airport capacity estimation process through an embedded metroplex flow pattern prediction model. While Figure 4-14 reveals the metroplex configuration is the second most important feature for predicting capacity, a natural question that arises is how the predictive performance of our model compares to a simplified model that does not account for metroplex effects. For this, an alternative Gaussian Process regression scheme is developed to predict arrival rates solely based on the current arrival rate and weather forecasts, i.e., metroplex state information is not included as a feature in the capacity estimation process:

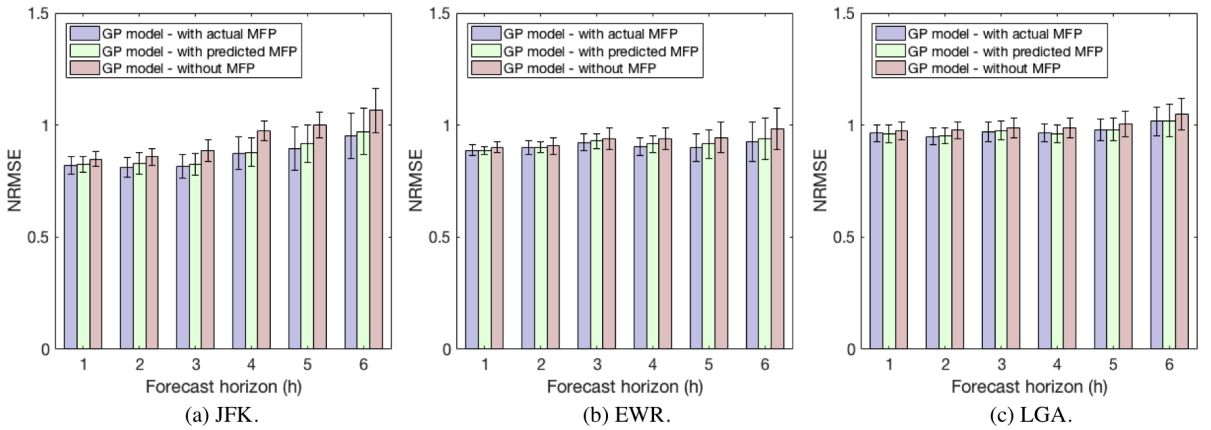
$$y_{k,t} = f(t, w_{k,t}, y_{k,t-1}) \tag{4.46}$$

where $y_{k,t}$ is the arrival rate output at time t for airport k and $w_{k,t}$ is a feature vector encompassing all forecast weather features at time t for airport k (as described in Section 4.3). It should be noted that headwind, tailwind and crosswind forecasts in the vector $w_{k,t}$ were replaced by wind speed and direction forecasts since there is no information about the metroplex configuration and, consequently, about the dominant arrival runway.

The predictive performance of the models is shown in Figure 4-16. It is observed that the two-layer prediction framework developed in this thesis outperforms the alternative model in predicting the capacity at the three New York airports throughout the forecast horizon. It

is also observed that the benefits are more expressive for JFK and for longer time horizons. This can be explained by the fact that JFK has a more diverse set of runway configurations that can be selected by ATC (from a capacity standpoint) and its throughput performance is affected more significantly by the metroplex configuration. Besides, because of the power of the current arrival rate feature (Figure 4-14), removing the metroplex configuration feature does not much impact the predictive performance for shorter time horizons (our model outperforms the alternative model only by a small amount for short look-ahead times of one hour). Yet, for longer time horizons, the inertia factor plays a smaller role as it is less likely that environmental conditions will remain the same and that the system will stay in the same state. As a result, predicting the metroplex state and including this information in the capacity estimation process contributes more to reduce the capacity prediction error.

We also evaluated how the uncertainty in the metroplex configuration state affects the performance of our model by replacing the predicted metroplex configuration with the actual metroplex configuration in the feature vector. The results are also shown in Figure 4-16. It is observed that, with deterministic knowledge about the true metroplex configuration state, the capacity prediction error tends to reduce, as expected. The impacts of the uncertainty about the metroplex configuration state on the capacity prediction error were found to be small, even for long time horizons. This highlights the value of the probabilistic metroplex flow pattern forecast and its good predictive performance, as shown in Figure 4-9.



(a) JFK. (b) EWR. (c) LGA.

Figure 4-16: NRMSE of alternative GP capacity prediction models for forecast horizons of 1 h to 6 h.

Comparison between data-driven and baseline capacity predictions

We also compared the data-driven capacity predictions with the baseline AARs retrieved from historical hourly airport reports provided by the FAA. Figures 4-17, 4-18 and 4-19 shows the MAPE values associated with the baseline capacity estimates and the capacity predictions generated by the GP model, for JFK, EWR and LGA, respectively. The MAPE values are presented aggregated for the full test dataset as well as disaggregated by type of day (with respect to the weather condition).

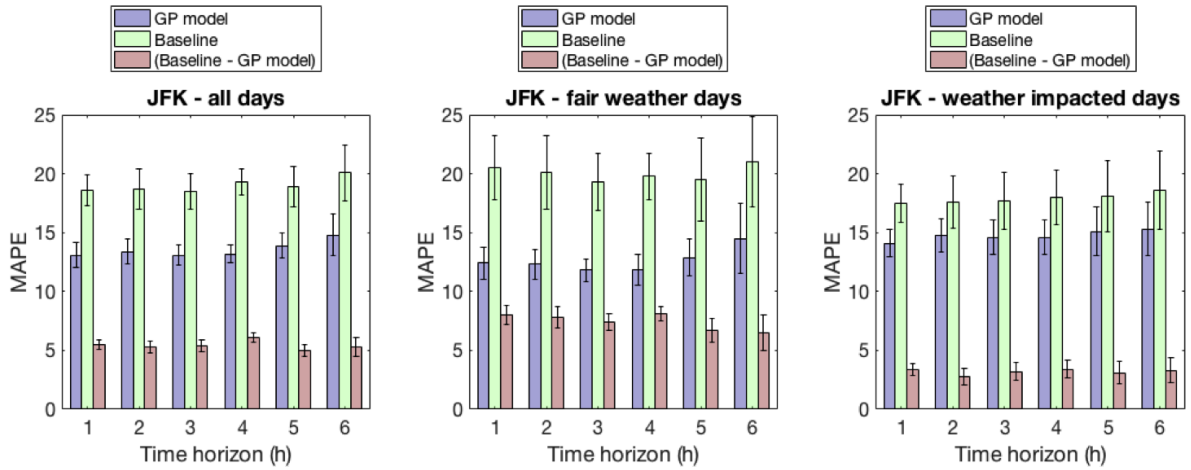


Figure 4-17: MAPE of GP model capacity predictions and baseline capacity estimates for JFK and forecast horizons of 1 h to 6 h.

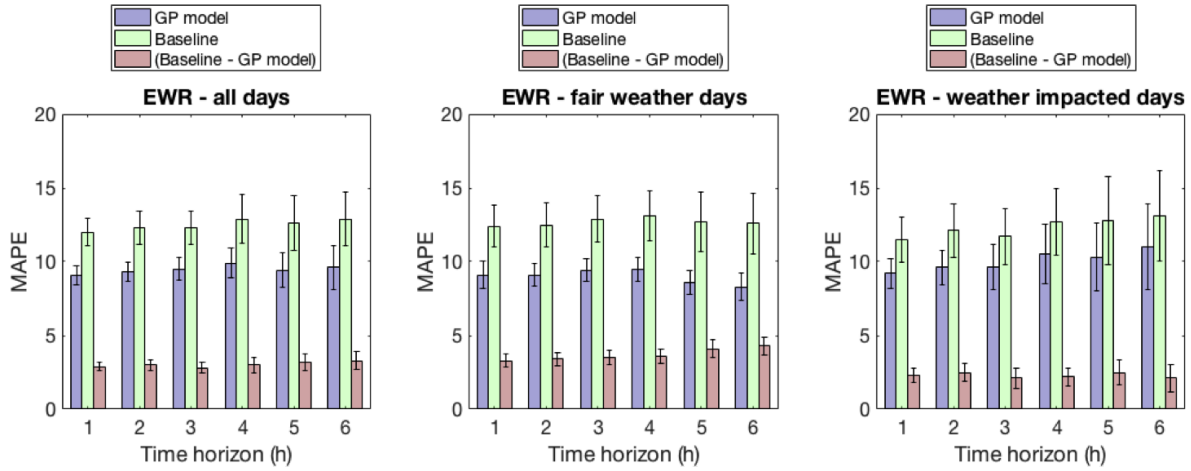


Figure 4-18: MAPE of GP model capacity predictions and baseline capacity estimates for EWR and forecast horizons of 1 h to 6 h.

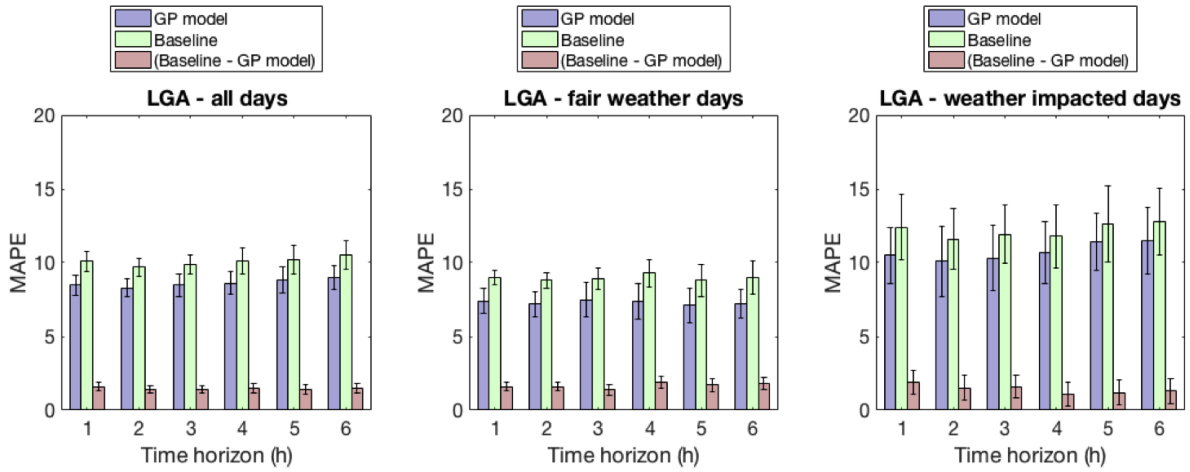


Figure 4-19: MAPE of GP model capacity predictions and baseline capacity estimates for LGA and forecast horizons of 1 h to 6 h.

For JFK, an average reduction of 5.4% in the prediction error was obtained with the data-driven capacity predictions generated by the GP model. The average MAPE value was 13.6% for the GP model predictions and 19.0% for the baseline capacity estimates. For fair weather days, a reduction of 7.4% in the prediction error was observed, whereas a smaller reduction of 3.2% was observed for weather impacted days. The histogram of the relative prediction error for individual observations is presented in Figure 4-20. It is observed that the distribution of the baseline prediction errors is slightly shift to the right of the distribution of the GP prediction errors for fair weather days, with more capacity estimates that are higher than the actual values. On the other hand, for weather impacted days, the distribution of the baseline prediction errors is slightly shift to the left of the distribution of the GP prediction errors, with more capacity estimates that are lower than the actual values.

For EWR, an average reduction of 3.0% in the prediction error was obtained with the data-driven capacity predictions. The average MAPE value was 9.5% for the GP model predictions and 12.5% for the baseline capacity estimates. Similarly to JFK, weather impacted days showed a smaller reduction in the prediction error (2.3%) than fair weather days (3.7%). The histogram of the relative prediction error for individual observations is presented in Figure 4-21. The distribution of the baseline prediction errors is also slightly shift to the right of the distribution of the GP prediction errors for fair weather days, whereas it is shift to the left and exhbitis a more frequent underestimation of capacity for weather impacted days.

Finally, for LGA, the average MAPE value was 8.6% for the GP model predictions and 10.1% for the baseline capacity estimates. The data-driven capacity estimation process generated an average reduction of 1.5% in the prediction error, the smallest value between the three airports. The reduction was 1.7% for fair weather days and 1.4% for weather impacted days. Figure 4-22 shows the histogram of the relative prediction error for individual observations. It is observed that the distributions of the baseline prediction error and of the GP prediction errors are more similar and narrower, with lower mass in the tails.

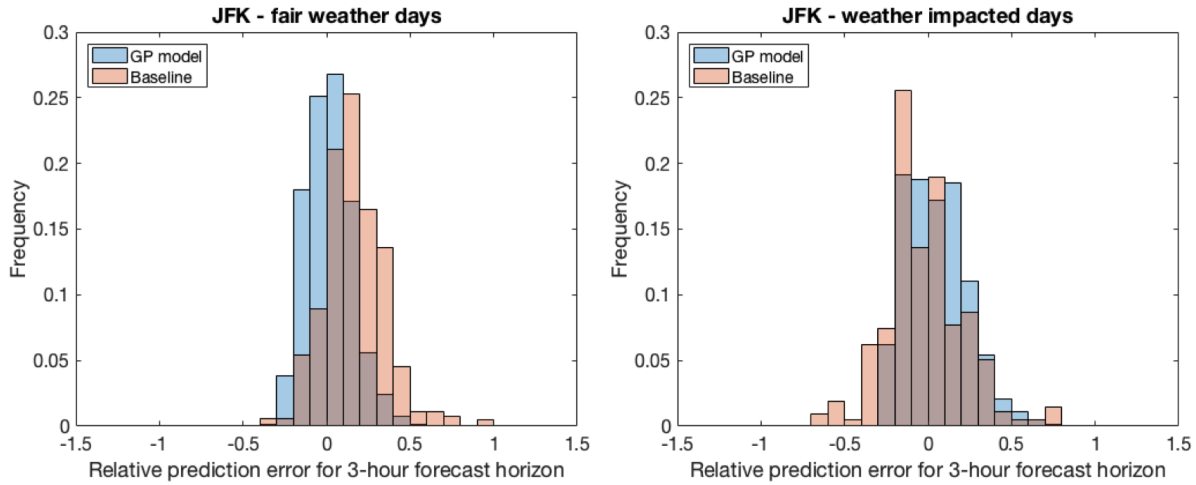


Figure 4-20: Histogram of prediction errors for GP model predictions (for 3-h look-ahead time) and baseline capacity estimates for JFK.

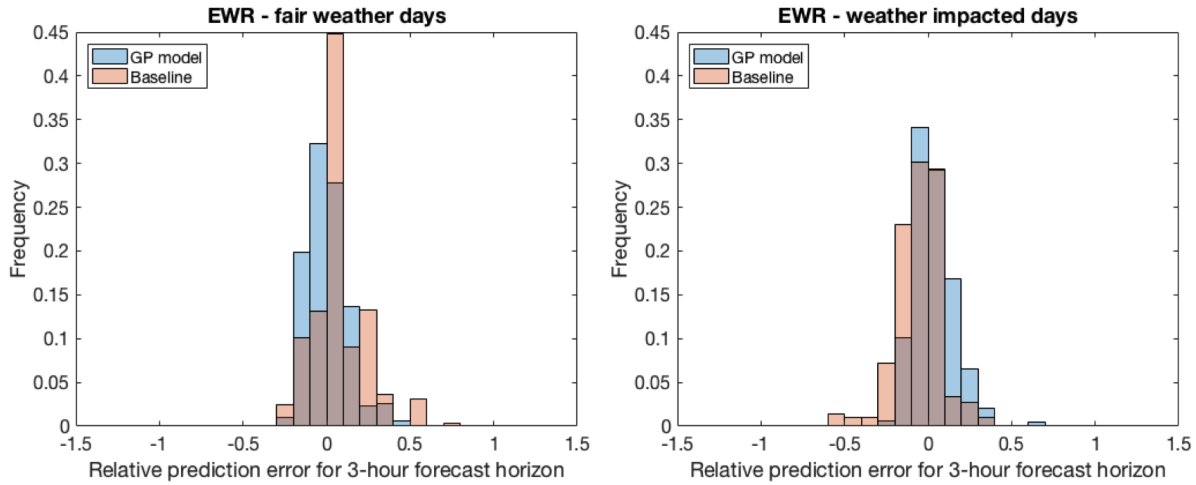


Figure 4-21: Histogram of prediction errors for GP model predictions (for 3-h look-ahead time) and baseline capacity estimates for EWR.

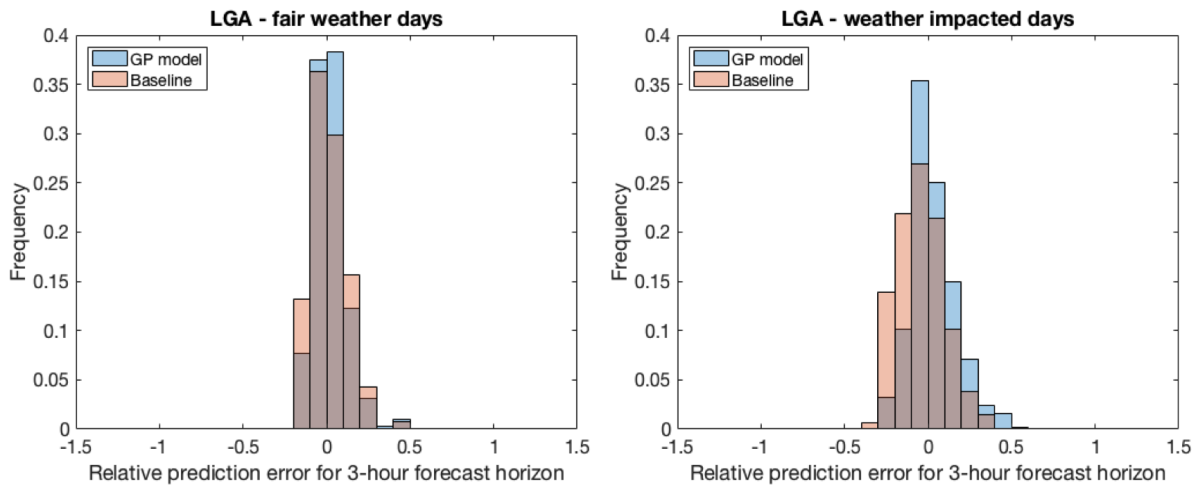


Figure 4-22: Histogram of prediction errors for GP model predictions (for 3-h look-ahead time) and baseline capacity estimates for LGA.

4.6 Prescription of Airport Arrival Rates

The last module of the metroplex configuration and airport capacity planning framework uses optimization methods in order to prescribe AARs for traffic flow management based on the capacity predictions generated by the regression models. We evaluate the impacts of the framework towards traffic flow management decision support by considering the planning of Ground Delay Programs (GDPs).

4.6.1 Ground Delay Programs

GDPs are the most common strategic TMI implemented is the NAS and the current mechanisms by which the GHP is solved in practice. A GDP can be initiated by the ATCSCC when there is a predicted demand-capacity imbalance at an airport and its plan contains the following attributes:

- *Duration:* it specifies the initial and end times for arrival capacity regulation; flights scheduled to arrive within the GDP duration are subject to ground delays.
- *Scope:* it specifies the set of airports/flights for which ground delays can be issued; typically, trans-continental and international flights are exempted from the program.

- *Airport Acceptance Rates (AARs):* They specify the number of arrival slots available at each time period (typically, hourly periods) throughout the program duration.

Given a GDP plan designed by the central authority, ground delays are initially computed to meet the planned arrival rates using the Ration-By-Schedule (RBS) principle (i.e., following the arrival order in the original flight schedules), and translated into a revised schedule in which each affected flight receives an Expected Departure Clearance Time (EDCT). In a subsequent step, flight operators are allowed to exchange arrival slots internally or with other flight operators under a CDM framework.

4.6.2 AAR Planning Model

The AAR planning model builds upon existing stochastic optimization models for the GHP, which aim to determine how much capacity to be allocated and how much ground delay to be assigned in order to minimize overall delay costs. Rather than modeling uncertain capacity behavior with scenario trees or theoretical probability distributions, the model accounts for uncertainty in the form of capacity profiles drawn from actual predictive distributions. In addition, it incorporates robustness goals, i.e., the decision-maker has the ability to select a desired level of robustness for the ATFM plan based on its relative valuation between efficiency and predictability. This is motivated by increased attention to predictability performance by the ATM community [145].

In order to formulate the model, we consider the implementation of a GDP during a planning horizon $T = \{T_i, ..., T_f\}$, where T is a set of hourly time periods, T_i and T_f are the initial and final hours of the GDP program. If a flight is scheduled to arrive at the airport during the program duration T, and its scheduled departure time is later than the time T_b at which the program is filed ($T_b \leq T_i$), it is considered in the capacity allocation process and becomes subject to ground delays. The number of affected flights scheduled to depart at time period s and arrive at time period i is given by N_{si}. We assume the demand is deterministically known. Airport capacity during the GDP duration is given by Q possible capacity profiles drawn from the predictive distributions generated by the airport capacity prediction model. Each capacity profile q is a time-series of predicted arrival rates

$\{M_{qj}, j = T_i, ..., T_f\}$. The proposed optimization model is presented below.

$$\min \quad \left[\sum_{s=T_b}^{T_f} \sum_{i=T_i}^{T_f} \sum_{j=i}^{T_f+1} c_g(j-i)x_{sij} + \frac{1}{Q} \sum_{q=1}^{Q} \sum_{j=T_i}^{T_f} c_a W_{qj} \right] + \alpha\theta \tag{4.47}$$

Subject to:

$$\sum_{j=i}^{T_f+1} x_{sij} = N_{si}, \quad \forall s \in \{T_b, ..., T_f\}, \quad \forall i \in \{T_i, ..., T_f\} \tag{4.48}$$

$$\sum_{s=T_b}^{T_f} \sum_{i=T_i}^{j} x_{sij} + W_{q,j-1} - W_{qj} \leq M_{qj}, \quad \forall j \in \{T_i, ..., T_f\}, \quad \forall q \in \{1, ..., Q\} \tag{4.49}$$

$$\theta \geq \sum_{j=T_i}^{T_f} W_{qj}, \quad \forall q \in \{1, ..., Q\} \tag{4.50}$$

$$W_{q,T_i-1} = 0, \quad \forall q \in \{1, ..., Q\} \tag{4.51}$$

$$x_{sij} \in \mathbb{Z}^+, \quad \forall s \in \{T_b, ..., T_f\}, \quad \forall i \in \{T_i, ..., T_f\}, \quad \forall j \in \{T_i, ..., T_f + 1\} \tag{4.52}$$

$$W_{qj} \in \mathbb{Z}^+, \quad \forall j \in \{T_i, ..., T_f\}, \quad \forall q \in \{1, ..., Q\} \tag{4.53}$$

$$\theta \in \mathbb{Z}^+ \tag{4.54}$$

The decision variables are x_{sij}, which is the number of flights originally scheduled to depart at time period s and arrive at time period i and rescheduled to arrive at time period j, and W_{qj}, which is the number of flights unable to land at time period j if capacity profile q materializes. The objective function is expressed in Equation 4.47 and seeks to find a balance between efficiency and predictability. The first term of the function is the cost of

planned ground delays and the second term is the expected cost of airborne delays, which is the average over all possible capacity profiles. Together, they account for the efficiency goal of minimization of overall delay costs. The unit cost of ground delay is c_g and the unit cost of airborne delay is c_a. In order to incorporate robustness, the third term introduces a penalty term for unpredictability as the maximum unplanned airborne delay (associated with the worst-case capacity profile). This robustness term has a weight α, which can be directly selected by the decision-maker depending on its relative valuation between efficiency and predictability. The higher the value of α, the higher the level of robustness and the lower the risk of unpredictable airborne delays. While we could indirectly introduce an unpredictability penalty by increasing the cost ratio c_a/c_g, we decided to treat separately the fact that airborne delays are less desirable because of higher operating costs and lower predictability. First, although we acknowledge that unpredictable airborne delays may generate additional costs due to their lower recovery potential (as flight operators are left with a shorter time horizon to revise flight schedules, aircraft routing, crew schedules and re-accommodate passengers in the event of a disruption, fewer options are available for recovery, which may end up increasing costs), the extent of this impact is difficult to quantify. Second, this modeling decision gives the decision-maker the ability to directly trade between efficiency and predictability with the robustness parameter α. Operational constraints are expressed in Equations 4.48 and 4.49 and they require that all flights must either arrive at the time period they were originally scheduled to arrive or at a later time period and that the number of arrivals cannot exceed capacity (flow balance constraint).

In order to allow revision of the plan throughout the planning horizon and take advantage of updated information and more accurate capacity forecasts, we consider the implementation of a rolling horizon approach. For example, considering that the plan can be revised n times at time periods $\{T'_k, k = 1, ..., n, T_i < T'_k < T_f\}$, the rolling horizon approach would be implemented as follows: first, the ground holding model would be solved for $T = \{T_i, ..., T_f\}$ and implemented for $T = \{T_i, ..., T'_1 - 1\}$, i.e., ground delays would be incurred as planned for all flights with controlled departure times earlier than T'_1; at the first revision time T'_1, the model would be re-solved for the updated planning horizon $T = \{T'_1, ..., T_f\}$, considering what was been implemented up to time period $T'_1 - 1$ and the updated capacity forecasts;

the same procedure would be repeated iteratively at each revision time T'_k.

Without loss of generality, we present below the formulation for the updated model to be solved at the first revision time T'. The updated cost function is expressed in Equation 4.55 and introduces an additional robustness term to account for the unpredictability that can be introduced by a plan revision. The first three terms are similar to those in Equation 4.47. The last term accounts for the unpredictability introduced by changes in ground delays. These changes are expressed in Equations 4.61, 4.62, 4.63 and 4.64, where x^*_{sij} is the optimal solution for the initial plan. If, for all flights initially scheduled to depart at time period s and arrive at time period i, the new plan results in more ground delays than what was previously planned when the model was solved at the previous step, y_{si} is positive and gives the total increase in ground delays assigned to these flights. On the other hand, if the new plan results in less ground delays than what was previously planned, z_{si} is positive and gives the total amount of ground delays that will not be incurred by these flights. This robustness term has a weight β, which can be directly selected by the decision-maker to control changes in ground delays during the revision. The higher the robustness parameter β, the lower the modification of the initial GDP plan. The underlying motivation for penalizing changes in ground delays during the revision is as follows: if, for a set of flights, the new plan results in more ground delays in expectation of degraded capacity at the destination airport, the ANSP may wish to be more adherent to the initial plan in order to give these flights the best chance to arrive at the time initially planned (especially if these flights are sensitive to the flight operator, for instance, in terms of connecting flights downstream); on the other hand, if the new plan results in less ground delays in expectation of improved capacity at the destination airport, the ANSP may also wish to be more adherent to the initial plan if resources have already been deployed to adapt to this plan, reducing the chance that a resource is released and, later on, has to be deployed again (after all, capacity is uncertain and additional airborne delays may be incurred).

Equations 4.56 - 4.58 are similar to Equations 4.48 - 4.50. In Equation 4.56, the updated demand N'_{si} is given by all flights that are still on the ground at the time of the revision and that have earliest possible departure time s and earliest possible arrival time i. If the previous plan is implemented exactly, the updated demand N'_{si} can be calculated with 4.59

and it is given by all flights with controlled departure times later than $T' - 1$, as determined by the previous plan, and that can now receive a revised ground delay assignment. Equation 4.59 aggregates the flights originally scheduled to depart from T' onwards with the flights originally scheduled to depart before T' but that were assigned a controlled departure time later than $T' - 1$ in the previous plan. Those flights are now considered as new flights scheduled to depart at T'. In Equation 4.57, the new term V_j is introduced to ensure that all flights that have taken off before T' and are anticipated to arrive at time period j are accounted for in the flow balance constraint. If the previous plan is implemented exactly, V_j is given by Equation 4.60. Still in Equation 4.57, M'_{qj} represent the updated capacity profiles. Finally, Equation 4.65 ensures that the airborne queue $W^*_{q,T'-1}$ at the beginning of the updated planning horizon is accounted for in the flow balance constraint.

$$\min \quad \left[\sum_{s=T'}^{T_f} \sum_{i=T'}^{T_f} \sum_{j=i}^{T_f+1} c_g(j-i) x'_{sij} + \frac{1}{Q} \sum_{q=1}^{Q} \sum_{j=T'}^{T_f} c_a W'_{qj} \right] + \alpha\theta + \beta \sum_{s=T'}^{T_f} \sum_{i=T'}^{T_f} (y_{si} + z_{si}) \quad (4.55)$$

Subject to:

$$\sum_{j=i}^{T_f+1} x'_{sij} = N'_{si}, \quad \forall s \in \{T', ..., T_f\}, \quad \forall i \in \{T', ..., T_f\} \quad (4.56)$$

$$\sum_{s=T'}^{T_f} \sum_{i=T'}^{j} x'_{sij} + V_j + W'_{q,j-1} - W'_{qj} \leq M'_{qj}, \quad \forall j \in \{T', ..., T_f\}, \quad \forall q \in \{1, ..., Q\} \quad (4.57)$$

$$\theta \geq \sum_{j=T'}^{T_f} W'_{qj}, \quad \forall q \in \{1, ..., Q\} \quad (4.58)$$

$$N'_{si} = \begin{cases} N_{si} + \sum\limits_{\substack{\{(l,m,n): l<T'; \\ l+n-m \geq T'; \\ m-l=i-s\}}} x_{lmn}, & s = T', \quad \forall i \geq s \\ N_{si}, & \forall s > T', \quad \forall i \geq s \end{cases} \quad (4.59)$$

$$V_j = \sum_{\{(s,i):s+j-i<T'\}} x_{sij}, \quad \forall j \geq T' \tag{4.60}$$

$$y_{si} \geq \sum_{j=i}^{T_f+1} (j-i)x'_{sij} - \sum_{j=i}^{T_f+1} (j-i)x^*_{sij}, \quad \forall s > T', \quad \forall i \geq s \tag{4.61}$$

$$y_{si} \geq \sum_{j=i}^{T_f+1} (j-i)x'_{sij} - \sum_{\substack{\{(l,m,n):l\leq T';\\ n\geq m;l+n-m\geq T';\\ m-l=i-s\}}} (l+n-m-T')x^*_{sij}, \quad s = T', \quad \forall i \geq s \tag{4.62}$$

$$z_{si} \geq \sum_{j=i}^{T_f+1} (j-i)x^*_{sij} - \sum_{j=i}^{T_f+1} (j-i)x'_{sij}, \quad \forall s > T', \quad \forall i \geq s \tag{4.63}$$

$$z_{si} \geq \sum_{\substack{\{(l,m,n):l\leq T';\\ n\geq m;l+n-m\geq T';\\ m-l=i-s\}}} (l+n-m-T')x^*_{sij} - \sum_{j=i}^{T_f+1} (j-i)x'_{sij}, \quad s = T', \quad \forall i \geq s \tag{4.64}$$

$$W'_{q,T'-1} = W^*_{q,T'-1}, \quad \forall q \in \{1,...,Q\} \tag{4.65}$$

$$x'_{sij} \in \mathbb{Z}^+, \quad \forall s \in \{T',...,T_f\}, \quad \forall i \in \{T',...,T_f\}, \quad \forall j \in \{T',...,T_f+1\} \tag{4.66}$$

$$W'_{qj} \in \mathbb{Z}^+, \quad \forall j \in \{T',...,T_f\}, \quad \forall q \in \{1,...,Q\} \tag{4.67}$$

$$y_{si}, z_{si} \in \mathbb{Z}^+, \quad \forall s \in \{T',...,T_f\}, \quad \forall i \in \{T',...,T_f\} \tag{4.68}$$

$$\theta \in \mathbb{Z}^+ \tag{4.69}$$

4.6.3 Analysis of Impacts on GDP planning

Test cases and assumptions

The AAR planning model described in Section 4.6.2 was used to optimize the traffic flow management plan for five test cases corresponding to historical GDP events at JFK (Table 4.2). The optimization problem was solved with Gurobi optimizer (version 6.5.0). The cost ratio between airborne and ground delay was assumed at 1.5 ($c_g = 1$ and $c_a = 1.5$), based on delay costs reported by the industry [146, 147]. The robustness parameters α and β were varied from 0 to 10 in small increments of 0.02 in order to investigate the trade-offs between efficiency and predictability. Finally, we assumed that the plan could be revised once and that the revision was performed in the middle of the planning horizon.

Table 4.2: Description of GDP events at JFK used as test cases.

Test case	Date	Planning horizon (h)
1	20130131	6
2	20130217	5
3	20130522	6
4	20131127	6
5	20140812	5

For each GDP event, 100 capacity profiles representing possible evolutions of capacity throughout the GDP planning horizon were drawn from the predicted arrival rate distributions generated by the prediction framework. The choice of number of capacity profiles was made considering trade-offs between variability in expected delay costs and computational time. Figure 4-23 illustrates these trade-offs for test case 1 (which was optimized with the simpler static AAR planning model). As the number of sampled capacity profiles increases, the variability in total delay costs decreases, while the computational time for obtaining the optimal solution increases. For more than 100 capacity profiles, the expected costs tend to converge, revealing that this number of capacity profiles is a reasonable choice for not much increasing the computational effort.

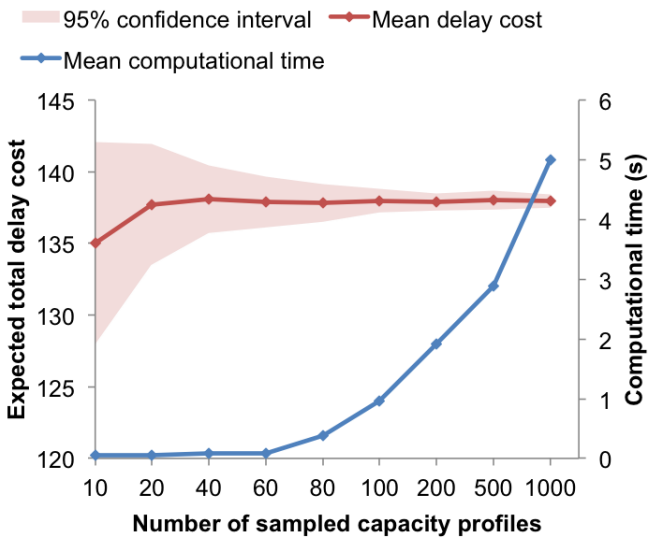


Figure 4-23: Expected delay costs and computational time as a function of the number of capacity profiles.

Impacts of robustness parameter α

We first assessed the trade-offs between efficiency and predictability by varying the robustness parameter α, which controls the impacts of the robustness term that penalizes unpredictable airborne delays, during the optimization of the GDP plan for each test case. For this analysis, the robustness parameter β was set to zero, i.e., changes in ground delays were not penalized during the revision.

The planned GDP costs obtained for different values of α are shown in Figures 4-24 and 4-25. The planned ground delay costs are shown in Figure 4-24, and the expected airborne delay costs are shown in Figure 4-25. As expected, as the decision-maker becomes less risk tolerant and increases the value of α, the robustness term more heavily impacts the objective function and the expected airborne delay costs decrease at the cost of increased ground delays. It is also observed that the expected airborne delay costs are very sensitive to changes in α for small values (less than 1).

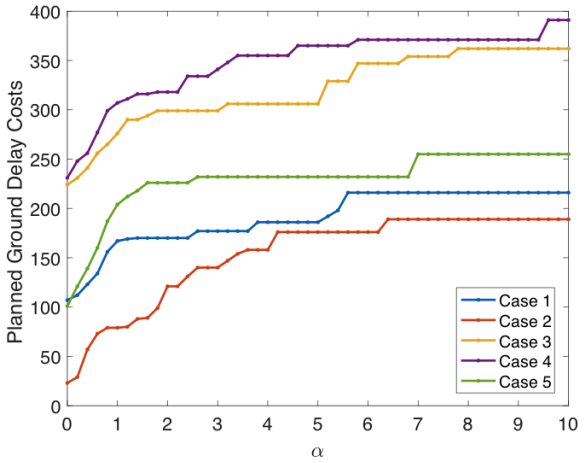


Figure 4-24: Planned ground delay costs as a function of robustness parameter α.

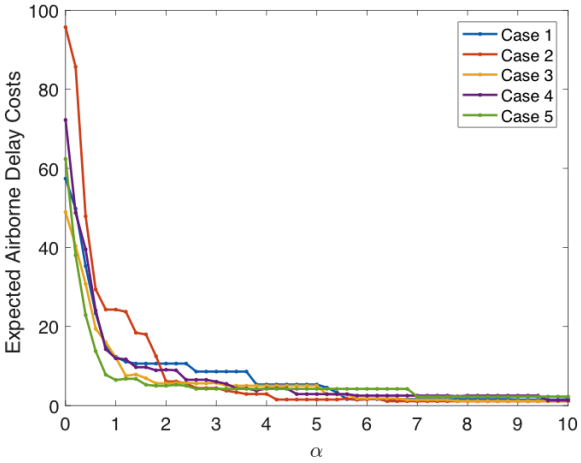


Figure 4-25: Expected airborne delay costs as a function of robustness parameter α.

Figure 4-26 shows how the optimized GDP plan performs under the actual capacity profiles. Specifically, it shows the sum of the planned ground delay costs and the actual airborne delay costs [1] as a function of α. When robustness is heavily accounted for in the objective function (large values of α), a decrease in efficiency is observed, as expected. Yet, for most of the test cases, it is also noted that reductions in the actual delay costs could be obtained with the introduction of the robustness term for smaller values of α. The average actual delay costs for all test cases is shown in Figure 4-27. For $\alpha < 1$, introducing robustness turned out to increase efficiency by protecting against imperfect capacity predictions. The minimum cost was observed for $\alpha = 0.6$.

[1] The actual airborne delay costs were calculated by running the AAR planning model with the optimized GDP plan solution frozen ($x_{sij} = x^*_{sij}$, where x^*_{sij} is the optimized ground delay solution) and with the predicted capacity profiles replaced by the actual capacity profile at the airport. In fact, this corresponds to a deterministic queueing model governed by the flow balance constraint given by Equation 4.49: for any time period $t \in T = \{T_i, ..., T'_1 - 1\}$, the number of departing flights and the number of ground-held flights matches the optimized GDP plan solution; if the number of flights planned to arrive at the airport is less than or equal to the actual capacity during time period t, all flights land; otherwise, the number of flights that land is equal to the actual capacity and the remaining flights are held in the air until the next time period; for any t from T' onwards, the process is repeated, but with the revised GDP plan solution. In summary, we refer to actual airborne delays as those associated with the actual capacity profiles, under the assumptions of the AAR planning model, and without any relation to historical delays incurred by the flights affected by the actual GDP events.

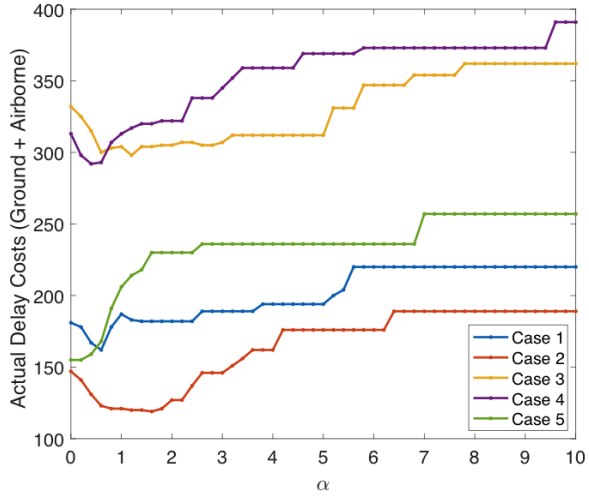


Figure 4-26: Actual delay costs as a function of robustness parameter α.

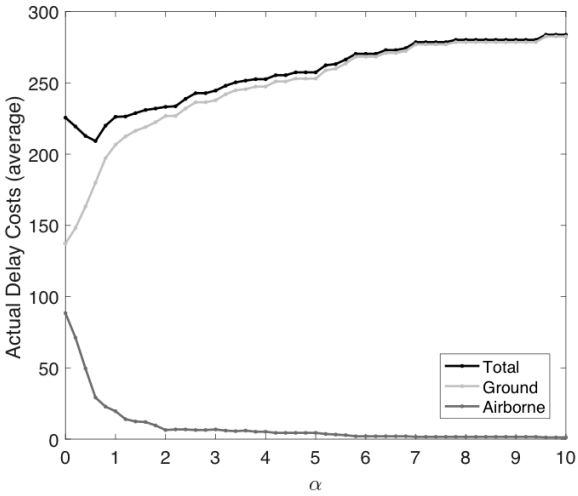


Figure 4-27: Actual delay costs (averaged over all test cases) as a function of robustness parameter α.

The main goal of incorporating robustness in the objective function for optimization of the GDP plan is increased predictability. Figure 4-28 shows the difference between the actual delay costs and the planned delay costs as a function of α. Indeed, the introduction of the robustness term increases GDP delay cost predictability. This is driven by a reduction in unexpected airborne delays, as shown in Figure 4-25.

Finally, Figure 4-29 shows the impacts of α on the optimized arrival rates, which are displayed as the corresponding percentile of the predicted capacity distributions. As α increases, the optimal arrival rate reduces, indicating that fewer flights should be allocated in order to reduce the risk of unplanned airborne delays. For each test case, the optimal arrival rate that minimizes the actual delay cost is also plotted. It is observed that it ranges from the 28^{th} to the 47^{th} percentile of the predicted capacity distributions.

Impacts of robustness parameter β

This section shows the results of the GDP plan optimization for various values of β, which controls the impacts of the robustness term that penalizes changes in ground delays during the revision. For this analysis, the robustness parameter α was set to zero.

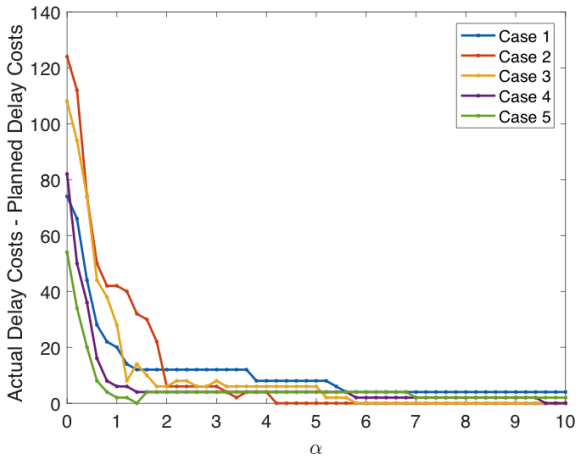


Figure 4-28: GDP delay cost predictability as a function of robustness parameter α.

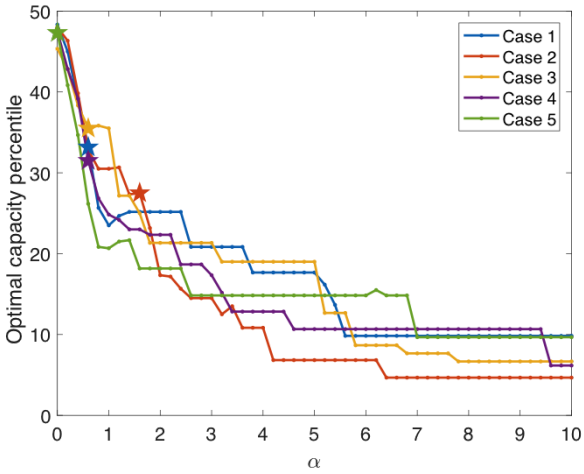


Figure 4-29: Optimized arrival rates as a function of robustness parameter α.

Figures 4-30 and 4-31 show the difference between the planned ground delays and the expected airborne delays for the revised GDP plan and the initial GDP plan. When $\beta = 0$, the highest changes in ground delays were observed, revealing that updated information was indeed leveraged to further optimize the plan during the revision. With the exception of test case 2, more ground delays were assigned during the revision. Indeed, Figure 4-31 shows that the updated information led to an increase in the expected airborne delays for all test cases. When changes in ground delays were "allowed" during the revision (small values of β), the model counterbalanced this increase in the expected airborne delays by assigning more ground delays in order to minimize the expected overall delay costs. The only exception was test case 2, as no changes in ground delays were observed despite the expected increase in airborne delays. Further analysis revealed that, for test case 2, all flights eligible to have an updated ground delay assignment had already been assigned the maximum possible ground delay in the original plan, as they were scheduled to arrive at the hour after the end of the program.

Figure 4-32 shows the planned delay costs associated with the revised and the initial GDP plans, averaged over all test cases. Figure 4-33 shows the actual delay costs resulting from the implementation of the revised and the initial GDP plans, averaged over all test cases. It is observed that the revision was useful to increase the predictability of the expected GDP delay costs, as the planned costs under the revised plan were closer to the actual delay costs.

Furthermore, when changes in ground delays were not penalized and were fully allowed during the revision ($\beta = 0$), an average increase in efficiency of 2% was observed.

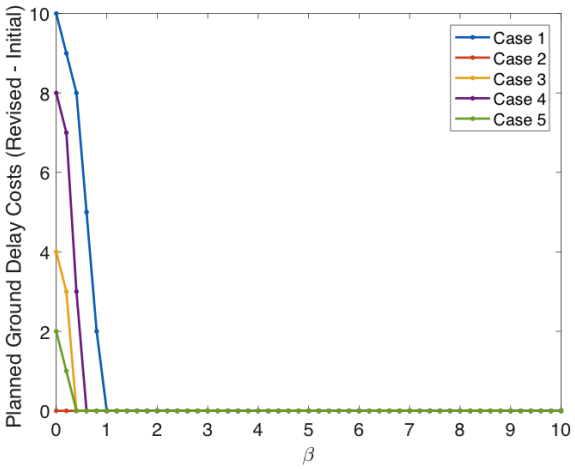


Figure 4-30: Difference between the revised and the initially planned ground delays as a function of robustness parameter β.

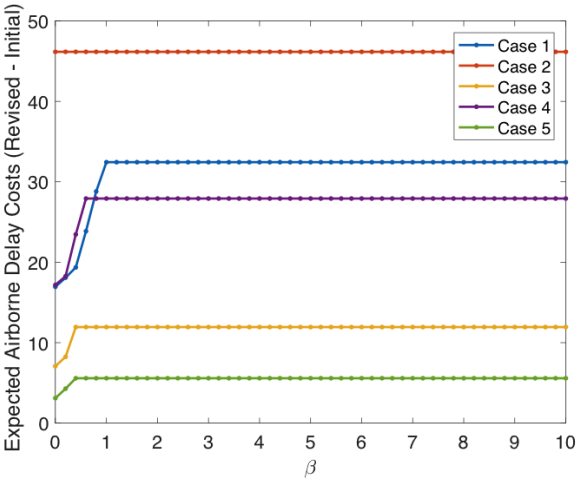


Figure 4-31: Difference between the revised and the initially expected airborne delays as a function of robustness parameter β.

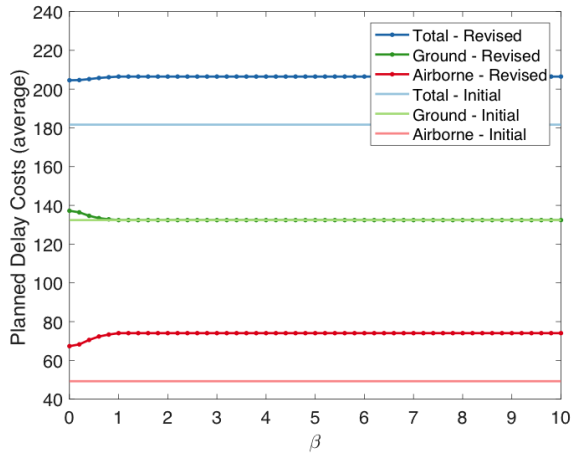


Figure 4-32: Planned delay costs for revised and initial plans as a function of robustness parameter β (averaged over all test cases).

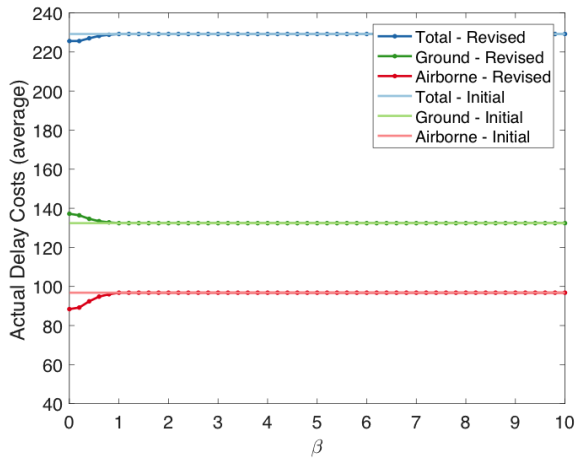


Figure 4-33: Actual delay costs for revised and initial plans as a function of robustness parameter β (averaged over all test cases).

Comparison between GDP costs resulting from data-driven and baseline AAR

The delay costs resulting from the use of the optimized AARs and the AARs actually implemented by the FAA for designing the GDP plan were compared. Figure 4-34 shows that,

with the exception of test case 5, lower GDP costs were achieved with the data-driven capacity predictions. When robustness was not accounted in the optimization of the GDP plan ($\alpha = 0$), a 2.4% reduction in the total GDP costs (sum of all test cases) was observed. When robustness was accounted for in the optimization of the GDP plan, with $\alpha = 0.6$ (observed value that contributed to increase efficiency on average), the reduction in total GDP costs was 9.7%. Figure 4-35 shows the MAPE for the capacity predictions generated by the GP model (averaged over the GDP planning horizon) for each test case. It is observed that, although test case 5 was the exception, it showed similar level of error as the other cases, suggesting that the AARs actually implemented on that date were closer to the true capacity at the airport.

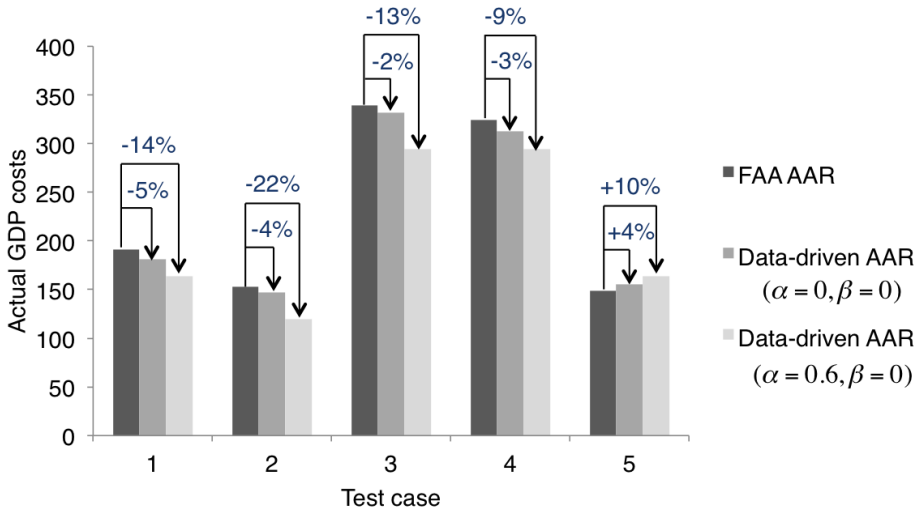


Figure 4-34: Comparison between GDP costs resulting from data-driven and baseline AAR.

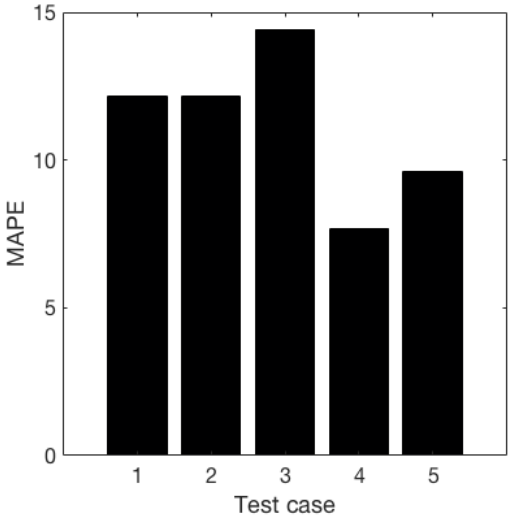


Figure 4-35: MAPE of GP model capacity predictions (averaged over planning horizon) for GDP test cases.

Chapter 5

Conclusions and Future Work

5.1 Summary

In this thesis, a flight trajectory data analytics framework was developed to provide a high-fidelity characterization of air traffic flows from large-scale aircraft tracking data. Through the application of machine learning methods to discover spatial and temporal trends in aircraft movement, the framework allows to automatically learn the airspace structure, assess the use of the airspace and identify patterns of usage of the airspace. For this, it includes three modules: (1) clustering flight trajectories at spatial scale to identify trajectory patterns, (2) trajectory classification to assess flight trajectory conformance against the learned airspace structure and identify air traffic flows, (3) clustering air traffic flows at temporal scale to identify traffic flow patterns. The framework was used to provide a detailed characterization of air traffic flows in the terminal area of multi-airport systems.

We first leveraged the knowledge generated by the framework to develop a systematic descriptive approach to analyze metroplex airspace design and use and to assess operational performance. Novel quantitative metrics were created to summarize metroplex performance in the areas of efficiency, capacity and predictability. The descriptive approach was demonstrated with a comparative analysis of terminal area air traffic operations in three representative multi-airport systems: New York, Hong Kong and Sao Paulo. The results revealed structural, operational and performance differences between these systems. We found that the New York multi-airport system presents the most complex airspace design and the most

dynamic flow behavior. Interestingly, it exhibits the best levels of trajectory lateral and temporal efficiency on average, yet the highest variability in operational performance. By contrast, the Hong Kong and Sao Paulo multi-airport systems show a relative simple and clean airspace design and lower variability in traffic flow behavior and performance. The results also highlighted different areas of action to be considered at each multi-airport system towards improving their air traffic operations. The Hong Kong metroplex route structure was found to be considerably less efficient, particularly for HKG, revealing opportunities for improvement in airspace design in order to increase trajectory lateral performance. The lower level of trajectory conformance and temporal efficiency observed for the Hong Kong and Sao Paulo metroplexes revealed opportunities for improvement in tactical traffic flow management towards better sequencing and scheduling of arriving flights. Sao Paulo was found to be the lowest throughput performing system, emphasizing the importance of actions to expand capacity at the individual airports. Finally, the higher level of airport interdependency and operational variability observed in the New York metroplex revealed opportunities for improving the airspace design in order to reduce flow interdependencies and provide more consistent all-weather operations. The particular influence of weather conditions on airspace use and traffic performance in the New York metroplex also highlighted the importance of integrating weather forecasting with decision-making in order to better anticipate weather conditions, translate them into operational impact and plan traffic flow management efficiently at this system.

Finally, in the last part of the thesis, we leveraged the knowledge generated by the flight trajectory data analytics framework to develop a data-driven approach for metroplex configuration and airport capacity planning towards improved decision support. The approach has two building blocks: predictive modeling for capacity estimation and prescriptive modeling for capacity allocation.

In the capacity estimation block, machine learning methods are used to translate weather forecasts into probabilistic forecasts of the metroplex configuration and airport capacity for strategic planning horizons. First, a traffic flow pattern prediction model was developed to deliver a probabilistic forecast of the traffic flow structure to be seen in the metroplex terminal airspace on an hourly basis. At any given time, the metroplex flow pattern reflects

a particular utilization pattern of runways and airspace, providing a complete picture of the multi-airport system configuration state. Three candidate traffic flow pattern prediction models were created using different supervised learning methods for multi-way classification. After performance evaluation, a Random Forests model was selected for traffic flow pattern prediction. For the New York multi-airport system, the model showed an average prediction accuracy of 83% for a short-term 1-hour forecast, 64% for a 3-hour forecast, and 52% for a 6-hour forecast. Despite the diminishing accuracy for long-term forecasts, it showed higher performance when compared with a static model that only considers the current flow pattern to predict the next hours. For a 6-hour forecast horizon, the performance gap was 13%. The probabilistic forecast was found to be useful, as the results indicated that, for many instances in which the model incorrectly predicted the traffic flow pattern, it was right on the second best option. Indeed, the prediction accuracy was 93% for a 1-hour forecast and stabilized at about 77% for longer look-ahead times, if the top two most likely flow patterns in the probabilistic forecast were considered instead of just the most likely one.

Second, an airport capacity prediction model was developed to deliver probabilistic arrival capacity forecasts on an hourly basis. The model takes as input the metroplex flow pattern forecast and weather features representing the forecast environmental conditions in the terminal area. Three different methods for machine learning were evaluated in a supervised learning regression scheme for arrival rate prediction: Bayesian Linear Regression, Random Forests, and Gaussian Process. The Gaussian Process model was chosen based on accuracy of point prediction and uncertainty quantification. An analysis of feature importance revealed that the metroplex configuration is the second most important factor influencing the arrival rates. Indeed, it was found that the two-layer prediction model outperformed a simplified model that does not account for metroplex system state in the capacity estimation process, especially for JFK and for longer forecast horizons. We also found that the data-driven capacity predictions obtained with the Gaussian Process model reduced the prediction error by 5.4% at JFK, 3.0% at EWR and 1.5% at LGA when compared with the baseline capacity estimates reported by the FAA.

Finally, in the capacity allocation block, optimization methods are used for AAR prescription and ground delay planning towards demonstrating the usability of the data-driven

capacity predictions and evaluating their impacts in the design of strategic TMIs such as GDPs. An AAR planning model was developed to prescribe optimal AARs that minimize the overall expected delay costs. The model has the following key features: (1) it is stochastic, i.e., capacity uncertainty is accounted for in the form of capacity profiles drawn from actual predictive distributions; (2) it enables revision of the plan in order to take advantage of updated information and more accurate predictions throughout the planning horizon; (3) it incorporates robustness goals, i.e., the decision-maker has the ability to select a desired level robustness for the TFM plan based on its relative valuation between efficiency and predictability. The model was used to optimize the traffic flow management plan for five test cases corresponding to historical GDP events at JFK. The results revealed that incorporating robustness in the design of the TFM plan not only helps increasing delay predictability, but can also lead to increased efficiency by protecting against errors in capacity predictions. The optimized AARs generated an overall reduction of 2.4% in the total GDP costs, if compared with the AARs actually implemented by the FAA. When robustness was accounted for in the optimization of the GDP plan, the observed reduction in total GDP costs was 9.7%.

5.2 Future Research Directions

The flight trajectory data analytics framework was demonstrated at the terminal area scale in this thesis. One clear direction for future research involves the use of the framework for characterization of air traffic flows at other scales. For instance, characterizing national-level air traffic flows can generate useful insights about en route traffic behavior for supporting en route air traffic management. One might be interested in exploiting routing patterns between origin and destination to diagnose en route flight inefficiencies, to better understand the factors that drive the selection of a particular route by flight operators, to discover feasible rerouting options or to assess the impacts of adverse weather conditions (such as convective weather) on route acceptability and capacity. At the sector level, characterization of traffic flow patterns may be useful for complexity management applications. One might be interested in exploiting patterns of air traffic control complexity in order to predict periods of low/high complexity and better inform staffing or sector reconfiguration decisions. While

the methods incorporated in the framework have general applicability, adaptations should be needed (e.g., data representation for clustering) to tailor the analysis for the intended application.

The predictive models developed in this thesis could benefit from higher quality information describing the input variables. For instance, traffic flow pattern predictions could be improved if actual demand information is used instead of flight schedules. In an actual implementation, this information could be retrieved from the Flight Scheduled Monitor of the TFMS. In the same way, capacity predictions could be improved with the use of different weather forecasts. Indeed, one particular limitation of the predictive models developed in this work is the use of short-term forecasts of convective weather impacts (ARSI). Ideally, the models should also incorporate features associated with long-term forecasts of convective weather impacts to improve the quality of strategic airport capacity predictions. In this direction, existing weather translation tools could be leveraged, such as the Traffic Flow Impact [148], which was designed to translate long-term convective weather forecasts into a measure of airspace permeability for strategic time horizons.

Within the airport capacity planning framework, one potential research direction is extending the capacity allocation model to account for differences in robustness goals between different flight operators, along the lines of decentralized approaches for traffic flow management, and investigating whether it is possible to prescribe an optimal time for revision of the traffic flow management plan based on the uncertainty in capacity forecasts.

Finally, this thesis was focused on evaluating the potential impacts of data-driven models towards supporting ATM decision-making using historical data, but a lot of work is required into actual decision-support tool development. In this direction, human-in-the-loop simulations could be performed in order to test the models for a wider range of scenarios and obtain user feedback. As mentioned previously, the predictive models could be enhanced with a richer set of features. Research efforts should also focus on how these tools can be designed to best leverage the strengths of analytics and computation to enable fast and reliable real-time implementation as well as to foster human engagement.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix A

Trajectory Clustering Analysis Results - New York Metroplex

Table A.1: DBSCAN parameter settings used for trajectory clustering.

Airport flow	**MinPts**	ε	**Number of clusters identified**
JFK Arrivals	6	1.4	20
EWR Arrivals	6	1.4	14
LGA Arrivals	6	1.2	16
JFK Departures	6	1.2	24
EWR Departures	6	1.3	13
LGA Departures	6	1.3	18

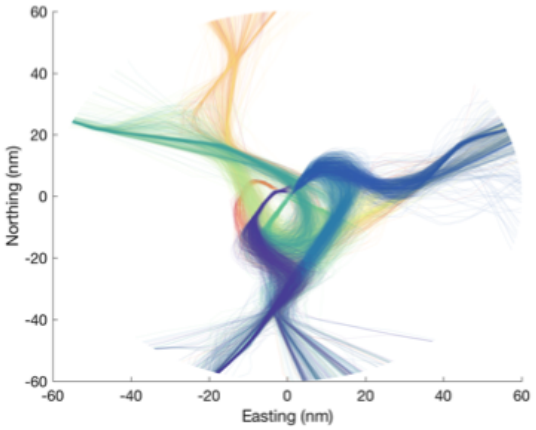


Figure A-1: JFK arrival trajectory clusters; each color represents one cluster.

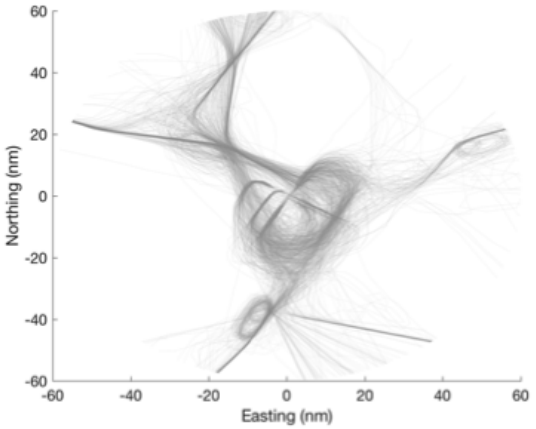


Figure A-2: JFK arrival trajectories labeled as noise.

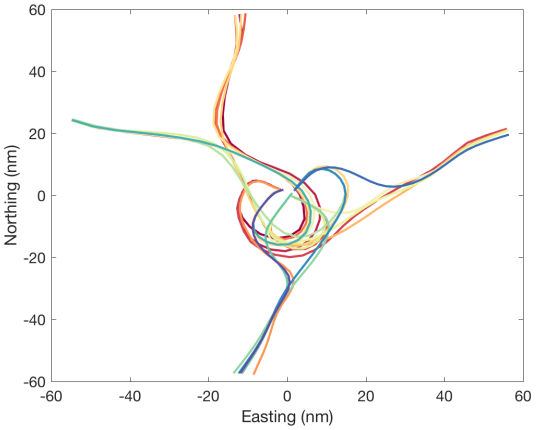


Figure A-3: Centroids of JFK arrival trajectory clusters.

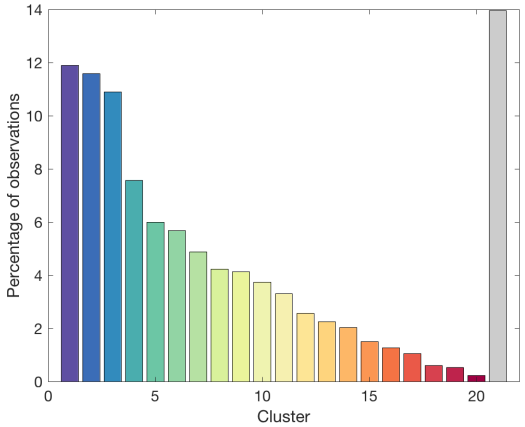


Figure A-4: Distribution of JFK arrival trajectories by cluster; grey bar represents the percentage of noise.

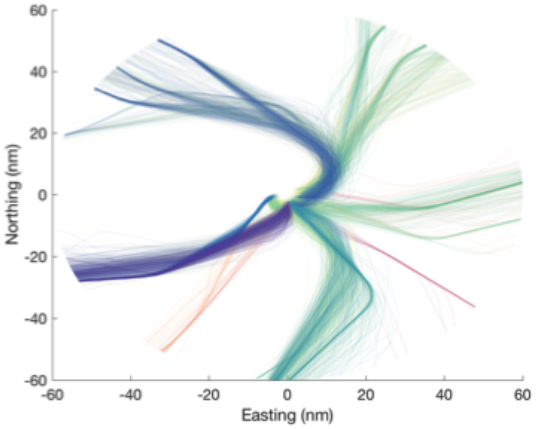


Figure A-5: JFK departure trajectory clusters; each color represents one cluster.

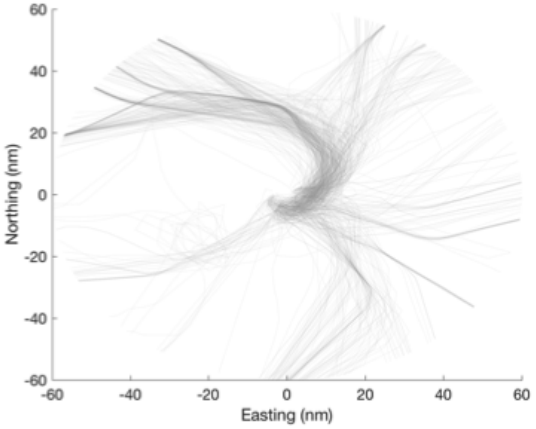


Figure A-6: JFK departure trajectories labeled as noise.

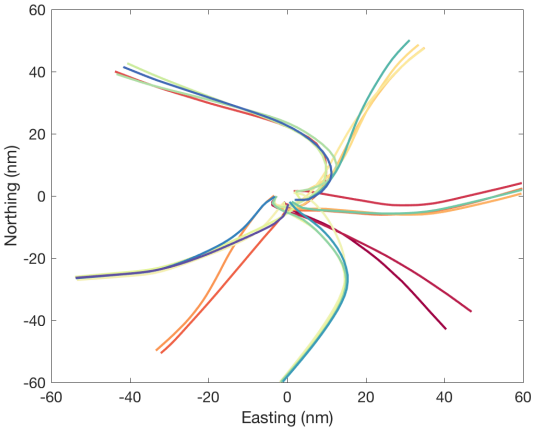


Figure A-7: Centroids of JFK departure trajectory clusters.

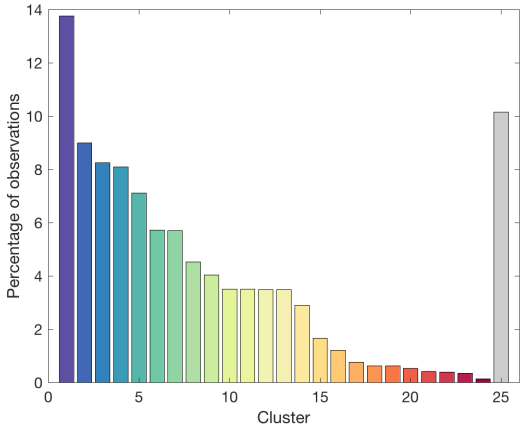


Figure A-8: Distribution of JFK departure trajectories by cluster; grey bar represents the percentage of noise.

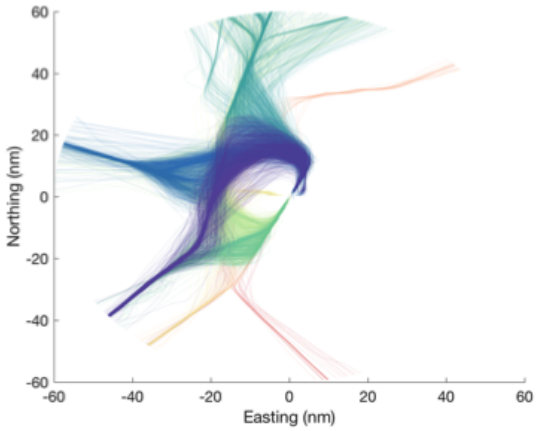


Figure A-9: EWR arrival trajectory clusters; each color represents one cluster.

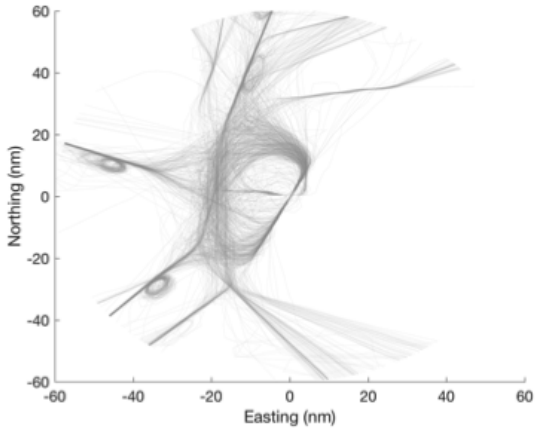


Figure A-10: EWR arrival trajectories labeled as noise.

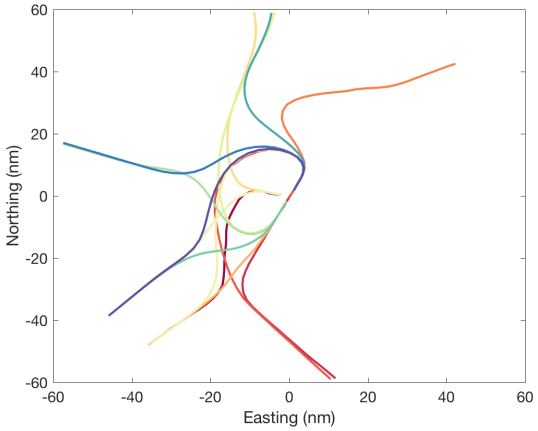


Figure A-11: Centroids of EWR arrival trajectory clusters.

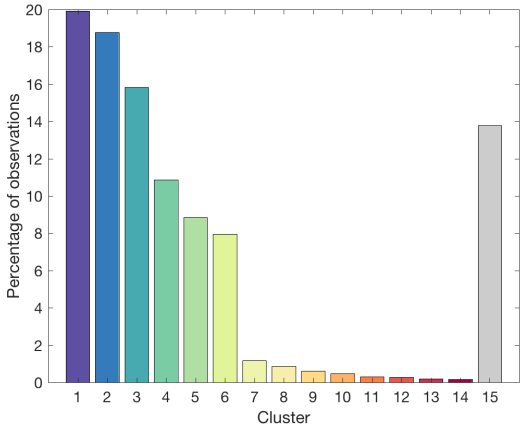


Figure A-12: Distribution of EWR arrival trajectories by cluster; grey bar represents the percentage of noise.

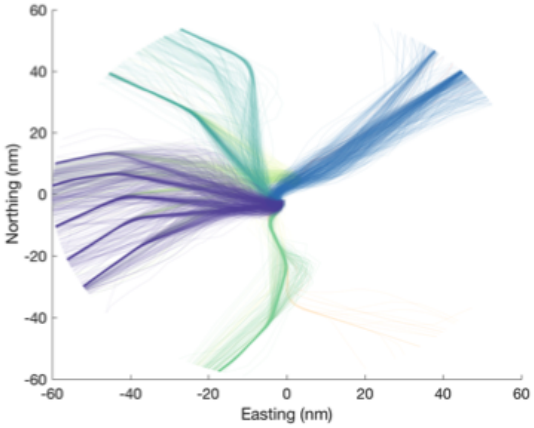


Figure A-13: EWR departure trajectory clusters; each color represents one cluster.

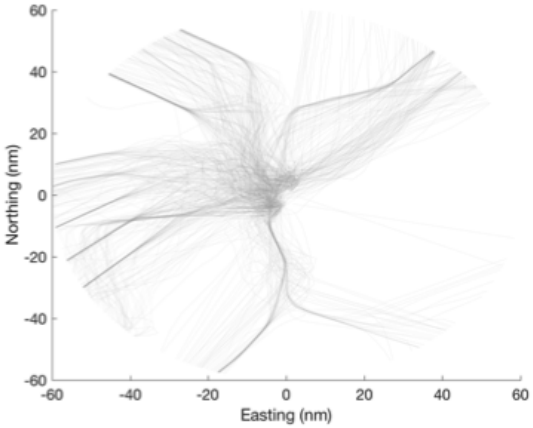


Figure A-14: EWR departure trajectories labeled as noise.

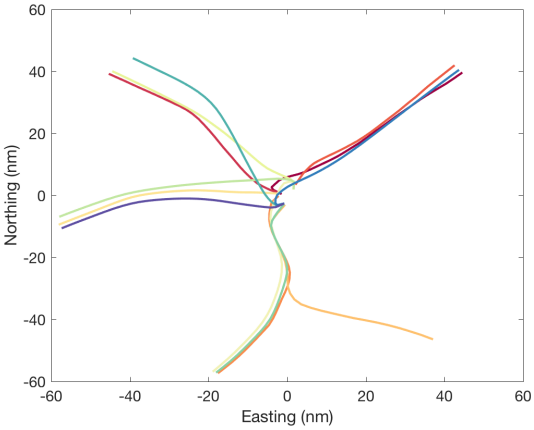


Figure A-15: Centroids of EWR departure trajectory clusters.

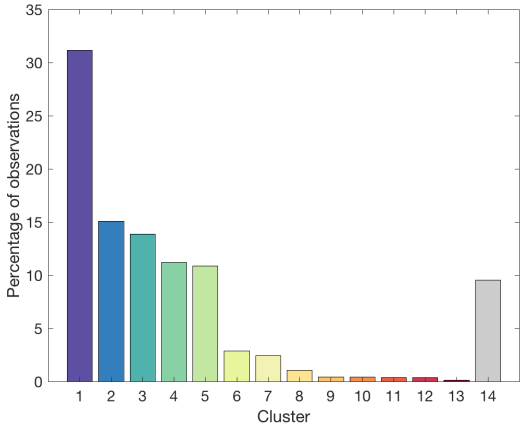


Figure A-16: Distribution of EWR departure trajectories by cluster; grey bar represents the percentage of noise.

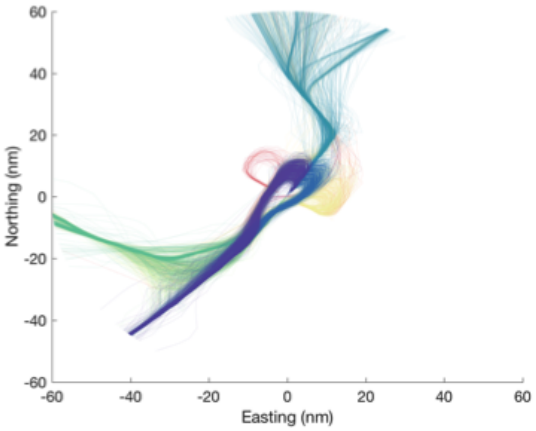


Figure A-17: LGA arrival trajectory clusters; each color represents one cluster.

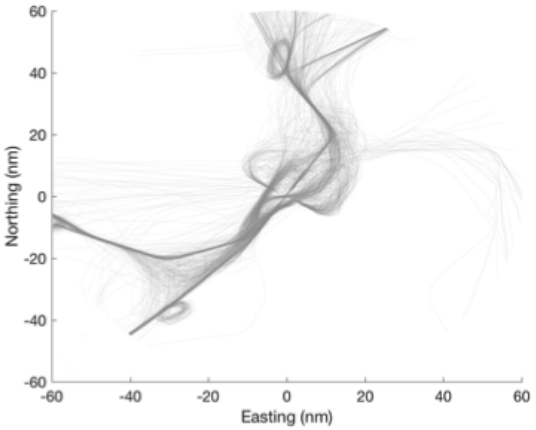


Figure A-18: LGA arrival trajectories labeled as noise.

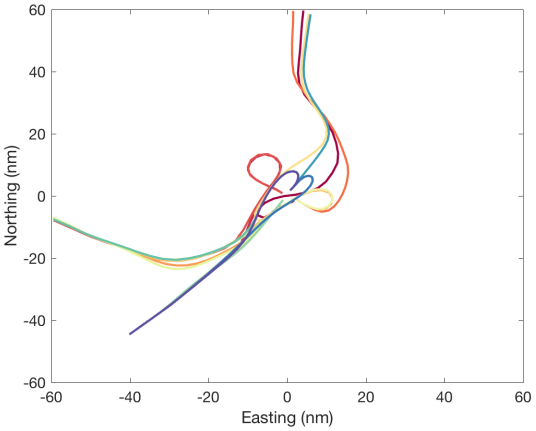


Figure A-19: Centroids of LGA arrival trajectory clusters.

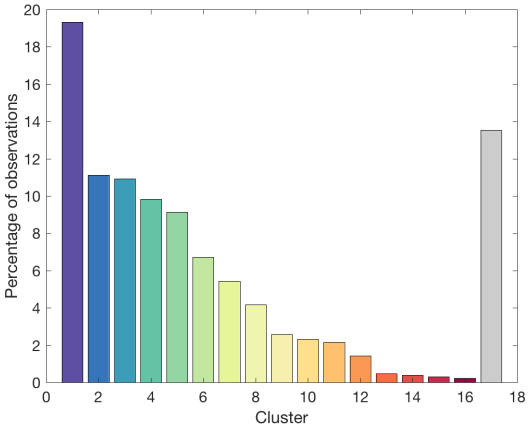


Figure A-20: Distribution of LGA arrival trajectories by cluster; grey bar represents the percentage of noise.

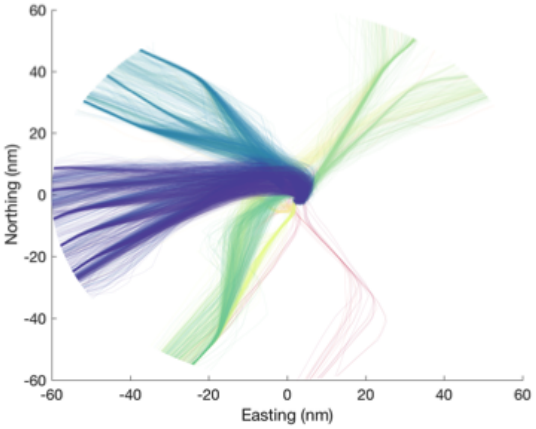


Figure A-21: LGA departure trajectory clusters; each color represents one cluster.

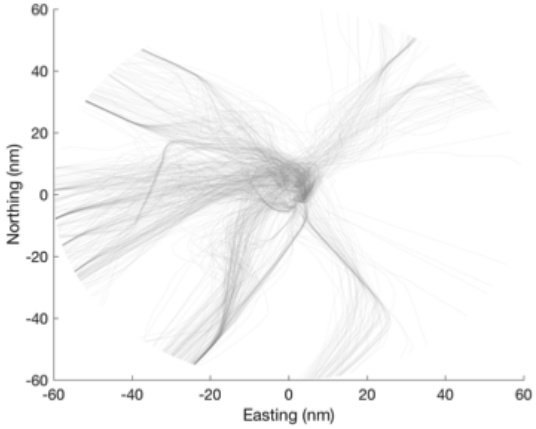


Figure A-22: LGA departure trajectories labeled as noise.

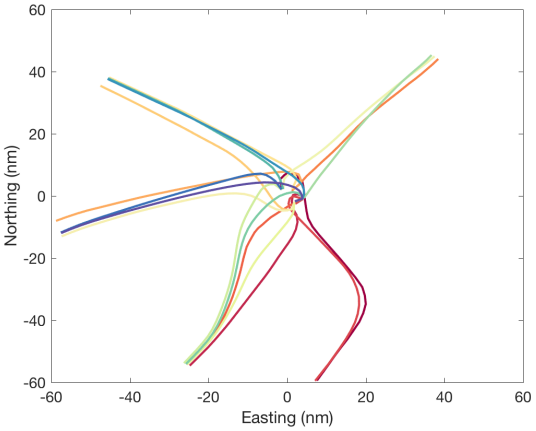


Figure A-23: Centroids of LGA departure trajectory clusters.

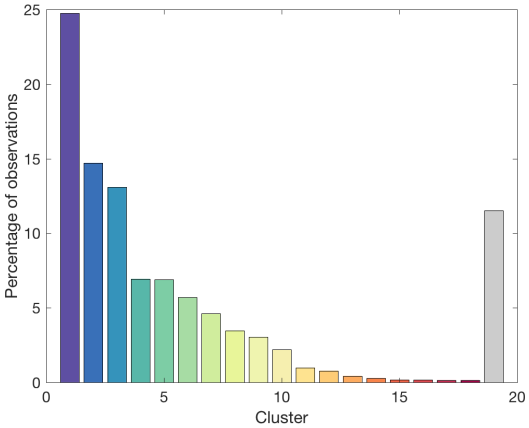


Figure A-24: Distribution of LGA departure trajectories by cluster; grey bar represents the percentage of noise.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix B

Trajectory Clustering Analysis Results - Hong Kong Metroplex

Table B.1: DBSCAN parameter settings used for trajectory clustering.

Airport flow	**MinPts**	ε	**Number of clusters identified**
HKG Arrivals	6	1.4	15
SZX Arrivals	6	1.4	8
MFM Arrivals	6	1.6	6
HKG Departures	6	1.4	10
SZX Departures	6	1.4	12
MFM Departures	6	1.4	6

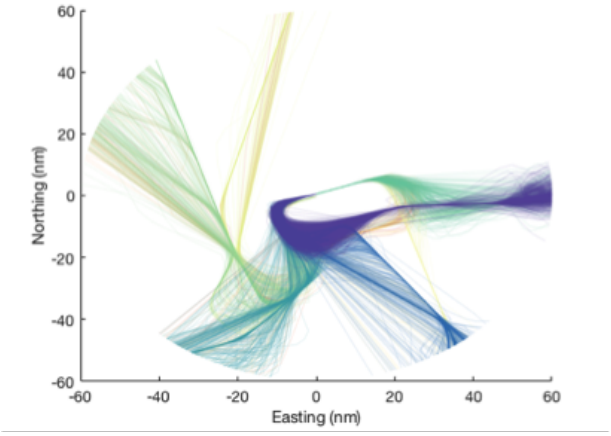


Figure B-1: HKG arrival trajectory clusters; each color represents one cluster.

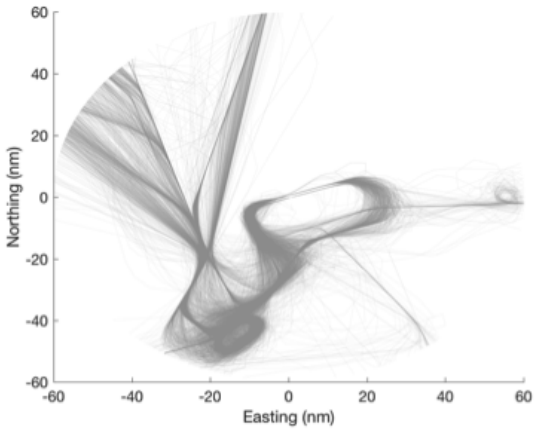


Figure B-2: HKG arrival trajectories labeled as noise.

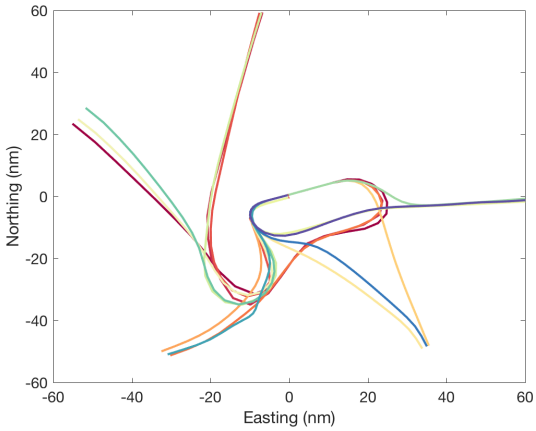


Figure B-3: Centroids of HKG arrival trajectory clusters.

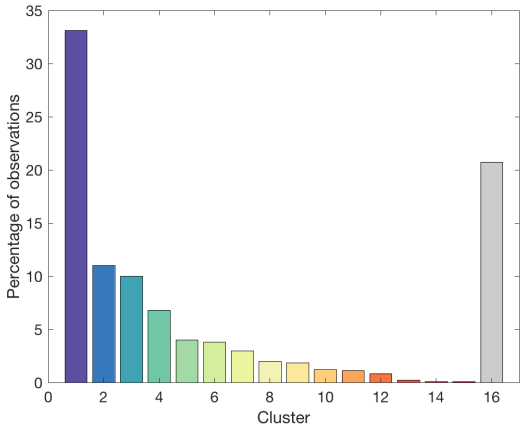


Figure B-4: Distribution of HKG arrival trajectories by cluster; grey bar represents the percentage of noise.

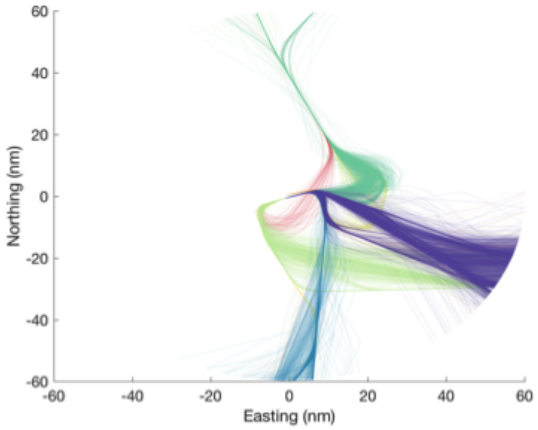


Figure B-5: HKG departure trajectory clusters; each color represents one cluster.

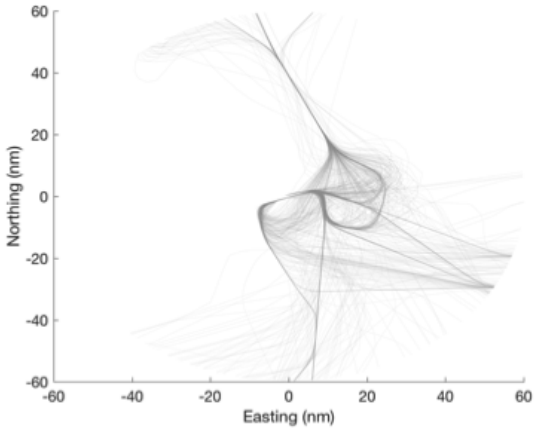


Figure B-6: HKG departure trajectories labeled as noise.

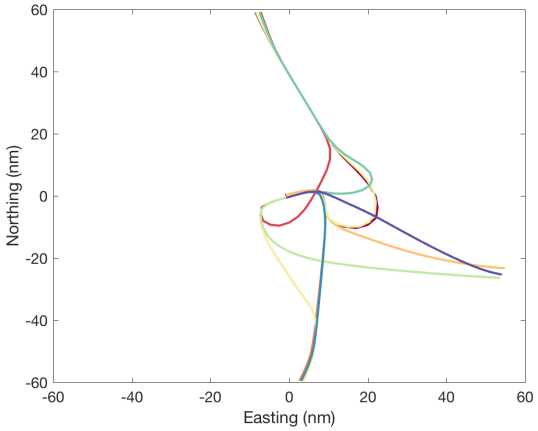


Figure B-7: Centroids of HKG departure trajectory clusters.

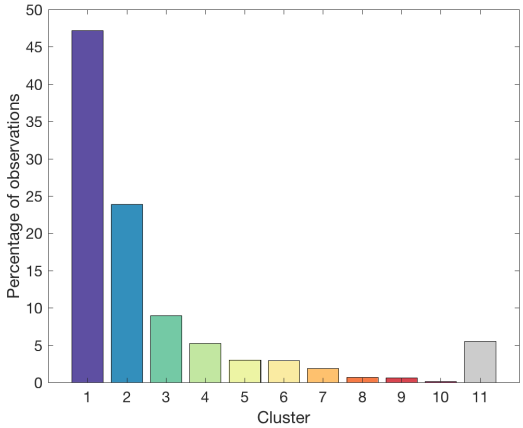


Figure B-8: Distribution of HKG departure trajectories by cluster; grey bar represents the percentage of noise.

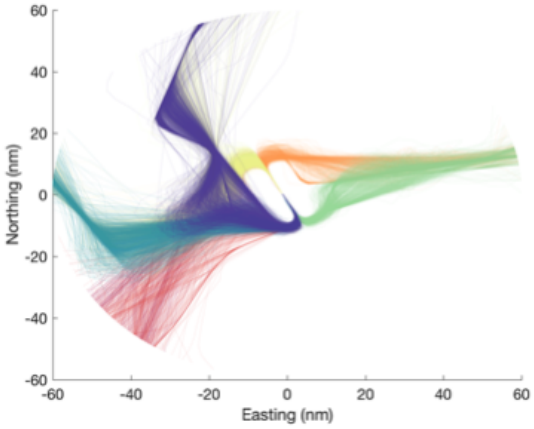


Figure B-9: SZX arrival trajectory clusters; each color represents one cluster.

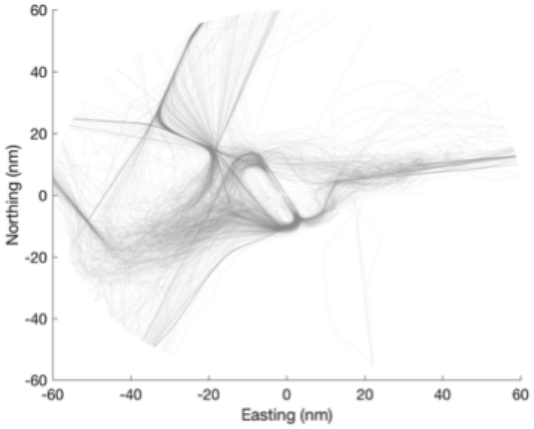


Figure B-10: SZX arrival trajectories labeled as noise.

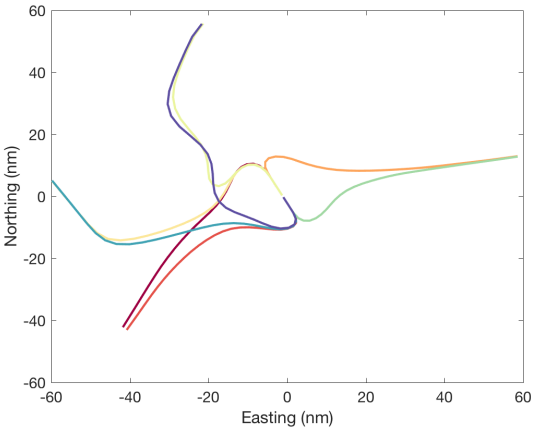


Figure B-11: Centroids of SZX arrival trajectory clusters.

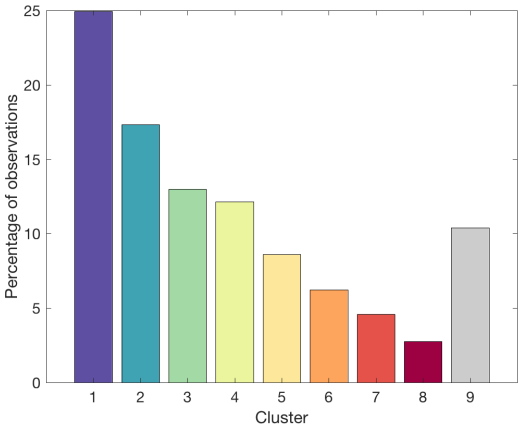


Figure B-12: Distribution of SZX arrival trajectories by cluster; grey bar represents the percentage of noise.

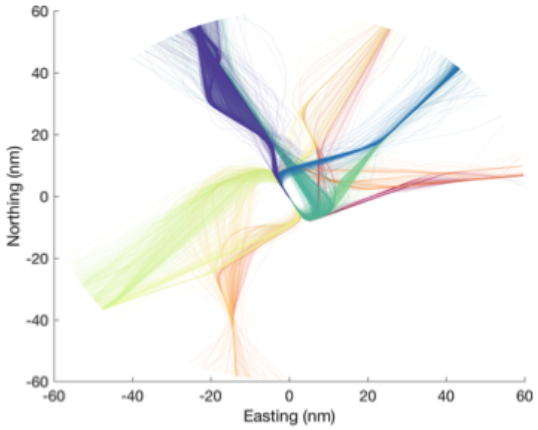


Figure B-13: SZX departure trajectory clusters; each color represents one cluster.

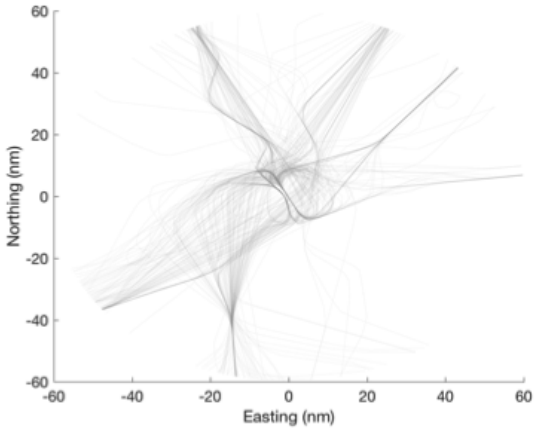


Figure B-14: SZX departure trajectories labeled as noise.

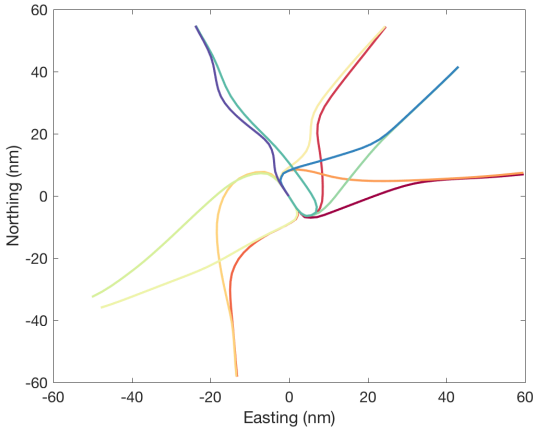


Figure B-15: Centroids of SZX departure trajectory clusters.

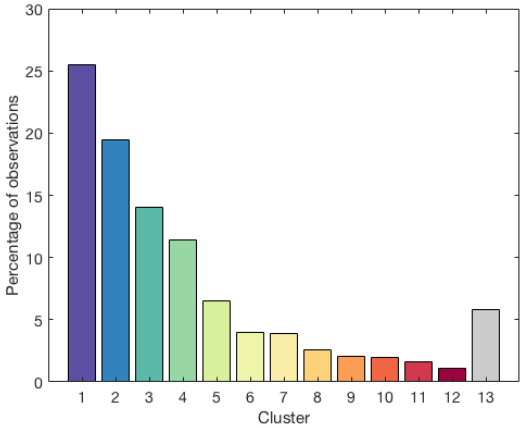


Figure B-16: Distribution of SZX departure trajectories by cluster; grey bar represents the percentage of noise.

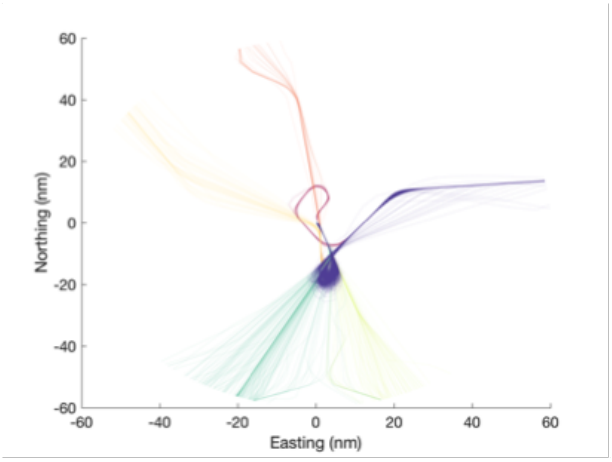


Figure B-17: MFM arrival trajectory clusters; each color represents one cluster.

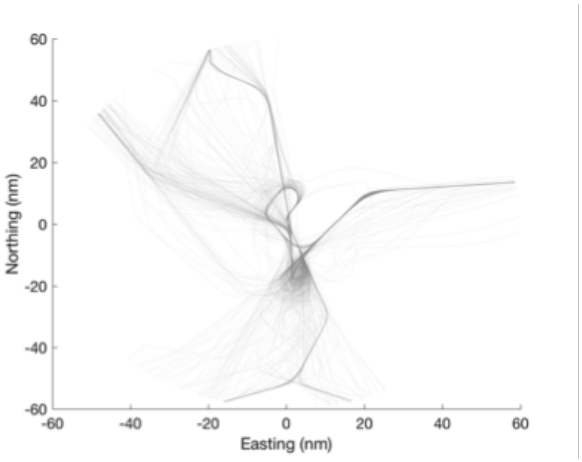


Figure B-18: MFM arrival trajectories labeled as noise.

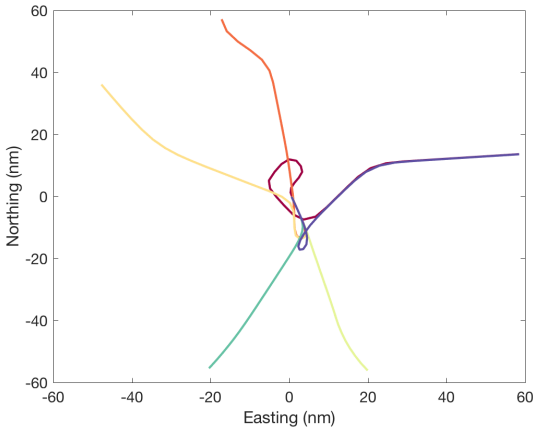


Figure B-19: Centroids of MFM arrival trajectory clusters.

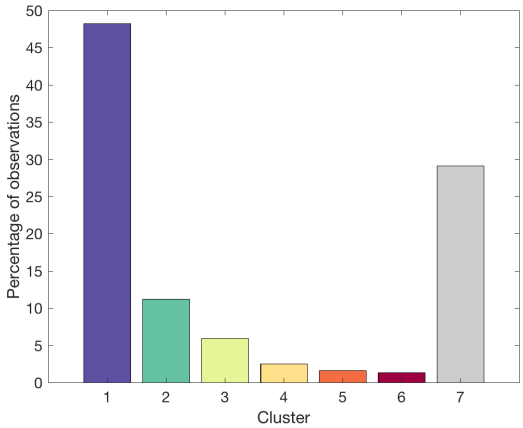


Figure B-20: Distribution of MFM arrival trajectories by cluster; grey bar represents the percentage of noise.

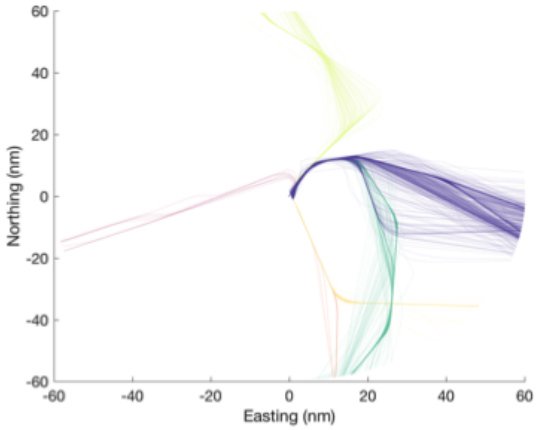


Figure B-21: MFM departure trajectory clusters; each color represents one cluster.

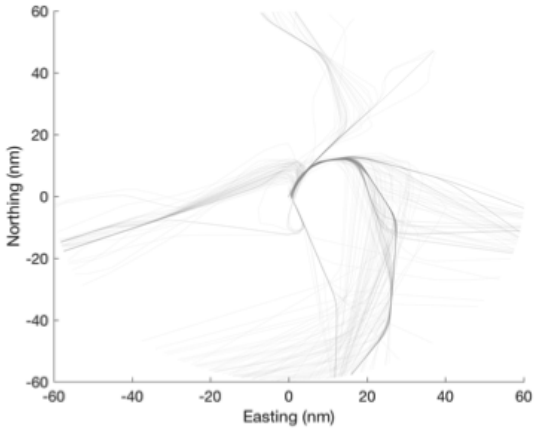


Figure B-22: MFM departure trajectories labeled as noise.

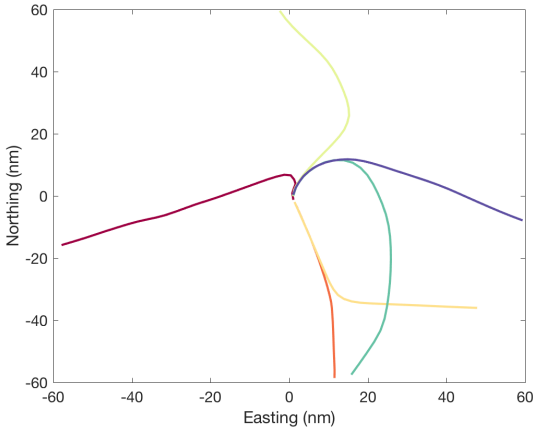


Figure B-23: Centroids of MFM departure trajectory clusters.

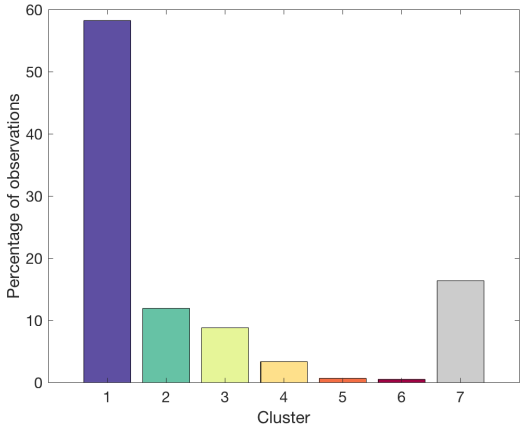


Figure B-24: Distribution of MFM departure trajectories by cluster; grey bar represents the percentage of noise.

THIS PAGE INTENTIONALLY LEFT BLANK

Appendix C

Trajectory Clustering Analysis Results - Sao Paulo Metroplex

Table C.1: DBSCAN parameter settings used for trajectory clustering.

Airport flow	**MinPts**	ε	**Number of clusters identified**
GRU Arrivals	6	1.4	10
CGH Arrivals	6	1.4	6
VCP Arrivals	6	1.4	8
GRU Departures	6	1	14
CGH Departures	6	1	10
VCP Departures	6	1.4	12

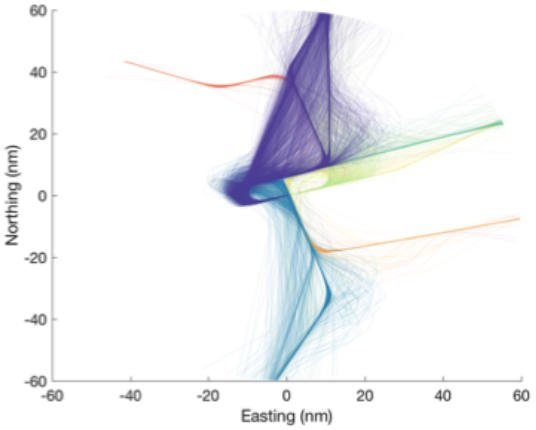


Figure C-1: GRU arrival trajectory clusters; each color represents one cluster.

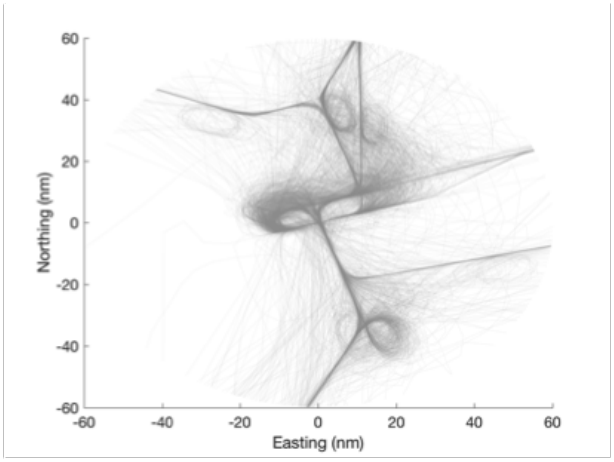


Figure C-2: GRU arrival trajectories labeled as noise.

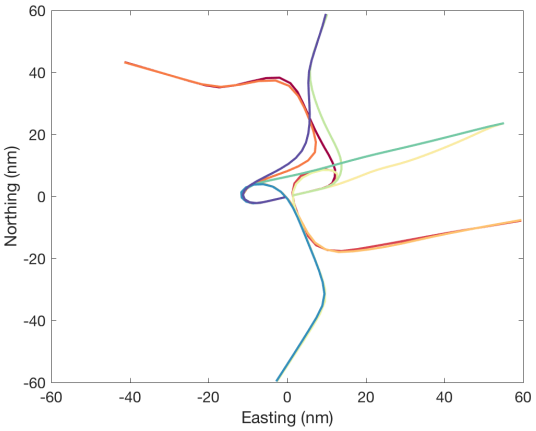


Figure C-3: Centroids of GRU arrival trajectory clusters.

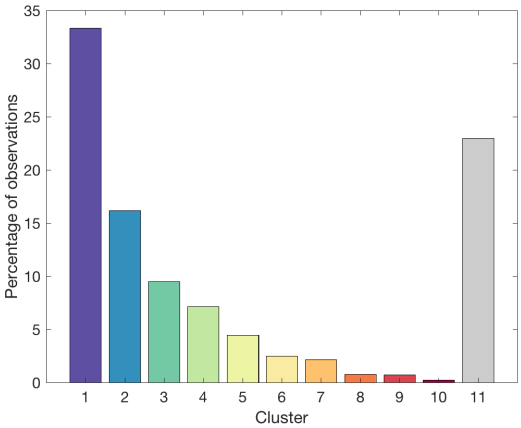


Figure C-4: Distribution of GRU arrival trajectories by cluster; grey bar represents the percentage of noise.

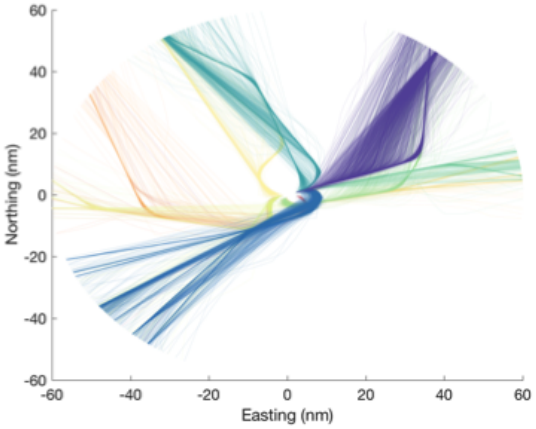


Figure C-5: GRU departure trajectory clusters; each color represents one cluster.

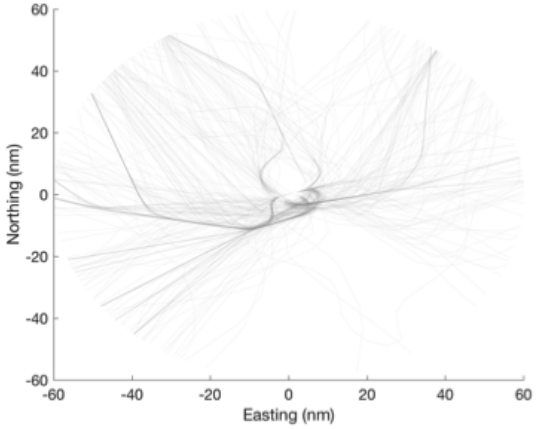


Figure C-6: GRU departure trajectories labeled as noise.

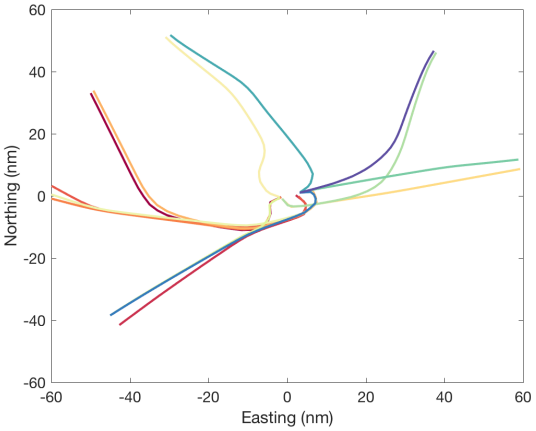


Figure C-7: Centroids of GRU departure trajectory clusters.

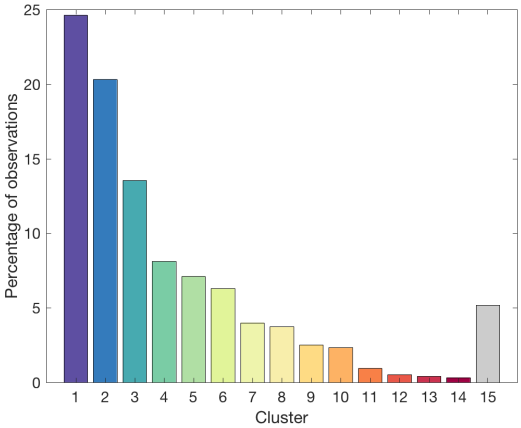


Figure C-8: Distribution of GRU departure trajectories by cluster; grey bar represents the percentage of noise.

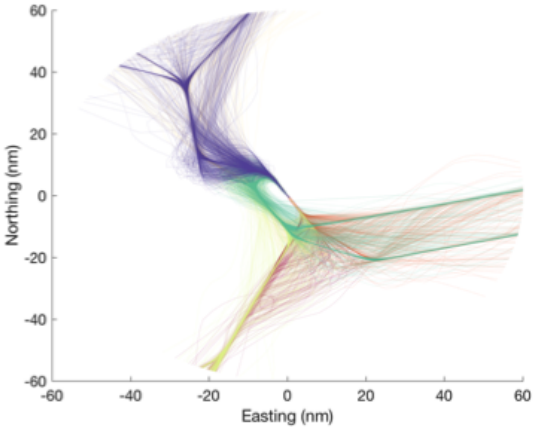


Figure C-9: CGH arrival trajectory clusters; each color represents one cluster.

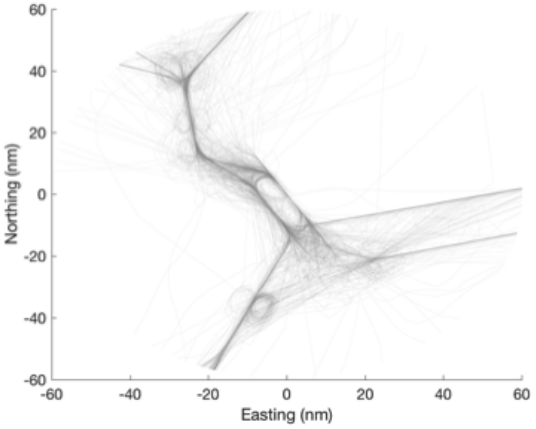


Figure C-10: CGH arrival trajectories labeled as noise.

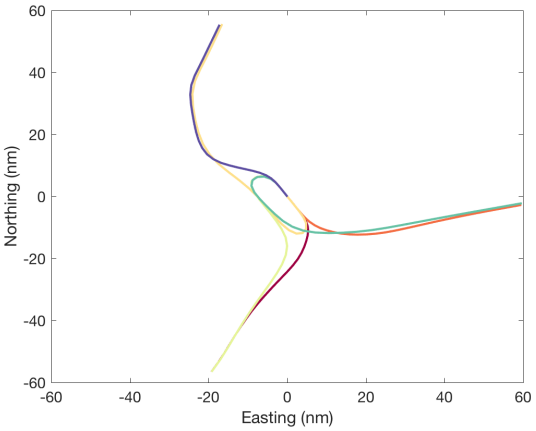


Figure C-11: Centroids of CGH arrival trajectory clusters.

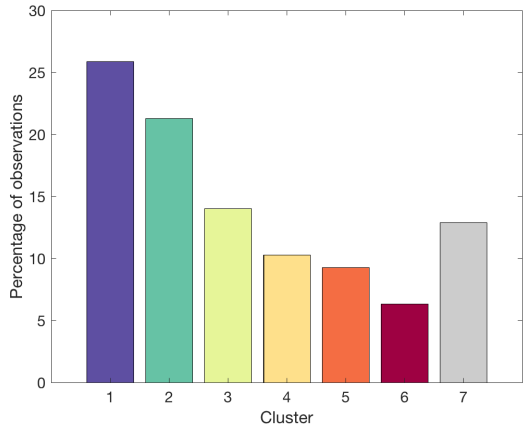


Figure C-12: Distribution of CGH arrival trajectories by cluster; grey bar represents the percentage of noise.

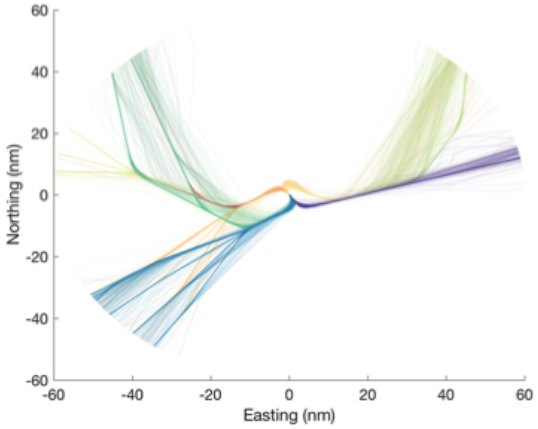


Figure C-13: CGH departure trajectory clusters; each color represents one cluster.

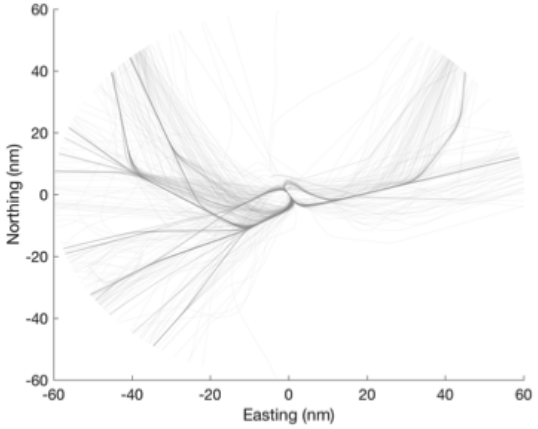


Figure C-14: CGH departure trajectories labeled as noise.

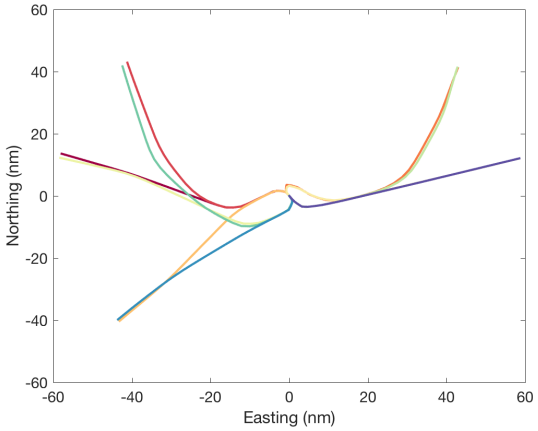


Figure C-15: Centroids of CGH departure trajectory clusters.

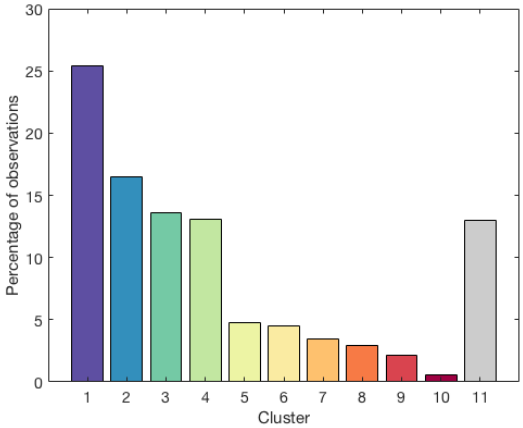


Figure C-16: Distribution of CGH departure trajectories by cluster; grey bar represents the percentage of noise.

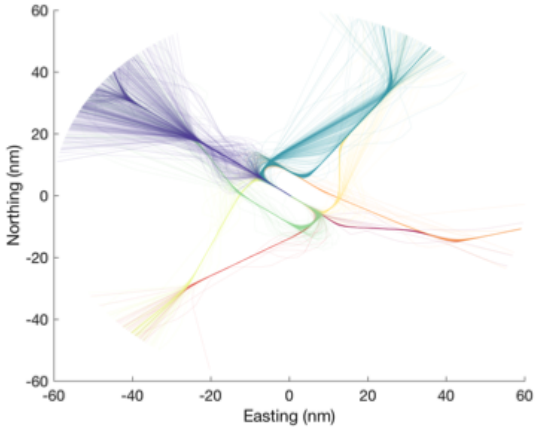


Figure C-17: VCP arrival trajectory clusters; each color represents one cluster.

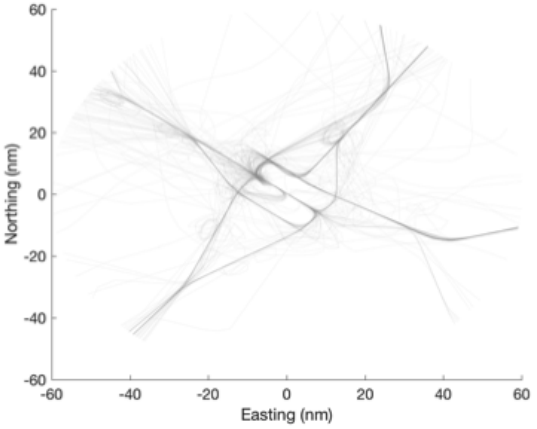


Figure C-18: VCP arrival trajectories labeled as noise.

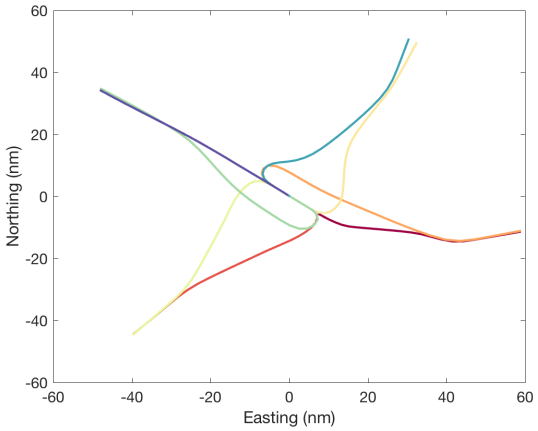


Figure C-19: Centroids of VCP arrival trajectory clusters.

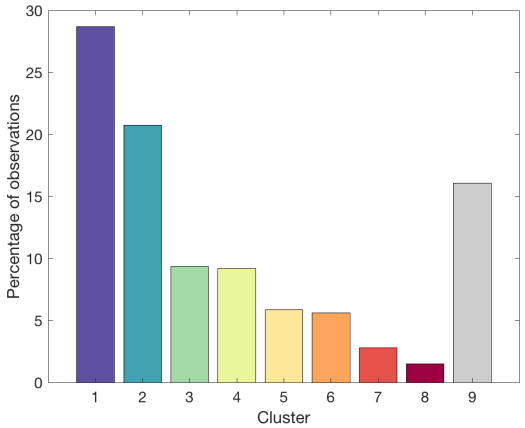


Figure C-20: Distribution of VCP arrival trajectories by cluster; grey bar represents the percentage of noise.

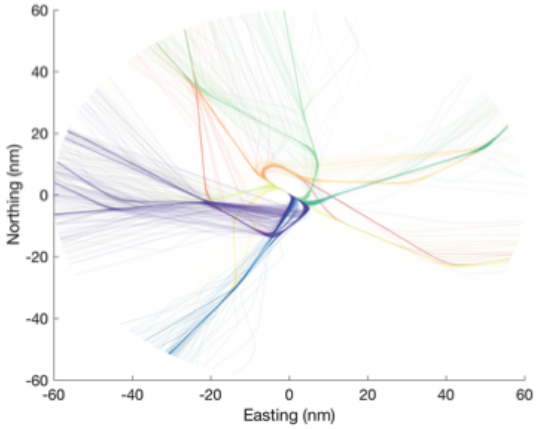


Figure C-21: VCP departure trajectory clusters; each color represents one cluster.

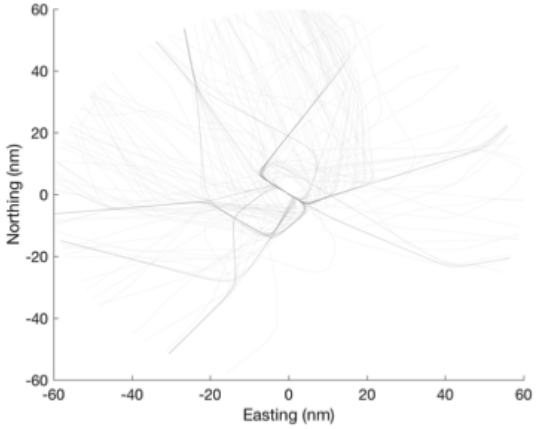


Figure C-22: VCP departure trajectories labeled as noise.

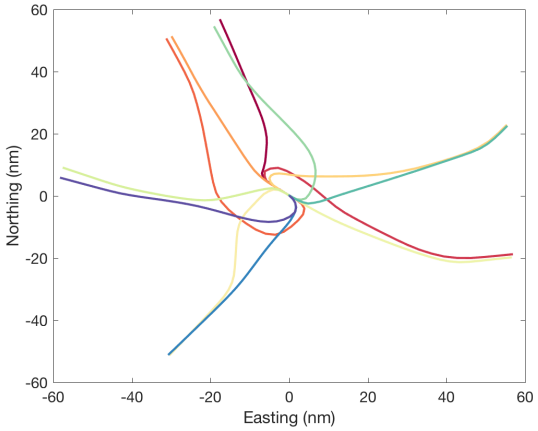


Figure C-23: Centroids of VCP departure trajectory clusters.

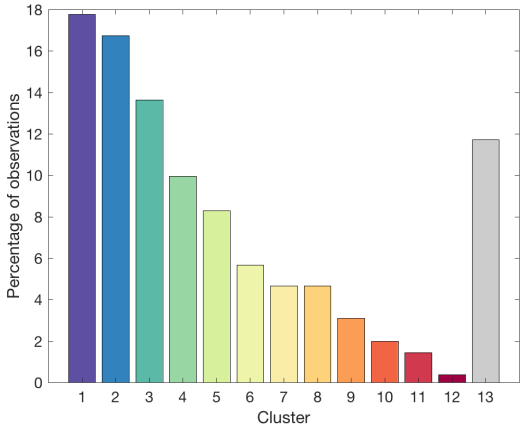


Figure C-24: Distribution of VCP departure trajectories by cluster; grey bar represents the percentage of noise.

THIS PAGE INTENTIONALLY LEFT BLANK

Bibliography

[1] J. Gudmundsson, P. Laube, and T. Wolle. *Encyclopedia of GIS*, chapter Movement patterns in spatio-temporal data, pages 726–732. Springer, 1st edition, 2008.

[2] New York Air Route Traffic Control Center. TRACON letters of agreement. `http://nyartcc.org/wiki/index.php/LGA_SOP`, (accessed 2017-06).

[3] J. P. B. Clarke, L. Ren, E. McClain, D. Schleicher, S. Timar, A. Saraf, D. Crisp, R. Gutterud, R. Laroza, T. Thompson, C. Cross, and T. Lewis. Evaluating concepts for operations in metroplex terminal area airspace. *Journal of Aircraft*, 49(3):758–773, 2012.

[4] International Air Transport Association. Fact Sheet: Industry Statistics. `http://www.iata.org/pressroom/facts_figures/Pages/index.aspx`, (accessed 2017-06).

[5] Air Transport Action Group. Aviation: Benefits Beyond Borders. `https://www.aviationbenefits.org/media/149668/abbb2016_full_a4_web.pdf`, (accessed 2017-06).

[6] International Air Transport Association. IATA forecasts passenger demand to double over 20 years. `http://www.iata.org/pressroom/pr/Pages/2016-10-18-02.aspx`, (accessed 2017-06).

[7] Federal Aviation Administration. FAA Aerospace Forecast - Fiscal Years 2016-2036. `https://www.faa.gov/data_research/aviation/aerospace_forecasts/media/FY2016-36_FAA_Aerospace_Forecast.pdf`, (accessed 2017-06).

[8] International Civil Aviation Organization. Global Air Navigation Plan for CNS/ATM Systems. `http://www.icao.int/publications/Documents/9750_2ed_en.pdf`, (accessed 2017-06).

[9] Bureau of Transportation Statistic. On-time performance - flight delays at a glance. `http://www.transtats.bts.gov/HomeDrillChart.asp?URL_SelectMonth=11&URL_SelectYear=2016`, (accessed 2017-06).

[10] Eurocontrol. Annual Network Operations Report 2015. `https://www.eurocontrol.int/publications/annual-network-operations-report-2015`, (accessed 2017-06).

[11] National Center of Excellence for Aviation Operations Research. Total Delay Impact Study. http://its.berkeley.edu/sites/default/files/NEXTOR_TDI_Report_Final_October_2010.pdf, (accessed 2017-06).

[12] Federal Aviation Administration. NextGen. https://www.faa.gov/nextgen/, (accessed 2017-06).

[13] Single European Sky ATM Research. Discover SESAR. http://www.sesarju.eu/index.php/discover-sesar, (accessed 2017-06).

[14] Transport Ministry of Land, Infrastructure and Tourism. CARATS: Collaborative Actions for Renovation of Air Traffic Systems. http://www.mlit.go.jp/common/000128185.pdf, (accessed 2017-06).

[15] Brazilian Department of Airspace Control. SIRIUS Program. https://www.decea.gov.br/sirius/?lang=en, (accessed 2017-06).

[16] International Civil Aviation Organization. A39-WP/304: China's Strategy for Modernizing Air Traffic Management. https://www.icao.int/Meetings/a39/Documents/WP/wp_304_en.pdf, (accessed 2017-06).

[17] Federal Aviation Administration. The Future of the NAS. https://www.faa.gov/nextgen/media/futureofthenas.pdf, (accessed 2017-06).

[18] Eurocontrol. European ATM Master Plan 2015: The Roadmap for Delivering High Performing Aviation for Europe. https://ec.europa.eu/transport/sites/transport/files/modes/air/sesar/doc/eu-atm-master-plan-2015.pdf, (accessed 2017-06).

[19] H. Hu, Y. Wen, T.-S. Chua, and X. Li. Toward scalable systems for big data analytics: A technology tutorial. *IEEE Access*, 2:652–687, 2014.

[20] G. Blackett. Analytics Network - OR and Analytics. https://www.theorsociety.com/Pages/SpecialInterest/AnalyticsNetwork_analytics.aspx, (accessed 2017-06).

[21] D. Bertsimas, A. O'Hair, and D. Pulleyblank. *The Analytics Edge*. Dynamic Ideas LLC, 1st edition, 2016.

[22] Commission of the European Communities. Air Traffic Management. http://aei.pitt.edu/1134/1/air_traffic_white_paper_COM_96_57.pdf, (accessed 2017-06).

[23] P. Belobaba, A. Odoni, and C. Barnhart. *The Global Airline Industry*. Wiley, 2nd edition, 2015.

[24] International Civil Aviation Organization. *Rules of the Air, Annex 2 of the Convention on International Civil Aviation*. ICAO, 10th edition, 2005.

[25] Federal Aviation Administration. AC 90-115 - Collaborative Trajectory Options Program (CTOP). https://www.faa.gov/regulations_policies/advisory_circulars/index.cfm/go/document.information/documentID/1024987, (accessed 2017-06).

[26] N. M. Smith, C. Brasil, P. U. Lee, N. Buckley, C. Gabriel, C. Mohlenbrink, F. Omar, B. Parke, C. Speridakos, and H. S. Yoo. Integrated demand management: Coordinating strategic and tactical flow scheduling operations. In *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016.

[27] Federal Aviation Administration. Air Traffic by the Numbers. https://www.faa.gov/air_traffic/by_the_numbers/, (accessed 2017-06).

[28] Federal Aviation Administration. Passenger boarding (enplanement) and all-cargo data for U.S. airports. https://www.faa.gov/airports/planning_capacity/passenger_allcargo_stats/passenger/, (accessed 2017-06).

[29] R. De Neufville and A. Odoni. *Airport Systems: Planning, Design, and Management.* McGraw-Hill, 2nd edition, 2013.

[30] P. A. Bonnefoy. *Scalability of the Air Transportation System and Development of Multi-Airport Systems: A Worldwide Perspective.* Ph.D. dissertation, Massachusetts Institute of Technology, 2008.

[31] L. Ren, J. P. B. Clarke, D. Schleicher, S. Timar, A. Saraf, D. Crisp, R. Gutterud, T. Lewis, and T. Thompson. Contrast and comparison of metroplex operations (an air traffic management study of Atlanta, Los Angeles, New York, and Miami. In *9th AIAA Aviation Technology, Integration, and Operations Conference*, 2009.

[32] A. D. Donaldson and R. J. Hansman. Capacity improvement potential for the new york metroplex system. In *10th AIAA Aviation Technology, Integration and Operations Conference*, 2010.

[33] A. R. Odoni. *Flow Control of Congested Networks*, chapter The Flow Management Problem in Air Traffic Control, pages 269–288. Springer-Verlag, Berlin, 1987.

[34] M. Terrab and A. R. Odoni. Strategic flow management for air traffic control. *Operations Research*, 41(1):138–152, 1993.

[35] O. Richetta and A. R. Odoni. Solving optimally the static ground-holding policy problem in air traffic control. *Transportation Science*, 27(3):228–238, 1993.

[36] M. Ball, A. R. Odoni, and R. Rifkin. Stochastic integer program with dual network structure and its application to the ground-holding problem. *Operations Research*, 51(1):167–171, 2003.

[37] O. Richetta and A. R. Odoni. Dynamic solution to the ground-holding problem in air traffic control. *Transportation Research Part A: Policy and Practice*, 28(3):167–185, 1994.

[38] A. Mukherjee and M. Hansen. Dynamic stochastic model for the single airport ground holding problem. *Transportation Science*, 41(4):444–456, 2007.

[39] Y. Liu and M. Hansen. Incorporating predictability into cost optimization for ground delay programs. *Transportation Science*, 50(1):132–149, 2015.

[40] P. Swaroop and M. Ball. Consensus-building mechanism for setting service expectations in air traffic flow management. *Transportation Research Record: Journal of the Transportation Research Board*, 2325:87–96, 2013.

[41] A. Evans, V. Vaze, and C. Barnhart. Airline-driven performance-based air traffic management: Game theoretic models and multicriteria evaluation. *Transportation Science*, 50(1):180–203, 2016.

[42] M. Ball, P. Swaroop, C. Barnhart, C. Yan, M. Hansen, L. Kang, Y. Liu, and V. Vaze. Service level expectation setting for air traffic flow management: Practical challenges and benefits assessment. In *12th USA/Europe Air Traffic Management Research and Development Seminar*, 2017.

[43] C. Yan, V. Vaze, and C. Barnhart. Airline-driven ground delay programs: a benefits assessment. *Transportation Research Part C: Emerging Technologies*, 89:268–288, 2018.

[44] D. Bertsimas and S. S. Patterson. The air traffic flow management problem with enroute capacities. *Operations Research*, 46(3):406–422, 1998.

[45] D. Bertsimas and S. S. Patterson. The traffic flow management rerouting problem in air traffic control: a dynamic network flow approach. *Transportation Science*, 34:239–255, 2000.

[46] D. Bertsimas, G. Lulli, and A. R. Odoni. An integer optimization approach to large-scale air traffic flow management. *Operations Research*, 59(1):211–227, 2011.

[47] A. Agustín, Alonso-Ayuso A., L. F. Escudero, and C. Pizarro. On air traffic flow management with rerouting. Part i: Deterministic case. *European Journal of Operational Research*, 219:156–166, 2012.

[48] A. Agustín, Alonso-Ayuso A., L. F. Escudero, and C. Pizarro. On air traffic flow management with rerouting. Part ii: Stochastic case. *European Journal of Operational Research*, 219:167–177, 2012.

[49] H. Balakrishnan and B. G. Chandran. Optimal large-scale air traffic flow management. `http://web.mit.edu/Hamsa/www/pubs/BalakrishnanChandran_ATFM.pdf`, 2014.

[50] P. K. Menon, G. D. Sweriduk, T. Lam, G. M. Diaz, and K. Bilimoria. Computer-aided eulerian air traffic flow modeling and predictive control. *AIAA Journal of Guidance, Control and Dynamics*, 29:12–19, 2006.

[51] B. Sridhar, T. Soni, K. Sheth, and G. B. Chatterji. Aggregate flow model for air-traffic management. *Journal of Guidance, Control, and Dynamics*, 26:992–997, 2006.

[52] D. Sun and A. M. Bayen. Multicommodity eulerian-lagrangian large-capacity cell transmission model for en route traffic. *Journal of Guidance, Control, and Dynamics*, 31(3):616–628, 2008.

[53] A. Marzuoli, M. Gariel, A. Vela, and E. Feron. Data-based modeling and optimization of en route traffic. *Journal of Guidance, Control, and Dynamics*, 37(6):1930–1945, 2014.

[54] A. Bombelli, L. Soler, E. Trumbauer, and K. D. Mease. Strategic air traffic planning with fréchet distance aggregation and rerouting. *Journal of Guidance, Control, and Dynamics*, 40(5):1117–1129, 2017.

[55] J. E. Beasley, M. Krishnamoorthy, Y.M. Sharaiha, and D. Abramson. Scheduling aircraft landings - the static case. *Transportation Science*, 34(2):180–197, 2000.

[56] H. Balakrishnan and B. Chandran. Algorithms for scheduling runway operations under constrained position shifting. *Operations Research*, 58(6):1650–1665, 2010.

[57] G. Solveling, S. Solak, J. P. Clarke, and E. Johnson. Runway operations optimization in the presence of uncertainties. *Journal of Guidance, Control, and Dynamics*, 34(5):1373–1382, 2011.

[58] M. C. R. Murça and C. Müller. Control-based optimization approach for aircraft scheduling in a terminal area with alternative arrival routes. *Transportation Research Part E*, 73:96–113, 2015.

[59] D. Bertsimas, M. Frankovich, and A. Odoni. Optimal selection of airport runway configurations. *Operations research*, 59(6):1407–1419, 2011.

[60] S. Rathinam, J. Montoya, and Y. Jung. An optimization model for reducing aircraft taxi times at the dallas fort worth international airport. In *26th International Congress of the Aeronautical Sciences*, 2008.

[61] W. Malik, G. Gupta, and Y. Jung. Managing departure aircraft release for efficient airport surface operations. In *AIAA Guidance, Navigation, and Control Conference*, 2010.

[62] I. Simaiakis, M. Sandberg, and H. Balakrishnan. Dynamic control of airport departures: algorithm development and field evaluation. *IEEE Transactions on Intelligent Transportation Systems*, 15(1):285–295, 2013.

[63] C. Verlhac and S. Manchon. Optimization of opening scheme. In *4th USA/Europe Air Traffic Management Research and Development Seminar*, 2001.

[64] M. Bloem, P. Kopardekar, and P. Gupta. Algorithms for combining airspace sectors. *Air Traffic Control Quarterly*, 17(3):245–268, 2009.

[65] D. Gianazza. Forecasting workload and airspace configuration with neural networks and tree search methods. *Artificial Intelligence*, 174(7):530–549, 2010.

[66] H. Trandac, P. Baptiste, and V. Duong. A constraint-programming formulation for dynamic airspace sectorization. In *AIAA/IEEE Digital Avionics Systems Conference*, 2002.

[67] M. Drew. Analysis of an optimal sector design method. In *27th Digital Avionics System Conference*, 2008.

[68] A. Basu, J. S. B. Mitchell, and G. Sabhnani. Geometric algorithms for optimal airspace design and air traffic controller workload balancing. In *10th Workshop on Algorithm Engineering and Experiments and 5th Workshop on Analytic Algorithmics and Combinatorics*, 2008.

[69] M. Xue. Airspace sector redesign based on voronoi diagrams. *Journal of Aerospace Computing*, 6(12):605–615, 2009.

[70] C. R. Brinton, K. Leiden, and J. Hinkey. Airspace sectorization by dynamic density. In *9th AIAA Aviation Technology, Integration and Operations Conference*, 2009.

[71] J. Li, T. Wang, M. Savai, and I. Hwang. Graph-based algorithm for dynamic airspace configuration. *Journal of Guidance, Control, and Dynamics*, 33(4):1082–1094, 2010.

[72] G. R. Sabhnani, A. Yousefi, and J. S. B. Mitchell. Flow conforming operational airspace sector design. In *10th AIAA Aviation Technology, Integration and Operations Conference*, 2010.

[73] M. Gariel, A. N. Srivastava, and E. Feron. Trajectory clustering and an application to airspace monitoring. *IEEE Transactions on Intelligent Transportation Systems*, 12(4):1511–1524, 2011.

[74] G. R. Sabhnani, A. Yousefi, I. Kostitsyna, J. S. B. Mitchell, and V. Polishchuk. Algorithmic traffic abstraction and its application to nextgen generic airspace. In *10th AIAA Aviation Technology, Integration, and Operations Conference*, 2010.

[75] C. A. Provan, L. Cook, and J. Cunningham. A probabilistic airport capacity model for improved ground delay program planning. In *AIAA/IEEE Digital Avionics Systems Conference*, 2011.

[76] D. M. Pfeil and H. Balakrishnan. Identification of robust terminal-area routes in convective weather. *Transportation Science*, 46(1):56–73, 2012.

[77] J. J. Rebollo and H. Balakrishnan. Characterization and prediction of air traffic delays. *Transportation Research Part C: Emerging Technologies*, 44:231–241, 2014.

[78] L. Li, M. Gariel, R. J. Hansman, and R. Palacios. Anomaly detection in onboard-recorded flight data using cluster analysis. In *30th Digital Avionics Systems Conference*, 2011.

[79] Y. S. Chati. *Statistical Modeling of Aircraft Engine Fuel Burn.* Ph.D. thesis, Massachusetts Institute of Technology, 2018.

[80] G. Antonini and J. P. Thiran. Counting pedestrians in video sequences using trajectory clustering. *IEEE Transactions on Circuits and Systems for Video Technology*, 16(8):1008–1020, 2016.

[81] Z. Fu, W. Hu, and T. Tan. Similarity based vehicle trajectory clustering and anomaly detection. In *12th IEEE International Conference on Image Processing*, 2005.

[82] S. J. Gaffney and P. Smyth. Trajectory clustering with mixtures of regression models. In *5th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1999.

[83] S. J. Gaffney, A. W. Robertson, P. Smyth, S. J. Camargo, and M. Ghil. Probabilistic clustering of extratropical cyclones using regression mixture models. *Climate Dynamics*, 29(4):423–440, 2007.

[84] J. G. Lee, J. Han, and K.Y. Whang. Trajectory clustering: A partition-and-group framework. In *Association for Computing Machinery's Special Interest Group on Management of Data Conference*, 2007.

[85] J. Kim and H. S. Mahmassani. Spatial and temporal characterization of travel patterns in a traffic network using vehicle trajectories. *Transportation Research Part C: Emerging Technologies*, 59:375–390, 2015.

[86] R. Wen, W. Yan, A.N. Zhang, N.Q. Chinh, and O. Akcan. Spatio-temporal route mining and visualization for busy waterways. In *IEEE International Conference on Systems, Man, and Cybernetics*, 2016.

[87] A. Eckstein. Automated flight track taxonomy for measuring benefits from performance based navigation. In *Integrated Communications, Navigation and Surveillance Conference*, 2009.

[88] F. Rehm. Clustering of flight tracks. In *AIAA Infotech@Aerospace Conferece*, 2010.

[89] M. Enriquez. Identifying temporally persistent flows in the terminal airspace via spectral clustering. In *10th USA/Europe Air Traffic Management Research and Development Seminar*, 2013.

[90] H. Arneson, A. Bombelli, A. S. Torne, and E. Tse. Analysis of convective weather impact on pre-departure routing of fights from fort worth center to new york center. In *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017.

[91] A. Bombelli, A. S. Torne, E. Trumbauer, and K. D. Mease. Automated route clustering for air traffic modeling. In *AIAA Modeling and Simulation Technologies Conference*, 2017.

[92] L. Song, C. Wanke, and D. P. Greenbaum. Predicting sector capacity for tfm decision support. In *6th AIAA Aviation Technology, Integration, and Operations Conference*, 2006.

[93] S. Sidiropoulos, K. Han, A. Majumdar, and W. Y. Ochieng. Robust identification of air traffic flow patterns in metroplex terminal areas under demand uncertainty. *Transportation Research Part C: Emerging Technologies*, 75:212–227, 2017.

[94] E. P. Gilbo. Airport capacity: Representation, estimation, optimization. *IEEE Transactions on Control Systems Technology*, 1(3):144–154, 1993.

[95] G. F. Newell. Airport capacity and delays. *Transportation Science*, 13(3):201–241, 1979.

[96] M. Ignaccolo. A simulation model for airport capacity and delay analysis. *Transportation Planning and Technology*, 26(2):135–170, 2003.

[97] P. B. Liu, M. Hansen, and A. Mukherjee. Scenario-based air traffic flow management: From theory to practice. *Transportation Research Part B: Methodological*, 42(7):685–702, 2008.

[98] G. Buxi and M. Hansen. Generating probabilistic capacity profiles from weather forecast: A design-of-experiment approach. In *9th USA/Europe Air Traffic Management Research and Development Seminar*, 2011.

[99] J. Cox and M. J. Kochenderfery. Probabilistic airport acceptance rate prediction. In *AIAA Modeling and Simulation Technologies Conference*, 2016.

[100] J. C. Jones, R. DeLaura, M. Pawlak, S. Troxel, and N. Underhill. Predicting and quantifying risk in airport capacity profile selection for air traffic management. In *14th USA/Europe Air Traffic Management Research and Development Seminar*, 2017.

[101] J. C. Jones and R. DeLaura. Predicting airport capacity in the presence of winds. In *17th AIAA Aviation Technology, Integration, and Operations Conference*, 2017.

[102] S. Houston and D. Murphy. Predicting runway configurations at airports. In *Annual Meeting of Transportation Research Board*, 2012.

[103] V. Ramanujam and H. Balakrishnan. Data-driven modeling of the airport configuration selection process. *IEEE Transactions on Human-Machine Systems*, 45(4):490–499, 2015.

[104] J. Avery and H. Balakrishnan. Predicting airport runway configuration: A discrete-choice modeling approach. In *11th USA/Europe Air Traffic Management Research and Development Seminar*, 2015.

[105] J. Avery. *Data-driven modeling of the airport runway configuration selection process using maximum likelihood discrete-choice models*. S.M. thesis, Massachusetts Institute of Technology, 2016.

[106] M. Vlachos, G. Kollios, and D. Gunopulos. Discovering similar multidimensional trajectories. In *18th International Conference on Data Engineering*, 2002.

[107] H. Zhu, J. Luo, H. Yin, X. Zhou, J. Z. Huang, and F. B. Zhan. Mining trajectory corridors using fréchet distance and meshing grids. In *14th Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 2010.

[108] A. K. Jain, M. N. Murty, and P. J. Flynn. Data clustering: a review. *ACM Computing Surveys*, 31:264–323, 1999.

[109] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31:651–666, 2010.

[110] M. Ester, H. P. Kriegel, J. Sander, and X. Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *2nd International Conference on Knowledge Discovery and Data Mining*, 1996.

[111] O. Arbelaitz, I. Gurrutxaga, J. Muguerza, J. M. Pérez, and I. Perona. An extensive comparative study of cluster validity indices. *Pattern Recognition*, 46:243–256, 2013.

[112] P. J. Rousseeuw. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53–65, 1987.

[113] E. Alpaydin. *Introduction to Machine Learning*. MIT Press, 2nd edition, 2010.

[114] L. Breiman. Random forests. *Machine Learning*, 45:5–32, 2001.

[115] V. Chandola, A. Banerjee, and V. Kumar. Anomaly detection: A survey. *ACM Computing Surveys*, 41(3):1–58, 2009.

[116] G. Shaffer and V. Vovk. A tutorial on conformal prediction. *Journal of Machine Learning Research*, 9:371–421, 2008.

[117] D. Devetyarov and I. Nouretdinov. Prediction with confidence based on a random forest classifier. In *6th IFIP WG 12.5 International Conference*, 2010.

[118] S. Bhattacharyya. Confidence in predictions from random tree ensembles. *Knowledge Information Systems*, 35:391–410, 2013.

[119] U. Johansson, H. Boström, T. Löfström, and H. Linusson. Regression conformal prediction with random forests. *Machine Learning*, 97:155–176, 2014.

[120] R. Laxhammar and G. Falkman. Conformal prediction for distribution-independent anomaly detection in streaming vessel data. In *1st International Workshop on Novel Data Stream Pattern Mining Techniques, ACM*, 2010.

[121] Port Authority of New York and New Jersey. Airport Traffic Report. `https://www.panynj.gov/airports/pdf-traffic/ATR2016.pdf`, (accessed 2018-01).

[122] Hong Kong International Airport. Air Traffic Statistics. `http://www.hongkongairport.com/eng/pdf/business/statistics/2016e.pdf`, (accessed 2018-01).

[123] Civil Aviation Resource Network. Transportation Statistics. http://i.carnoc.com/detail/385649, (accessed 2018-01).

[124] Macau International Airport. Passenger Figure and Movement Statistics. http://www.macau-airport.com/en/media-centre/facts-figures/statistics-passengers, (accessed 2018-01).

[125] GRU Airport. Operational Information. https://www.gru.com.br/en/institutional/sobre-gru-airport/operational-information, (accessed 2018-01).

[126] Aeroportos Brasil Viracopos. Statistics and Publications. http://www.viracopos.com/institucional/estatisticas-e-publicacoes/, (accessed 2018-01).

[127] Infraero Aeroportos. Statistics. http://www4.infraero.gov.br/acesso-a-informacao/institucional/estatisticas/, (accessed 2018-01).

[128] Airports Council International. Annual Traffic Data. http://www.aci.aero/Data-Centre/Annual-Traffic-Data, (accessed 2017-11).

[129] Federal Aviation Administration. Faa JO Order 7210.3Z, Facility Operation and Administration. http://www.faa.gov/documentLibrary/media/Order/7210.3Z_FAC.pdf, (accessed 2018-01).

[130] P. Glaab, R. Tamburro, and P. Lee. Analysis of the capacity potential of current day and novel configurations for new york's john f. kennedy airport. In *16th AIAA Aviation Technology, Integration, and Operations Conference*, 2016.

[131] Civil Aviation Department of the Hong Kong Special Administrative Region. Air Traffic Control Facilities. https://www.cad.gov.hk/english/cad_division.html, (accessed 2018-01).

[132] FlightRadar24. Live Air Traffic. https://www.flightradar24.com/, (accessed 2018-01).

[133] International Civil Aviation Organization. *Doc9854-AN/458: Global Air Traffic Management Operational Concept*. ICAO, 1st edition, 2005.

[134] R. Koenker and K. Hallock. Quantile regression. *Journal of Economic Perspectives*, 15:143–156, 2001.

[135] Federal Aviation Administration. JFK, EWR and LGA traffic management tips. https://www.fly.faa.gov, (accessed 2017-06).

[136] Civil Air Navigation Services Organization. Recommended Key Performance Indicators for Measuring ANSP Operational Performance. https://www.canso.org/recommended-key-performance-indicators-measuring-ansp-operational-performance, (accessed 2018-01).

[137] Federal Aviation Administration. 14 CFR Part 1 - Definitions and Abbreviations. https://www.ecfr.gov/cgi-bin/retrieveECFR?n=14y1.0.1.1.1, (accessed 2018-04).

[138] M. Robinson, R. DeLaura, and N. Underhill. The Route Availability Planning Tool (RAPT): evaluation of departure management decision support in New York during the 2008 convective weather season. In *8th USA/Europe Air Traffic Management Research and Development Seminar*, 2009.

[139] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, 1st edition, 1995.

[140] J. Platt. Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods. *Advances in large margin classifiers*, 10(3):61–74, 1999.

[141] T.-F. Wu, C.-J. Lin, and R. C. Weng. Probability estimates for multi-class classification by pairwise coupling. *Journal of Machine Learning Research*, 5:975–1005, 2004.

[142] N. Meinshausen. Quantile regression forests. *Journal of Machine Learning Research*, 7:983–999, 2006.

[143] C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. MIT Press, 1st edition, 2006.

[144] J. Vanhatalo, J. Riihimäki, J. Hartikainen, P. Jylänki, V. Tolvanen, and A. Vehtari. Gpstuff: Bayesian modeling with gaussian processes. *Journal of Machine Learning Research*, 14:1175–1179, 2013.

[145] International Civil Aviation Organization. Doc 9750-AN/963: 2016-2030 global Air Navigation Plan. https://www.icao.int/airnavigation/Documents/GANP-2016-interactive.pdf, (accessed 2018-04).

[146] J. Ferguson, A. Q. Kara, K. Hoffman, and L. Sherry. Estimating domestic US airline cost of delay based on European model. *Transportation Research Part C: Emerging Technologies*, 33:311–323, 2013.

[147] Eurocontrol. Standard inputs for Eurocontrol cost benefit analyses. https://www.eurocontrol.int/sites/default/files/publication/files/standard-input-for-eurocontrol-cost-benefit-analyses-2015.pdf, (accessed 2017-12).

[148] M. P. Matthews, M. S. Veillette, J. C. Venuti, R. A. DeLaura, and J. K. Kuchar. Heterogeneous convective weather forecast translation into airspace permeability with prediction intervals. *Journal of Air Transportation*, 14(2):41–54, 2016.