

Analysis and Optimization of Terminal Area Air Traffic Control Operations

by

C.S. Venkatakrisnan

S.B., Management, M.I.T. (1986)
S.M., Operations Research, M.I.T. (1989)

Submitted to the Sloan School of Management
in Partial Fulfillment of
the Requirements for the Degree of
Doctor of Philosophy in Operations Research

at the

Massachusetts Institute of Technology

February, 1991

© C.S.Venkatakrisnan 1991

The author hereby grants to MIT permission to reproduce and to
distribute copies of this thesis document in whole or in part.

Signature of Author _____
Sloan School of Management
January 11, 1991

Certified by _____
Arnold Barnett
Professor of Operations Research and Management
Thesis Supervisor

Certified by _____
Amedeo Odoni
Professor of Aeronautics and Astronautics and of Civil Engineering
Thesis Supervisor

Accepted by _____
Amedeo Odoni
Co-Director, Operations Research Center

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JAN 11 1991

LIBRARIES

Analysis and Optimization of Terminal Area Air Traffic Control Operations

by

C.S. Venkatakrishnan

Submitted to the Sloan School of Management

on January 11, 1991

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Operations Research

Abstract

This thesis examines two issues relating to terminal area air traffic control operations: the time intervals between aircraft at landing, and the sequencing of aircraft in the near terminal area. The study is motivated by the goals of advanced terminal area air traffic control, which include tighter spacing of aircraft for landing and better sequencing for them.

The first part of the thesis develops a model for the Landing Time Intervals (LTIs) between aircraft, determining what factors affect them and quantifying these effects. The model is based on a large collection of data on air traffic flow in the terminal area for Boston. Two factors that affect the mean LTI significantly are identified. These are the runway configuration and the weight classes of the lead and trail aircraft comprising the LTI. A model for the LTIs as a function of these two variables is developed on 10 data sets taken in heavy traffic under a variety of operating conditions. This model is then validated on a variety of statistics (mean, standard deviation and 25th-percentile) on completely new data. The validation results do not lead to a rejection of the calibration model on any statistic.

The second part of the thesis studies the aircraft sequencing problem. The LTI model is used to provide constraints for the spacing between aircraft. Three successively restrictive models of the operating environment and constraints to sequencing are developed. The first model is the static, which assumes complete knowledge of all entry times into the terminal area, a lower bound on arrival times at the runway, and that aircraft can be held in stacks (hence no upper bound). The second model is the dynamic with fixed time windows, which assumes that the existence of an aircraft is known only upon its entry into the terminal area. Further, that aircraft can be delayed only within the terminal area by path length variation and speed control. This delay can be used at any time, and provides a fixed upper bound to the landing time. The third model is also dynamic but with shrinking time windows: it incorporates the fact that, as an aircraft nears the runway, lost opportunities for delay cannot be regained.

The problems are solved on each of six data sets using a Dynamic Programming algorithm for a vehicle routing problem with time windows. In the static problem, the optimal sequence is computed for all aircraft at once. The dynamic algorithms re-solve, upon the entry of each new aircraft, for the optimal sequence for those aircraft in the terminal area. There is no look-ahead, hence the algorithms for the dynamic problem are heuristics. The three models are solved with three types of LTI constraints and three types of objective functions. For all models, and for all types of objectives and LTI constraints, there are indications of some gains to be had from better sequencing. The number of overtakes, especially within stream ones, is also relatively small.

Thesis Jointly Supervised by:

Arnold Barnett	Professor of Operations Research
Amedeo Odoni	Professor of Civil Engineering and Aeronautics and Astronautics

Acknowledgements

According to Indian thought, the teacher is central to the success of any learning endeavor. I am indeed fortunate to have had two of the very best Gurus, Prof. Arnold Barnett and Prof. Amedeo Odoni, as thesis advisors. Their intellectual guidance has moulded this thesis, and their personal concern and encouragement have seen it through to completion. In particular, I am thankful to Prof. Barnett for initiating me into the world of applied statistics and his impressively methodical and systematic approach to problems. I am grateful to Prof. Odoni for imparting to me some of his deep knowledge of civil aviation, and his fine judgement and incisive insight into technical issues. I would also like to acknowledge the very helpful comments of Prof. Stephen Graves, the third member of my thesis committee, and to thank Prof. Robert Simpson for useful discussions on the Aircraft Sequencing Problem. Finally, I owe a deep personal debt of gratitude to Prof. Alvin Drake, teacher, guide, and friend. He has given me unstintingly of his advice and assistance, and I am thankful indeed for having known him and worked under him.

I would like to thank the members of Group 42, M.I.T. Lincoln Laboratory, for introducing me to the problems studied in this thesis, for providing radar data and much technical assistance, and for their financial support. In particular, I am grateful to Dr. Jerry Welch and Dr. Steven Eoswell, for their close interest in my work and their many valuable suggestions. I am also thankful to the M.I.T. Center for Transportation Studies and the United Parcel Service Foundation for awarding me a fellowship during 1989-1990.

I am grateful to my many colleagues and friends at the O.R. Center. In particular I would like to thank Christopher Athaide, humbly, for many stimulating discussions, technical and otherwise; Karla Ballman, for her friendship, and for patiently advising me on many problems, real and imagined; Arnout Eikeboom, especially for introducing me to the probabilistic theory of Blackjack; the Franco-Belgian mafia, present and past, for their companionship and excellent desserts; Ramasubramanian Gopalan, pour de nombreuses tasses de café et conversations; Murali Kodialam, for many enjoyable arguments; Tim Magee, for deep discussions on life and the state of O.R.; George Polychronopoulos, for many late night rides home; Kalyan Talluri, for our daily morning tête-à-têtes; Mostafa Terrab, for enjoyable and relaxed lunch-time discussions; and all my tennis and squash partners.

I would also like to thank some of my roommates over the years: Suresh Jagannathan for, among other things, a particularly effective lecture before my general exams; Sumanth Kaushik, for many years of close friendship, and for keeping my zest for fine food alive; and Narendra Subramanian for his immensely enjoyable company at all times.

My greatest debts are, however, to my first and foremost Gurus, my parents. Their deep affection, guidance and numerous sacrifices have enabled me to pursue my academic studies to their fullest. Along with my sister, Gowri, and brother-in-law, Sanjiv, they have seen me through my graduate career and Ph.D. thesis with caring, encouragement and support. I can never thank my family enough.

Dedication

For My Parents:

மாத்ரு தேவோ பவ, டித்ரு தேவோ பவ

Contents

1	Introduction	12
1.1	The Problem of Air Traffic Congestion	12
1.2	Possible Solutions for Air Traffic Congestion	13
1.2.1	En route Flow Management	13
1.2.2	Enhancing Capacity in the Near-Terminal Area	14
1.3	The Contributions of Our Thesis	16
1.4	Outline of Thesis	18
2	Air Traffic Control in the Terminal Area	20
2.1	En route Operations and the TRACON	21
2.2	Arrival Operations in the Boston Terminal Area	22
2.3	Operations on Final Approach	26
2.3.1	Flying Weather and Flying Rules	26
2.3.2	Navigation on Final Approach	30
2.4	Runway Operations	31
2.4.1	Parameters for Choosing Runway Configurations	31
2.4.2	Simultaneous Independent Landings	34
2.4.3	Landings on a Single Runway	36
2.4.4	Interaction between Arrival and Departure Traffic	36
2.5	Traffic Management Procedures	37
3	A Model for Landing Time Intervals	41
3.1	Overview	41
3.1.1	The Statistical Problem	41
3.1.2	Contents of Chapter	42
3.2	Data for Calibrating the Model	43
3.2.1	Time Periods when Data was Collected	43
3.2.2	Collection of Data by Radar	45
3.2.3	Examples of Data Sets	48
3.3	Pre-processing of LTI Data	50
3.3.1	Adjusting Radar LTIs	50
3.3.2	Defining Locally Heavy Traffic	51
3.3.3	Factors Not Included in the LTI Model	55

3.4	Calibrating the Model for LTIs	58
3.4.1	Preliminaries	58
3.4.2	Overall Summary of Data	59
3.4.3	Is the Mean LTI Independent of Runway Configuration	62
3.4.4	The Sequence Category and the Mean LTI	66
3.4.5	Does the LTI Variance Depend on Runway Configuration or Sequence Category?	69
3.4.6	Homogeneous Groupings of Mean LTIs—Preliminaries and Data Review	72
3.4.7	Grouping Sequence Categories with Homogeneous Means	75
3.4.8	The Effects of Runway Configuration on Mean LTIs	82
3.4.9	Implication of Results for Throughput	84
3.4.10	The Calibration Model for LTIs	85
3.5	Validating the LTI Model	89
3.5.1	Summary of Validation Data	89
3.5.2	Comparing Statistical Estimates from Calibration and Validation	92
3.5.3	Aggregating the Results: A Bootstrap Analysis	96
3.6	Summary and Discussion	100
4	The Aircraft Sequencing Problem	104
4.1	Preliminaries	104
4.2	Literature Review	107
4.3	The Static Model of the ASP	109
4.3.1	Objectives for the ASP	109
4.3.2	The LTI Constraints	112
4.3.3	A Model for the Static ASP	113
4.3.4	Algorithm to Solve the Static Model	115
4.4	DASP-1: A Dynamic Model with Fixed Time Windows	119
4.4.1	The DP Algorithm to Solve DASP-1	120
4.5	The Dynamic Model with Shrinking Time Windows	123
4.5.1	The Algorithm for DASP-2	124
5	Application of Sequencing Algorithms	126
5.1	Framework for Analysis	127
5.2	Sequencing Using the Static Model	131
5.2.1	The Mean Calibration LTIs as Constraints	131
5.2.2	The Effects of Varying the LTI Constraint	137
5.2.3	The Effects of Varying the Objective Function	140
5.3	Sequencing Using the Dynamic Algorithm with Fixed Time Windows	144
5.4	The Dynamic Algorithm with Shrinking Time Windows	149
5.4.1	Differential Delays to Northern and Southern Aircraft	151
5.4.2	Varying the Objective Functions and Constraints	153
5.5	Implementation of Sequencing	154
5.5.1	Workload and Overtakes	155

6 Conclusion	159
6.1 The Model for Landing Time Intervals	159
6.1.1 Review of Results	159
6.1.2 Applications and Future Work	162
6.2 The Aircraft Sequencing Problem	163
6.2.1 Review of Results	163
6.2.2 Future Work	166
6.3 Final Remarks	167
A Summary of Data Sets Used for Calibration	168
B Descriptions of Statistical Tests	174
B.1 Difference of Means Test	174
B.2 The C-test	175
B.3 Wilcoxon Rank-Sum with Blocked Treatments	175
B.4 A Rank-based Test for Equality in 25th Percentiles	176
B.5 The Bootstrap Method	180
C Calculating Times-to-fly from Fix to Runway	182
C.1 Preliminaries	182
C.2 The TTF From SCUPP to 4R	186
C.3 The TTF from Providence to 4R	191
C.4 Estimating the Other TTFs	193
C.4.1 The TTF for Jets from BRONC	193
C.4.2 The TTFs for Props from the Northern Fixes: EXALT, KHRIS, LOBBY	193
C.4.3 TTFs For Props From the Southern Fixes: FREDO and WOONS	195
D Computing Shrinking Time Windows	198
D.1 Introduction	198
D.2 Shrinking Windows for Aircraft from the Northern Fixes	200
D.3 Shrinking Windows for Aircraft from the Southern Fixes	203
D.4 Modifying the "Nearest-Neighbor" Technique	205

List of Tables

2.1	Categories of Flying Weather	28
2.2	FAA Separation Standards (in nm)	29
2.3	Standard Runway Configurations in Boston	33
2.4	AAR's By Runway Configuration for Boston	38
3.1	Number of Landings on 2-17-89 by Runway and Time	49
3.2	Number of Aircraft Landing on 33L on 2-17-89 by Sequence Category	49
3.3	Number of Landings on 3-30-89 by Runway and Time	50
3.4	Mean and Standard Deviation of LTIs by Runway Configuration and Date	60
3.5	Average Hourly Landing Rate on Runways for Each Data Set	64
3.6	Mean and Standard Deviation of LTIs by Runway Configuration Grouping	65
3.7	FAA Separation Standards (in nm) for IFR Weather	66
3.8	Mean and Standard Deviations of LTIs by Sequence Category	67
3.9	Mean and Standard Deviation of LTIs by Sequence Category Grouping	68
3.10	Mean LTI by Sequence Category and Data Set	73
3.11	Number of LTIs by Sequence Category and Data Set	74
3.12	Mean Base and Big LTIs with Normalized Differences	76
3.13	Mean Group A and B Big LTIs with Normalized Differences	79
3.14	Lj/S and Lp/S LTIs by Data Set	81
3.15	Results of Difference of Means Tests for LTI Types: Single and Multiple Runways in Use	83
3.16	Parameter Estimates for Calibration Model	87
3.17	Calibration and Validation LTIs by Sequence Category/Runway Configu- ration	91
3.18	Means LTIs From Calibration and Validation Data	93
3.19	Standard Deviations of LTIs From Calibration and Validation Data	94
3.20	25th Percentiles of LTIs From Calibration and Validation Data	95
3.21	Computations for P_i^M , P_i^D and P_i^R	98
5.1	Data Sets used for Sequencing Study	128
5.2	Formulary LTIs (in seconds)	129
5.3	Calibration LTIs (in seconds)	130
5.4	Cumulative Delay (minutes) Results for the Static Algorithm	132
5.5	Mean Position Shift Between Actual and Assigned Static Sequences	133

5.6	Delay of Actual Sequence with Mean LTIs	134
5.7	Standard Deviation in Delay for the Static Algorithm	136
5.8	Cumulative Delay Reduction (mins.) per Aircraft for Static Algorithm . . .	137
5.9	Optimal Cumulative Delays with Different LTI Constraints to the Static Problem	138
5.10	Example of Effect of Different LTI Constraints on Optimal Sequence	139
5.11	Comparison of Optimal Static Sequences with Different Objective Functions	141
5.12	Average Delay by Objective Function and Weight Class	142
5.13	Comparison of Minimum Cumulative (mins.) From DASP-1 and Static Sequences	145
5.14	Average Delay by Weight Class for Static and DASP-1	147
5.15	Performance of Static and DASP-1 Algorithms for Dollar and Passenger Delay Objective Functions	148
5.16	Comparison of Minimum Cumulative Delay (mins.) for All Algorithms . . .	150
5.17	Average Delay for Aircraft from Northern and Southern Fixes	152
5.18	Comparison of Optimal DASP-2 Sequences with Different Objective Functions	154
5.19	Overtaking with Sequencing Algorithms	156
A.1	Number of Landings on 9-13-87 by Runway and Time	169
A.2	Number of Landings on 10-11-87 by Runway and Time	169
A.3	Number of Landings on 5-13-88 by Runway and Time	170
A.4	Number of Landings on 5-18-88 by Runway and Time	170
A.5	Number of Landings on 3-30-89 by Runway and Time	171
A.6	Number of Landings on 5-31-89 by Runway and Time	172
A.7	Number of Landings on 6-09-89 by Runway and Time	172
A.8	Number of Landings on 6-15-89 by Runway and Time	172
A.9	Number of Landings on 6-28-89 by Runway and Time	173
B.1	Number of Rejections of H_0 in 1000 Tests	180
C.1	Variations in TTF for the SCUPP Approach to 4R	189
C.2	Variations in TTF for Northern Prop Approaches to 4R	195
C.3	Nominal TTF and Bounds for Jet Arrivals	196
C.4	Nominal TTF and Bounds for Prop Arrivals	196

List of Figures

2.1	Boston TRACON and Entry Fixes	23
2.2	Jet Approach Routes to Runway 4R at Boston	24
2.3	Prop Approach Routes to Runway 4R at Boston	25
2.4	Final Stages of a Descent From the Long Side (Source: Sorensen [1990]) . .	27
2.5	Logan Airport and its Runways	32
3.1	Hourly Demand for Operations at Logan Airport	44
3.2	Histogram of LTIs on 4R on 10-11-87	54
3.3	Bootstrap Distribution P^*O	99
C.1	Jet Approach Routes to Runway 4R at Boston	183
C.2	Prop Approach Routes to Runway 4R at Boston	184
C.3	Flight Path from SCUPP to 4R	188
C.4	Flight Path from Providence to 4R	190

Chapter 1

Introduction

1.1 The Problem of Air Traffic Congestion

Flight delays induced by air traffic congestion are a source of great inconvenience to travellers. The root of the problem, quite simply, is that the air traffic system has had to handle increasingly higher levels of traffic, without major additions to the infrastructure. In the United States, the number of passenger enplanements on commercial air carriers has almost doubled in the last 15 years from 240 million in 1977 to over 440 million in 1988 [12]. During the same period, the number of scheduled departures each year from all American airports has increased from approximately 5 million to over 6.1 million. Yet, not one new major airport has been built since 1974, when the new Dallas-Fort Worth airport commenced operations.

In Europe, the situation is equally worrisome. The number of passenger enplanements there has been growing at 5-6%¹ since 1975 and is expected to double by the end of this decade.² Further, it is estimated³ that congestion in Europe is costing airlines and passengers some \$5 billion per annum, compared with \$2 billion for the U.S.

The greater congestion from increasing levels of passenger traffic has been compounded by the development in recent years of the hub-and-spoke system among airlines. In such

¹All figures cited are for non-Warsaw Pact countries. Traffic is likely to be higher (and congestion worse) than predicted, particularly in Germany, with the opening of Eastern Europe to business and pleasure travel.

²"Aviation Brief", *The Economist*, February 27, 1988.

³See *The Economist*, April 14, 1990

a network, airlines use a few airports as the focal points of their operations (hubs). Their schedules are designed so that a large number of flights from many cities (spokes) arrive at approximately the same time at the hub, then passengers and baggage are redistributed according to respective destinations, and the flights return along the spokes, again at roughly the same time. Thus, in the U.S., the 30 large hub airports currently handle about 70% of passenger enplanements [12], and passengers there have been experiencing mounting, well-publicized delays.

1.2 Possible Solutions for Air Traffic Congestion

The increasing severity of air traffic congestion, and fears of commercial aviation gridlock, have stimulated much thought on how best to tackle this problem.⁴ These discussions focus on two basic parts of the aviation system: operations at airports (and associated near-terminal areas) and en route air traffic flow management. These foci reflect the congestion points in the Air Traffic Control (ATC) system: the airports, and the en route centers which handle heavy traffic—essentially high density intersections in the sky.

1.2.1 En route Flow Management

The en route flow management problem has tactical as well as strategic aspects (see Odoni (1987)). In general, the tactical aspect concerns itself with control decisions with respect to aircraft already in the air. These include:

1. “Metering” of traffic—regulating the rates of aircraft flow through specific en route control centers (also called waypoints).
2. En route re-routing—modifying the flight path to bypass congested en route areas.
3. En route speed control—reducing the speed at which aircraft would normally travel the airways between waypoints.
4. High-altitude holding and path-stretching manoeuvres to avoid doing the same at greater expense at low altitudes.

⁴See, for example, ‘Aviation Experts Warn of Gridlock at U.S. Airports’, *The New York Times*, June 19, 1988.

In Europe, this aspect of the problem becomes more difficult because airspace is not under a single ATC authority. Rather, European airspace is a patchwork of 22 national systems, with 44 control centers. By contrast, the U.S. controls twice as much airspace in a unified system with 20 centers ⁵.

The strategic problem in flow management is to develop optimal ground holding strategies for aircraft at origin airports given uncertain capacities (due to weather, for example) at destination airports (or, less often, at waypoints). Ground holding is much cheaper than air holding, because fuel burn is reduced, and unnecessary congestion at the destination is eliminated. Thus the strategic problem is to schedule departures under stochastic demand, dynamically as instantiations of demand become known, and across an interacting network.⁶ The flow management problem is proposed by Odoni (1987) and mathematical models are discussed in detail by Andreatta and Romanin-Jacur (1987), and Terrab (1990).

1.2.2 Enhancing Capacity in the Near-Terminal Area

The flow management problem, however, has also to be viewed in the context of the critical bottleneck in the ATC network: capacity in the near-terminal area of airports. The near-terminal area is generally a region of radius 30-40 miles around an airport, and is the most congested portion of the system. Its landing capacity is determined by how efficiently incoming aircraft can traverse the congested airspace for landing, as well as the physical capacity of the airport to land them.

Suggestions for improving congestion here fall into four categories:

1. Reorganizing the existing schedule by encouraging, with reduced ticket prices, passengers to fly in off-peak periods, for instance, late at night.
2. Rationing the use of existing capacity either by increasing landing fees during the peak periods of traffic, or limiting the number of landing slots available.
3. Physically increasing the landing capacity at cities by expanding existing airports (adding runways) or by building new airports altogether.

⁵See "Chaos in the Air", from *The Financial Times* reproduced in *World Press Review*, January, 1990

⁶There are network interactions because, for example, delaying a Boston-Chicago flight now may also delay its continuation to Los Angeles.

4. Using existing airport capacity more efficiently by improving air traffic flow and control in the airspace of the near-terminal area.

Altering ticket prices is difficult since these are determined by the airlines themselves, who would change the schedule only if they were to perceive an underexploited market of passengers who prefer night travel to daytime delays. It is as yet unclear that such a market exists. Solutions that ration existing capacity suffer on two counts. First, there are questions about the legality of peak period pricing. Such proposals for Logan Airport in Boston were successfully challenged in court recently, by small aircraft owners, on the grounds that they were discriminatory. More importantly, however, rationing merely re-allocates existing capacity, albeit helping to move flights from peak periods to less congested hours. But rationing does not *per se* increase capacity during peak periods.

Expanding existing airports, or building new ones, would in principle be an effective way to alleviate congestion, but it is fraught with enormous financial, political and environmental concerns. For example, in Boston, even if money were allocated for constructing the one new runway for which there is room, any such move would likely be opposed by residents in communities adjacent to the airport worried about noise pollution. Although there are many expensive expansion programs underway in the world, only 3 new big airports are currently being built: in Denver, Munich and Osaka. All are replacements for existing facilities. The one at Denver is to be completed in the mid-1990's, after over a decade of controversy, at an estimated cost of \$1.7 billion [6].

An interesting variant of the above approach would have new airports built in uncongested parts of the country, such as the Midwest, to serve primarily as waypoints. This aims to counter the effects on traffic of the hub and spoke system. Since a small number of big airports are used as transit points for many customers of a single airline, some of the congestion at major airports is 'artificial'. It is not associated with origins and destinations there. Building waypoint airports is, however, opposed by airlines who say that they depend on traffic originating or destined to their hubs for an important fraction of their revenues. Moreover, they are not keen to have to reinvest in building infrastructure at new airports (for baggage handling and other services). In Denver, for instance, United

Airlines and Continental Airlines strongly resisted moving from the old airport to the new one for precisely this reason.

Given the problems associated with the first three approaches to improving airport capacity, much attention has been focussed on the fourth: efficient use of existing airport capacity by improving air traffic flow and control in the near-terminal area. The aim is to combine the latest in electronic tracking and guidance techniques with the methodological tools of Operations Research and systems analysis for better flow control. These measures may be implemented more quickly and cheaply than the other proposed solutions. An important such effort, the Terminal Air Traffic Control Automation (TATCA) project has been initiated by the Federal Aviation Administration (FAA) of the United States. This project is being carried out at Lincoln Laboratory, M.I.T.

1.3 The Contributions of Our Thesis

Current research on terminal air traffic automation focuses primarily on the control of arriving aircraft, where there appears to be much scope for reducing delays in the terminal area [32]. We study two forms of inefficiency that often result from the complexity of landing operations in the near terminal area, and which are of concern to planners [3]. These are:

1. Large gaps between aircraft in the arrival stream; and
2. Suboptimal sequencing of aircraft in the terminal area for landing.

The first is caused by imprecise control and/or poor planning, and the second by the computational and physical complexity of the problem. Our thesis focuses on these two problems. In this section, we motivate and discuss them in general terms.

First we consider the question of large gaps between aircraft in the arrival stream. The landing process involves merging disparate streams of aircraft and funneling them to one or two runways for landing. The time intervals between successive aircraft landing on a single runway (the Landing Time Intervals or LTIs) are a proxy for the throughput of the

runway. For a given airport runway, if \overline{LTI} is the mean LTI between any two aircraft,

$$\frac{1}{\overline{LTI}} \simeq \text{Average arrival rate at runway.} \quad (1.1)$$

Thus the higher the LTIs, on average the less efficient the terminal area operations.

At the same time, however, there is a lower bound on the LTIs from safety considerations. The *distance* between successive aircraft has to be above certain minima specified by the FAA. These minima, which we shall discuss in detail later, are based on the weight-classes⁷ of the aircraft involved.

A primary aim of an automated air traffic control system is to assist a controller with the task of maintaining separations and devising better landing sequences. Currently, controllers do not have accurate means of judging the actual distance separations between aircraft and sequencing is done manually. TATCA plans for a more advanced method known as Dynamic Time-Based Planning. Suppose that we knew what the Desired LTI (DLTI) was between any given pair of aircraft: a time separation that achieves minimum separations perfectly. Further, suppose that there is a group of N aircraft in the terminal area. The Dynamic Time Based Planning method, given the desired time intervals (DLTIs) between any pair of aircraft at the runway, and the current positions of all the aircraft in the terminal area, would compute a landing sequence for the N aircraft and advise the controller on the speed and directional headings to give each aircraft all the way to landing. These instructions would also ensure that FAA distance separations requirements are not violated anywhere along the approach. Maintaining precise control would be aided by accurate radar constantly monitoring the progress of aircraft and a computer calculating the separations.

Knowing what LTIs currently are in practice (both in mean and distribution) tells us what the collective wisdom and experience of air traffic control regards as acceptable. Even though ATC today may maintain separations imprecisely, the actual separations reflect LTIs that controllers and pilots feel comfortable with. Ignoring these values could result in suggested LTIs that the users of the system may feel uncomfortable working with. Thus it

⁷A categorization of aircraft in terms of their weights—a term to be discussed later.

would be useful to know what LTIs are like currently. One should also study how the LTIs vary under different operating conditions. Thus we would understand how controllers and pilots respond in practice to different runway configurations and weather conditions for example. Such an empirical model of LTIs is the first major aim of this thesis.

If an advanced air traffic control system were to "know" the desired time interval at the runway for two given aircraft (DLTI), the next question is how best to sequence aircraft in the near-terminal area for landing. The question arises because we discover that, for example, when a Boeing 747 lands after a Boeing 727, the LTI is about 95 seconds, but if the sequence were reversed, the LTI would increase by about 27 seconds. Hence if there are 5 aircraft to be sequenced, four of which are B727's and the other a B747, it would be best to land the B747 last. This example illustrates the potential advantages to be gained from optimal sequencing of aircraft. The problem is known as the Aircraft Sequencing Problem (ASP).

While there clearly are gains from better sequencing, it is not apparent that they can be easily achieved. The terminal area is, after all, a very dynamic environment, with new aircraft entering all the time. It is also very constricted: aircraft cannot be delayed at will, and sequences cannot be arbitrarily reordered. Thus, given these constraints, how much benefit might be actually accrue from better sequencing into the terminal area. This question is the second main focus of this thesis.

1.4 Outline of Thesis

We begin, in Chapter 2, with a survey of air traffic control, with emphasis on operations in the near terminal area. We study the layout of Boston airport and its runways; the operation of the airport and the runway configurations used in different weather conditions; and the typical paths aircraft follow upon entering the terminal area up to landing. In doing so, we try to provide a qualitative understanding of ATC in the near-terminal area. This knowledge is important because it motivates much of the ensuing work.

Chapter 3 details the development of a model for LTIs. We obtain data on the LTIs

under a variety of airport operating conditions. The data are for Logan Airport in Boston. This is one of the largest data collections of its kind ever undertaken. We study what factors significantly affect the LTIs, and quantify their effects. While we focus on the mean LTI in developing our model, we also study other aspects of the LTI distribution, such as the standard deviation, and the 25-th percentile (for reasons to be explained later). We calibrate our equations on 10 data sets, and then validate them on 8 completely different ones. We discuss the insights we get into terminal area operations, and also how the current distribution of LTIs provides insights into what the DLTIs might be.

Chapter 4 begins the study of the Aircraft Sequencing Problem (ASP). We motivate the problem and discuss its importance in the overall scheme of advanced ATC for the near terminal area. We review the literature on the subject and present algorithms (for the static and dynamic cases) for our version of the problem. We develop algorithms for the ASP that aims to find the landing sequence that minimizes delays in the terminal area, subject to constraints on aircraft delay and manoeuvrability that result from congested airspace.

In Chapter 5, we present the computational results of our study of the sequencing problem. We quantify the comparison between the actual sequence used by ATC in certain data sets with the sequence proposed by our algorithm. We see that there are gains from optimal sequencing, but that these diminish with the types of constraints we impose on how much aircraft can be resequenced and/or delayed in the terminal area.

Chapter 6 recapitulates the important conclusions of our investigation into the two main problems studied in this thesis. We also discuss extensions and areas for future research.

Chapter 2

Air Traffic Control in the Terminal Area

This thesis focuses on two specific aspects of Air Traffic Control (ATC) operations in the terminal area: Landing Time Intervals (LTIs) between successively landing aircraft on a given runway, and the benefits of optimal sequencing of aircraft for landing. Our work first requires, however, an understanding of the ATC process, especially those procedures used in the near terminal area. This is a rather difficult task not only because of the sheer enormity and complexity of the system, but the dearth of material that describes the system, or even parts of it, in a general and unified manner. Most written material is for specialists and thus tends to be very focused on operational minutiae. But we found two documents to be of particular help: Moore (1989) and Mundra (1989). The latter is an excellent description of terminal area operations in general, and the former a detailed study of operations at Boston.

Since we are interested in modelling and optimizing near terminal area operations at Boston, we focus our description in this chapter on such operations there. The description of operations at Boston is in general representative of those at other facilities. We also describe briefly en route operations and their interaction with terminal area ATC. These are not, strictly speaking, necessary to understand this thesis, but they provide a perspective for the role of terminal area operations in the larger control process.

2.1 En route Operations and the TRACON

There are two types of ATC facilities: terminal area and en route. Terminal area facilities include air traffic control towers (simply called towers) as well as Terminal Radar Approach CONTROL facilities (TRACONs). Towers control the movement of aircraft in the very final stages of landing (typically the last five miles), movement on the ground, and also the first minute or so after takeoff. Thus when an aircraft departs it is under the control of the tower until takeoff and then is handed over to the TRACON. Conversely, at the other end of the flight, the plane enters the terminal area under the control of the TRACON and lands guided by the tower.

The TRACON is a region of airspace typically extending up to 40 miles¹ around and 10,000 feet above the airport. Though they handle departures as well as arrivals, the TRACONs are usually referred to as approach control facilities. Busy TRACONs also contain an airspace called the Terminal Control Area (TCA). This is a subset of the TRACON and contains within itself the airspace controlled by the tower. At any time, only aircraft with specific equipment may enter the TCA. Once within the TCA, all aircraft are provided with "positive control" by the TRACON or tower, as the case may be.

Between its origin and destination TRACONs, an aircraft is controlled by one or more Air Route Traffic Control Centers (ARTCC, or simply *en route center*). These control the cruise phase of the flight.² A TRACON can serve more than one major airport: the New York TRACON located at Westbury, Long Island, serves John F. Kennedy, La Guardia, and Newark airports (thus towers). That the TRACON and the tower need not be located at the same place calls attention to the fact that control in the TRACON, as in the ARTCC, is done by radar. Tower operations, while they use radar, are primarily visual.

¹Unless explicitly stated otherwise, in this thesis we shall follow the convention of air traffic control in referring to all distances in nautical miles and all speeds in knots.

²We ignore the short flights which originate at a satellite airport, and land at a major one, staying entirely within one TRACON zone. Such a flight for Boston could, for example, be from Hanscom Field to Logan.

Traffic operations within ARTCC's are quite different from those within the TRACON. We are concerned mainly with ATC in the terminal area, and shall only discuss TRACON and tower operations in detail. In particular, we shall focus on operations in Boston,³ since we are studying data from there. There should, however, be strong similarities with operations at other facilities.

2.2 Arrival Operations in the Boston Terminal Area

TRACONs, as mentioned earlier, handle both arrivals and departures. Arriving traffic funnels in from higher speeds and altitudes for landing and departing traffic fans out to higher speeds and altitudes for the cruise phase. In busy TRACON areas, these two types of traffic are completely segregated by dedicating specific three-dimensional corridors of airspace for them. The particular corridors used depend on the runways in use. The focus of the TATCA project is on arriving traffic and we shall focus on ATC procedures for arrivals only.

A flight in the terminal area lasts about 10-20 minutes. Figure 2.1 shows the TRACON of Boston with entry fixes.

The entry fixes (or arrival fixes) are the points at which en route aircraft are handed over to the TRACON by the ARTCC.⁴ The Boston TRACON has a vertical airspace limit of 14,000 feet and can be approximately enclosed in a circle of radius 30 miles centered at Logan Airport. The TCA for Boston is one of the busiest in the nation.

The designated arrival fixes for jet aircraft into Boston are SCUPP, BRONC and Providence (PVD). These aircraft are typically at 11,000 feet and travelling at 250 knots at this handover point. Propeller aircraft (known as "props") enter at other fixes (LOBBY, KHRIS, EXALT, FREDO and WOONS) typically at a lower altitude and speed: 6,000 feet and 190-210 knots respectively. Figures 2.2 and 2.3 show arrival flight paths in the TRACON for landing at runway 4R, for jet and propeller aircraft respectively.⁵

³Unless otherwise stated, the tower and airport we refer to is Logan Airport.

⁴The ARTCC for this region is located at Nashua, New Hampshire.

⁵These arrival flight paths are also known as "approaches", the last segment of which, about five miles long, is termed the "final approach". We shall explain the layout of the airport and the naming convention

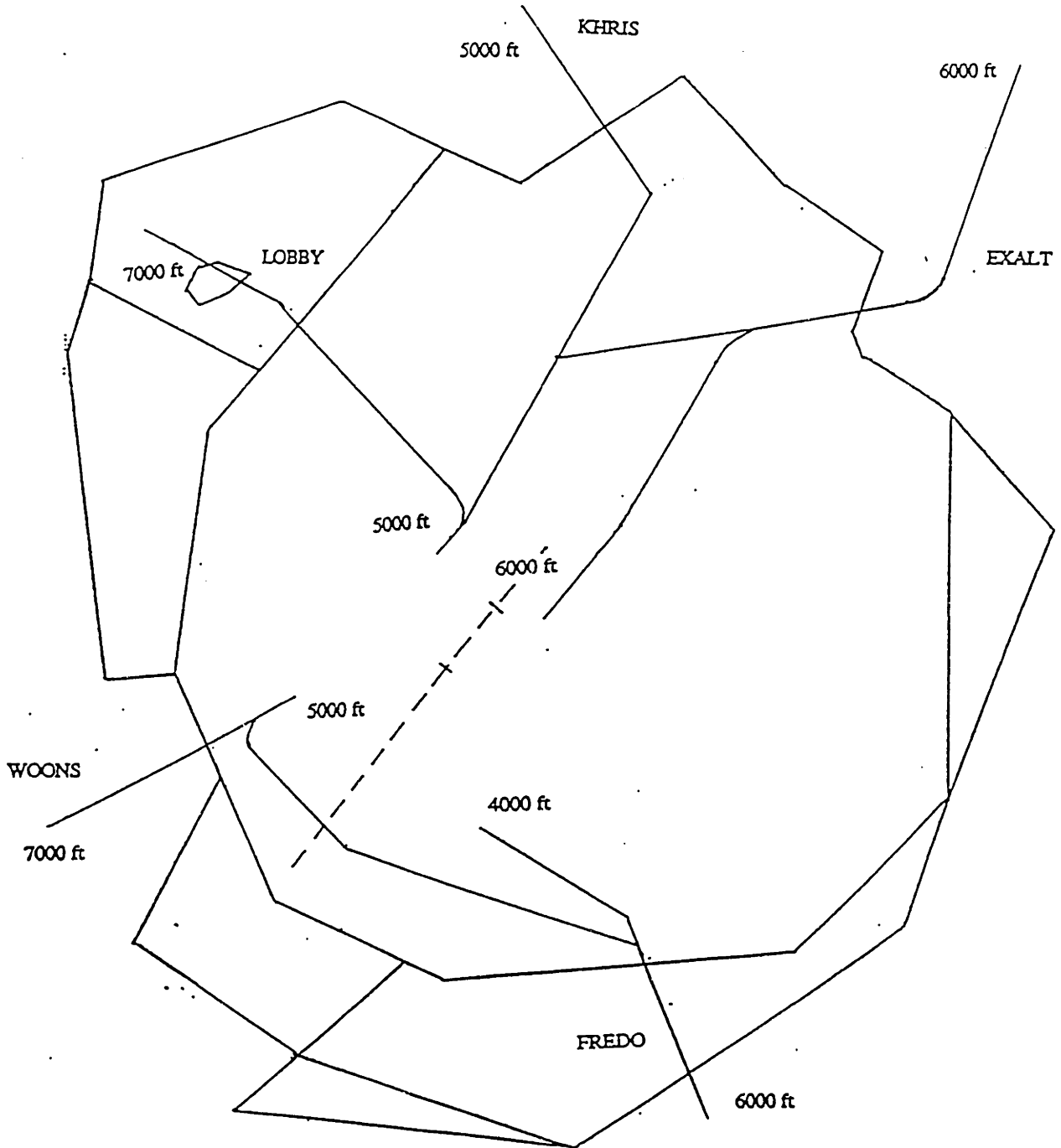


Figure 2.1: Boston TRACON and Entry Fixes

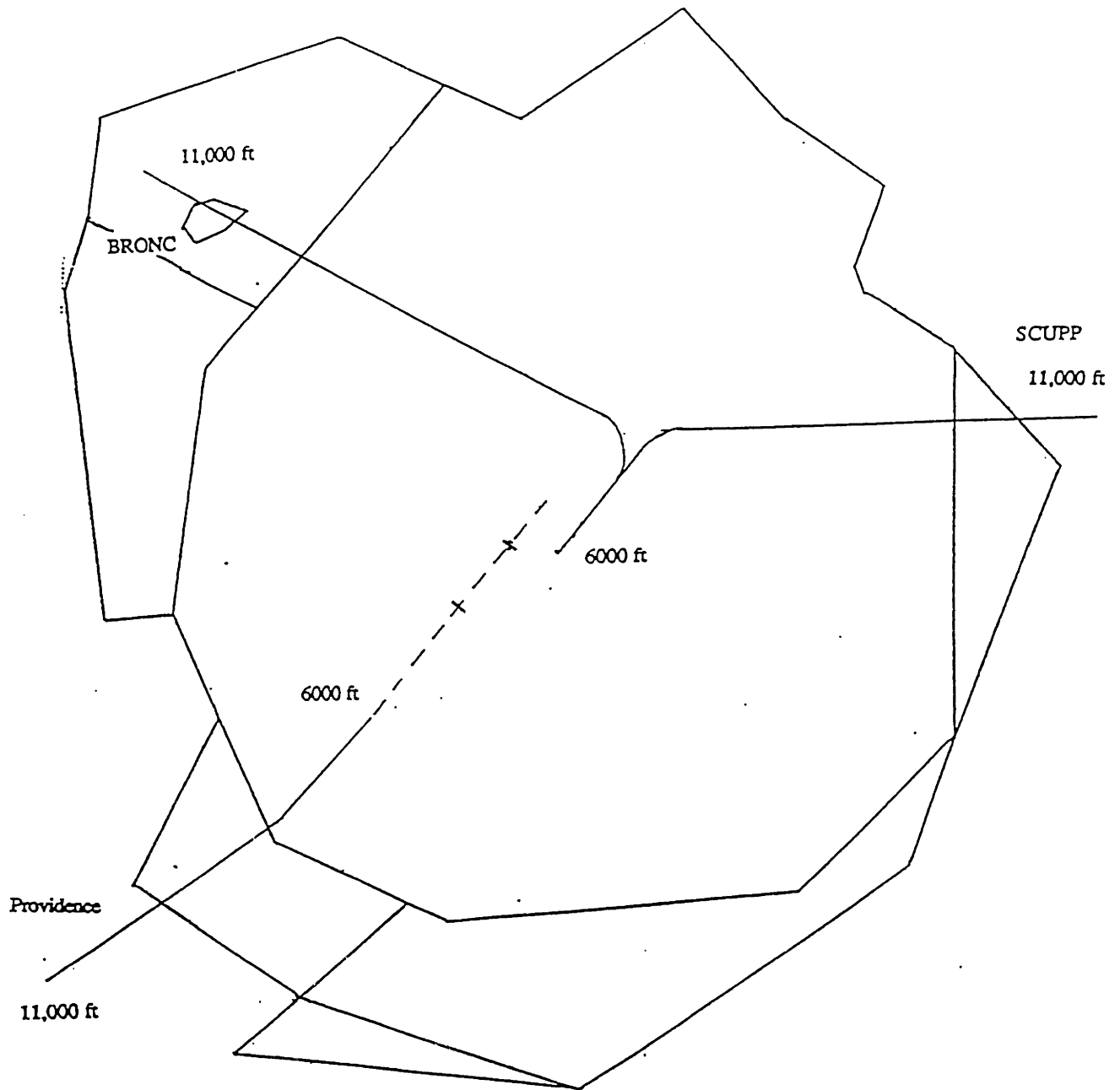


Figure 2.2: Jet Approach Routes to Runway 4R at Boston

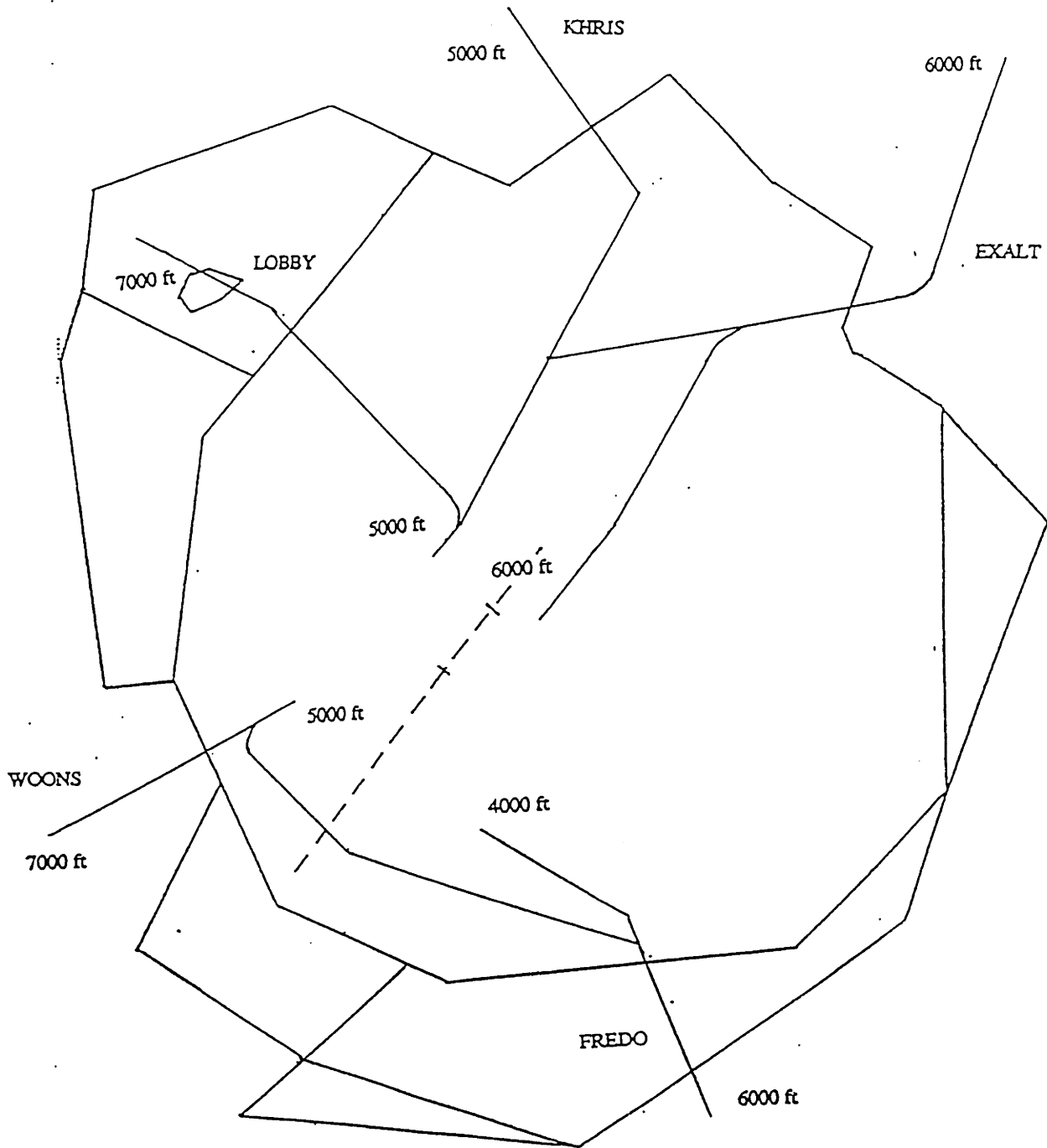


Figure 2.3: Prop Approach Routes to Runway 4R at Boston

The arrival airspace is composed of two subspaces: the pattern space and the sequencing space. The pattern space consists of the final stages of descent for an aircraft: the downwind leg, base leg, and final approach. Figure 2.4 magnifies this section of flight for landings from SCUPP to 4R. The downwind leg (and the subsequent turn to the base leg) exists only for the “long side”, *i.e.*, for approaches from those fixes farthest from the runway threshold. SCUPP and BRONC are on the long side for landings on 4R (see Figure 2.2). Also the downwind leg is travelled (as denoted by the name) along the direction of the wind. This is because all landings and takeoffs must be into the wind.

The sequencing space is the space other than that included in the pattern space. This is the section of the TRACON where aircraft can be delayed, by being made to slow down or to fly circuitous routes (path stretching), before they are merged into a single file for landing. These methods for delay, and what time margins they allow, are discussed in great detail in Chapter 4, where we study the Aircraft Sequencing Problem.

We have illustrated the typical flight paths in the TRACON, for jets and props when both are landing at runway 4R in Boston. But there are many possible approaches depending on the particular runways accepting landings. These approaches share the basic elements of the downwind and base legs and the final approach as described earlier. In the next section, we describe the rules governing operations on final approach for landing.

2.3 Operations on Final Approach

Operations on final approach are greatly dependent on flying weather. Thus we begin by discussing the categorization of flying weather and how it affects landing operations. Then we shall describe navigation on final approach.

2.3.1 Flying Weather and Flying Rules

Flying weather is categorized into two ranges: Visual Meteorological Conditions (VMC) and Instrument Meteorological Conditions (IMC). Corresponding to these conditions,

for runways later.

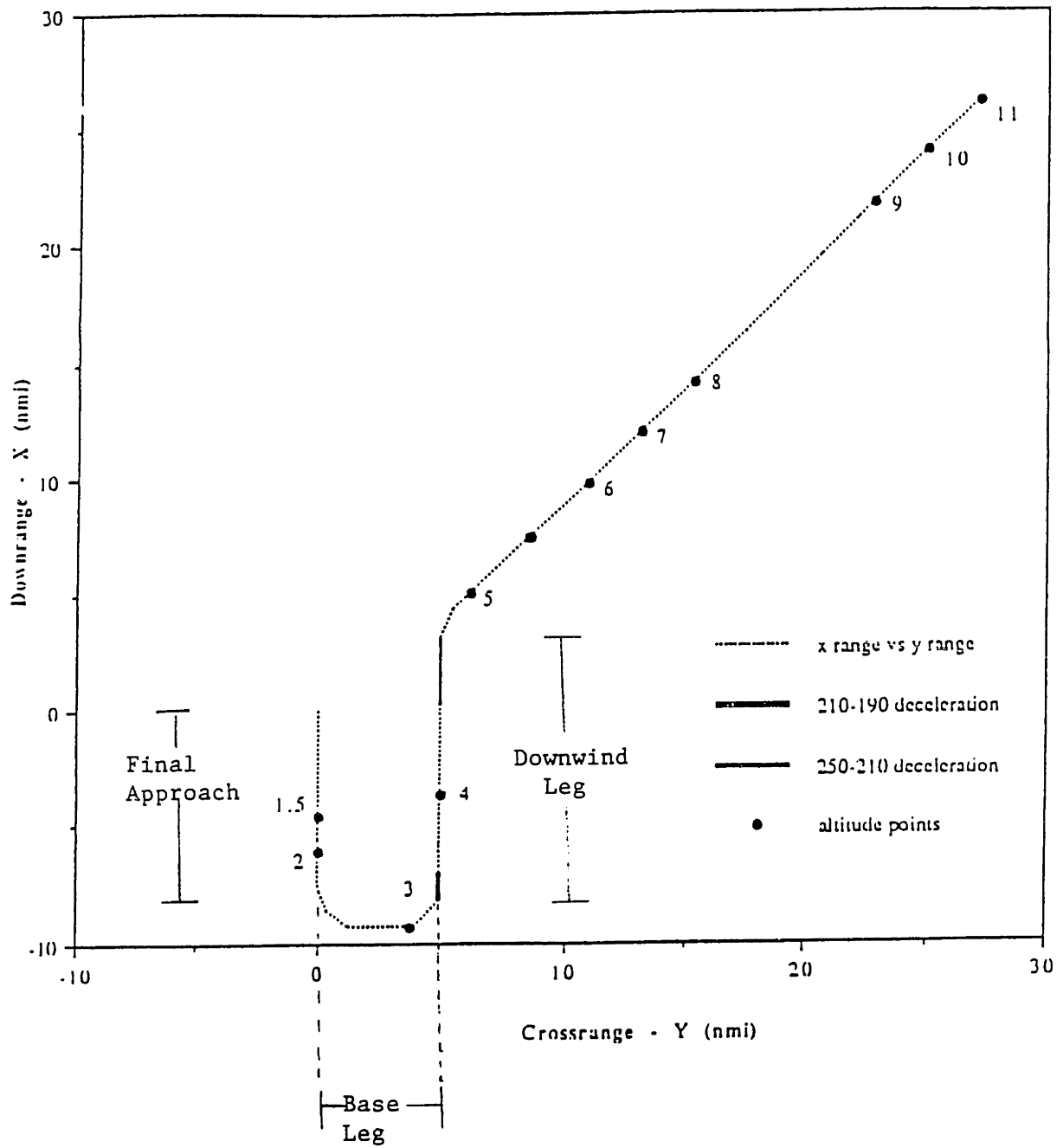


Figure 2.4: Final Stages of a Descent From the Long Side (Source: Sorensen [1990])

there are flight rules known as Visual Flight Rules (VFR) and Instrument Flight Rules. In general, VFR are used in VMC and IFR in IMC. The correspondence may, however, depend on the controllers, as we discuss shortly. These weather categories, and their subdivisions, are determined by the cloud ceiling and the visibility at ground level at the airport. The details are given in Table 2.1. Thus we see, for example, that VFR-1 requires a minimum visibility of 5 *statute* miles and/or 2,500 feet. Further, that VFR-2/IFR-1 describes the weather when the visibility is between 1 and 5 statute miles and/or the ceiling between 800 and 2,500 feet.

Weather Category	Minimum Visibility (statute miles)	Minimum Ceiling (feet)
VFR-1	5	2,500
VFR-2/IFR-1	1	800
IFR-2	0.5	200
IFR-3	0.25	100
IFR-4	0	0

Table 2.1: Categories of Flying Weather

The FAA has established separation standards between aircraft. These separation standards are to avoid two dangers: collision and hazards due to wake-vortex. To avoid collision danger, a minimum base separation of 2.5 nm is usually applied between any two aircraft. There is then an additional wake-vortex separation depending on the weights of the lead and trail aircraft. Wake-vortices represent turbulent flow of air, and the phenomenon has not yet been modelled precisely. It is known to depend on the winds, the temperature and, importantly, the weight of the aircraft. Just as a motor boat leaves a bigger wake behind it than a canoe, so too does a B747 leave a bigger wake than a Cessna. The FAA has defined incremental wake-vortex separations based on the weights of the lead and trail aircraft. The weight of the lead aircraft determines the strength of the wake, and the weight of the trail how much it might be affected by the wake.

The FAA has defined three classes of planes [29], in terms of their Maximum Gross Take-off Weight (MGTOW):

Weight-class of Trail Aircraft

		Heavy	Large	Small
Weight Class of Lead Aircraft	Heavy	4	5	6
	Large	2.5	2.5	4
	Small	2.5	2.5	2.5

Table 2.2: FAA Separation Standards (in nm)

Heavy - MGTOW \geq 300,000 lbs.

Large - 12,500 \leq MGTOW < 300,000 lbs.

Small - MGTOW < 12,500 lbs.

Heavy aircraft include the B747, B767, A300, DC10, DC8 and L1011. The Large aircraft comprise a wide range, from a small executive jet or a larger turboprop like the ATR42, to the B757 and the DC9. Small aircraft consist primarily of small piston engined aircraft and the smallest turboprops like a Beech 99. These are usually referred to as General Aviation Aircraft (GAA).⁶

Based on these three weight-classes, the FAA has established standards (given in Table 2.2) for the minimum separation (as measured by the radar distance) between successive aircraft in a flight path. Thus a B747 trailing a B727 is required to be just the base 2.5 nm behind the lead aircraft; if the sequence were reversed, the B727 would have to be an additional 2.5 miles behind for wake-vortex reasons. Thus an overall separation of 5 miles would be required.

These standards are mandatory in IMC; in VMC the tower may enforce them if it so wishes. When separation standards are in effect, aircraft are deemed to be flying under Instrument Flying Rules (IFR). In good VMC, however, the tower may offer aircraft the opportunity to maintain visual separation from each other. Should the pilots accept (and they usually do) the IFR standards cease to apply: the pilot is free to exercise his judgement in the matter. Planes flying under visual separation are said to be following

⁶GAA are defined as those flights which are not scheduled commercial flights or military ones.

Visual Flight Rules (VFR). Wake-vortex phenomena are not eliminated in VMC, but collision avoidance is in the hands of the pilots.

Visual separation may be provided whenever:

1. The controller sees both aircraft and visually maintains separation between them;
OR
2. One aircraft sees the other and maintains visual separation from it.

In addition, the controller must be in communication, or capable of being in instantaneous communication, with one of the aircraft in question.

2.3.2 Navigation on Final Approach

In IMC, aircraft land with the assistance of the Instrument Landing System (ILS). The ILS consists of two signals emitted from the ground: a localizer and a glide slope indicator. Each signal is received by a separate instrument in the cockpit. The localizer tells the pilot whether he is travelling in the correct direction to intercept the centerline of the runway. The glide slope indicator enables him to descend at the correct slope, typically 3° . Major carriers (i.e. airlines), as a matter of policy, use the ILS even in clear weather.

In addition to the ILS system, there are other navigational aids on the final approach. These are the "markers" or "landing fixes": radio beacons along the final approach path. The first such fix encountered by landing aircraft is the "outer marker". The outer marker is 5 miles from the runway threshold and along the extended centerline of the runway. Aircraft align themselves to the localizer and glide slope at this point. This is also typically the point at which the TRACON hands over control of the aircraft to the tower. The next fix is the "middle marker", at 2 miles from the runway and also along the extended centerline. This is typically the "decision height" (about 600-700 feet) at which aircraft decide whether to land or to do a "go-around" or a "missed-approach". In Boston, there is a third marker known as the "inner marker" at 0.5 miles from the runway. In addition to these navigational aids, there are of course the ground lights at the airport.

In the next section, we describe Logan Airport, its layout, and elements of surface and tower operations germane to our study.

2.4 Runway Operations

2.4.1 Parameters for Choosing Runway Configurations

Figure 2.5 shows a map of Logan airport and its runways. A runway is numbered by the nearest 10° of its orientation from magnetic north. The orientation of the runway is the direction along which an aircraft using it departs or lands. Thus an aircraft departing on runway 9 is heading approximately 90° , or due east. If it were departing in exactly the opposite direction, it would be taking off on runway 27. When there are parallel runways, the letter "L" or "R" is suffixed to the runway number according to whether the runway is to the left or right of the pilot on final approach. Thus a pilot landing on heading 40° will find 4L to the left of 4R. There are two pairs of parallel runways at Logan 4R/L (22L/R) and 33R/L (15L/R), and a single one, 27 (9), that intersects 4R and 33L. All runways except 33R (15L) are long enough to accept large jets for landing. In practice, though, the very biggest jets do not land on 4L or 22R, due to noise considerations. In Boston only 5 runways are equipped with ILS: 4R, 15R, 22L, 27 and 33L. Further, only 4R is equipped with additional equipment to enable landings in IFR-III conditions

A runway configuration is the combination of runways used at a given time for departures and landings. For example, at Boston, a standard configuration uses runway 9 for departures and 4R and 4L for landings and departures. Runway configurations are determined by wind, weather and the physical layout of the runways. All aircraft must land or takeoff into the wind. Winds of less than 3 knots, called "calms", are considered to have no effect on landings. In general, the FAA levies no specific requirement on the maximum crosswind and lateral components for runway selection. Accepting a runway is the pilot's responsibility. But, in practice, tail winds need to be less than 5 to 10 knots, and crosswinds less than 20 knots. The exact ranges depend on whether the runway surface is wet or dry: the wetter it is, the more difficult to brake. If winds exceed these limits, pilots are unlikely to land and ATC is unlikely to propose the configuration in question.

In good VMC, as long as runway centerlines of parallel runways are more than 1,200 feet apart, both may be used for landings. This is satisfied in Boston, where the runways

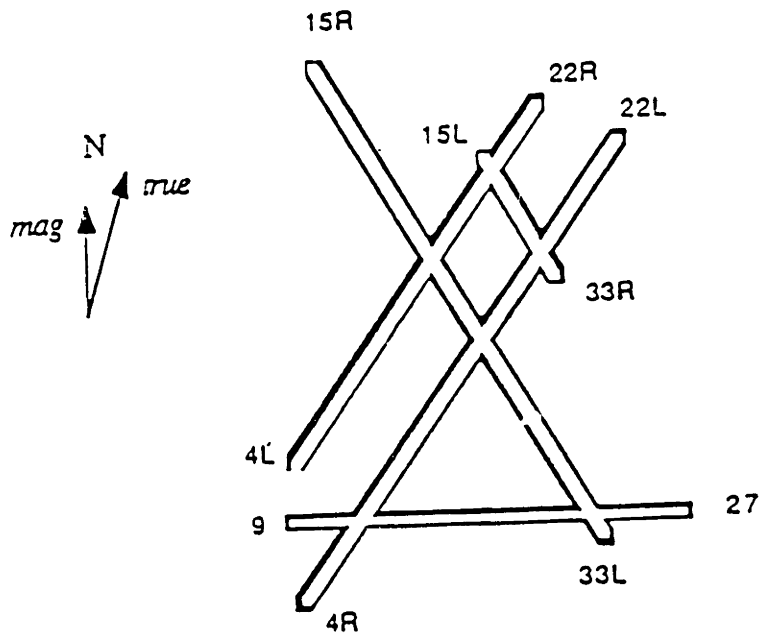


Figure 2.5: Logan Airport and its Runways

in each set of parallel runways are 1,800 feet apart. The restrictions on landings at intersecting runways is that the intersection point is at least 6,000 feet from the threshold. This is true for runways 22L and 27 in Boston. In IMC, simultaneous independent ILS approaches are allowed only if the centerlines of parallel runways are at least 4,300 feet apart. This condition is not satisfied at Boston. Intersecting runway operations in IMC are only conducted at a few airports, and then only with high ceilings. They are not performed at Boston. In marginal IFR/VFR conditions there are dependent parallel runway operations that we shall describe in detail later.

Weather	Arrival	Departure
IFR-1 or Better	4R*/L*	9
IFR-2	4R*	9
IFR-3	4R	9
IFR-1 or Better	22L*/R*	22R
IFR-2	22L*	22R
VFR-2 or Better	22L*/27	22R
VFR-1	33L*/33R	27
VFR-2/IFR-1	33L*	27
IFR-2	33L*	33L
VFR-1	15R*/15L	9
VFR-2/IFR-1	15R*	9
IFR-2	15R*	15R

Table 2.3: Standard Runway Configurations in Boston

Given all these restrictions, Table 2.3 gives the standard arrival and departure configurations for Boston. When two runways are used for landing, one is designated the primary and the other the secondary. The primary usually accepts most jet aircraft and the secondary all prop aircraft and occasional jets. Thus in Table 2.3, 4R/L denotes that 4R is the primary runway and 4L the secondary. 27/22L, where the secondary runway also regularly accepts jets, is an exception to this rule that we shall discuss later. An asterisk appended to the arrival runway denotes that it may also be used for takeoffs. If this arrival runway is not a main departure runway, such use is marginal, as we discuss.

In terms of the way in which arrival traffic is handled, we can divide the runway con-

figurations in Boston into three groups: two runways accepting simultaneous independent landings; two runways accepting dependent parallel landings; single runway accepting landings. These groups are not mutually exclusive: as we shall show, there can be simultaneous independent arrivals on 4R/L as well as dependent ones. We now describe how arrival traffic is handled in each of these groups of runway configurations.

2.4.2 Simultaneous Independent Landings

There are five landing configurations accommodating simultaneous independent landings on two runways: 4R/L, 22L/R, 15R/L, 33L/R and 27/22L. As mentioned earlier, this type of configuration obtains in VFR or marginal IFR only. When there are simultaneous independent landings, usually the jet and propeller traffic land on the primary and secondary runways respectively. They enter from different fixes and are merged into two landing streams at the outer markers of their respective runways. There are some exceptions to this general rule. When 15R/L or 33L/R are used for landings, the secondary runway (15L or 33R respectively) is so short (2,500 feet) that only the smallest of propeller aircraft land there. Thus the landing stream on the primary includes larger propeller aircraft. The other exception is the 27/22L combination.

According to air-traffic controllers, the operating "rules" that they follow for configuration 27/22L are as follows:

1. In general, land jets on 27 and props on 22L.
2. Try to land traffic from Providence and Cape Cod on 27 and the rest on 22L. This facilitates an easier approach for these aircraft.
3. If a jet can land on 22L (which is close to the terminal) and stop short of 27 (all but the biggest can do this), then it may do so.
4. A prop (from Cape Cod) may land on 27, if its doing so does not slow down a jet approaching 27, or divert to 22L.

These rules are stated roughly in order of the importance accorded them by air-traffic controllers. But given their conflicting objectives, and the flexibility for landings on 22L, the variation in relative loads between the primary and secondary runways can be quite

high. Further, different controllers may distribute the same traffic mix differently between these two runways.

Dependent Parallel Landings

Sometimes the flying weather worsens to the point that simultaneous independent landings are not possible, yet the weather is not so bad that only a single runway must be used. In these circumstances there is usually some form of dependent landings on two runways.

The most limited form of dependent landings is the "sidestep". The sidestep is used when all planes are making an ILS approach to one runway (e.g. 22L). At approximately the middle marker, propeller aircraft ask for, or are offered, the parallel runway (22R, in this case) if they confirm that the field is in sight. If clearance is given and accepted, the plane "sidesteps" or "transitions" to land on the parallel runway. This is typically the only form of dependent parallel landings in use with 22L/R, 15R/L and 33L/R. Also 27/22L is used only in good VFR conditions.

With 4R/L in use, however, there is a more involved form of dependent landings.⁷ This involves two separate converging ILS approaches operating at the same time. One approach is to runway 4R when the aircraft plan to land on 4R, and the other is to 15R where the aircraft plan to break out of the clouds in time to see the airport and circle to land on 4L. This manoeuvre requires a minimum cloud ceiling of 1,000 feet and a minimum visibility of 3 miles. All jets make the ILS approach to 4R. Props from the south also land on 4R, but the remaining prop traffic makes the ILS approach to 15R to circle and land on 4L. This approach is denoted as 4R/15Rc4L.

The sidestep is not so much designed to increase capacity as to make it convenient for planes to land closer to the terminal. This is particularly true if the primary runway is 4R or 22L. Since there is a single final approach path, the only capacity increase is by flying jets slightly closer to props on the final approach, in anticipation of the latter stepping to the side runway. The 4R/15Rc4L approach, however, uses the secondary runway in a planned way and thus does serve to increase capacity.

⁷See Crone and Moore (1990).

2.4.3 Landings on a Single Runway

Finally, when the weather deteriorates to a point where the ceiling is below 1,000 feet, or the visibility below 3 miles, only a single runway is used for landings. Simultaneous, independent ILS landings are not possible at Boston because the runways are less than 4,300 feet apart. Thus props and jets are merged into a single stream at the final approach fix and all land on the same runway. During such poor weather, however, some smaller prop planes that usually travel to or from New Hampshire and Maine, may not fly.

When weather or wind changes necessitate a change in runway configuration, the decision is coordinated between the tower and the TRACON. The TRACON supervisor must determine the exact point in time in the traffic stream where subsequent arrivals will land in the new configuration. This can involve detecting a natural gap in the landing traffic or creating one. Then the two separate streams must be handled differently by controllers.

2.4.4 Interaction between Arrival and Departure Traffic

The level of interaction between arrival and departure traffic depends on the particular configuration used. There is least interaction when one runway is completely dedicated to departing traffic. Examples are 9 (with arrivals on 4R/L), 27 (with arrivals on 33L/R) and 22R (with arrivals on 22L and 27). In such cases, especially with intersection runways, when an aircraft is close to landing, the departure awaits the landing and takes off instantly after touchdown.

The next higher level of interaction occurs when one of the arrival runways also hosts many departures. Examples are 22R (when 22L/R is in use), 4L (when 4R/L is in use) and 33L (when 33L/R is in use). The latter two landing/take-off configurations are relatively rare. When 22R accepts departures with 22L/R in use, there is a rule of thumb among controllers that an arrival is permitted for every three departures. The highest level of interaction is when the arrival and departure runway is the same. This happens only in the extreme case of 4R accepting all arrivals and departures.

2.5 Traffic Management Procedures

One important element has been missing from our description so far. While we have discussed how aircraft are controlled in the near terminal area, we have not discussed the broader question of traffic management. By traffic management, we mean three questions:

1. How does the TRACON decide on the number of aircraft to accept?
2. How is this decision coordinated with traffic managers at the en route center and further upstream?
3. How are aircraft delayed in the system if there is more traffic desiring to land at the TRACON than it can handle?

In this section we shall discuss these points.

At a national level, traffic is managed by the Central Flow Control Facility in Washington, D.C. The CFCF has the responsibility to limit congestion in the national airspace. Among the policies it follows is to dictate ground holds for aircraft because of bad weather or congestion at destination airports. The problem of assigning these delays optimally is the flow management problem that we discussed earlier.

The CFCF works closely with a network of Traffic Management Units (TMUs) in the ARTCCs. The TMU is responsible for inter-ARTCC traffic flows as well as flow from the ARTCCs to individual TRACONs. The parameter that the TRACON sets for the TMU is the Arrival Acceptance Rate (AAR). This is determined to be the number of arriving aircraft that can be served at the airport in each hour. The AAR is not a maximum, however, as we discuss shortly. The AARs for Boston were computed by the FAA in 1981. The overall airport AAR is apportioned by primary and secondary runways in use, based on the runway configuration in operation. The AAR is supposed to make allowance for normal operating changes like runway changes, missed approaches etc.

The AARs for runway configurations at Logan are given in Table 2.4. Whenever a primary landing runway supports only the occasional departure, it has an AAR of 34 planes/hour. Depending on the number of departures that it accommodates, the AAR drops to 30 (when 15R is used in this way) or 26 planes/hour (likewise for 33L). The

Arrival Runway(s)	Departure Runway	Airport Arrival Rate		
		Total	Primary	Secondary
4R/L	9	58	34	24
4R/15Rc4L	9	44	34	10
4R	9	34	34	-
22L/R	22R	44	34	10
22L	22R	34	34	-
27/22L	22R	60	34	26
33L/R	27	44	34	10
33L	27	34	34	-
33L	33L	26	26	-
15R/L	9	44	34	10
15R	9	30	30	-
Single Runway Arrive and Depart		26	26	-

Table 2.4: AAR's By Runway Configuration for Boston

AAR for a secondary runway is typically 10 if it supports departures. This is also the figure for 33R (15L), even though it does not support departures, because it is a short runway and accepts only very small aircraft. If the secondary runway is long and accepts no departures, the AAR goes up to 24 (for 4L when 9 is the main departure runway) or 26 (for 22L when 27 is the primary and 22R the main departure runway). We see thus that 4R/15Rc4L has an AAR of $34 + 10 = 44$, whereas 4R with sidesteps has an AAR of only 34.

The center cannot alter a given AAR, though the figure can be “negotiated” with the manager of the TRACON. The TMU monitors the TRACON traffic to make sure that it is neither starved nor overloaded. The way in which the TMU responds to the AAR depends on the actual demand for landing at the airport. There are three typical responses.

Free Flow: When there is relatively little demand (see definition for “metering”) for landings at the airport, aircraft are allowed to enter the TRACON at will.

Metering: If the demand exceeds the AAR for more than 20 minutes, or if the overall AAR, drops below 40 per hour, the center “meters” arrivals into the TRACON.⁸

⁸See FAA Order ZBW 7210.482A of 1987 issued by the Boston Center.

This regulates the number of aircraft entering each hour to be no more than the AAR. The rule is not hard and fast, however. If, for instance, the AAR is x , and x aircraft have landed in the first 50 minutes of the hour in question, it is not that aircraft will be forbidden to land in the remaining 10 minutes. Rather, if the center persistently feeds in more than the TRACON specifies, the latter would complain.

In-trail Restrictions: In addition to metering (or sometimes as a proxy for it) the center can increase the minimum separation between aircraft in-trail, from the normal 10 miles (in force outside the terminal area) to something higher. This effectively reduces the arrival rate into the TRACON.

In-trail delays, ground holds and metering are all means of transferring the terminal area delays upstream in the system. Sometimes, however, there is a need to delay aircraft at the TRACON itself. This could be due to congestion in the terminal area that was not anticipated far enough ahead to commence metering. Such unanticipated congestion could, for example, be due to weather changes.

Delays in the TRACON are induced in two ways: path stretching and speed control within the TRACON, and holding at the entry fixes. The former is useful for short delays, no more than 10-20 minutes depending on the approach path. Path-stretching involves making the plane fly a circuitous route in the terminal area, deviating from the typical approach paths, depicted in Fig. 2.2. Speed control usually implies slowing an aircraft in order to increase the time that it takes to traverse the approach path. If longer delays are required, the aircraft are made to "circle" or "hold" at the entry fix. Typically the holding pattern is a racetrack-like loop which takes jets⁹ about 4 minutes to complete. These loops are one above the other separated by 1,000 feet in altitude. Usually at each fix there are up to 6 aircraft circling at heights ranging from 6,000 feet to 11,000 feet. When a landing place becomes available, the bottommost aircraft leaves the holding stack, and all the others move down one level each, leaving space for a newcomer at the topmost level.

⁹Props are rarely held.

We shall pay close attention to these methods of delay when we study the Aircraft Sequencing Problem, in Chapters 4 and 5. Before that, however, we shall develop, in the next chapter (3), a model for the Landing Time Intervals between aircraft. That model, as we shall see, will draw a great deal upon the understanding of ATC procedures and practices that we have tried to achieve in this Chapter.

Chapter 3

A Model for Landing Time Intervals

3.1 Overview

3.1.1 The Statistical Problem

This chapter details the development of a model for Landing Time Intervals (LTIs), as they have been observed at Logan Airport. We study LTI data collected under heavy traffic conditions and on the primary runway, when the airport is most stressed (as we shall explain).

We wish to examine three important statistics of the LTI distribution: the mean, the standard deviation, and the 25th percentile. The mean LTI is important because of its direct bearing on the airport landing rate. It is related to the throughput of a runway by the approximate formula

$$\frac{1}{LTI} \simeq \text{Average arrival rate at runway.} \quad (3.1)$$

The standard deviation gives an indication of the fluctuation in LTIs (about the mean) and some idea of how “tightly controlled” LTIs are.

The statistical problem with respect to the mean LTI and the standard deviation is to identify salient factors (such as runway configurations, weather conditions etc.) that significantly influence them and to quantify the effects of these factors. Our aim is to

develop an approximate equation for the mean LTI in terms of the factors that significantly affect them. In this chapter we shall discuss the series of statistical procedures used to identify these factors, and to estimate their effects. We shall discover that the mean varies significantly with respect to these factors, but that the standard deviation does not. Hence the LTI equation is developed in terms of the mean values.

The third statistic we examine is the 25th percentile. This statistic has an important bearing on the prescriptive question about LTIs which will arise from the descriptive model: in an advanced ATC system, how might one specify minimum time separations between aircraft. We shall argue that the 25th percentile of observed LTIs might be acceptable to controllers and pilots since they go below it a quarter of the time anyway. Thus it approximates LTI values with which controllers and pilots feel comfortable operating, or at least accept presently.

The statistical problem includes developing a model and then testing for its correctness. Thus we have a two part exercise: calibrating a model on a given amount of data, and then validating it on data completely different from those on which it was calibrated. Describing the calibration and validation will be the focus of the rest of the chapter.

3.1.2 Contents of Chapter

In the next section (3.2), we describe the data sets that we collected for the calibration exercise explaining how and when they were gathered. Then, in section 3.3, we discuss the conventions governing the pre-processing of data. These include explaining why we study landings only in “heavy-traffic” conditions and, furthermore, only those on the primary runway. We also discuss the factors which might have an important effect on LTIs (and thus will be used in the model), as well explaining those that might not.

In section 3.4, we describe the statistical tests used in developing the calibration model. As mentioned, and will be discussed again, these focus on the mean LTI. We identify the variables (e.g. runway configuration, sequence category) that have a significant effect on the mean LTI. We try to develop groupings of these variables within which LTIs are homogeneous in their mean values. These groupings lead to a calibration model for LTIs

which we specify at the end of Section 3.4. In section 3.5, we focus on the exercise to validate the model we have developed for LTIs. Thus we describe the new data collected, and the statistical procedures used for validation.

Section 3.6 concludes this chapter with an assessment of the implications of the model for LTIs. This includes both what insights we have gained into near terminal area air traffic operations, and thoughts on the application of a descriptive model for prescriptive purposes in an advanced ATC system.

3.2 Data for Calibrating the Model

3.2.1 Time Periods when Data was Collected

The descriptive model is meant to study the LTIs under a variety of weather and runway conditions. It is also important, for reasons we discuss, to focus on heavy traffic conditions. Thus a major task was to organize the data collection effort so that it could best meet these needs. In order to capture different runway and weather conditions, we would plan a particular data gathering session (there were 18 in all) typically on a single day's notice at most. We would listen to daily weather forecasts, and find out the runway configuration being used at Logan. The aim was to collect as diverse a sample of these variables as possible. At the same time, it was also important to focus on those conditions that are frequently occur at the airport (e.g. 4R/L for landings).

It is also important to collect data only in heavy traffic conditions: that is when the runways and airspace are most congested and the data would best reflect the stressed operating conditions which TATCA most hopes to alleviate. In light traffic LTIs would be large because of sparse traffic, and other variables such as the sequence category would presumably have little effect on them. The airspace and runways would be uncrowded and conditions in general not of immediate interest to TATCA.

Broadly, we may consider a heavy traffic period as one during which the airport is operating close to capacity. One "first-order" approximation of a heavy traffic period is suggested by the Arrival Acceptance Rate (AAR)—the ATC system's definition of

runway capacity. As discussed in Chapter 2, this is roughly the rated capacity of each operating runway in a given landing configuration. Thus we attempted broadly to satisfy the requirement of heavy traffic conditions by aiming to collect data when the actual arrival rate of aircraft at a runway was at or above its corresponding AAR. Even so, however, there were periods when the traffic was locally sparse. In the next section we discuss this problem and explain how we develop a criterion to eliminate LTIs from periods with locally sparse traffic.

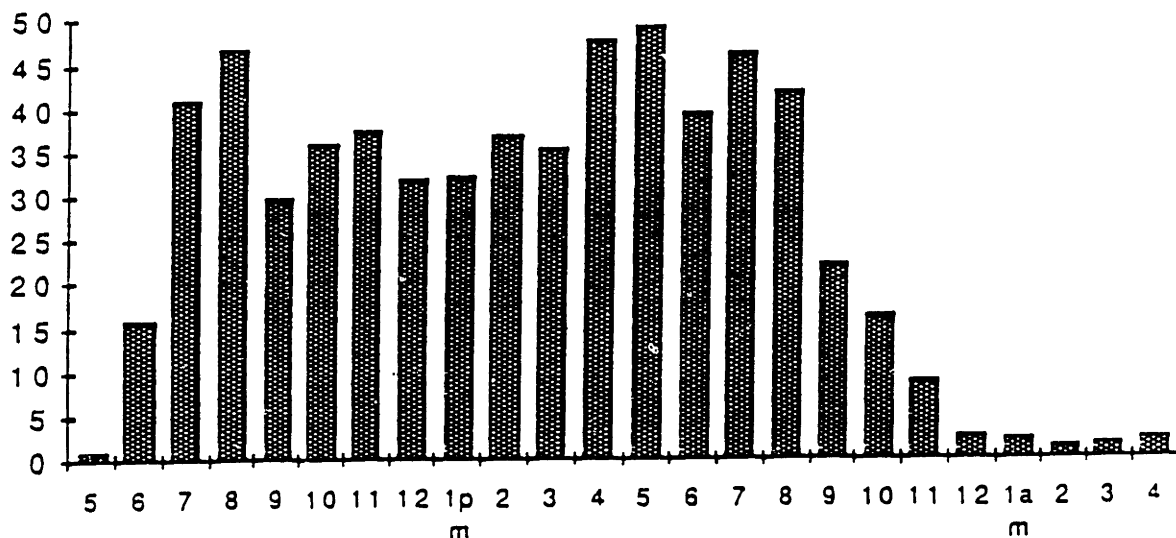


Figure 3.1: Hourly Demand for Operations at Logan Airport

To satisfy the first criterion for heavy traffic, we needed to collect data from times of day with the greatest landing traffic. The notion being that during these periods, the airport would be operating at or above the AAR. Figure 3.1 shows the average hourly demand for landing operations at Logan Airport (see Abundo (1990) [1]). There are roughly two peaks in demand, in the morning between 7:00 a.m. and 9:00 a.m., and a more sustained one in the evening from 4:00 p.m. to 9:00 p.m. The early morning peak

is due to arrivals of commuter flights from the New England area, and shuttle flights from Washington, D.C. and the New York metropolitan airports. The evening rush is also caused by shuttle flights and commuter aircraft, in addition to trans-Atlantic flights and many middle-distance flights (e.g. Chicago-Boston) that are scheduled for this time because it is convenient for passengers.

Based on this traffic pattern, we focussed our data collection efforts in the evening rush period. Data was gathered in two ways: by radar and by hand, as we shall discuss shortly. We tried to capture a representative sample in terms of weather conditions and runway configurations. It was impossible to capture all runway configurations—for example, we never saw 4R used alone for both landings and takeoffs. We collected data beginning approximately at 4:00 p.m. (for the start of the evening rush) and lasting three to four hours. The starting time and the duration often varied because of the availability of radar operators at Lincoln Laboratory.¹ Thus sometimes the weather conditions and runway configurations were of a type we wished to observe, but the radar would be available only until 6:00 p.m. We would thus begin at 3:00 p.m. even though the evening rush would commence later. At the other end, we never collected data after 9:00 p.m., when the traffic would perceptibly slacken.

3.2.2 Collection of Data by Radar

As mentioned, we collected data by two methods: radar and by hand. Collecting data by radar is an involved process, but it yields accurate results. This is especially true since Lincoln uses a very advanced civilian radar (a “Mode-S” facility) for the purpose. The procedures followed for each data collection were as follows. First the radar was turned on for the duration of the data collection. During this time all aircraft within a radius of 60 miles of the *radar facility* were tracked. This adequately covers the TRACON area of Boston. Thus all aircraft were tracked for the portion of their flights above 400 feet. Below this altitude the radar cannot detect aircraft. The tracking method was the common “ $\alpha - \beta$ ” algorithm (see Blackman [4]). A computer attached to the radar performed the

¹There was a setup time for any data collection, especially if by radar, as most of the calibration ones were.

tracking and stored the tracks of each aircraft on magnetic tape. A four hour data session typically generated 80 Mbytes of information!

The tracks stored on magnetic tapes were, at the next stage of the process, filtered through the GATEX program to obtain the LTI values. The GATEX program takes a series of tracks, stored for each plane as the locus of its flight over the region monitored by radar. Each locus is expressed as a series of coordinates on a cartesian system with respect to the region monitored. Note that the locus is 4-dimensional, giving the path of a plane in terms of (x, y, z) coordinates as well as time. The program includes as parameters certain "gates", lines or circles, whose start and end coordinates, or equations and centers, respectively, are given. These gates are of infinitesimal width and infinite height. Thus a gate that is a line segment should be thought of, in 3-dimensional space, as a plane perpendicular to the surface of the earth, and the circular gate as a cylinder with its base on the surface of the earth.

The circular boundary used is:

- **cir_OUTER**, referred to as 'Circle Outer', or the outer circle: a circle of radius 30 nm, centered at the intersection of runways 4 and 15. This approximately covers the region of the TRACON.

The line segments are located at points on the straight line landing approach to a runway (an extension of the runway centerline) and are thus defined for each runway. They are:

- **The Outer Marker**: A segment 1 nm in length perpendicular to the relevant runway and at a distance of 5 nm from its threshold.
- **The Middle Marker**: A segment ~ 0.5 nm in length, perpendicular to the relevant runway at a distance of 2 nm from the threshold.

Given the flight path of each plane that traverses the TRACON during the time period of interest, GATEX determines the time, altitude, speed and exact position (important for the circular marker) at which the plane crossed each of the gates. Note that a plane

may not cross all the gates: an aircraft overflying Logan is unlikely to cross the middle marker of the landing runway.

There are two crossing times of interest to us. The first is the crossing time of the middle marker. This is usually the point closest to the runway (only two miles away) at which the radar starts losing the aircraft. The intervals between successive crossings of the middle marker of a runway can be taken as being approximately equal to the LTIs. The other crossing time of interest for each plane is at the outer circle. The difference between the crossing time at the outer circle and at the middle marker is approximately equal to the time each aircraft spends in the TRACON. This time span, termed the time-to-fly, is of great importance when studying the sequencing problem in Chapters 4 and 5.

The radar at Lincoln identifies each aircraft by its transponder code. This is the same code used by controllers. The final step of each data collection is to identify the weight-class of each aircraft corresponding to the tracks filtered by GATEX. For this we use information stored in the TRACON computer identifying aircraft by transponder codes. For each transponder code (and hence each radar track) we can find the flight number, type and weight-class of the aircraft referred to.

Our original plan was to collect all data by radar. But the process, while accurate, was very involved and not absolutely necessary for studying LTIs. When data was collected by radar, it took at least a month for the raw tracks to be processed by the GATEX program and for the requisite information to be collated for analysis. The other problem was that radar collection always depended on the availability of both equipment and personnel. This was often difficult at short notice, such as if there were specific runway configurations in use which we had not seen previously.

Thus it became necessary to supplement this radar data by hand-collected observations. Among the 10 data sets used for calibration, only three were taken manually, but all the validation data were obtained manually, by the author. When collecting data by hand, the observer notes the time at which each aircraft crosses the runway threshold. The flight number is identified by simultaneously listening to the radio conversation between the pilot and the control tower. The aircraft type (and weight-class) is obtained either

this way, or by referring to the schedule in the *Official Airline Guide* for the fortnight in question. Obviously, when collecting data by hand, one cannot get crossing times at points other than the runway. Thus LTIs taken by radar are measured at the middle marker, while those taken by hand are measured at the runway threshold. It is hence important, for the sake of consistency, to convert one series of measurements to the other. In the next section (3.3) we will discuss how we do so.

3.2.3 Examples of Data Sets

Our initial data collection for calibration purposes was comprised of 10 data sets. 3 were collected by hand and 7 by radar. They spanned a period slightly less than two years, between September, 1987, and July, 1989, but the bulk were collected in 1989. Hence the data collection covered a variety of seasonal patterns of traffic over a relatively long time span. The primary seasonal effect of traffic, as discussed in Chapter 2, is to change the relative composition of traffic, in terms of Large, Small and Heavy aircraft [20]. Hence the seasonal effect would change also the relative proportions of LTIs by sequence category.

In this subsection, we describe two data sets, one obtained in VFR and the other in IFR conditions. We hope thus to give a flavor of the amount of data available as well as the types of operating conditions we studied. Appendix A contains similar descriptions of the remaining 8 data sets used in the calibration model.

2-17-89: Data were collected by radar for 4 hours from 4:00 p.m. onwards. There were 158 landings in the period. The conditions were fine VFR with visibility better than 5 statute miles and the cloud ceiling higher than 5,000 feet. The wind, however, fluctuated frequently. It was initially from 290° at 10 knots, and then changed abruptly at 6:10 p.m. to 16° at 8 knots. The runway configuration for landings changed correspondingly from 33L/R to 4R/L. At 7:30 p.m. the winds changed again, now to 336° at 5 knots. The runway configuration consequently reverted to 33L/R. We refer to the period 4:00 p.m. - 6:10 p.m. as the first period of use for 33L/R, and 7:30 p.m. - 8:00 p.m. as the second period. As mentioned in Chapter 2, 33R is a very short runway (2,500 ft. long) and accepts only the smallest of aircraft for landings. The AAR was 34/10 when 33R/L was

in use, and 34/24 otherwise. As Table 3.1 shows, the primary runway always received more arrivals than its AAR, but the secondary was rarely stressed.

Time (P.M.)	Total Number of Landings	Runway			
		33L	33R	4L	4R
4:00 - 5:00	44	41	3	0	0
5:00 - 6:00	39	37	2	0	0
6:00 - 7:00	42	7	2	12	21
7:00 - 8:00	33	16	1	5	11

Table 3.1: Number of Landings on 2-17-89 by Runway and Time

Further, Table 3.2 shows the number of aircraft by sequence categories for the two periods when 33L/R was in use. One observes a predominance of Large and Heavy aircraft in the traffic mix. Even if there are Small aircraft in the traffic mix, our data tend to exclude them if the conditions are good enough to have parallel landings, as they were on this day.

3-30-89:

Data were collected by radar between 3:00 p.m. and 6:00 p.m. under very poor weather conditions. A thunderstorm front with driving rain extended 100 statute miles from SW to NE; moreover, it was stationary in the region of the airport. The cloud ceiling was 500 feet, thus weather conditions were IFR-2. The number of arrivals is given in Table 3.3. For the first 25 minutes, 15R was used for landings. It accommodated 15 aircraft. In the remaining 2.5 hours, 4R was used for landings, and it accommodated only 64 of them.

Second Aircraft in Sequence

		First Period			Second Period		
		H	L	S	H	L	S
First Aircraft in Sequence	H	2	13	0	0	1	0
	L	14	53	1	0	12	1
	S	0	1	0	0	1	1

Table 3.2: Number of Aircraft Landing on 33L on 2-17-89 by Sequence Category

There were periods of as long as 20 minutes without a landing. On a few occasions, pilots declined to land, and had to be held at an entry fix. This was the worst weather encountered in any data set.

Time (P.M.)	Total Number of Landings	
	15R	4L
3:00 - 4:00	15	12
4:00 - 5:00	0	24
5:00 - 6:00	0	28

Table 3.3: Number of Landings on 3-30-89 by Runway and Time

3.3 Pre-processing of LTI Data

In this section we discuss some conventions established for studying LTI data. These are of three types:

1. Translating LTIs collected by radar to make them consistent with those collected by hand.
2. Deciding which LTIs come from (locally) heavy traffic conditions and which do not.
3. Determining what variables will be explicitly included in the calibration model.

When discussing the second convention, we will also explain why we focus on LTIs on the primary runway. We consider each convention separately in the following subsections.

3.3.1 Adjusting Radar LTIs

Given that the data was collected by two different methods, radar and hand, it is necessary, for the sake of consistency, to translate crossing times at the runways to those at the marker, or vice-versa. We chose to focus on runway landing times in the expectation that doing so would yield more accurate results, given that the velocity of each aircraft when it crossed the middle marker is known. The conversion of crossing times at the middle marker to times at the runway proceeds as follows.

Let t_m be the (known) time at which an aircraft crosses the middle marker, and t_r the (unknown) time at which it crosses the runway. We know the velocity v at which the aircraft crosses the middle marker, which is 2 nm from the threshold of the runway. Assuming that this velocity is constant for the last two miles, we may estimate

$$t_r = t_m + \frac{2}{v} \quad (3.2)$$

Using (3.2) we adjust LTIs from the radar data sets to make them comparable to those from the hand collected sets. The average LTI for those computed at the middle marker was 101.49 seconds. After the adjustment this mean was 101.65 seconds, only one-sixth of a second different. The maximum absolute difference was attained when there was the greatest difference in speed between the lead and trail aircraft comprising the LTI. This largest (absolute) change in estimated LTIs was 24 seconds, when the aircraft velocities differed by 64 mph (the lead travelling at 170 mph and the trail at 106 mph).

3.3.2 Defining Locally Heavy Traffic

As noted earlier, LTIs from light traffic conditions are not particularly useful for our purpose, because the airport is not stressed. In order to focus on heavy traffic conditions, we collected data during the period in the evening of high demand for landings. We need to fine-tune this data further, however.

Studying LTIs on the Primary Runway Only

The immediate implication of wishing to concentrate on heavy traffic conditions is that we shall restrict our study to landings on the primary runway.

We notice in our summary descriptions (section 3.2 and Appendix A) of the data sets that the secondary runway (except for the 27/22L combination) never accepted more than 10 landings per hour, whatever its AAR. This results in part from the type of traffic diverted to such runways and in part from the way the secondary is used by the controllers. The latter is especially true when, as with the 22L/R combination, the secondary is used as a main departure runway as well. We mentioned in Chapter 2 that a rule of thumb for

22L/R is that each landing on 22R must be succeeded by 3 departures (at least). Even otherwise, usually only propeller aircraft land on the secondary runway (except, again, for 27/22L). These come to the airport under "free-flow" and do not strain the system, on different landing streams (lower and slower), from different fixes, and do not merge into the same final landing stream. Further, arrivals on the secondary, are not constrained in terms of separations between individual aircraft: they are so far apart that there is scarce danger of them coming too close to one another. We should emphasize, however, that the use of a secondary runway is important. The existence of a secondary runway accepting landings relieves the primary of traffic, and can have a substantial effect on the LTIs there, as we shall see later.

Studying LTIs Under 200 seconds Only

Focussing now on LTIs on the primary runway (when two were used), we would like to establish some standards for what constitutes the heavy traffic conditions under which we shall study LTIs. We use two methods to do this. One is to collect data only when the airport faces greatest demand, the other is to eliminate locally sparse periods of traffic from consideration.

The airport experiences the most demand for landings on week day evenings, between 4:00 p.m. and 9:00 p.m. We tried to collect data during such periods, when the actual arrival rate was greater than or equal to the AAR. There were, however, exceptions. But these were of three kinds:

1. When the weather was very bad (as on 3-30-89) and traffic was held outside the TRACON or on the ground;
2. When 27/22L was used and the overall capacity was very high, and could be used differently by different controllers;
3. When conditions were fine, but there was merely a small shortfall in demand for the airport (as between 4:00 p.m. and 5:00 p.m. on 10-11-88, or between 4:00 p.m. and 6:00 p.m. on 5-31-89).

On all these days, however, there were rather long periods when the traffic was "locally heavy". Thus there were long stretches with many aircraft landing frequently (we will

quantify this later) followed by long gaps, typically toward the last third of the hour. Thus the broad criterion of choosing the evenings assures long periods of “locally heavy” traffic; but we need to have a finer grain definition of “locally heavy” traffic conditions under which we will study LTIs.

While the broad criterion relating heavy traffic to the AAR comes from normative notions in ATC, the finer grain definition of heavy (or locally heavy) traffic comes from the data. We studied histograms of the LTIs on the primary runway for each configuration used on each day. Figure 3.2 is a typical example of such a histogram. It depicts the distribution of the LTIs for landings on runway 4R (there was no secondary runway then) on 10-11-87. Typically, the smallest LTIs on any given day are slightly below 50 seconds. At the other end of the scale, they are, on occasion, over 300 seconds. The bulk of the LTIs are, however, below roughly 180 seconds on most days.

In the histograms studied, the 200 second mark typically constitutes a break between the main mass of data points and the outliers. Typically, no more than 5% of the data lie to the right of this. The exception is when 27 is the primary, and the secondary (22L) absorbs an uncommonly large fraction of all landings. Note that, on the secondary runway, most LTIs are larger than 200 seconds. Hence we define LTIs of 200 seconds or less to have arisen from locally heavy traffic. Admittedly, the choice of 200 seconds as a cutoff point is arbitrary. However, given the somewhat clear delineation of LTIs in the histograms into those less than 200 sec. and those greater, it is a logical choice.

The physical interpretation of this cutoff is that LTIs above 200 seconds are assumed to be large because of paucity of traffic, rather than an overly safety-conscious pilot, or an ATC miscalculation. On the other hand, an LTI of 150 seconds, though it may be larger than average, is assumed to have occurred in heavy traffic. This would be even though the the previous LTI and successive one may have been larger than 200 seconds. This is, however, a very infrequent case: typically there would be at least five to six aircraft between the gaps greater than 200 seconds. The notion being that the traffic was locally dense enough that the LTI could have been smaller; but the LTI was high either due to systematic factors, such as separation requirements, or to chance fluctuations. There is,

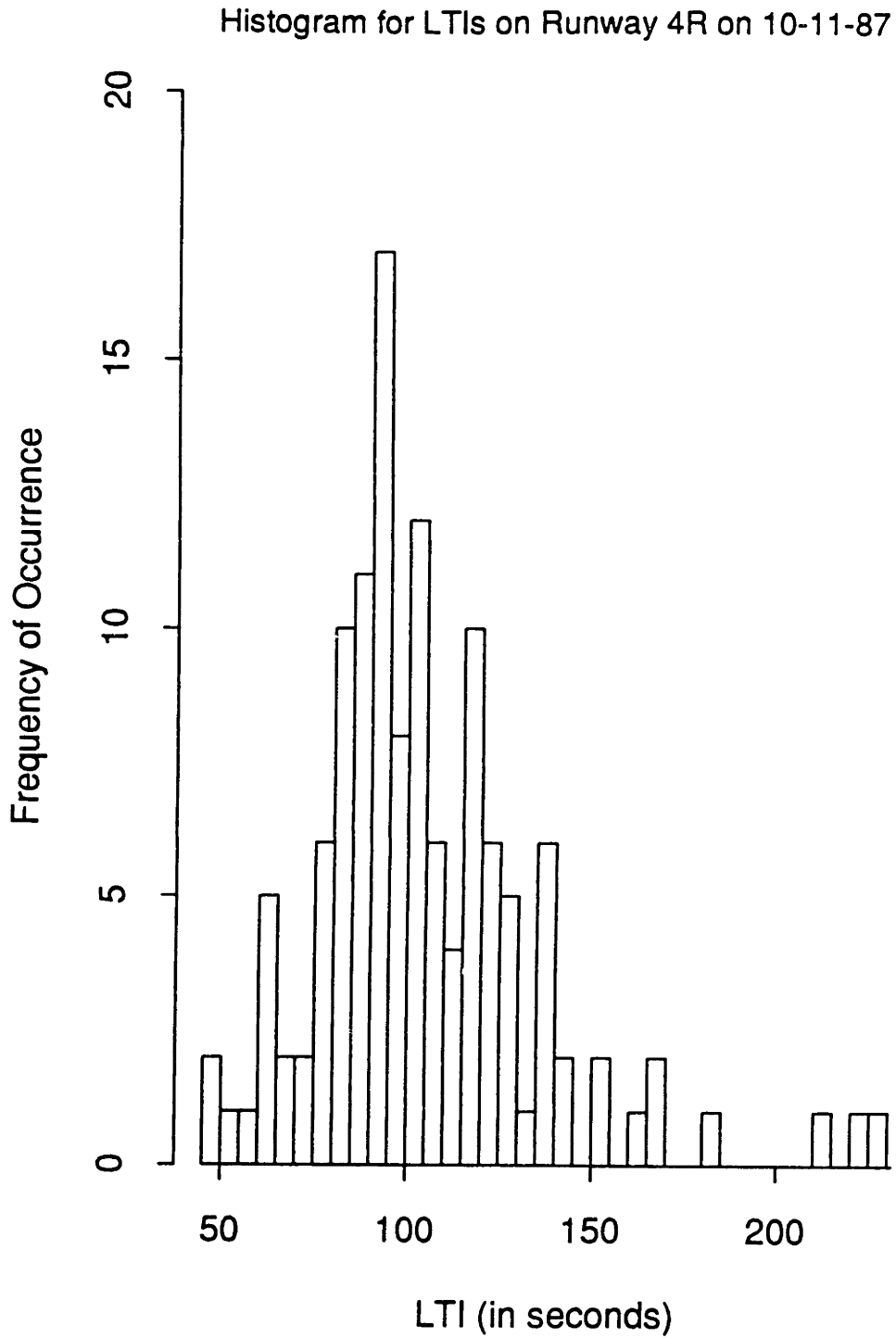


Figure 3.2: Histogram of LTIs on 4R on 10-11-87

of course, always the chance of a false positive error: a 150 second LTI could have been due to a pilot being merged into the final approach stream late rather than anything else. Thus the cutoff tries to capture physical realities but may not do so accurately.

Finally, we note that there are occasions when a single runway is used for landings, but an aircraft (invariably a prop) requests to land on the parallel runway because it is closer to the terminal. Thus a plane on an ILS approach to 4R might request “a transition” to 4L. If such a request is granted, it would move over. In our analysis we ignore LTIs elongated by such a transition. Thus if there three successive aircraft, and the middle one transitions, we ignore the LTI between the first and the third. Typically this LTI is above 200 seconds, though it does not really reflect locally light traffic, as we have described it. Further, if we use radar tracking and the aircraft made the transition after it crossed the inner marker, we also ignore the LTIs between the first and second and the second and third aircraft. This is a relatively infrequent occurrence, however.

3.3.3 Factors Not Included in the LTI Model

We have already mentioned some variables we believe could have important effects on LTIS. These are sequence category (equivalently the weight-classes of the lead and trail aircraft), weather and runway configuration. There are many other factors that can affect the LTI, but which we do not include explicitly in our model: some that are subsumed in those variables already mentioned, others presumed to have marginal effects on LTIs, and still others for which obtaining accurate measurements and data was too difficult.

We discuss below other factors we do not consider explicitly in our LTI model, and explain why not.

Departures: In Section 2.4.4, we had discussed the levels of interaction between arrival and departure traffic. Essentially, there is very little interaction between arrivals on the primary runway (which we study) and any form of departure traffic. The only exceptions are if there is only one runway in use for BOTH arrivals and departures, or if there is an occasional departure from the primary. The former condition is found only in extremely poor weather, and we never observed it in our data collection. The latter case,

occasional departures from the primary, takes place relatively infrequently—rarely more than one or two an hour. These involve aircraft that wish a long runway to depart (e.g. a fully loaded B747 prefers to depart on 4R rather than 9). In that case, the controller waits for a gap in the arrival traffic, or he creates one, to permit the takeoff. Such a gap is well over 200 seconds, and is not included in our data. Hence we can ignore the impact of departure traffic on the analysis of LTI data.

Season: We assume the season of year not to be a factor significantly affecting LTIs. The main effect of seasons is to change the mix of aircraft in terms of the relative proportions of Large, Heavy and Small aircraft. There is, for example, a marked increase in the proportion of Small aircraft landing at Boston during the summers. This is because pleasure travel to Northern New England is more common then. Hence seasonality will primarily change the relative proportions of certain sequence categories among the LTIs. We shall be incorporating the sequence category as an important variable in our study of LTIs, and thus do not need to consider the effects of seasonality explicitly. Further, the season is obviously related to the weather, but the weather variable is included in the runway configuration.

Identity of Aircraft: The FAA rules classify aircraft by weight-class, as discussed earlier. The weight-class variable is important in determining separations and naturally will be included in the study. Knowledge of the specific type of aircraft (e.g. DC9) or carrier (e.g. Delta) seems unlikely provides extra information. ATC in the terminal area pays no attention to these factors.

Entry Fix: The entry fix determines the time the aircraft spends in the terminal area, as also its position in the landing sequence. Hence it is important in our study of sequencing. The LTI model, however, focusses on the very end of the final approach path, at the runway threshold. By this stage, separations are determined by the factors on the ground (weather and runway configurations) and the identity of the landing aircraft. Hence the entry fix is not really of importance when studying LTIs.

Aircraft Speeds on Final Approach The groundspeeds on final approach of the lead and trail aircraft comprising an LTI, and the actual distance separation between

them, determine the LTI. The airspeed of an aircraft on final approach decreases from about 170 knots at the outer marker to a recommended landing speed just before the threshold. The recommended landing airspeed depends mostly on the the weight class of the aircraft and marginally on the wind on the ground. The weight class of the aircraft affects its stall speed.² The recommended landing airspeed for an aircraft is 1.3 times its stall speed. If there are gusting winds, the landing speed is increased by half the range of the gusts. Thus if winds are gusting from 10 to 20 knots, the landing airspeed rises by 5 knots. The groundspeed of an aircraft is the sum of its airspeed and the component of the wind velocity along its direction of travel.

We have used radar measurements of groundspeeds in computing inner marker crossing times for certain aircraft. However, we do not use speeds explicitly in a model for LTIs, because we have these measurements only when we collected data by radar, or about 7 out of 18 total data sets. Further, we cannot make estimates of groundspeeds for other data sets, based on assumed standard landing airspeeds. Then we would need data on wind velocities, which as we discuss next, are also difficult to obtain.

Wind: The wind as we mentioned affects the speeds at which aircraft travel. It is also present at every stage of the aircraft's descent. The tower provides information about winds velocities at the ground, but the problem is to find out the velocities of winds aloft. This is little relationship between the two: winds even a few hundred feet up can differ radically in magnitude and direction from those on the ground.³ Indeed, listening to radio communications, one sometimes hears the control tower asking pilots about wind conditions on final approach.

To include winds as a factor explicitly, we would have to consider both winds on the ground and those aloft. The important effect of wind velocity at ground level is that it determines the runway configuration used for landings. This we consider as a separate variable in our LTI study. The question then is whether one can compute winds aloft and,

²Technically, the stall speed of an aircraft is the minimum speed at which the lift (from the aerofoil, or "Bernoulli", effect) on the aircraft equals the sum of its weight plus the downward component of the drag. It depends on the weight of the aircraft (including passengers and fuel) and on such factors as altitude, atmospheric pressure, etc.

³W. Hollister, Private Communication.

if so, how.

There are ways of estimating winds aloft from radar tracks when aircraft are turning upon entry into the terminal area, but not on final approach. There is no measuring station near Boston, that would provide measurements at various altitudes. The nearest are at Albany, NY, Portland, ME, and Chatham, MA. Approximating the winds over Boston from these measurements would not really be meaningful. Hence we cannot include winds aloft in our model (in combination with airspeeds).

3.4 Calibrating the Model for LTIs

3.4.1 Preliminaries

We discussed, in the preceding sections (3.1-3.3) how we gathered data, and the conventions and procedures used to prepare them for analysis. We focus on LTIs less than 200 seconds obtained on the primary runway. Having thus pared the data we are ready to commence developing a descriptive model for LTIs. As mentioned previously, the model development has two stages: calibration and validation. In this section, we shall devote ourselves to the former.

The aim of the calibration exercise is to develop a reasonable model for the LTIs. We have already mentioned the variables (runway configuration, weather, and sequence category) that we think might affect LTIs. In calibrating a model, our aim is to evaluate whether such variables affect the important statistics of the LTI distribution, and to quantify these effects. In doing so, we shall be focusing on the mean LTI. This is because the average LTI is intuitively appealing and it is of interest to air-traffic controllers, since that is the variable that they try to control.

Although we develop our model in terms of the mean, we are also interested in two other statistics of the distribution of LTIs: the standard deviation and the 25th percentile. The standard deviation is indicative of the spread of the LTI description around the mean. The 25th-percentile, we will argue, is useful for prescriptive purposes for an advanced ATC system. Further, we assume a priori that if the mean LTI is the same under a given set

of circumstances (runway configurations or sequence categories) then so is the standard deviation and the 25th percentile.⁴

Indeed, this a priori assumption at the calibration stage serves to highlight the important relationship of the calibration to the validation exercise. We develop a model which, based on some statistical tests, we believe to be reasonable. These tests are themselves motivated by our understanding of the ATC process (vide our arguments for why an aircraft's entry fix is unimportant but its weight-class is). We have, however, relative latitude in developing such a model because we are going to verify its validity with new data different from those on which it was calibrated. Thus we may detect any relationships that were artifacts of the calibration data and not truly representative of some underlying physical process.

We hope to develop a model that accurately captures the effects of important variables in the ATC process, yet is relatively compact. Hence an important issue in the ensuing discussion is the tradeoff between compactness and accuracy. Thus in a nutshell our strategy will be to test for variables which have significant effects on the mean LTI. Then we shall develop groupings of circumstances, with respect to the variables studied, that have homogeneous mean LTIs.⁵ We shall compare the LTIs within each such grouping in the calibration data to the corresponding ones in the validation data. These comparisons, however, will not merely be in terms of means, but also the standard deviations and the 25th percentiles, the other statistics of interest.

3.4.2 Overall Summary of Data

As already mentioned, we have collected calibration data on 10 days. Often, however, there was more than one landing configuration used within a single collection period. We shall henceforth refer to data for a particular landing configuration on a particular day as a single data set. The 10 days of calibration data yielded 15 such data sets. Table 3.4

⁴Moreover, as we shall discuss, the standard deviation is not significantly affected by the variables identified above.

⁵We shall use the term "homogeneous mean LTIs" not to denote that the mean LTIs are literally identical, but that they are essentially equal, from a statistical standpoint, any difference being attributed to random variation.

summarizes the mean and standard deviation of LTIs for each data set.

Date	Runway	# Obs.	Mean	Std. Deviation
9-13-87	4R	96	102	28
"	15R	19	101	14
10-11-87	4R	90	105	35
5-13-88	22L/R	107	104	28
"	27/22L	18	115	33
5-18-88	4R 91		92	28
2-17-89	33L/R 92		93	31
"	4R/L	25	114	33
3-30-89	4R	55	107	31
5-31-89	4R/15Rc4L	69	101	33
"	15R	14	99	33
"	4R	27	92	39
6-09-89	4R	88	99	36
6-15-89	4R	95	95	31
6-28-89	27/22L	51	113	33

Table 3.4: Mean and Standard Deviation of LTIs by Runway Configuration and Date

The data sets varied in size from 14, when 15R was used on 5-31-89, for a short interval, as a ship passed through the channel behind 4R, to 107, when 22L/R was used for $3\frac{1}{4}$ hours continuously on 5-13-88. We note that we are only considering LTIs of 200 seconds or less. Thus we mean that there were 107 such LTIs on 22L on 5-13-88; the actual number of landings was 118.

The mean LTI ranged from 92 sec. on 5-31-89 (with 4R in use) to 115 sec. on 5-13-88 with 27/22L in use. This suggests that the average time interval between aircraft is

between a minute and a half and two minutes. The smallest average value occurs when a single runway was used for landings, and all aircraft had to land on it. By contrast, the largest mean LTI arose when two long parallel runways are used for landings, with the secondary able to accommodate jets as well as props. Under such circumstances, the primary runway is strained relatively little. We should note that the mean LTI presented is rarely the reciprocal of the actual hourly landing rate, since some period(s) of light traffic (gaps larger than 200 seconds) are invariably present in each hour.

The lowest LTI in the data is about 50 seconds; at the other end of the scale, the highest LTI is 200 seconds, our cutoff point. Incidentally, it should be pointed out that the right tail of the distribution is not in general a continuum up to 200 seconds. Rather, it typically ends at about 170-180 seconds, and there would be some isolated LTIs bigger than that.

The standard deviation of the LTIs is somewhat high, typically around 30 seconds. The high value of this number is explained in part by the long right tail of the distribution. There are relatively few LTIs below 60 seconds, and relatively many above 120 seconds. Part of the reason for the long right tail is, as we discover later, that LTIs from sequence categories which require larger separations based on wake-vortex considerations, are on average bigger than those from sequence categories which do not. Thus some of this right tail can be explained by systematic factors, and the rest may be random variation. There could also be the false-positive error mentioned earlier, that some LTIs were big due to sparse traffic, but were below the cutoff of 200 seconds and thus were included in our data analysis. Thus a long right tail to the distribution leads to a high standard deviation in LTIs. There are some standard deviations, however, that are different from the rest: one of 14 seconds (on 15R on 9-13-87) and one of 39 seconds (4R on 5-31-89). Later, in Section 3.4.5, we explore the differences among standard deviations in the various data sets.

3.4.3 Is the Mean LTI Independent of Runway Configuration

A starting hypothesis could be that the mean LTI on the primary runway is independent of any factor of the ATC system: of the runways used, of the weather conditions, of the weight-classes of the lead and trail aircraft. Thus a null hypothesis could be that any differences in the mean LTI can be explained by random variation alone. Stated mathematically, the null hypothesis can be expressed as the mean LTI being equal to a constant, independent of all else:

$$\overline{LTI} = K. \quad (3.3)$$

To see whether the data support such an extreme hypothesis, we begin with some general tests for whether the runway configurations and the sequence categories have any effect on the mean LTI. In this subsection, we focus on runway configurations.

Our null hypothesis is that runway configuration has no effect on the mean LTI.⁶ A priori there are arguments both for and against such a null hypothesis. On the one hand, we are focussing only on LTIs in heavy traffic on the primary runway only. The traffic on this runway consists largely of scheduled arrivals. When there is a secondary runway in use, a large portion of its landings are props, many of which are unscheduled General Aviation aircraft. These may not have landed at all when a single runway was in use, because they might have cancelled their trips in the IFR weather. Thus the landings on the secondary runway might not increase the LTIs on the primary. Hence we might think that the particular runway configuration does not affect the LTIs on the primary runway.

On the other hand, however, it seems logical to expect at least some form of interdependence between LTIs on the primary runway and whether or not there are landings on a secondary runway. From the point of view of traffic, some of the props on the secondary runway are scheduled flights, which might otherwise land on the primary runway in bad weather. From an operational point of view, in some types of marginal IFR conditions, parallel landings do occur (e.g. 4R/15Rc4L). These parallel arrivals are often staggered, so that the aircraft may keep each other in view. Hence all these might lead to longer

⁶The null hypothesis says nothing else has an effect, either, but we look at the runway configuration only, for the moment.

LTI on the primary.

Thus to test whether the runway configuration has any effect at all on the mean LTI, it seems logical to group the runway configurations according to the use of a secondary runway. We can partition runway configurations into two groups: (I) with only a single landing runway in use, and (II) with two landing runways in use. However, we have to be somewhat careful in assigning runway configurations to these groups. The standard VFR multiple runway configurations (22L/R, 27/22L, 4R/L) clearly belong in Group II. The secondary is fully used and accepts many landings. The standard IFR single runway configurations (22L, 4R, 33L, and 15R) all belong in group I.

There are three configurations, however, where the assignment is not clear. One is 4R/15Rc4L, and the others are 33L/R and its counterpart 15R/L. The reason that we expect the existence of a secondary runway to have an effect on mean LTI on the primary is because it accepts a substantial number of aircraft that would otherwise have landed on the primary. Hence this should be the motivating criterion in assigning the three runway configurations that are exceptions to either Group I or Group II.

Table 3.5 lists the actual hourly rate at which aircraft land on the primary and secondary runways for each data set. It also lists the AAR (Airport Arrival Rate) for each runway in the configuration in use. Some points become quite clear from this table. When the secondary runway is 22L (with 27 as primary) it accommodates the most aircraft—17 per hour on one occasion and 15 per hour on another. When 4L or 22R are the secondaries, they accommodate fewer, 12 per hour and 7.8 per hour respectively.

Let us consider the two runway configurations 4R/15Rc4L and 33R/L. Both of these have an AAR of 34/10, but the secondary runway is not used nearly as much for the latter combination as the former. When 4R/15Rc4L is used, we have an actual average landing rate on the secondary of 7.3 aircraft/hour, very close to the lesser of those for 22L/R. On the other hand, the actual average landing rate for 33R is only 3 aircraft/hour. This is most likely because 33R is so short that very few aircraft can land on it. Hence, in terms of actual use (i.e. the amount of traffic handled by the secondary runway), 4R/15Rc4L is more like to a standard two runway configuration, and 33L/R like to a single runway

Date	Runway Config.	Primary		Secondary	
		Observed	AAR	Observed	AAR
9-13-87	4R	36.3	34	-	-
"	15R	33.6	30	-	-
10-11-87	4R	35	34	-	-
5-13-88	22L/R	34	34	7.7	10
"	27/22L	32	34	15	26
5-18-88	4R	37.7	34	-	-
2-17-89	33L/R	37.8	34	3	10
"	4R/L	24	34	12	24
3-30-89	4R	25.6	34	-	-
5-31-89	4R/15Rc4L	28	34	7.3	10
"	15R	30	30	-	-
"	4R	34.7	34	-	-
6-09-89	4R	34	34	-	-
6-15-89	4R	36.7	34	-	-
6-28-89	27/22L	24.3	34	17	26

Table 3.5: Average Hourly Landing Rate on Runways for Each Data Set

configuration.⁷ Hence, we shall consider Group I to comprise all single runway landing configurations, as well as 33L/R. Group II shall comprise all VFR multiple runway landing configurations except 33L/R, but it shall include 4R/15Rc4L.

Grouping	# of Obs.	Mean	Std. Dev.
I	662	98.3	31.3
II	283	107.5	31.7

Table 3.6: Mean and Standard Deviation of LTIs by Runway Configuration Grouping

We wish to test the null hypothesis that runway configuration grouping has no effect on the mean LTI. We have presented arguments for and against this hypothesis, and a statistical test would help us resolve whether or not the simple model of (3.3) is viable. Table 3.6 gives the mean and standard deviation for LTIs in groups I and II. While the standard deviations of the two types of LTIs are roughly equal, their means differ. When two runways are used, the mean LTI is about 9 seconds higher than when only one is used. The question is whether this is a statistically significant difference or not.

To determine whether the difference in mean LTIs is statistically significant, we use the Difference of Means Test (see Appendix B) on the data in Table 3.6. We have that the test statistic $Z = 4.41$, and $\Phi(Z) \sim 1.0$. Thus the probability of a more extreme result given the null hypothesis is essentially 0, and we reject the null hypothesis. This indicates that runway configuration has a significant effect on the mean LTI.

It is interesting to note the direction of the difference between mean group I and group II LTIs. On average, the group I LTI is less than the group II LTI. The explanation is that when two runways are used traffic that would otherwise have landed on the primary now lands on the secondary. Referring again to Table 3.5 we can see that gross aircraft arrival rates would indicate the same. The landing rate on the primary is always at or above the AAR when only a single runway, or 33L/R, is used. The sole exception is 4R on 3-30-89. But that was a day of very bad weather, when many aircraft were being held on

⁷15R/L which is the analogue of 33R/L for aircraft from the opposite direction would also be considered a single runway configuration, were there data on it.

the ground at their origins. On the other hand, when two runways were in use (excepting 33L/R), the landing rate at the primary was at most equal to the AAR. But typically it was less. Thus the data on landing rates on the runway corroborate the test result: when a secondary runway is in use, there are fewer landings on the primary, and the mean LTI on the primary is less than otherwise. We should emphasize, however, that the total number of landings for the airport is higher when two runways are in use than when only one is in use.

3.4.4 The Sequence Category and the Mean LTI

We see, therefore, that the mean LTI is not a constant, independent of all factors: the runway configuration has a significant effect on it. The other major factor that can have a significant effect on the mean LTI is the sequence category. In this section, we explore whether indeed it does so or not.

The rules of ATC require different separations on final approach depending on the weight-classes of the lead and trail aircraft, i.e. the sequence category of the LTI. The separation is based on two considerations: collision avoidance and wake-vortex. There is a required minimum separation of 2.5 nm between any two aircraft to avoid collisions. Also, to prevent an aircraft from overturning due to wake-vortex induced turbulence, an extra separation is required whenever the lead aircraft is a Heavy (regardless of the type of trail aircraft) and whenever the lead is a Large and the trail a Small. These minimum separations, as we discussed in Chapter 2, are mandatory in IFR but not in VFR. Below is the required separation “matrix” reproduced from Chapter 2.

		Weight-class of Trail Aircraft		
		Heavy	Large	Small
Weight Class of Lead Aircraft	Heavy	4	5	6
	Large	2.5	2.5	4
	Small	2.5	2.5	2.5

Table 3.7: FAA Separation Standards (in nm) for IFR Weather

There are 9 possible sequence categories. The bulk of LTIs fall into three of them: H/L, L/H and L/L. We saw this in Section 3.2.3 in a typical example with landings on 33L/R on 2-17-89. We see it again in Table 3.8, which gives the means and standard deviations of LTIs by sequence category. The number of aircraft in sequence categories with at least one Large aircraft is much greater than in the rest. In particular, L/L LTIs are more numerous than all the others combined.

Sequence Category (with min. sep.)	Number of Observations	Mean	Standard Deviation
H/S (6 nm)	14	130	33
H/L (5 nm)	118	121	28
H/H (4 nm)	30	102	23
L/S (4 nm)	55	90	29
L/H (2.5 nm)	119	94	29
L/L (2.5 nm)	530	99	31
S/H (2.5 nm)	12	92	46
S/L (2.5 nm)	62	93	32
S/S (2.5 nm)	5	101	42

Table 3.8: Mean and Standard Deviations of LTIs by Sequence Category

The mean LTIs range from 90 seconds, when a Large is followed by a Small, to about 130 seconds, when a Heavy is followed by a Small. The H/S interval requires the longest distance separation (6 miles) and we are not surprised to see it have the biggest LTI. On the other hand, L/S LTIs require a minimum separation of 4 miles and it is somewhat surprising to see them have the smallest mean.⁸ Nevertheless, it does appear that those sequence categories requiring only the base separation of 2.5 miles—L/H, L/L, S/L, S/H, and S/S—are in general smaller on average than those that require additional distance separations.

We could test the null hypothesis that all sequence categories have identical mean LTIs. The standard method for such a multiple (more than 2) sample problem is the F-test from Analysis of Variance (ANOVA). That test requires, however, that all the samples

⁸As we shall discuss, this can be explained by the special way in which propeller aircraft merge into the final approach stream.

have equal variances. We could test to see if the homoscedasticity assumption is satisfied or not. But if we do not reject a hypothesis of equal variances, we may not really be sure that the variances were equal. Even if we do reject such a null hypothesis, we might have to use a nonparametric test (*e.g.* Kruskal-Wallis) instead of ANOVA. That in turn would have the drawback of not being exactly a test for equal means.

Since our purpose at present is just to determine if sequence category has any effect on the mean LTI or not, we propose a simpler test. Rather than considering all sequence categories (SC) separately, we construct a particularly simple division of them into two groups, the “Base” and the “Big”. The Base group includes those sequence-categories requiring no additional wake-vortex separations. The sequence categories in the Base group require a 2.5 n.m. separation between aircraft in normal IFR (or VFR) conditions. The Big group contains those requiring a larger separation (L/S, H/H, H/L and H/S). This division is practical and it is conceptually simple, similar to the division of runway configurations into groups I and II. If the sequence category at all affects the mean LTI we should notice it in such a bipolar division. We emphasize in this connexion that later we will explore further subdivision in terms of individual sequence categories.

We test the null hypothesis that Base and Big LTIs (*i.e.* LTIs from the Base SC and Big SC groupings respectively) have the same mean, against the alternate that Big LTIs have a higher mean. The test used is the Difference of Means Test presented earlier. Table 3.9 gives the mean and standard deviation for Base and Big LTIs.

SC Group	# of Obs.	Mean	Std. Dev.
Base	728	98.0	31.1
Big	217	111.4	31.5

Table 3.9: Mean and Standard Deviation of LTIs by Sequence Category Grouping

The mean Base LTI is 98 seconds, about 13 seconds less than the mean Big LTI. There are many more of the former LTIs (728 vs. 217) because of the predominance of Large aircraft in the traffic mix. Using the Difference of Means test on this data, we have $Z = 5.5$, and $\Phi(5.5) \sim 1$. Thus $P(\text{False Rejection}|H_0) \sim 0$ and we reject the null

hypothesis.

The result says that, on average, the LTIs corresponding to sequence categories requiring additional wake-vortex separations are larger than for those that do not require such extra separations. The indication seems to be that such additional separation is maintained even in VFR conditions, when it is not required. Of course, we have not yet tested for this explicitly, and shall do so later.

Hence the tests in this section have shown that two important variables—sequence category and runway configuration—indeed have a significant effect on the mean LTI, as we might have anticipated. In the next few subsections, we shall explore the effects of these two variables further. We wish to develop groupings of homogeneous LTIs (with respect to the means), in terms of these two variables. Before we do so, however, we shall examine the variance of the LTIs. Does it also differ by runway configuration, or by sequence category?

3.4.5 Does the LTI Variance Depend on Runway Configuration or Sequence Category?

We have thus far concentrated on the mean LTI because, as we have discussed, we shall build the calibration model with respect to the mean. However, it would also be interesting to explore whether LTIs have equal variances, across runway configurations or sequence categories. In this subsection we pursue this point.

Runway Configuration and the Variance in LTIs

The first question is whether the variance in LTIs depends at all on the runway configuration. One might argue, for instance, that certain runway configurations are easier for controllers to “manage” than others and that the variance of LTIs for that configuration would be less. Such a configuration might be 27/22L where all aircraft have long flight paths in the terminal area. The longer the flight paths, the easier it might be for controllers closely to monitor the intervals between aircraft. On the other hand, a landing configuration like 4R/L is to some degree “driven” by the arrivals at Providence, who get

priority for landing. Thus there may be greater variance in the LTIs for this configuration than for 27/22L.

In Table 3.4, we listed the mean and standard deviations of LTIs by data set. Although there are some extreme values, the bulk of the standard deviations hover around 30. On the other hand, we are prompted by the above considerations to test the null hypothesis of no relationship between the variance in LTIs and the runway configuration in use. There are parametric and nonparametric procedures for this. The parametric procedures are all very sensitive to departures from normality (see Miller (1986) [19]). Hence using them might require further testing for normality in each data set. The nonparametric tests themselves require assumptions of equal medians. Thus we may have to test for this in each data set. Either way, one could get a chain of tests, each imperfect and all dependent on one another.

To avoid both these types of tests (parametric and nonparametric), we use a robust test by Levene (*op. cit.*) for equal variances. Consider that we have I groups of data, with n_i elements in the i th group, $i = 1, 2, 3, \dots, I$. Let y_{ij} be the j th element in the i th group. We wish to test if σ_i^2 are equal for $i = 1, \dots, I$. The test by Levene is to treat the values $z_{ij} = (y_{ij} - \bar{y}_i)^2$, $j = 1, \dots, n_i$ as if they are independently, normally distributed under H_0 . The usual ANOVA procedure is then applied to the z_i 's to see if they have significantly different means or not. Clearly the z_{ij} do not satisfy the assumptions imposed on them. Within a given sample, they are not independent because of the common \bar{y}_i . The data are not normally distributed. Nevertheless, simulation studies have indicated that the test is quite robust in spite of these drawbacks.

Applying the Levene test to the standard deviations (actually the variances) presented in Table 3.4, we accept the null hypothesis of equal variances at the 5% level. The probability of false rejection, given that the null hypothesis is true, is computed to be 0.18. The result says that there is no evidence of the LTIs being more narrowly distributed about the mean for one runway configuration than for the other, the observed differences being quite likely due to random fluctuations alone. Thus it does not appear to be the case that some runway configurations LTIs are more tightly controlled than on others.

There was another issue that was raised when we first discussed Table 3.4. That was to examine the extreme values of the observed standard deviations. The 14 second standard deviation comes from a data set of 19 points. There appear to be no particularly small LTIs and no particularly large ones. All are between 80 and 130 seconds. Further, the mix of traffic includes sequence categories requiring wake vortex separations as well as those not requiring any. Hence this may just be a data sample with relatively smooth operations and precise “control” over LTIs.

The large standard deviation (39 seconds) comes from a data set with both a disproportionate number of small LTIs (less than 70 seconds) and a disproportionate number of very large ones (above 160 seconds). Overall, in the data, 15% of LTIs are below 70 seconds and 6% above 160 seconds. In this data set, of 27 LTIs, 9 (or 33%) are below 70 seconds, and 3 (or 11%) above 160 seconds. One (speculative) explanation is that the 3 large LTIs which were roughly evenly spaced through the data set, could have been due to mistakes on the part of ATC or the pilots.⁹ This may have then induced shorter than average LTIs to keep a high throughput at the runway.

Sequence Categories and Variances in LTIs

Another question with regard to the variances in LTIs is whether they differ according to the sequence category. Again, the argument may be based on differences in the accuracy of air traffic control. It could be that some types of LTIs (such as H/S, for example) may be more closely controlled because of the need for minimum separations. On the other hand, all aircraft separations may be getting the same attention (from pilots and controllers) and variations in LTIs purely random.

Table 3.8 gave the mean and standard deviation of LTIs by sequence category. Again the deviations were typically around 30, though two (S/H and S/S) were 42 and 46 respectively. In these cases, however, there were very few data points in the samples. The Levene test applied to the data does not indicate a significant difference in the variances by sequence category. The conditional probability of false rejection of the null hypothesis

⁹Typically, such errors are from imprecise metering into the TRACON from the center, setting up gaps in the sequence of planes.

of no difference in variances is 0.25.

As we mentioned previously, we are assuming a priori that the standard deviation and the 25th-percentile of the LTI distribution are homogeneous for those circumstances for which the means are equal. Indeed, with respect to the standard deviation, this subsection has shown that it is essentially the same across all runway configurations and sequence categories. From the next subsection, we continue our focus on the mean LTI, and try to find groupings of sequence categories and runway configurations within which it is homogeneous.

3.4.6 Homogeneous Groupings of Mean LTIs—Preliminaries and Data Review

In Section 3.4.3, we considered whether the runway configuration, or the sequence category, had a significant effect on the mean LTI. We discovered a marked effect in each case. In that analysis we had constructed gross divisions of LTIs by these two variables. Thus we had grouped sequence categories in two ways—Big (requiring additional wake-vortex separation) and Base (not requiring such separation)—and in two ways by runway configuration—those with landings on a single runway and those with landings on two runways.¹⁰

While this bipolar division was sufficient to answer the question of whether the sequence category variable was important or not, it need not necessarily represent the finest grain grouping of LTIs by sequence category. In particular, there are some aspects of ATC operations which are not captured by such a coarse model. For example, the sequence categories comprising the Big group of LTIs require differing minimum separations.

Our aim, thus, is to “fine tune” these groupings of LTIs by sequence category and runway configuration. But this fine tuning is not without a tradeoff. A very finely grained model, in the extreme case an “atomic” one, which treats each individual sequence category and runway configuration as different from the next, may be more accurate, but it is not

¹⁰33L/R is considered to be a single landing runway configuration, since the secondary is very short and hardly accepts landings.

compact or even practical. (Of course, such a model might also be less accurate than the original.)

Date-Runway	Big SCs				Base SCs				
	H/H	H/L	H/S	L/S	L/H	L/L	S/H	S/L	S/S
9-13-87 4R	104	125	132	93	92	102	58	93	-
9-13-87 15R	96	110	-	-	96	101	-	91	-
10-11-87 4R	102	137	-	113	96	97	61	103	56
5-13-88 22L/R	108	116	181	72	103	101	151	144	-
5-13-88 27/22L	138	117	90	-	137	111	-	-	-
5-18-88 4R	117	115	148	79	78	96	60	78	-
2-17-89 33L/R	92	124	-	110	97	84	-	106	104
2-17-89 4R/L	105	113	-	-	121	113	-	-	-
3-30-89 4R	133	120	-	120	104	100	-	108	-
5-31-89 4R/15Rc4L	95	106	127	85	85	101	110	112	83
5-31-89 15R	-	76	-	-	72	102	-	106	169
5-31-89 4R	-	-	-	76	99	95	-	68	-
6-09-89 4R	68	125	121	87	84	105	90	107	94
6-15-89 4R	80	116	106	83	80	97	-	86	-
6-28-89 27/22L	100	134	-	-	98	113	120	99	-

Table 3.10: Mean LTI by Sequence Category and Data Set

It is impractical to have a very finely grained model partly because we have insufficient data for each sequence category and each runway configuration. Table 3.10 lists the mean LTI by sequence category for each data set. Although we have a fairly extensive data collection, we notice that they are not evenly distributed across sequence categories or runway configurations. For instance, L/L LTIs are very well represented, as we had noticed when discussing individual data sets in Section 3.2.3. But S/H, H/S and S/S LTIs are totally absent in some data sets. Indeed, the small aircraft are included mostly on days with single runways operations. This is not surprising, since they are usually diverted to the secondary runway on other days. Further, as we see from Table 3.11, even when we do have aircraft from sequence categories involving Small or Heavy aircraft, they are few in number: there is never more than 1 S/S LTI in any data set, given that there are any at all.

Date-Runway	Big SCs				Base SCs				
	H/H	H/L	H/S	L/S	L/H	L/L	S/H	S/L	S/S
9-13-87 4R	4	12	2	7	16	47	1	7	-
9-13-87 15R	1	4	-	-	3	9	-	2	-
10-11-87 4R	6	17	-	7	12	38	3	6	1
5-13-88 22L/R	1	12	1	1	11	79	1	1	-
5-13-88 27/22L	1	2	1	-	2	12	-	-	-
5-18-88 4R	3	7	4	14	7	38	2	16	-
2-17-89 33L/R	2	12	-	2	14	58	-	3	1
2-17-89 4R/L	1	4	-	-	2	17	-	-	-
3-30-89 4R	2	10	-	3	11	27	-	2	-
5-31-89 4R/15Rc4L	2	9	2	3	11	35	1	5	1
5-31-89 15R	-	2	-	-	2	8	-	1	1
5-31-89 4R	-	-	-	2	1	22	-	2	-
6-09-89 4R	2	11	1	8	8	47	3	7	1
6-15-89 4R	2	7	3	8	9	57	-	9	-
6-28-89 27/22L	3	9	-	-	8	29	1	1	-

Table 3.11: Number of LTIs by Sequence Category and Data Set

There is also some nonuniformity in the way that data are distributed across different runway configurations. For example, 4R is the single landing runway configuration for which we have a preponderance of data. Its counterpart 22L, which ought to be similar from an ATC point of view, is not present as all. Further, we have data on 33L/R for about 3 hours, but none on 15R/L.

Thus it is not practical to consider sequence categories and runway configurations at atomic levels, and to compute mean LTIs for every combination of these two atomic levels. Nor indeed is it desirable, at an "aesthetic" level, even if it were practical. We want to develop a model that is reasonably compact, without too many variables. We wish to consider groupings of sequence categories, and runway configurations, which are broad enough that we have enough data to make reasonably stable estimates of mean LTIs; at the same time, the groupings should be different enough that they reflect "to the first order" differences in the effects of ATC procedures on the mean LTI.

There is, however, a technical problem with grouping the LTIs in terms of the indi-

vidual effects of two variables: runway configuration and sequence category. We wish to examine the first order effects of these variables and do not presume any interaction. The technical problem is easily seen. For example, suppose we wished to construct groupings of runway configurations with homogeneous mean LTIs. We cannot just compute the overall mean LTI for each runway configuration, and then decide which ones can be grouped together. The problem with this approach is that the overall mean LTI for some runway configuration also depends on the mix (in terms of sequence categories) of the aircraft that landed there: the mean for this configuration may be larger simply because there are many more Big LTIs. Suppose, on the other hand, that we wished to compute groupings of mean LTIs by sequence categories. Here we cannot just compute the mean for each sequence category (across all runway configurations) and then group the means. The corresponding problem is that the mean L/L LTI (for example) may well differ by the runway configuration used.

These two instances amply illustrate the need to develop a method for studying the effect of one variable (either runway configuration or sequence category) while holding the other constant. Thus we must fix the sequence category effects when studying the runway configuration, and vice-versa. We shall present a statistical technique which aims to handle this issue. We begin in the next subsection (3.4.7) with developing sequence category groupings and in the following one (Section 3.4.8) with groupings of runway configurations.

3.4.7 Grouping Sequence Categories with Homogeneous Means

Are Base LTIs Consistently Smaller in Mean than Big LTIs?

We have already determined that sequence category has an effect on the mean LTI. The test we used, however, was relatively simple. It considered all Base LTIs (regardless of runway configuration) and likewise all Big LTIs and compared their means. However, as we have discussed, this does not adjust for the possible effect of runway configuration on the mean LTI. Hence, we wish to compare the mean Base and Big LTIs for all data sets and see if a consistent difference is observed across them. For the present, we are

not particularly concerned whether the mean Base LTI in a given data set is significantly different from the mean Base LTI in some other data set (we shall explore that later). Rather, whether the mean Base LTI from a particular data set is significantly different from the mean Big LTI in the same data set. Moreover, is this true for all data sets.

In any one data set, we could use a Difference of Means test on the Base and Big LTIs, to produce a value z_i . But what we need to do is to combine the results from each data set, so as to be able to make an aggregate statement on the overall data. We develop a test statistic

$$C = \frac{\sum_i^N z_i}{\sqrt{N}} \quad (3.4)$$

which has a $N(0, 1)$ distribution, and we wish to test the null hypothesis that $C = 0$.¹¹

Date-Runway (V=VFR, I=IFR)	Mean LTI		z value
	Base	Big	
9-13-87 4R (I)	94	113	-2.46
9-13-87 15R (I)	98	107	-1.09
10-11-87 4R (I)	95	124	-3.99
5-13-88 22L/R (V)	102	116	-1.43
5-13-88 27/22L (V)	114	115	-0.11
5-18-88 4R (I)	88	101	-1.59
2-17-89 33L/R (V)	88	118	-4.2
2-17-89 4R/L (V)	114	112	0.28
3-30-89 4R (I)	107	122	-2.19
5-31-89 4R/15Rc4L (I)	99	103	-0.32
5-31-89 15R (I)	102	76	2.02
5-31-89 4R (I)	92	75	1.56
6-09-89 4R (I)	99	106	-0.40
6-15-89 4R (I)	95	98	-0.44
6-28-89 27/22L (V)	110	125	-1.54
			Sum = -16.26

Table 3.12: Mean Base and Big LTIs with Normalized Differences

Table 3.12 lists the mean Base and Big LTIs, for each data set. We see that the mean Base LTI is smaller than the mean Big LTI for 12 out of 15 cases, and often considerably so. The normalized differences, the z_i from the C-test described earlier, are also quite high

¹¹For details, see the section on the C-test in Appendix B.

(negative). The value of $\sum z_i$ is -16.26 . Dividing this by $\sqrt{15}$ since we use 15 data sets, we find that $C = -4.2$. Using the normal tables, the p-value of the test is 1.3×10^{-5} . Thus we emphatically reject the null hypothesis that the mean Base and Big LTIs are equal for all data sets.

This result is very interesting because it points out the fact that regardless of runway configurations and, equivalently, the weather conditions, the mean Big LTI is considerably larger than the mean Base LTI. This is true in IFR, when the extra wake-vortex separations, leading to larger LTIs, are required, as well as in VFR conditions, when they are not. Thus, planners for an advanced ATC system cannot blithely assume that the extra wake-vortex separations will not be followed in VFR because they are not mandatory. Rather, as the data indicate, there is a natural caution among pilots to observe these rules, despite good weather and being able to see the preceding aircraft.

Decomposition into Homogeneous Groups

The Base LTIs all come from sequence categories requiring the standard 2.5 nm separation. Hence they can be treated as a homogeneous grouping. On the other hand, it is easy to see why the Big LTIs may not constitute a homogeneous group by sequence category. Of the constituent sequence categories (H/H, H/L, H/S, and L/S) only two (H/H and L/S) have the same required minimum distance separations (4 nm.). The other two, H/L and H/S require minimum separations of 5 and 6 nm respectively. Thus the combination of larger distance separation requirements and slower speeds of trail aircraft may make the average H/L or H/S LTI bigger than the average H/H LTI.

We wish to test whether the mean LTI for the various sequence categories comprising Big LTIs are significantly different from each other. We wish to make some aggregate statement for all data sets, about any such difference. The problem is somewhat difficult to solve for the data from Table 3.10 because of the missing values in many "cells". For example, there were no H/S LTIs observed on 3-30-89. If we had all cells complete, we could perform a two-way analysis of variance with each cell containing the mean LTI for that cell.

The model would be given as

$$LTI = \mu + \alpha_i + \beta_j + \epsilon \quad (3.5)$$

Thus there is an overall mean (μ), with additive effects for each row (α_i) and each column (β_j) and an error term (ϵ). There is no interaction term. We would test the null hypothesis that there there was no column effect on the mean LTI. This model would require normality and homoscedasticity, both of which may not be serious constraints. The main problem is, however, that some cells are empty.

One way around it could be to restrict ourselves to the data sets where each row has at least one sample value in each sequence category. The problem with this is that we do not use a lot of information. Thus we would (for Big LTIs) be ignoring 5 data sets with no H/S values. For the sake of the calibration exercise there may be a simpler approach.

We devise a test procedure that does not require data in each sequence category for the Big LTIs. We begin with a simple bipartite division of the sequence categories. We consider the two groupings identified above:

- A: Requiring a minimum separation of 4 miles.
- B: Requiring a minimum separation of more than 4 miles.

The reason for this bipartite division is to test in a simple way whether significant difference exists in an extreme case. If no such difference is found then the existing grouping may be considered homogeneous—as for Base LTIs; if significant difference is found, then we can consider smaller groupings of sequence categories. The null hypothesis is that the two groups have equal means. The alternate is that group B LTIs have the larger mean. We might expect a larger mean for these LTIs, because they require a larger separation distance.

Using these groupings of sequence categories, we can use the C -test described earlier. The null hypothesis is that group A LTIs have the same mean as group B LTIs, for all the data sets. Since we are agglomerating sequence categories, we shall be using more information than we would have with the two-way ANOVA. For this reason, and for the

absence of parametric assumptions about the data, one can argue that this method is more appealing than the standard ANOVA.

Date-Runway	Mean LTI		z value
	A (4nm sep)	B (> 4 nm sep)	
9-13-87 4R	97	126	-3.35
9-13-87 15R	96	110	*
10-11-87 4R	108	137	-2.46
5-13-88 22L/R	90	121	-1.46
5-13-88 27/22L	138	108	*
5-18-88 4R	86	127	-3.88
2-17-89 33L/R	101	124	-1.94
2-17-89 4R/L	105	113	*
3-30-89 4R	125	120	0.33
5-31-89 4R/15Rc4L	89	109	-2.02
5-31-89 15R	-	76	-
6-09-89 4R	83	124	-3.35
6-15-89 4R	82	113	-2.53
6-28-89 27/22L	101	126	-1.45
			Sum = -22.11

Table 3.13: Mean Group A and B Big LTIs with Normalized Differences

Table 3.13 lists the mean LTI for groups A and B above, for each data set. We see that the mean group B LTI is larger than the mean group A LTI for 9 out of 10 cases, and considerably so each time. The normalized differences,¹² the z_i from the C-test described earlier, are also quite high (negative). The overall result is not in itself unexpected, because we have already seen, when comparing Base and Big LTIs, the effect of extra separations on LTIs. The value of $\sum z_i$ is -22.11 . Dividing this by $\sqrt{10}$ since we use 10 data sets, we find that $C = -7$. Using the normal tables, the p-value of the test is ~ 0 . Thus we emphatically reject the null hypothesis that the mean group A and group B LTIs are equal.

Again, it is appropriate to emphasize on the implications of our results so far: we have found that the requirement of extra separation due to wake vortex effects has a significant effect on the mean LTI. Not only is this effect seen when comparing Base LTIs and Big

¹²The * denotes insufficient sample size to compute a variance, i.e. one sample value.

LTIs, but also at a more detailed level within Big LTIs. Further, the effect has been tested across many data sets, spanning both VFR and IFR conditions. *Thus aircraft maintain larger separations not just in IFR, when they are required to do so, but also in VFR when no such requirement exists.* This is a strong indication of a natural cautiousness among pilots and has important implications for an advanced ATC system which would try to stipulate and maintain certain LTIs.

Since we have evidence of the extra distance separation making a difference to the mean LTIs, we treat the groups H/L and H/S separately. But what of H/H and L/S? Both of these require a minimum separation of 4 nm, and we would consider them to be homogeneous, as we did the Base sequence categories. There is a difference here, however. The Large group is very broad, comprising both propeller and jet aircraft. The two types of aircraft merge differently into the final stream for landing. Typically, the propeller aircraft (Small and Large) enter the terminal area at lower altitudes and fly at slower speeds within the terminal area. They are merged into the final landing stream at around the outer marker. Usually, according to many air traffic controllers,¹³ the propeller planes are merged at sharp angles and relatively high speeds into the gaps. This is especially so if the previous aircraft was relatively smaller.

Any difference induced in this landing process would be especially pronounced if we compare the difference within L/S LTIs depending on whether the lead aircraft is a propeller plane (denoted L_p) or it is a jet (denoted L_j). Thus we wish to test if the mean L_j/S and mean L_p/S LTIs are equal for all data sets.

Table 3.14 gives the mean and number of L_p/S and L_j/S LTIs by data set. As we see, there are very few data points, and the normality approximation used for the C -test would really not be justified. Hence we use a nonparametric procedure: the Wilcoxon rank-sum procedure for blocked treatments (see Lehmann (1978) [17], and Appendix B). This essentially represents a method which, like our C -test, combines the results from many experiments. It examines differences in distribution between L_j/S and L_p/S LTIs and does not strictly test for difference in means. But we are using the nonparametric

¹³ And as the author also corroborates, based on many hours of personal observation.

Date	Lj/S LTIs		Lp/S LTIs	
	# Obs.	Mean	# Obs.	Mean
9-13-87 4R	7	93	-	-
10-11-87 4R	7	113	-	-
5-13-88 22L/R	1	72	-	-
5-18-88 4R	11	83	3	65
2-17-89 33L/R	1	128	1	92
3-30-89 4R	2	136	1	87
5-31-89 4R/15Rc4L	2	86	1	84
5-31-89 15R	-	-	-	-
5-31-89 4R	-	-	2	76
6-09-89 4R	7	92	1	49
6-15-89 4R	6	82	2	85
6-28-89 27/22L	7	93	-	-

Table 3.14: Lj/S and Lp/S LTIs by Data Set

tests to see if there is reason to treat these two types of LTIs differently. So the procedure is sufficient for our purposes.

We use the Wilcoxon procedure for blocked treatments to test the null hypothesis that Lj/S and Lp/S LTIs are identical. We denote the the Lp/S LTis to be the treatments. The value of W_s , the test statistic, is computed to be 3.1. Further, $E(W_s) = 4.5$ and $Var(W_s) = 0.69$. Using the normal approximation for $\frac{W_s - E(W_s)}{Std.Dev(W_s)}$, the p-value of this test is $\Phi\left(\frac{3.1-4.5}{0.69}\right) = \Phi(-2.1) = 0.01$. Hence we can reject the null hypothesis of homogeneity. The Lp/S LTIs appear to be significantly smaller than the Lj/S LTIs. The likely explanation is that propeller aircraft are merged differently, being brought in at sharp angles and relatively high speeds into the main stream of aircraft.

Given this fact, Lp/S and Lj/S LTIs ought not to be treated homogeneously. Further, we shall also consider the H/H sequence category separately. Hence we do not treat the Big LTIs as a homogeneous group: the sequence categories H/H, H/L, H/S, and Lj/S and Lp/S being treated separately. The reason is that some of the sequence categories require and increased distance separation, and some reflect different ATC operations procedures for merging into the final stream of landing aircraft.

We have thus derived the grouping of sequence categories: Base, H/H, H/L, H/S, Lj/S and Lp/S. The next step is to examine whether the type of runway configuration used affects the mean LTI within any one of these sequence categories. That is the aim of the next subsection.

3.4.8 The Effects of Runway Configuration on Mean LTIs

In this subsection, our aim is to see whether the mean LTI for each of the sequence categories (or for the Base grouping) is affected by the type of runway configuration in use. When first studying the effects of runway configurations, we had considered them in two groups depending on the existence of a secondary runway. Group I is comprised of runway configurations with a single landing runway, and also the 33L/R configuration. Group II comprises runway configurations where there is a secondary runway, including 4R/15Rc4L, but excepting 33L/R.

We have concluded then that the effect of a secondary runway is to lighten the load of the primary by accepting some of the traffic that would otherwise land on it. Thus the overall mean LTI on the primary is significantly larger when a secondary runway is in use than when one is not. Now we wish to investigate more closely whether such an effect is present for each of the homogeneous sequence category groupings that we have developed.

The homogeneous groups of sequence categories that we have are: Base (not requiring any additional wake-vortex separation), H/H, H/L, H/S, Lj/S and Lp/S. The question is next whether (and how) these vary with the two groupings of runway configurations. Consider a particular grouping of LTIs given above, for example H/H. Does the H/H LTI have the same mean when a single runway is used for landings as when two runways are used for landings? Likewise for all other groupings of LTIs.

Of course, we need to explain why the aggregate differences in mean LTIs by runway configuration ought (or ought not) to be measured at the level of individual sequence categories or their groupings. Essentially, the answer would lie once again with traffic. The Base separations, based on collision avoidance, are the smallest, and are most likely to reflect the effects of traffic. On the other hand, the Big LTIs tend to be larger than the

Base ones, because of the need to maintain additional wake vortex separation. Thus these are less likely than the Base LTIs to be significantly affected by fluctuations in traffic. For example, a Base LTI of 90 seconds (for example) in heavy traffic may increase to 100 seconds if there is sparser traffic. On the other hand, the H/L LTI at 120 seconds, continues to remain above the level “dictated” by the sparser traffic, and may be unlikely to change.

Hence we examine the null hypotheses of no difference in the mean LTI between group I and group II runway configurations, for the Base case and for each sequence category of the Big case. We do not study the Lp/S case, because there was only one data point with a secondary runway in use. We can study these null hypotheses rather easily using difference of means tests, for which we have sufficient data. The results are given in the following table.

LTI Type	Group I Mean (# Obs)	Group II Mean (# Obs)	P Value
H/H	100.8 (22)	105.5 (8)	0.68
H/L	122.7 (81)	118.6 (31)	0.24
H/S	130 (9)	129 (5)	0.48
Lj/S	94.7 (41)	81.2 (3)	0.05
Base	94.4 (499)	105.8 (229)	~1

Table 3.15: Results of Difference of Means Tests for LTI Types: Single and Multiple Runways in Use

Here we see that the explanation we initially offered is borne out by the data. Those LTIs that have relatively large means (H/H, H/L and H/S) are not significantly affected by the change in runway configuration. The Lj/S LTI is affected, but opposite to how we would have expected it. But there are so few data points when two runways are used, that the result is not really meaningful. Finally, for the Base LTI, we see a marked and statistically significant difference of 11 seconds between the two types of runway configurations. Thus it does appear that those LTIs (essentially the Base case) that would be sensitive to traffic have different mean values depending on the runway configuration grouping. The others do not. To repeat, when there are two runways in use, the mean

Base LTI increases on the primary runway, but the other types of LTIs do not change significantly in mean.

3.4.9 Implication of Results for Throughput

Let us consider the implication of these results. We saw first that the mean Base LTI *on the primary* is higher when there are two runways used than when there is one. Further, two runways are typically used in VFR weather, and a single runway is used in IFR weather.¹⁴ Hence the data indicate that the mean Base LTI on the primary is lower in IFR weather than it is in VFR weather. In other words, the primary runway has a higher throughput in IFR than it does in VFR. Of course, we must emphasize, the airport as a whole has a larger throughput when two runways are used than when one is used, as can be easily seen from Table 3.5.

The reason for the higher mean Base LTI in VFR than in IFR, is that the primary runway possibly does not see enough traffic in VFR, and it is relatively unstressed. We consider Table 3.5 once again. The AAR is an estimate of the capacity of the runway. However, it appears to be a conservative measure: the average landing rate is always above it on the primary when a single runway is used. Only then does it appear that the airport is capacitated, or at least closer to it than otherwise. Hence, when two runways are used the primary is underused, relative to its own use in IFR. But if one were to schedule arrivals to the airport in terms of the maximum VFR capacity, there would be severe congestion in IFR. There would only be one runway to handle all the aircraft scheduled in VFR, and the runway and the terminal area would be overloaded. It remains to be seen whether, if there is increased pressure on the primary in VFR, it would have a higher landing rate than in IFR.

Thus the effects of scheduling aircraft with the IFR runway configuration as a bottleneck (eminently reasonable) leads to the lower mean LTI on the primary in IFR. The interesting question is whether the conventional wisdom in terms of throughput holds for

¹⁴There is an exception: Group I contains arrivals for 33L/R which is an IFR combination. But the results for difference in mean LTI by runway configuration groupings remain valid even if we ignore data from 33L/R.

airports (such as Denver or Atlanta) where there are parallel operations in IFR. There, according to the conventional wisdom, one would expect to see the mean LTI in IFR be larger than its corresponding VFR value.

3.4.10 The Calibration Model for LTIs

At the outset, we stated that the aim of the statistical exercise was to determine which factors significantly affect the mean LTI and to quantify these effects. We have discovered that the standard deviation in LTIs does not vary significantly with sequence category or runway configuration, but the mean does. The analyses so far provide the framework for a model for LTIs, based on the calibration data.

We have identified the following groupings based on sequence categories: the Base LTI, H/H, H/L, H/S, Lj/S and Lp/S. Further, we have divided runway configurations into two groups:

I When a single landing runway, or 33L/R, is in use.

II When two landing runways are used, excepting 33L/R.

We note there are three runway configurations that have not been observed in the data collection: 15R/L, 22L alone for landings, and 4R for landings and takeoffs. The last of these two is extremely rare, and is found in the worst weather prior to shutting down the airport. We do not consider it in our analysis. The first, 15R/L, is analogous to 33L/R in terms of the way the secondary runway is (barely) used. Hence we will consider it to be a configuration of Group I. Similarly, 22L alone is similar to 4R alone, and we consider it also to be in Group I.

We observed that the mean Base LTI differed significantly according to the runway configuration grouping, but none of the others did. Thus the variables in the LTI model, groupings of LTIs with homogeneous means (also to be referred to as LTI types), are:

1. Base-I: Base LTIs from Group I of runway configurations.
2. Base-II: Base LTIs from Group II of runway configurations.
3. H/H LTIs.

4. H/L LTIs.
5. H/S LTIs.
6. Lj/S LTIs.
7. Lp/S LTIs.

We are interested in three statistics about LTIs: the mean, the standard deviation and the 25th percentile. The mean, as we have discussed, is a measure of throughput, and the standard deviation a measure of how "tightly controlled" the LTIs are. These two statistics provide information about the LTI distribution as it currently is. But one aim of the LTI model is also to help us plan for what the LTIs should be like in the future, under an advanced Air Traffic Control system.

A major aim of advanced ATC is to increase the throughput of the airport. The mean LTI, a proxy for the throughput, thus has to be lower in the future than it is now. At the same, time, however, one cannot set a target that is impractical. The impracticality of a target is determined by the users and operators, and final arbiters, of the ATC system: the pilots and air traffic controllers. Let us assume, for the purposes of the following discussion, that any target LTI (the DLTI, or Desired Landing time Interval) can be met with precision by the computers operating the system. There are lower physical limits on the DLTI: minimum distance separations have to be met, and aircraft cannot be brought in at very higher speeds, or they will not be able to land. But there may be values of the DLTI that are above the physical lower bounds, but which are still unacceptably low for controllers and/or pilots.

The fundamental question then becomes, what value of the DLTI is low enough to increase throughput substantially, and yet high enough so as not to worry controllers and pilots. This is where the empirical distribution of LTIs can be of use. The 25th percentile of the LTI distribution (for each grouping of LTIs, as given above) is a viable candidate for the DLTI. Currently, the controllers and pilots go below this value 25% of the time. Hence they must not find it intolerably low. At the same time, it is considerably smaller than the current mean LTI. Clearly, there are other values, such as the 30th or 35th percentiles, that could also be used. We recognize, thus, that the 25th percentile is not

the only candidate for the DLTI, but propose it for its intuitive appeal—it is the median of the lower half of the distribution. We must emphasize that the 25th percentile serves purely an illustrative purpose: in no way is the development of the calibration model dependent upon it.

Statistic	Base-1	Base-II	H/H	H/L	H/S	Lj/S	Lp/S
Mean	94	106	102	121	130	94	75
Std. Dev	30	32	23	28	33	30	21
25-perc	73	83	86	104	113	76	63

Table 3.16: Parameter Estimates for Calibration Model

Table 3.16 gives the unbiased estimates for each of the three statistics for the homogeneous groupings of LTIs. The Mean LTI we have studied in detail. We see that the Lp/S is the smallest on average—75 seconds. The explanation is that propeller planes are merged in sharply and fast into the main stream of landing aircraft. The Lj/S has the same mean (94 seconds) as the Base-I LTI. This is somewhat surprising considering that the Lj/S (like the Lp/S) has a required separation of 4 nm. But once again, the explanation lies in the way the Small (propeller) aircraft merge for landing. The H/* LTIs are all on average larger than the Base-I or L/S LTIs. The explanation is that they require larger minimum distance separations. Further, the mean H/H LTI (102 seconds) is less than the mean H/L LTI (121 seconds) which is in turn smaller than the mean H/S LTI (130 seconds). This is in the order of their (increasing) minimum distance separations: 4, 5 and 6 miles respectively. Finally, we see that the Base-II LTI is 12 seconds higher than the Base I LTI (105 seconds vs. 94 seconds), because traffic is much lighter on the primary when a secondary runway is used.

The standard deviations do not vary much across the groupings. The majority are between 28 and 33 seconds. H/H and Lp/S are smaller (23 and 21 seconds respectively) but the difference is not statistically significant, as we discussed in Section 3.4.5. Finally

the 25th percentile estimates are smaller than the respective means (typically by about 20 seconds), as we would expect. They also are related to one another in magnitude in much the same way as the means are.

We have made unbiased estimates for each of the statistics of interest. Assume we have n data points, x_1, \dots, x_n . The estimate for the mean is the sample average (\bar{x}). The estimate for the standard deviation is, likewise, the sample standard deviation:

$$\sigma(x) = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (3.6)$$

For the 25-th percentile, we linearly interpolate between the i th and $i + 1$ th order statistics, where the i th order statistic is the one just below the 0.25 quantile. Thus, $\frac{i-0.5}{n} \leq 0.25 \leq \frac{i+0.5}{n}$. In the event of a normal distribution, the estimates for the mean and standard deviation are the maximum likelihood ones as well. Interestingly, for the 25th percentile, the normal approximation for the 25th percentile (0.67 standard deviations less than the mean) is relatively close to the estimates we have made, typically within 1 second. For the H/S LTIs, however, with relatively few sample points, the difference is over 6 seconds.

It is desirable to verify the accuracy of all these estimates, because of what we can construe about the entire approach we have used in model development.

The procedure used to get these groupings of sequence categories was based on identifying salient influential factors from ATC operations, and testing for their effects. The procedure is by no means the only one that could have been used, and these groupings not necessarily the only ones allowed by the data. Our aim is to get a “first-order” approximation of the effects of runway configuration and sequence category on the mean LTI. We were driven by the desire to have a simple model, yet one that captures important numerical, and conceptual, differences among LTIs. There were many groupings we could have tried: for example H/L and H/S may well not be significantly different in mean. By the same token, existing groupings may even be split up: the data could permit breaking the Base group into subgroups. But not only can this continuous splitting into smaller and smaller groups produce diminishing returns, in terms of accuracy, but it vitiates the intuitive appeal of a compact model. Thus we consider our calibration model to be an

attempt to capture the important effects of sequence category and runway configuration on the mean LTI.

As this discussion indicates, the calibration model is not without qualifications. From the very start, we have had to make assumptions (e.g. when translating middle marker crossing times to crossing times at the runway). Our goal is that the model, and the estimates of the statistics, be “robust”: able to be replicated by a subsequent study, and not a function of the data or approach we have taken. So we need a validation procedure. We wish to develop new estimates for the statistics, but on completely different data. For example, we would compare the mean, standard deviation and 25th percentile for the Base-I LTIs from the calibration model with their counterparts from the validation data. We would do this for every sequence category/runway configuration variable in the model (as listed above).

A validation exercise that results in estimates for the statistics above that are not significantly different from their counterparts in the calibration model will reassure us of the accuracy of the calibration procedure. It will not confirm that it is the only correct model. Rather, that the numerical estimates developed from this approach are corroborated by new data. On the other hand, estimates that are significantly different will, at the very least, call for an explanation. They may even, depending on the disparities, provide sufficient cause to reassess the approach we have used. The next (and final) section (3.5) of this chapter, will describe the validation of the LTI model.

3.5 Validating the LTI Model

3.5.1 Summary of Validation Data

The final stage of the LTI model is validation. We have developed a model, identifying the important variables, and computed estimates for the statistics of interest. Now the goal is to see these results corroborated by completely different data. We wish to see how well the validation data fit the calibration model. By validation, we mean the following: we will compute estimates for the statistics of interest, based on the new data this time.

Then we will compare the estimates from the old and new data, examining whether any differences are significant or not. We will do so for each estimate of each statistic for each variable of interest. We will aggregate these various comparisons to produce an overall statement on the proximity of the old and new estimates.

Hence we performed a new data collection in early 1990. This data collection, for reasons of ease and speed, was done entirely by hand, by the author. Again, the intentions were the same as before: we focussed on the heavy traffic period in the evenings (roughly from 3:00 p.m. - 8:00 p.m. The data was collected from an observation room in the Control Tower of Logan Airport. The observer would notice the time that each aircraft crossed the runway threshold and, using a radio tuned to the frequency used by controllers and pilots on final approach, would ascertain the identity of the aircraft. Thus the weight class of the aircraft could be determined, if the aircraft's identity were not easily distinguishable visually. The intervals between successive crossings of the runway constitute the LTIs.

We tried to gather data for as wide a range of weather conditions as possible, and did not intentionally concentrate on any one runway configuration. On one of the days, 3-30-90, the weather was almost as bad as that on the worst day in the calibration data, 3-30-89.¹⁵ There was snow mixed with rain, making for very poor visibility on this day. On the other hand, on a number of days (all those with 27/22L, and with 33L/R) the visibility was greater than 10 miles and the ceiling higher than 5,000 feet.

We study, as before, landings only on the primary runway. We collected 8 data sets from March 19, 1990 to May 4, 1990. The number of LTIs observed is 686, compared to 945 for calibration. The data collection spanned 24 hours, of which 8.5 hours were in IFR and the rest in VFR. Of the 13.5 hours in VFR, however, 4 were with 33L/R in use. Thus there were 12.5 hours of data with Group I runway configurations, and 11.5 with Group II configurations.

Table 3.17 gives the estimates of the three statistics of interest for each LTI sequence category/runway configuration type. A qualitative comparison of the estimates between calibration and validation shows a relatively close similarity. For example, for the mean

¹⁵The very same day of the previous year!

Statistic	Data	Base-1	Base-2	H/H	H/L	H/S	Lj/S	Lp/S
# of Obs.	Cal.	499	229	30	118	14	44	11
	Val.	327	235	22	89	1	10	2
Mean	Cal.	94	106	102	121	130	94	75
	Val.	93	103	106	128	-	101	-
Std. Dev.	Cal.	30	32	23	28	33	30	21
	Val.	26	31	29	26	-	22	-
25th Perc.	Cal.	73	83	86	104	113	76	63
	Val.	75	83	88	110	-	85	-

Table 3.17: Calibration and Validation LTIs by Sequence Category/Runway Configuration

LTI, the maximum difference for any of these categories is 7 seconds (for H/L LTIs and Lj/S ones). For the Standard deviation, the maximum difference is 8 seconds (for the Lj/S LTIs), and for the 25th percentile, the maximum difference is 9 seconds (for Lj/S LTIs). Of course, we note that we have relatively few Lj/S observations, so the the variance of the estimate values would be high anyway. Typically, however, differences are on the order of two to three seconds. Hence the calibration and validation estimates do not appear to diverge sharply. We shall explore this question more formally shortly.

One striking aspect of the data, is the paucity of Small aircraft: Lj/S has very few, and the sequence categories H/S and Lp/S have almost none. In fact, of the 686 aircraft in the validation data, only 13, or 2% were Small. In the calibration data, 74 out of 945, or 8% were Small. The great difference is probably because of the times of year in which the data were collected. The calibration data were mostly collected in the Summer: 7 out of 10 data sets were in the months of May-September. Small aircraft are more frequent in the Boston area in the Summer months, because of increased pleasure travel to northern New England.

Commencing in the next subsection, we perform rigorous statistical tests on each mean, standard deviation, and 25th percentile estimate, in each sequence category/runway configuration grouping (LTI type), for the calibration and validation data sets. We can

test the sample means of two data sets for equality using the difference of means test, two standard deviations using the Levene test, but we have not yet introduced any such test for equality in the 25th percentiles. We have developed a test for this particular purpose and present it in Appendix B. In brief, the test pools the two data samples that are being tested, and computes the overall 25th percentile of the pooled sample. Then the test examines whether the proportion of observations in any one of the samples that is above the pooled 25th-percentile is “very different” from 25% or not.

3.5.2 Comparing Statistical Estimates from Calibration and Validation

We now test the estimates for each of the calibration and validation statistics of the LTI, for all sequence category/runway configuration combinations. We begin with the mean LTI, and compare the calibration and validation estimates for each type of LTI. Then we shall make an aggregate statement of how the means compare in calibration and in validation. We shall repeat this procedure with the standard deviation and then with the 25th percentile. Finally, we combine the results for each of the statistics to make an overall statement of closeness of fit on all three dimensions.

The Mean of the LTIs

For a given type of LTI (e.g. Base-I) we compare the calibration and validation estimates of mean using a Difference of Means test. The null hypothesis for each type of LTI is that there is no difference in calibration and validation means. Not only do we want to do this for every type of LTI, but we want to combine the results to obtain some overall estimate of how well the calibration and validation data compare in means. To do this, we combine the individual results from the difference of means tests to get the C -statistic that we have presented earlier (a sum of the normalized differences). Then we perform a test on this C value. Here the null hypothesis is that there is no difference in calibration and validation means for all the LTI types.

Table 3.18 gives the mean calibration and validation LTI for each sequence category and runway configuration combination. It also gives the normalized differences in the

LTI Type	Calibration	Validation	z_i	$1 - P$ -value
Base-I	94	93	0.59	0.72
Base-II	106	103	0.84	0.80
H/H	102	106	-0.58	0.71
H/L	121	128	-1.79	0.96
Lj/S	94	101	-0.86	0.79
				Sum = -1.8

Table 3.18: Means LTIs From Calibration and Validation Data

means, denoted z_i , consistent with previous use. We also give the $1 - P$ -values of the test ($\Phi(|z_i|)$). We do not test Lp/S and H/S LTIs because there were so few in the validation data, as we have discussed. We see that for each LTI type, except one (H/L), the means are not significantly different at all. Indeed they are remarkably close for the biggest group: Base-I LTIs. For the H/L LTI, the p-value of the test was 0.96. This is, however, still below the rejection level of 0.975 for a two-sided test. Combining the results, we get that $\sum z_i = -1.8$. Dividing by $\sqrt{5}$, since there are 5 comparisons, we have $C = -0.8$, and $\Phi(-0.8) = 0.21$. Hence we do not reject the null hypothesis that validation and calibration means are equal for all LTI types, the observable differences being explained in terms of sampling variation.

The Standard Deviation of the LTIs

The next statistic of interest is the standard deviation. Table 3.19 presents the standard deviation for calibration and validation data, for each type of LTI. We use the Levene's Test for the null hypothesis of equality in standard deviations. That test, as we discussed, involves an ANOVA (in its many sample version) on transformed data. For each particular type of LTI (e.g. Base-I) we have two sample standard deviations, from validation and from calibration. For the two sample case, Levene's test simplifies to a t -test on transformed data (see Miller (1986) [19]). Hence Table 3.19 also presents the t statistic (denoted t_i) for test on LTI type i , with the number of degrees of freedom.

Except in one case (Base-I LTIs) the standard deviations are not significantly different

LTI Type	Calibration	Validation	t_i (d.o.f.)	$1 - P$ -value
Base-I	30	26	1.96 (824)	0.975
Base-II	32	31	0.21 (462)	0.585
H/H	23	29	-1.22 (50)	0.885
H/L	28	26	0.85 (205)	0.815
Lj/S	30	22	0.80(52)	0.784
				Sum = 2.62

Table 3.19: Standard Deviations of LTIs From Calibration and Validation Data

in the validation and calibration data. For the Base-I LTIs, the raw difference in standard deviation is 4 seconds, but it is significant. The t -statistic is 1.96, and the P -value ($\text{Prob}(t \leq |t_i|)$) is 0.025, which is just at the 5% rejection level. However, we need to consider all the LTI types, and make an aggregate statement of the differences in standard deviations.

The question then is how to combine the results, since each individual test statistic has a t distribution and not a normal one. Let T denote the sum of the individual t statistics, i.e. $T = \sum t_i$. The distribution of T is obviously the convolution of individual t distributions, with their respective degrees of freedom. This convolution is not a commonly known distribution, but for the samples we have, we can use normal approximations.

It is a well known fact that the t distribution for a large number of degrees of freedom tends to a standard normal. Larsen and Marx (1981) [16] give a table comparing the tails of the normal and t distributions for various numbers of degrees of freedom (pp. 296). From that table, at 50 degrees of freedom (the lowest we have) the difference at the 95th percentile is on the order of the second decimal place. Hence, even for our smallest sample sizes, we can approximate $t_i = z_i$. Hence, T should be distributed as $N(0, \sqrt{5})$, since we have 5 data sets. So we have that $\sum t_i = 2.62$, hence $T(\sqrt{5})^{-1} = 1.17$ and the overall P -value is $1 - \Phi(1.17) = 0.12$. This is well within the standard 5% rejection region, and we cannot reject the null hypothesis that calibration and validation standard deviations are equal for all LTI types.

The 25th Percentile

The final statistic of interest is the 25th percentile. We have already discussed a method (the Percentile Test) to test the null hypothesis that the 25th percentiles of two distributions are equal. Now, as for the mean and standard deviation, we compare the 25th percentiles of each LTI type from the validation and calibration samples.

LTI Type	Cal. (# Obs)	Val. (# Obs)	# Val. below Pooled val.	1 - P-value
Base-I	73 (499)	75 (327)	76	0.88
Base-II	83(229)	83(235)	60	0.56
H/H	86(30)	88(22)	6	0.63
H/L	104(118)	110(89)	19	0.97
Lj/S	76(44)	85(10)	2	0.68
				Sum = -3.01

Table 3.20: 25th Percentiles of LTIs From Calibration and Validation Data

Table 3.20 gives the 25th percentiles for each LTI type from calibration and validation data respectively. It also provides the $1 - P$ -value for each test, as well as the number of the validation data points below the pooled 25th percentile. We note that this is usually quite close to one-quarter of the validation sample. Quantitatively, again, there is no significant difference between the calibration 25th percentiles and the validation ones. Except in one case: H/L LTIs. This is not surprising, because there was a similar difference in means of about the same significance level. The P-value is 0.03, which is barely below the 5% level of significance (for a two-sided test).

We also need, however, to compute an aggregate measure of how the closely the 25th percentiles compare in calibration and in validation for all the LTI types. By similar arguments to ones we have used, $\sum p_i$ (where $p_i = 1 - P$ -value) is distributed as $N(0, \sqrt{5})$. We have that $\sum p_i = 3.01$, and the overall P-value is given by $1 - \Phi(\frac{3.01}{\sqrt{5}}) = 1 - \Phi(1.35) = 0.09$. While this is the most extreme of the overall P-values we have computed so far, it is still quite well within the acceptance region of the test. So we cannot reject the null hypothesis that the 25th percentiles are equal for all LTI types in calibration and in

validation.

Thus for each of the statistics of importance—mean, standard deviation and 25th percentile—we found no significant differences between the calibration and validation data. This lack of evidence in itself is encouraging: observed differences in estimates are quite likely due just to random variation in sampling. Before we make a final assessment, however, we would like to have an overall statement about how well the validation data fit the calibration model. That is the aim of the next subsection.

3.5.3 Aggregating the Results: A Bootstrap Analysis

We wish to combine the results of the tests on the three statistics. Unfortunately, there is a problem: the tests are correlated. For example, if the means of two data sets are not significantly different, then it is more likely than otherwise that the 25th percentiles would not be different either. The two estimates are from the same data sets and are correlated. So similarity or difference in one would tend to imply the same in the other.

The difficult problem in any statistical test is to compute the distribution of the test statistic. When we combine the results of some given (N) statistical tests, we usually wish to compute a new test statistic (say S) based on the some function (say f) of the individual ones (say $s_i, i = 1, \dots, N$). Hence we would usually have that $S = f(s_1, \dots, s_N)$ and we want to compute the distribution of S . If these individual statistics s_i are i.i.d, the problem can be simple. In the C-test, for example, we just added the test statistics— $f(s_1, \dots, s_N) = \sum_1^N s_i$ —and the distribution of the new test statistic (S) was a convolution of distributions of the old ones (s_i). On the other hand, combining the results of dependent tests is far more involved. Unless one knows precisely the form of the dependence between the s_i in question, the distribution of S is difficult to compute.

One way to skirt the problem is to use simulation: more precisely, the Bootstrap method, due to Efron [10] (see Appendix B). It is a tool to help us compute the distribution of the test statistic for a combination of dependent tests. Consider the comparison of calibration and validation means. It consists of aggregating the results of individual difference of means tests between calibration model and validation data, for different LTI

types. There are five LTI types (Base-I, Base-II, H/H,H/L, Lj/S), which we index by $i, i = 1, \dots, 5$. Each of these difference of means tests produces a normal statistic z_i , and a probability value (actually $1 - P$ -value) of $\Phi(|z_i|)$. Let P_i^M be the probability value of the i -th test of difference of means, where $i = 1, \dots, 5$. Likewise, we have P_i^D representing the probability values for the tests of equality in standard deviation, and P_i^R the values for the tests of equality in 25th percentiles. A value of P_i^M (or P_i^D or P_i^R) greater than 0.975 connotes a rejection of the null hypothesis of equality in the appropriate statistic at a 5% level, using a two-sided test.

We have combined the results of the individual tests for a given statistic to test whether the validation data matched the calibration model for that statistic. Now we wish to combine the results for these individual statistics in turn to make a global statement. The null hypothesis H_0 that we wish to test is that the validation data equal the calibration model for all LTI types and for all three statistics of interest. A natural choice for the overall test statistic P^O is

$$P^O = \sum_{i=1}^5 (P_i^M + P_i^D + P_i^R). \quad (3.7)$$

Ideally, if we knew the distribution of P^O , we would compare the actual value P_A^O that we have computed, from the given calibration model and the collected validation data, with this distribution. A higher value of P^O connotes a greater difference between the validation and calibration. Hence the rejection region of P^O would be on the right of the distribution, and we would reject H_0 at a 5% level if $Prob(P_A^O \geq P^O) = 0.05$.

Unfortunately, we do not know the distribution of P^O . However, we can develop an estimate for it using the Bootstrap method. The method is as follows. Under the null hypothesis, the validation data is identical to the calibration data in these three measures. If we were to obtain two samples of data, chosen with replacement, from the calibration sample, we would have two data sets that satisfy the null hypothesis: they arise from the same distribution. Then we could perform the series of tests on estimates for the statistics for the sequence category/runway configuration combinations, and obtain a sample point on the distribution of P^O , following (3.7). If we were to repeat this many times, we would obtain a sample distribution P^{*O} of P^O . Then if one were to perform the actual tests

on the validation data, and obtain a statistic, P_A^O , we could compare its value with the sample distribution of P^O , as discussed.

Hence the bootstrap method that we use is:

1. For each LTI type i , we compute P_i^M, P_i^D and P_i^R by performing the appropriate tests on two bootstrap data sets of the same size as the original calibration sample (appropriately to capture the variation in the test statistic). These bootstrap data sets are obtained by sampling with replacement from the original calibration data.
2. Compute the value of $P^O = \sum_{i=1}^5 (P_i^M + P_i^D + P_i^R)$.
3. Repeat (1) and (2) 1000 times.

	Base-I	Base-II	H/H	H/L	Lj/S	Row sum
Test						
Mean	0.72	0.80	0.72	0.96	0.80	$P_i^M = 4.00$
Std. Dev.	0.97	0.58	0.89	0.80	0.79	$P_i^D = 4.03$
25th Perc.	0.89	0.56	0.63	0.97	0.68	$P_i^R = 3.73$
						$P_A^O = 11.76$

Table 3.21: Computations for P_i^M, P_i^D and P_i^R

So we compute the sample distribution P^{*O} and see where on it P_A^O lies. Table 3.21 gives the value of the P_i^M, P_i^D and P_i^R . The numbers are derived from the results presented in Tables (3.18-3.20). Figure 3.3 gives the distribution of the Bootstrap distribution P^{*O} .

The actual value $P_A^O = 11.76$ falls on the 16th percentile of P^{*O} . Thus the chance of a more extreme result given H_0 is 0.84. We are therefore far from rejecting the notion that the validation data do not equal the calibration model on all the three statistics of interest. Hence the validation data is consonant with the calibration model individually on the three statistics of interest, and in an aggregate sense with all of them.

In conclusion, we do not have great reason to fear for the incorrectness of the various techniques and approximations and assumptions we have had to make in developing the calibration model. Of course the caveat of any statistical model applies here: we do not know that the model is correct in any absolute sense; all we do know is that we developed

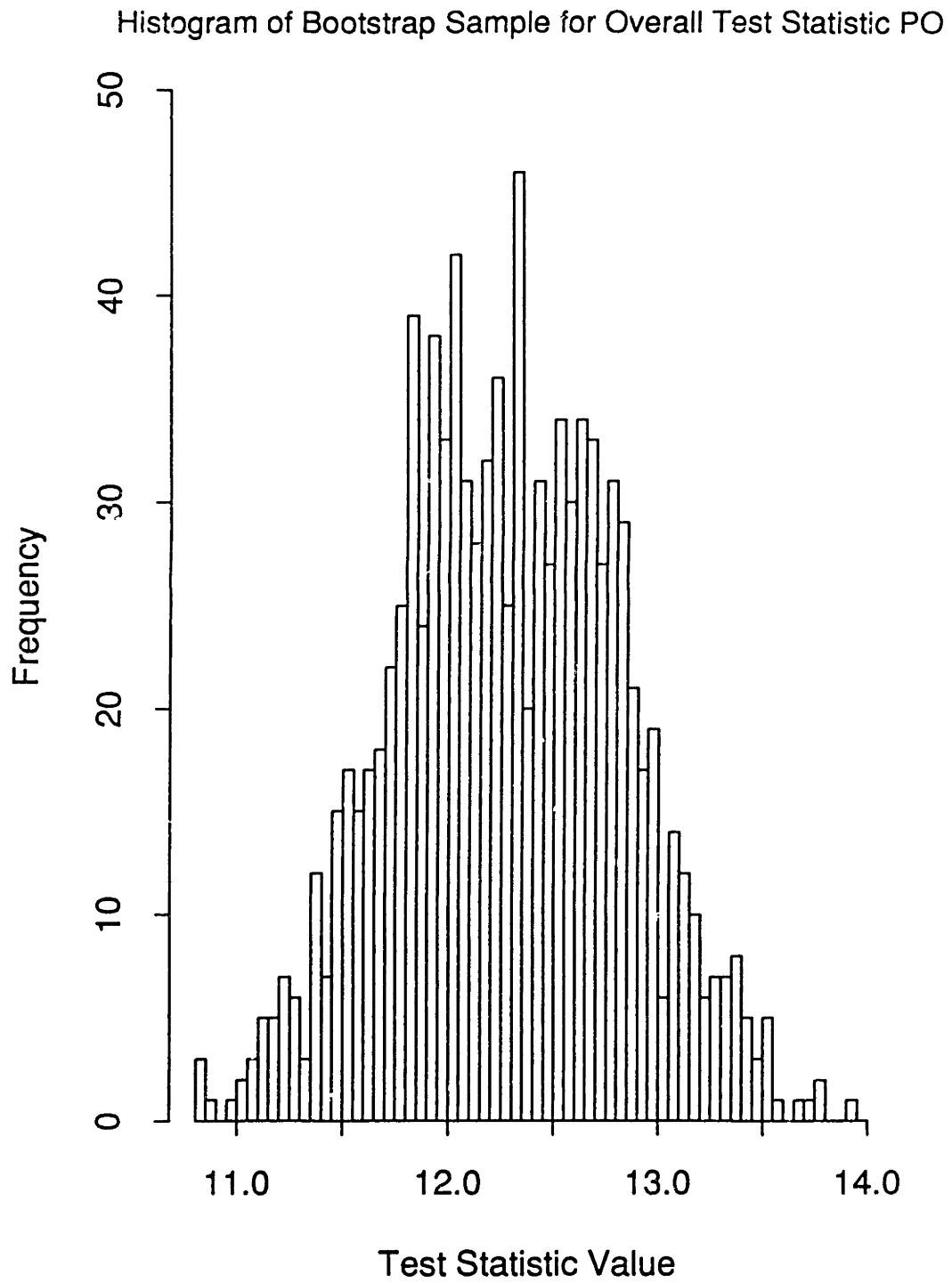


Figure 3.3: Bootstrap Distribution P^{*0}

a model that we felt reasonable, and fresh data did not contradict it in any significant way.

3.6 Summary and Discussion

This chapter has presented one of the two major topics of the thesis: developing a model for LTIs. At the outset, we stated that our aim was to identify those factors that affect the LTI and to quantify these effects. We focussed on the mean LTI because the standard deviation does not fluctuate significantly with respect to operating conditions. We adopted a two stage approach: calibrating the model on one collection of data and validating it on another. We did not find any significant difference between the validation data and the calibration model. Hence, our final model for LTIs is the calibration model presented in Table 3.16.

We have identified two factors that have a significant effect on the mean LTI. These are the runway configuration (which is correlated to the weather) and the sequence category of the LTI. We have two very interesting findings with respect to these variables. The first is that the mean LTI on the primary runway, when two of them are used for landings, is on average larger than it is when only one is used. The explanation lies in the distribution of traffic. As we showed, when two landing runways are used, roughly 8-10% of overall traffic is diverted to the secondary runway, and the reduced strain on the primary leads to longer average LTIs there. By contrast, in bad weather, much more traffic lands on the primary, rendering the LTIs there shorter on average. As we noted, of course, the IFR single runway operations constitute the bottleneck in airport operations and thus we cannot schedule more planes in VFR to take advantage of relatively underused capacity. Such a schedule would overwhelm the airport in IFR. Further, it remains to be seen whether if the primary were as stressed in VFR as in IFR how the mean VFR LTIs would relate to the IFR ones.

The second important finding of the study is that the extra wake vortex separations based on weight-class are followed in VFR even when they are not mandatory. We discover that the mean LTI for sequence categories not requiring additional separation is signifi-

cantly less than it is for those requiring it. Furthermore, this is true in general across all days. This affirmation of the safety consciousness of pilots has an important cautionary effect for the implementation of an advanced Air Traffic Control system. In a system of the future, one cannot expect pilots to ignore in VFR the extra wake-vortex separations that they maintain in IFR.

We partitioned LTIs into 7 homogeneous types: groupings of sequence categories and runway configurations. These are Base-I, Base-II, H/H, H/L, H/S, Lj/S and Lp/S. The Base grouping comprises those sequence categories requiring only the 2.5 nm IFR separation. The mean Base LTI differs according to whether one or two landing runways are in use. The other LTIs are considered separately, either because they require extra minimum distance separations (H/L and H/S), or because there are different air traffic control procedures used for them (H/H, Lj/S and Lp/S).

Naturally, this particular division of LTIs into types is not the only one that need be supported by the data. A given set of data may satisfy many, differing, hypotheses. Nevertheless, we have tried to arrive at groupings that are sufficiently few in number as not to make for an unwieldy model, yet sufficiently different as to capture the important effects of the operational factors at work. The point is that we have derived a model using a reasonable procedure. The model was based on statistical testing of hypotheses motivated by differences we perceived would arise from ATC operations.

We determined three statistics to be of importance and estimated them for each LTI type. The statistics are the mean, the standard deviation, and the 25-th percentile. The last of them is meant as a value that one might use in an advanced ATC system of the future. We compute estimates of these statistics from the calibration data for each LTI type.

Given the fact that many assumptions went into the LTI model, and, had we proceeded differently, that it could have taken a different shape with different estimates of the statistics of interest, it was important to test the model. So we developed a procedure to validate the model. This procedure involved, first, a completely new data collection, somewhat smaller than the first, but large enough to permit statistical testing. We found

that this new data lacked in Small aircraft, and had to test only 5 LTI types, omitting Lp/S and H/S. We tested the validation data with the calibration model for each LTI type for each statistic. We found no reason to reject the null hypothesis of equality for any the respective statistics. Further, we performed a bootstrap analysis to get an overall "judgement" on the issue. Again, we found that there was no reason to reject a null hypothesis of no difference on all counts: mean, standard deviation and 25th percentile. Indeed, the chance of a more extreme difference than that which we observed, given the null hypothesis, is 84%. Thus we are reassured of the accuracy of the calibration enterprise and model.

Of course, the whole modelling effort is a means to some particular ends. In this case, we have obtained some insights into current ATC operations. But the larger purpose of the exercise is to help with the ATC system of the future, one of the important components of which is the desired LTI at the runway. We have proposed the 25th percentile. It is a value which 25% of the LTIs, of the respective types, go below currently anyway. Hence, one can argue that if controllers and pilots are relatively comfortable with it now, they ought to be remain so with a new system that accurately "achieves" it.

But the new air traffic control system involves more than just technological advances helping controllers to do more efficiently that which they already do. Beyond this, such a system will also suggest changes in current operations, if such changes are for the better. One important problem, which is solved currently in the controllers' minds, is how to sequence aircraft in the terminal area.

Thus far we have considered the question: given aircraft A and B, landing in that order, on some particular runway configuration, what is the LTI between them (in terms of mean and standard deviation). Moreover, what might one want it to be (the 25th percentile). The next important question is: why do A and B (and all other aircraft) land in the order that they do (i.e. that we have observed)? Is it self-evident that this is the best, or only, order? Is there a better one?

The question we have introduced, is known as the Aircraft Sequencing Problem. Continuing the empirically motivated analysis of our thesis, we ask the following question:

given reduced mean LTIs (operating at the current 25th percentiles) in the ATC system of the future, what benefits can be realized by optimal sequencing. This is the second important topic in our thesis. We study this question, in the next two Chapters (4 and 5). In Chapter 4, we develop a model for aircraft flow in the terminal area, and an algorithm for sequencing. We apply these, in Chapter 5, to air traffic data at Boston, to see how much current operations can be bettered.

Chapter 4

The Aircraft Sequencing Problem

4.1 Preliminaries

In the introduction to this thesis, we said that we would investigate two types of inefficiencies that often occur in terminal area ATC operations: large gaps between landing aircraft, and suboptimal sequencing of aircraft for landing. In Chapter 3, we considered the first problem, developing a model for LTIs, and prescribing values that might be used in an advanced ATC system. In this chapter, we commence studying the second inefficiency, by examining the Aircraft Sequencing Problem (ASP).

Jet aircraft landing at Boston usually enter the terminal area at 3 entry fixes and are merged into one stream for landing on a single runway. The air traffic controller who does the sequencing for landings and merging for final approach is known as the final vector controller.¹ At present, the final vector controller does not know in advance when each aircraft will arrive: he learns of the arrival at approximately the point of its entry into the TRACON. The controller is then usually committed to land the aircraft without returning it to the entry fix for holding. Typically he does not alter the aircraft's position in its own landing stream (from its particular fix). The procedure is in general first-come-first-serve (FCFS), though it may in individual cases depart from this, at the discretion of the controller.²

¹In light traffic conditions there is usually only one.

²Dear (1976) distinguishes between First-Come-First-Serve at the System Entrance (FCFSSE) and the same with respect to expected arrival times at the runway (FCFSRW). FCFSSE and FCFSRW differ

A natural question, then, is whether one can sequence aircraft more efficiently than controllers do currently with FCFS. But first we need to explain why there might be any benefit at all to better sequencing. In our calibration model, we estimated that the Base-I LTI is 94 seconds on average, whereas the mean H/L LTI is 121 seconds. Consider then the following set of aircraft, in FCFS order (at the runway): L-L-H-L-L. Let us assume that we have perfect air traffic control, with the actual LTIs being exactly equal to their respective mean calibration values. In terms of LTIs, the 'cost' of this sequence is the sum of the individual mean LTIs, and is $94 + 94 + 121 + 94 = 403$ seconds. Clearly, however, by reversing the order of the two aircraft in an H/L LTI, we "save" $121 - 94 = 27$ seconds. So the best landing sequence for this set is L-L-L-L-H, at a cost of $94 \times 4 = 376$ seconds. Thus there are benefits to sequencing because different sequence categories have different mean LTIs associated with them.

But this very simple example also illustrates some of the problems involved with better sequencing. We took an aircraft that was third in line for the runway, and pushed it back to the end of the queue. In reality, however, this would involve delaying the given aircraft, by path stretching or speed control, speeding up those behind it, and then re-inserting the delayed aircraft back in the landing sequence. In congested airspace with overworked controllers this might be a problem. Moreover, we solved the problem given a set of 5 aircraft. Suppose the 6th to enter were also a Heavy, and were sequenced behind the first Heavy to create the order L-L-L-L-H-H, at cost $376 + 130 = 506$ seconds. (The H/H LTI is 130 seconds on average). This sequence in fact is inferior to the FCFS one: L-L-H-L-L-H, at cost $403 + 94 = 497$ seconds. Thus the entry of new aircraft into the system may make some 'locally' optimal sequences 'globally' suboptimal. Finally, as we have seen, Large aircraft form a disproportionately large fraction of the aircraft mix. Clearly, there is not much of a sequencing problem if we do not have a rich mix of aircraft. This, then, is another concern: are there enough aircraft of different weight classes, arriving close enough to each other, for there to be a sequencing problem.

because two aircraft that enter the terminal area at the same instant may take different times to transit the area and land at the runway. Dear notes that FCFSRW is more frequently used in practice and in Air Traffic Control research.

This discussion captures in a microcosm the problem with sequencing in the terminal area. There are, clearly, potential gains from better sequencing. The question, however, is whether in the congested terminal airspace, with restricted room for manoeuvring, and in a dynamic context, with new aircraft always arriving, though, perhaps, in a poor mix, we can realize many of these benefits. This is the problem that we shall study.

It is important to emphasize that the aim of this study is to assess what the benefits may be of “optimal” sequencing over FCFS. We shall devise an algorithm for sequencing and we shall use it on data sets that we have collected using the radar at Lincoln. For various data sets we have the actual entry time (and hence order) of aircraft into the terminal area as also their actual landing times (and order). Since we are focussing on sequencing within the terminal area only, we pose the following question: given the entry times of aircraft into the terminal area (and other information to be discussed later) what would be the optimal landing sequence? Then we contrast the optimal sequence with the actual one in which the aircraft landed, and compare the two sequences with respect to various objective functions to be specified. Although part of our study requires that we develop an algorithm for aircraft sequencing, the problem is not algorithmic *per se*. Rather, the question is to ascertain if benefit can accrue from better sequencing and, if so, how much.

Our approach differs from previous attempts in two important ways. The first is that ours is the most detailed attempt to incorporate the realistic operating constraints of the terminal area. This will be elucidated in subsequent discussion. The second difference is that we test our algorithm on extensive, detailed and accurate data for an airport. Hence we aim to improve on previous work by developing more accurate models for the problem, and testing them on realistic data.

We begin, in Section 4.2, with a review of the literature on the ASP. As we discuss, the problem has been studied only in a simplified operational environment. Previous researchers have not given as much attention as we will to both the dynamic nature of the problem and the restricted scope for reordering sequences. Specifying both the objective functions for the ASP and modelling the constraints is a major task. We will discuss in

Section 4.3-4.5, 3 models for the ASP and algorithms for solving them. These models represent successively more detailed representations of the operating environment of the TRACON.

We use Dynamic Programming algorithms to solve the three different models of the ASP. These algorithms are a modification of an approach by developed by Psaraftis (1978) [25, 26]. As we discuss, there is a “core” algorithm for the static ASP, with a fixed number of aircraft. We shall present this core algorithm, and show how we use it (as a subroutine) for the two dynamic versions of the ASP, with new aircraft regularly entering the terminal area. In Chapter 5, we discuss our implementation of the algorithms for Boston.

4.2 Literature Review

The ASP has been studied in the academic literature for some time. There are three major papers, by Dear (1976) [8], Psaraftis (1978) [25, 26], and Trivizas (1987) [35]. In each of these models, in the static problem, there are N aircraft which are already present on holding stacks outside the terminal area. Each aircraft can land at any time, and the problem is to find the sequence that minimizes (for example) throughput. There are dynamic versions of the problem studied by Dear, where the composition of the set of aircraft changes with time. But again, each aircraft, it is assumed, can be held in stacks.

Dear [8] examines the dynamic sequencing problem. He studies various objective functions for the sequencing problem (such as maximize throughput, minimize total delay, minimize maximum delay) and compares their performance with respect to FCFS. He also introduces the notion of Constrained Position Shifting(CPS). According to this, for reasons of practicality, an aircraft’s position in the recommended landing sequence can differ from its position in the initial arrival sequence only by an amount less than or equal to the Maximum Position Shift (MPS), a given parameter. Thus if an aircraft arrives in position 7 in the initial sequence, and $MPS=3$, it must be between positions 4 and 10 in the final landing sequence. Every time a new aircraft joins the stream (at say position N)

Dear assumes the optimal sequence is frozen prior to position $N - MPS$. He reoptimizes the remaining portion using complete enumeration. The notion of practicality imbued in the MPS is important. A landing sequence reoptimized on the arrival of a new aircraft may have this latest entrant assigned first place in the landing stream. This would be impossible to achieve in practical terms. The notion of equity is, of course, in preventing the possibility of a single aircraft being delayed excessively.

Psaraftis [25, 26] considers only the static case of the ASP, and develops a dynamic programming (DP) formulation (and solution technique) both with and without CPS constraints. He notes that this problem is quite similar in structure to the Travelling Salesman Problem and the Vehicle Routing Problem. Hence it is no surprise that the DP algorithm in general requires time that is exponential in the size of the problem (the number of aircraft under consideration). Psaraftis also discusses some characteristic properties of the optimal sequence, including “clustering” of aircraft by weight-class. For example, let the times between aircraft be directly proportional to the distance separation requirements between aircraft, and let the objective be to minimize the landing time of the last aircraft. An optimal sequence is to land the Small aircraft first, then the Large ones and finally the Heavy ones.

Trivizas (1987) studies the stage-state structure of the dynamic program, and develops, using parallel processors, fast algorithms to find the next best aircraft to land at each stage. Although the time for this search is made linear by Trivizas (instead of the usual polynomial complexity, as we shall discuss later) the storage costs go up tremendously. But as we suggest later, the size of this problem in normal circumstances is not so prohibitively large as to require necessarily the complex data structures and search techniques of Trivizas.

Outside the academic literature, there has also been some work on the ASP. Magill et. al. (1988) [18] refer to heuristics they use in designing an arrivals flow management system for Heathrow Airport, London. More recently, Thompson (1990) [34], considered the problem in the context of Boston. He randomly generates traffic patterns for three arrival fixes, with sequence category mixes representative of those found in Boston. He

uses clustered sequences, similar to those of Psaraftis, to study the effects of various formulations of the constraints. These formulations reflect, among other things, different rules restricting the amount of overtaking that can take place within a single stream and across different streams. These restrictions on overtaking are important practical considerations for any sequencing algorithm, and we shall ourselves refer to them later.

From the next section onwards, we shall present our development of the DP approach to the ASP. The Dynamic Programming algorithm presented by Psaraftis provides the algorithmic framework for our own method. But our models of the terminal airspace are much more complex. We do not always assume that holding stacks are in use: aircraft are usually delayed (or expedited) only by path length variation and speed control. Further, we study a dynamic problem, with new arrivals into the terminal area. Psaraftis' formulation lacked a time dimension. Moreover, we do not simulate the performance of our algorithms: we test them on real data for Boston. Working with such data requires, as will be clear, that detailed attention be given to control practices in the terminal area. The next section describes the models that we develop for the ASP.

4.3 The Static Model of the ASP

In this section, we formulate the static model of the ASP, and develop a dynamic programming algorithm (a modification of Psaraftis') to solve the problem. In the following two sections (4.4-4.5) we build on the static model, developing dynamic models, with fixed and variable time windows respectively. We shall also present modifications to the static algorithm to enable us to solve the dynamic problems. Before we do so, however, we consider some common aspects to all three models: the objective functions (Section 4.3.1), and the LTI constraints (Section 4.3.2).

4.3.1 Objectives for the ASP

What is the objective of the sequencing problem? This is a very difficult issue. Two frequently used measures of system performance are throughput and delay. Both have their

merits. Throughput is a simple and intuitive measure and relates readily to conventional measures of capacity, like the AAR. On the other hand, with better sequencing, although we mean to increase throughput, we do not wish to delay any aircraft unduly. This is important, moreover, because there are different costs of delaying aircraft with different numbers of passengers. Thus because we want to find sequences that do not penalize some individual aircraft severely, we model our objective in terms of delay.

By delay, we mean the time difference between an aircraft's preferred landing time and its actual landing time. When an aircraft enters the TRACON at a given fix, there is an earliest possible time at which it could land, assuming that there were no traffic in the terminal area. We shall term this landing time as the preferred landing time. Hence the objective of the sequencing problem, simply stated, will be to minimize the cost of delays in the terminal area.

The next step is to develop a quantitative measure for "cost of delays"? Here there are many choices. We can just use delays for each aircraft, or delays to the passengers in the aircraft, or, more tangibly, losses due to extra fuel spent. Hence an algorithm for the aircraft sequencing problem can seek to minimize any of the following objective functions:

1. The total time delay suffered by the aircraft in the set being considered. For each aircraft, we define its time delay to be the difference between its earliest possible landing time (preferred landing time) and its actual landing time.
2. The total time delay suffered by the passengers in the set of aircraft being considered.
3. The total monetary cost of delaying the set of aircraft being considered. Thus, for a particular type of aircraft A there may be a cost function $C_A(t)$ in terms of the delay t . This function would measure the dollar costs of fuel consumption.

We note that these objectives correspond to minimizing average aircraft delay, average passenger delay and average cost of delay respectively.

The first objective is intuitive and uncomplicated, but perhaps too much so. It does not recognize that 5 minutes of delay to a fully loaded B747 are more "costly" in some sense than the same amount of delay to a Cessna with one-hundredth as many passengers. Hence, the second objective is more realistic and factors in the passenger-minutes of delay per aircraft. We would have to specify the average number of passengers in aircraft of

each weight-class. We could reasonably approximate this by considering a Beechcraft 99 as a typical Small aircraft, an ATR-42 as a typical Large prop, a Boeing 737 as a typical Large jet, and a DC-10 as a typical Heavy one.

A third formulation could not only treat different aircraft differently, but could use a monetary cost of delay to reflect expenditure on fuel. The third objective function that we use does this. In it, we also consider the fact that the disutility function of delay is not necessarily linear. As Terrab (1990) [33] states, in his discussion on numerical cost functions for delay, there is often a risk-averseness among consumers that requires that larger costs be valued disproportionately higher than smaller ones. Hence, he uses a convex cost function for delay. Likewise, our monetary cost function for delay is convex. By this we do not mean to say that fuel expenses increase nonlinearly with time spent in the air. Rather, that there is a risk aversion on the part of airline companies to bearing higher expenses for terminal area delay, and hence we have a nonlinear valuation of such costs.

Terrab assumes cost functions such that $\Delta_J(i)$, the cost of delaying an aircraft of type i for a J th time period, given that it has been delayed for $J - 1$ time periods, is given by

$$\Delta_J(i) = C(i, J) - C(i, J - 1) = C(i, 1)(1 + \alpha)^{J-1} \quad (4.1)$$

where $C(i, J)$ is the cost of delaying a type i aircraft for J continuous time periods. The parameter α is the relative cost increase due to delaying a flight an additional time period (in our case one minute).

From this specification, we get a cost function

$$C(i, J) = C(i, 1) \frac{(1 + \alpha)^J - 1}{\alpha} \text{ for } \alpha \neq 0 \quad (4.2)$$

and we have that

$$C(i, J) = J * C(i, 1) \text{ if } \alpha = 0, \quad (4.3)$$

corresponding to a linear cost function. Since we use non-negative α our cost functions for delay are convex.

Terrab uses ground holding costs for the first hour of \$400, \$1,200 and \$2,000 for Small, Large and Heavy aircraft respectively. These costs are just based on operating

expenses and do not include costs to passengers. Terrab also finds that, using a value of $\alpha = 2$, which implies that ground holding costs double every hour, achieves a reasonably low overall holding cost and does not delay individual aircraft for unduly long periods. Finally, we consider air holding costs to be 1.6 times ground holding costs.³ We will consider using $\alpha = 3$ since air holding costs have a higher rate of growth (increasing disutility for delay) than ground holding costs. Further, we need to convert the hourly figures for $C(i, J)$ and α into figures for minutes, a time period more suited to planning in the near terminal area.

We note that from the ratio of 1.6 between air holding costs and ground holding costs, the values of the former for the first hour are \$640, \$1,920 and \$3,200 for Small, Large and Heavy aircraft respectively. Let us consider time periods of minutes. We wish to find values of $C(i, 1)$ and α such that

$$C(i, 60) = \begin{cases} \$640 & \text{if } i = \text{Small} \\ \$1,920 & \text{if } i = \text{Large} \\ \$3,200 & \text{if } i = \text{Heavy} \end{cases} \quad (4.4)$$

Terrab shows that when using minutes, we can compute α_{min} the relative increase in cost per minute from α_{hr} the relative cost increase per hour from

$$\alpha_{min} = (1 + \alpha_{hr})^{\frac{1}{60}} - 1. \quad (4.5)$$

Substituting $\alpha_{hr} = 3$, we have that $\alpha_{min} = 0.023$, or that the cost increases by 2.3% each minute. Hereafter we drop the subscript and refer to α_{min} as α . Finally we may compute $C(i, 1)$ from $C(i, 60)$ using (4.2) to get

$$C(i, 1) = \begin{cases} \$4.98 & \text{if } i = \text{Small} \\ \$14.96 & \text{if } i = \text{Large} \\ \$24.92 & \text{if } i = \text{Heavy} \end{cases} \quad (4.6)$$

4.3.2 The LTI Constraints

Having specified the objective functions for the ASP, we now move on to model the constraints. In this work, we aim to provide a much richer description of the ASP than

³Amedeo Odoni, personal communication.

previous models, because we want to capture more closely the practical issues that surround it. We develop three models of terminal area ATC, which successively constrain to higher degrees the controller's freedom to manoeuvre aircraft. We shall focus on each of these models separately, in later sections. First, however, we discuss the constraints common to all models: the LTIs.

The relationship of the LTI model to the sequencing study is that the former supplies the values of the LTI constraints used in the latter. The LTI constraint applies to aircraft at the threshold: the time separation between them is equal to, at the minimum, some Desired LTI (DLTI). We assume that the ATC system is advanced enough to control aircraft so that this DLTI can be achieved precisely. As we discuss, we shall compute the optimal sequences for different values of the DLTI. We shall use the mean LTIs from calibration, the 25th percentiles from calibration, and finally the LTIs that would result assuming standard velocities for aircraft and strict adherence to separation requirements.

We emphasize that in our model the LTI is a constraint on the minimum separation. Since we have an accurate model (to be discussed) of the aircraft's time-to-fly in the terminal area, sometimes an aircraft may be physically incapable of achieving a given LTI. This is likely in periods of light traffic when aircraft are entering the TRACON relatively infrequently. What might happen then is that a particular aircraft, even if speeded up through the TRACON as fast as possible, would still land further behind its predecessor than the LTI. In such a case, the algorithm would assume that the minimum physically possible LTI is the separation achieved by ATC.

4.3.3 A Model for the Static ASP

We begin, in this subsection, with the simplest model for the ASP. This is a static model. We assume that there are N aircraft to be sequenced, and that we know in advance the entry time of each aircraft into the terminal area. Typically, aircraft enter the Boston terminal area from three jet fixes and five propeller fixes. They are then merged in to land on a particular runway. Given the entry time of an aircraft into the TRACON, there is a time window within which it must land. Hence, associated with each aircraft is a time

window, with lower and upper bounds representing, respectively, the earliest and latest times at which it can land.

The earliest time at which an aircraft can land depends on its entry time, and the shortest time taken to fly from its particular entry fix to the runway. The lower bound of the time window is hence the earliest possible time at which it could land, if there were no traffic at all in the terminal area. This is the preferred landing time. The upper bound, for the static model is essentially infinite. This is because we assume that holding is permitted in this model. An aircraft can be held in a stack for up to one hour. Given the time frame of the sequencing problem, and the cost functions, which will make it very expensive to hold an aircraft for that long, the amount of time a plane can be held on stack is almost never a binding constraint.

The time that the aircraft takes to transit the terminal area is known as its time-to-fly. The time-to-fly for an aircraft is itself dependent on the entry fixes and on how much the aircraft can be delayed within the terminal area. Aircraft can be delayed in the terminal area by speed control and path stretching, as we discussed in Chapter 2. In Appendix C, we present the technical details for calculating the minimum and maximum times-to-fly for aircraft entering the various fixes for Boston and landing on 4R. Our study focusses on landings on 4R alone because the time-to-fly estimates that we have are only for that runway. Given the times-to-fly, one can easily calculate time windows for an aircraft. If t_{min} and t_{max} are the minimum and maximum times-to-fly for a given aircraft that enters the TRACON at time T , and that can be held at a stack for time H , then the time-windows for landing are $(T + t_{min}, T + H + t_{max})$. We partition the time-to-fly interval into the portion resulting from speed control and that from path length variation.

Our study is based on data in that it compares, with respect to a given objective function, the landing sequence suggested by the dynamic programming algorithms with that achieved in practice on the days in question. The data we need for such a study includes the entry time and entry fix of each aircraft into the terminal area (and from this the upper and lower bounds for its landing time), along with the actual landing times (or order) for the set of aircraft.

For the static model, then we consider a data set of N aircraft. We are given the time windows within which each aircraft must land. We then compute the optimal landing sequence (and assign landing times) so that the LTI constraints are satisfied and each aircraft lands within its time window. The optimal sequence is the one minimizing delays in the terminal area (for any one of the objective functions that we have proposed), for the N aircraft.

4.3.4 Algorithm to Solve the Static Model

In this section, we present the Dynamic Programming (DP) formulation for the static ASP with holding stacks (and time windows).

The dynamic programming method that we present is essentially the one used by Psaraftis (1978), with the modification that we associate time windows with each aircraft, within which it must land. The dynamic programming method takes advantage of the fact that the LTI depends only on the weight-classes of the lead and trail aircraft. We assume that in general there are a fixed number N of aircraft, comprised of p weight-classes. We subdivide the Large weight class according to whether the aircraft is a jet or a prop. Hence we have that $p = 4$ (H,Lp,Lj and S).

The stage of the DP is the number of aircraft landed so far. At a given stage, the state of the algorithm is defined by $(L, K) = (L, k_1, \dots, k_p)$, where k_j is the number of aircraft in category j that have landed so far, and the current aircraft landed is of category L. We define $K^{MAX} = (k_1^{MAX}, \dots, k_p^{MAX})$ to be the number of aircraft in each category for the given problem instance. At the final stage, all of these have been landed. Within each category $j \in \{1, \dots, p\}$, we assume an implicit ordering of aircraft in terms of increasing order of lower bounds of their time windows. The reasons for this will be clear later.

We assume that all the aircraft are being held in stacks, from which they are removed for landing. The earliest landing time, a_i , for each aircraft i , is how soon it could land given its present position (or holding stack). This would depend on the weight-class of the aircraft and the runway currently in use for landings. Since all aircraft are in stacks, the

latest landing time b_i for each aircraft i is for practical purposes without limit.⁴ Hence $b_i = \infty, \forall i \in \{1, \dots, N\}$. Thus corresponding to the set of aircraft are a set of time windows given by the variable $W = \{(a_1, b_1), \dots, (a_N, b_N)\}$. All the aircraft in a given holding stack will have the same lower bound for the time window. As we shall see, however, with holding in use, the time windows play a role only in the sequencing of the first few aircraft from the stacks. The general formulation with time windows is far more useful for the dynamic models which we study later.

The algorithm uses dynamic programming to find the optimal⁵ landing sequence. The method used is forward recursion. We do not proceed by stages, rather by states in lexicographic order. Given the structure of this problem, broken by categories instead of aircraft, this is an easier way. The problem with the stages approach is that, for a given stage, there are many states associated with it. And the number of states associated with a stage also varies with the stage. Thus we have to write algorithms for computing all possible states associated with a stage.

At each state $(L, K) = (L, k_1, \dots, k_p)$, we are sequencing an aircraft of category L such that $k_L > 0$ (i.e. there is some positive number of aircraft left to sequence). We wish to find the optimal sequence of aircraft from $(L_0, 0, \dots, 0)$, where L_0 represents a dummy zeroth aircraft, to the current state. We have already computed the optimal sequence to all of the predecessor states of (L, K) (from the DP principle of optimality). Hence we wish to find the predecessor state of (L, K) that gives the least cost sequence to (L, K) .

Bearing this brief verbal description in mind, we define the key variables of the DP as follows:

- L , the category of the aircraft currently being landed.
- $V_x(L, k_1, \dots, k_p)$ the optimal value of the sequence leading from the start of the sequencing problem ($k_1 = \dots = k_p = 0$) to the current state (L, k_1, \dots, k_p) . Z is a set of cost functions; cost functions have the identical form for all aircraft in a given category (weight-class).
- $t_{x,L}$ the LTI between aircraft of categories x and L . These are the constraints to the ASP.

⁴As we discussed earlier, an aircraft can be held in stacks for 1 hour, but this is a nonbinding constraint.

⁵As we will discuss, the algorithm is not optimal for the dynamic problem

- $T(L, k_1, \dots, k_p)$ is the time at which the aircraft from category L lands at the current state. This sequence has value $V_z(L, k_1, \dots, k_p)$.
- The decision variable x , the identity of the category of aircraft to land before the current one (from category L). We see that $x \in X = \{i : 1 \leq i \leq p, 0 < k_i \leq k_i^{MAX}\}$
- $S_i = \{1, \dots, k_i\}$ the set of aircraft which are in category $i, 1 \leq i \leq p$ which have not yet landed.
- $S_i^W = \{(a_i^1, b_i^1), \dots, (a_i^{k_i}, b_i^{k_i})\}$ the set of time windows for the aircraft in set S_i .
- F_i the "best" aircraft to choose from category i . F_i has time windows (a_{F_i}, b_{F_i}) . The choice will be based solely on the time windows, and will be discussed in detail later.
- $D(L, x, k_1, \dots, k_p)$ the time at which the aircraft from L would land if it were sequenced after the one from x , given the current state of the system.
- $C_z(t)$ the cost, given objective z of delaying an aircraft by t minutes; i.e. of landing this aircraft t minutes after its preferred landing time.

We begin our discussion of the algorithm by noting that if we choose x , the predecessor state was K' where

$$k'_i = \begin{cases} k_i - 1 & \text{if } x = i \\ k_i & \text{otherwise} \end{cases} \quad (4.7)$$

From the time window constraints, the aircraft F_x lands at its lower bound (even if LTI constraints were not binding), below the upper bound, or at ∞ , if the upper bound is violated. Hence

$$D(L, x, k_1, \dots, k_p) = \begin{cases} a_{F_L} & \text{if } T(x, k'_1, \dots, k'_p) + t_{x,L} < a_{F_L} \\ \infty & \text{if } T(x, k'_1, \dots, k'_p) + t_{x,L} > b_{F_L} \\ T(x, k'_1, \dots, k'_p) + t_{x,L} & \text{otherwise} \end{cases} \quad (4.8)$$

Following the DP principle of optimality, we wish to find the value of x that minimizes the cost of the sequence up to state (L, K) . This is given by

$$V_z(L, K) = \min_{x \in X} C_z(D(L, x, K) - a_{F_L}) + V_z(x, K') \quad (4.9)$$

In this equation, the first term on the right hand side is the cost of delaying the best aircraft from category L by the difference between its actual landing time and the lower

bound (or preferred landing time).⁶ The second term, $V_z(x, K')$ is the optimal value of all decisions taken from to $k_1 = \dots = k_P = 0$, the initial state, to the predecessor state K' .

If the best predecessor category to choose is x^* , we have that

$$T(L, K) = D(L, x^*, K) \quad (4.10)$$

and that

$$V_Z(L, K) = V_Z(x^*, K') + C_z(D(L, x^*, K) - a_{F_L}) \quad (4.11)$$

We also have as boundary conditions $V_Z(L, 0, \dots, 0) = 0$, for all cost functions Z and all choices of initial landing category $L \in 1, \dots, P$. This, since if $k_1 = \dots = k_P = 0$ we have not sequenced any aircraft.

Nearest-Neighbor Search for Best Aircraft to Land from a Given Category

We have as yet omitted to specify precisely how we sequence aircraft within a given category. That is, given a category L and S_L , the set of aircraft yet to be landed from that category, how we choose F_L , the best of these aircraft to land. The choice is based on the time windows S_L^W of the aircraft in S_L . Let the time windows in S_L^W be listed in non-decreasing order of their lower bounds: thus $a_i \leq a_j$ for any $i < j : i, j \in \{1, \dots, k_L\}$. Since, if holding is permitted, the upper bounds are essentially ∞ as mentioned earlier, the best choice is the aircraft with the smallest lower bound (breaking ties arbitrarily). This is easily seen. Consider two aircraft m and n from category x , with $a_m < a_n$. If $a_m, a_n \leq T(L, K') + t_{x,L}$ then $F_x = m$ or n will give the same landing time from (3). In any other event, the landing time for aircraft m will be less than that for n from (3). On the other hand, if the upper bounds are not ∞ , this "nearest-neighbor" technique is not necessarily optimal. It may not be best to choose the aircraft with the smallest lower bound. We shall return to this point later, when discussing the dynamic versions of the ASP.

As we noted earlier, all the aircraft in a given stack have the same lower bound for their time windows, and all aircraft have an upper bound on their time windows

⁶From the way that the time window constraints operate this delay is always non-negative: an aircraft cannot land before the lower bound of its time window.

of ∞ . This means that the time window concept is not really essential, except at the initial stages, to the static heavy traffic model. This is because after the stage when $T(L, K) > \max_{i \in \{1, \dots, N\}} a_i$, all aircraft landings will fall within their respective time windows. Nevertheless, this concept will be very useful in following sections where we study the light traffic model and also when we consider the dynamic version of the ASP. First, however, we must consider the complexity of the DP solution technique to the static model.

Complexity Requirements

We note that there are P categories of aircraft in all; let M be the maximum number of aircraft in any category. Then the arrays $V_Z(L, K)$, $D(L, x, K)$ and $T(L, K)$ have to be accessed $3P \prod_{i=1}^P (M) = O(PM^P)$ times. Within a given category, searching for the best aircraft is constant time since they are stored by increasing lower bound of the time windows. Thus both complexity and storage is $O(PM^P)$. This is polynomial in the number of aircraft in a category but exponential in the number of categories.

In practice, with holding, there are no more than about 18 aircraft at any given time, since the number of fixes is three and each holds no more than 6 aircraft.

4.4 DASP-1: A Dynamic Model with Fixed Time Windows

The static model is unrealistic in two respects. First, we do not know in advance the entry times of the aircraft that are going to arrive in any time period. In reality, though controllers may be familiar with arrival schedules, sequencing is very much a dynamic task. And the final vector controller knows of the arrival of an aircraft for sure only when it is handed over to him. This typically happens when it is at the boundary of the TRACON. Thus upon the entry of each new aircraft, the previous landing sequence may be suboptimal, and a new one have to be computed again.

The second unrealistic assumption in the static model is that of holding stacks. In

recent years, anticipatory ground holding has been used extensively (recall our discussion of flow management strategies in Chapter 1) as a cheaper and safer alternative to (air) holding stacks at the TRACON. Indeed, in our 10 calibration data sets, air holding was used only on one day, when there was very bad weather (3-30-89). Further, the aim of air-traffic controllers is to use holding stacks as a last resort. Thus aircraft are delayed as much as possible by path stretching and speed reduction, and sent to a holding stack only if they need to be delayed yet more. In Appendix C, we compute the maximal delays within the terminal area for aircraft entering from various fixes.

The DASP-1 model studies this more realistic scenario of TRACON operations. The dynamic problem is solved by reconsidering the landing sequence each time a new aircraft enters the terminal area. There is no look-ahead: the algorithm does not try to predict or even use some probabilistic information about the type of new aircraft to enter, or its entry time. Further, there are some restrictions in computing the new landing sequence. An aircraft cannot be pulled out of its place in the original sequence if it is “too close” to the runway, and its position is frozen. We shall later define where we decide to freeze the landing sequence. Moreover, the absence of holding stacks means that the time windows are much narrower than they used to be for the static problem.

We should note that the restricted time windows are a proxy for the MPS constraints. The MPS constraints forbid an aircraft to be shifted by more than a certain amount from its position in the FCFS sequence. The narrower time windows function in a similar way. Since an aircraft is constrained to land within its time window, it can only be advanced or delayed by a fixed amount, and the scope for departing from the FCFS sequence is limited. It is also important to stress that this model still does permit overtaking within streams and across them.

4.4.1 The DP Algorithm to Solve DASP-1

The aim in DASP-1 is again to find the best sequence (minimum delay) subject to the LTI constraints, and to each aircraft landing within its time windows. Naturally, if there is no feasible solution for the problem above, then holding stacks have to be used. Further,

whenever there is a new arrival, the lower bound of the time windows increase for some aircraft: they have been delayed so much that they can no longer land at the runway as early as they originally could have.

In general we solve the dynamic problem by re-solving the static case, upon each new entry, for those aircraft within the terminal area (excepting those who are too close to landing). Hence our approach to the dynamic problem is to hold a certain portion of the assigned landing sequence at each stage fixed and only reoptimize the rest. We accomplish this by specifying a lead time which is the minimum time prior to its assigned landing time that an aircraft must be notified by ATC. Thus an appropriately chosen lead time will ensure that all aircraft can be given their final assignments, say, 15 nm from the runway (or 5 minutes before landing)⁷ prior to landing.

Let us assume again that there are N aircraft to be sequenced. We solve the static problem for those within the terminal area at the beginning (assume the first five). Then with each successive entry, $i, 6 \leq i \leq N$, at time t_i , we re-solve the static sequence for those that have not yet landed by time t_i or are not too close to the runway. We note further that although the concept of this lead time is introduced in the context of making the Dynamic ASP tractable in real time, it is also a practical restraint. Like the MPS notion, it admits that the landing order of aircraft cannot be arbitrarily changed at any time.

In this problem, as we mentioned, holding stacks are not in use, and the aircraft time windows will not have upper bounds of ∞ . Further, the upper and lower bounds will in general not be the same for each aircraft from a given category, because, in general, they will be at different positions on their path to the runway. Next we discuss the modifications to the static ASP algorithm that are necessitated by this version of the dynamic problem. There are essentially two: how to choose the best aircraft from each category, and how to re-solve the dynamic problem each time a new aircraft enters the system. We assume that the N aircraft to be sequenced are fixed and located in the terminal area, or even in some larger region, as long as the time windows are known with some precision.

⁷This value is changed later, in DASP-2, on the suggestion of controllers.

As in the static algorithm, in DASP-1, we choose the best aircraft in any category the one having the smallest lower bound for the landing time of that category. In theory, because of the finite upper bound, this nearest neighbor method makes the algorithm a heuristic and not necessarily optimal. Consider two aircraft m and n from some given category, with $a_m < a_n$, but $b_m > b_n$. In other words, m has the smaller lower bound but the higher upper bound. The nearest neighbor method would choose to land m before n , but it is conceivable that at some later point, the upper bound b_n for n would not be satisfied when b_m , that for m , would have. Hence this current choice may lead to an infeasible one later. In practice, though, as we see in Chapter 5, this does not happen.

There is, however, another reason for which the algorithm for DASP-1 is a heuristic, and that is because it uses no look-ahead for the dynamic problem. Upon the entry of each new aircraft, we resequence those within the terminal area that are not too close to the runway. The re-optimization works as follows:

1. Given an initial set of N aircraft wishing to land, solve the ASP (choosing F_x as described above) producing an assigned landing sequence $Q = \{q_1, \dots, q_N\}$ with assigned landing times $T = \{\tau_1, \dots, \tau_N\}$.
2. Whenever a new aircraft enters the system, at time Θ say, partition the set Q into the permanent and changeable portions, Q_P and Q_C respectively. Therefore $Q_P = \{q_i : \tau_i - \Theta < D\}$ and $Q_C = \{q_i : \tau_i - \Theta \geq D\}$, where D is the leadtime. Re-solve the Static ASP for all aircraft in Q_C . The time windows of these aircraft are readjusted to reflect the fact that they have been delayed so much that they cannot land as early as was once thought. Let $M = \max_{q_i \in Q_P} \tau_i$; this is the last immutable landing time and let it be of category c_M . Then we change the lower bounds of the time windows for aircraft in Q_C as follows:

$$a'_i = \begin{cases} a_i + t_{c_M, c_i} & \text{if } a_i < M \\ a_i & \text{otherwise} \end{cases} \quad (4.12)$$

We note that our algorithm merely re-solves the problem with each new entering aircraft. There is no look-ahead. The algorithm does not try to anticipate what kind of aircraft the next entrant might be, and when it might enter. Even if there were limited probabilistic information used, the algorithm would, in theory, perform better. We also note that the dynamic algorithm is an extension that Psaraftis had also suggested in his paper [26].

4.5 The Dynamic Model with Shrinking Time Windows

The model DASP-1 proposed above is much more realistic than the static formulation. Nevertheless, according to some Air Traffic Controllers, it still does not capture fully the restrictions on sequencing in effect in today's airspace.

The main objection that controllers have to DASP-1 is that the margin for manoeuvring aircraft shrinks rapidly as they approach the runway. In the DASP-1, we assumed that the upper bound of the time window remained fixed as the aircraft moved closer to the runway.⁸ Physically, this means that—until the aircraft's position in the landing sequence gets frozen—the original maximum amount of delay is always usable. But these maximum amounts were computed assuming that speed was reduced early and the aircraft was given a particularly long route to fly at the outset (for Providence) or a long downwind leg (for SCUPP).

If we forego the opportunity to use a given amount of delay, we can no longer avail of it. Consider, for example, Figure C.3, which shows the landing path for SCUPP. Once we cross the first point where we could have reduced speed from 250 knots to 210 knots, we can no longer take advantage of that early speed reduction. Even with regard to path stretching, one cannot easily pull a plane out of the landing sequence at the last minute and send it out on a long sidetrip. Therefore Air Traffic Controllers have suggested that DASP-1 be altered to reflect the decreasing room for delay and manoeuvre (or, conversely, expedition) as an aircraft nears the runway. We term this changing upper and lower bound on the time window for landing as a “shrinking” time window.

DASP-2 is the most restrictive version of the ASP that we will study. In addition to the shrinking time window restriction, it also incorporates far more restrictive constraints on maximum delays in the terminal area, as we shall discuss shortly. In the next section, we will present algorithms for the ASP. We begin with the simple static model. The algorithm for this is the basis (major subroutine) for the dynamic algorithms, where the constraints are different.

It is important to note that in some sense all three models are idealized examples.

⁸The lower bound could, however, change.

There are many different modelling situations: one could have, for instance, a dynamic problem with holding in use, and so on. But these three examples capture three different situations which are interesting for various reasons. The static problem with holding does not occur often in practice. But it is a version of a situation which has been studied much in the aviation literature. Further, it serves as an important benchmark. It is the least restrictive model of airspace operations, and the one with most information. Thus it gives an upper bound (perhaps loose) on the gains from sequencing (or, equivalently, a lower bound on terminal area delays).

The differences between DASP-1 and DASP-2 are more than just technical: they represent two different conceptions of terminal area operations. The first is an approximation of what we think advanced ATC might be like. Active control (especially monitoring compliance with instructions) would be done mostly by computers and not human controllers. Further, though the terminal area would be restricted, radar and tracking techniques will be accurate. Thus an aircraft could be brought out of a landing stream and manoeuvred around the terminal area with some freedom. On the other hand, today, without advanced ATC, not all these options are feasible. Hence DASP-2 represents what flexibility controllers have today, and DASP-1 what they might have in the future. Both DASP-1 and DASP-2 assume, however, that the LTIs can be maintained accurately.

4.5.1 The Algorithm for DASP-2

DASP-2 represents a major modification to DASP-1 by incorporating the new set of shrinking time windows. Deriving these new constraints is somewhat involved, but the constraints themselves can be incorporated into the DASP-1 algorithm quite easily.

We present a general description of reformulating the constraints here; all technical details are discussed in Appendix D. Let us assume that we know the position of each aircraft in the terminal area at time T_i , when a sequencing decision is made. Then at time T_{i+1} a new aircraft enters the terminal area and a new landing sequence has to be computed. In the interim, all the aircraft within the terminal area have moved closer to the runway. Some of these will not be resequenced, as their positions in sequence are

frozen. Of those that are not frozen, the amount they can be further delayed or speeded up depends on their positions at time T_{i+1} .

In Appendix D, we describe how we try to approximate the position of each aircraft at each point that we have to recompute the sequence. We base the estimate on the time it entered the terminal area, and how much it has been delayed in there. Further, we try to see how the delay has been achieved, whether by speed reduction and/or path control, and how much of each. We also redefine the point at which we freeze the landing sequence. Previously, we had frozen the landing sequence when the aircraft were 300 seconds from landing. This is roughly at 15 miles from the runway threshold. On the suggestion of air-traffic controllers, however, in DASP-2 we increase this to the point when aircraft have 20 nm left to fly to the runway along the nominal path. Given two aircraft, A and B, it is possible that A is beyond the decision point, but will be sent on a long downwind leg and will land after B, which is still before its decision point. In such a case, B's position in sequence is also frozen *de facto*.

The dynamic portion of the algorithm is similar to DASP-1 in that it employs re-optimization (with the sequence frozen according to different criteria) and nearest-neighbor techniques to determine the next best aircraft to land. Hence the algorithm for DASP-2 is also a heuristic, like that for DASP-1.

In conclusion, we have developed three models and algorithms for different versions of the Aircraft Sequencing Problem. These algorithms are intended to study the benefits of terminal area sequencing over the current FCFS (in general) method that is used. In Chapter 5, we test how well these sequencing algorithms perform, comparing them to the actual sequence in which the aircraft landed in various data sets.

Chapter 5

Application of Sequencing Algorithms

In this chapter, we apply our sequencing algorithms to actual air traffic data for Boston. Our aim is to quantify any benefits that may accrue from sequencing, and to study how they come about. To that end, we shall implement the algorithms we have developed, on different data sets, comparing the resulting delays with the actual delays on the days in question. We shall also examine closely the effects of varying objective functions and constraints to the problem, as mentioned in Chapter 4. Moreover, we shall pay close attention to the differences between the sequences suggested by the algorithm and those actually used by controllers. In this chapter, we discuss in detail this application of aircraft sequencing.

We begin, in Section 5.1, with some preliminary discussion of the data we have, and the general framework within which we shall perform our analysis. In the following three sections (5.2-5.4), we study the performance of algorithms for the three problem types: the static, DASP-1 and DASP-2. For each of 5 data sets, we shall contrast the performance of these different algorithms with each other, and the actual sequences used by controllers. Our aim is to gauge how different restrictions on terminal area operations may affect any gains from better sequencing. We conclude, in Section 5.5, with a discussion of our findings and their implications.

5.1 Framework for Analysis

One important attribute of our study is that it is based on actual traffic data for Boston airport. The information that we need for our algorithm includes the time of arrival of each aircraft into the terminal area, as well as its landing time (and thus the sequence in which all aircraft land). An aircraft's arrival into the terminal area is indicated by its crossing of a 30 nm circular marker centered at the airport. This crossing time can be known only for data collected by radar, and our study has to be limited to these. Moreover, as we mentioned in Chapter 4, we have calculated the time-to-fly bounds only for landings on 4R (or 4R/L). Hence we are further limited to data sets with these landing configurations.

Despite these restrictions, we have 4 relatively large data sets which we can use for our study. A fifth, collected on 3-30-89, also had landings on 4R, but weather conditions were so bad that day that sequencing is unlikely to have been used, and it is unrealistic to use this data set in our analysis. Two other data sets, collected on 5-31-89 and 6-09-89, had to be split into two groups. On 5-31-89, there was an intervening period with 15R in use for landing, hence we omit this period from consideration, and consider the other portions separately. On 6-09-89, there was a brief period when our radar broke down, and we do not have data during this interruption. Again, we consider the two portions before and after the interruption separately.

Table 5.1 lists the data sets used in the sequencing study, with the number of aircraft that landed, and the runway configuration used. We have data sets of varying sizes, and with both single and multiple runways. Of the two very large data sets—5-31-89a and 6-15-89—one is with a single runway in use and the other is with a multiple runway configuration.

For each data set, we have the entry times of aircraft into the TRACON, the fixes at which they entered, the actual landing sequence and the actual landing times. The algorithms will, in different ways depending on the problem type, compute an “optimal” landing sequence. We aim to compare the delay suffered by aircraft in the terminal area, under the sequence suggested by the algorithm, and the one actually used by the

Data Set	Landing Config.	# Landings
2-17-89	4R/L	31
5-31-89a	4R/15Rc4L	75
5-31-89b	4R	28
6-09-89a	4R	38
6-09-89b	4R	40
6-15-89	4R	95

Table 5.1: Data Sets used for Sequencing Study

controllers. We shall do so for each data set.¹

We shall only be sequencing aircraft approaching for the primary runway, even when two runways are being used. This is in keeping with the focus on the primary runway in Chapter 3. Then the reason for ignoring the secondary runway was that the number of landings on it are few and far between, and the LTI's quite large and not very meaningful. Similarly, there is usually no real sequencing problem for the secondary runway (except possibly 22R in 27/22R, but we do not study that here). The aircraft for the secondary runway (mostly props) come in under free flow and land, without really affecting the sequence on the primary. Traffic is so sparse that there may not often be alternative sequences to FCFS.

In Chapter 4, we discussed the formulation of three possible objective functions to the ASP: minimize time delay, passenger delay, or the dollar (fuel) cost of delay. Delay, for a given aircraft, is denoted by the difference between its assigned (or actual) landing time and its earliest possible landing time. Likewise, we can use three possible sets of values for the LTI constraints. The first, natural, choice for the LTI constraints, is to use the mean values from the calibration model. These reflect the average LTI values under current circumstances. Hence, if we assume that the ATC can achieve these values perfectly, we are considering a future with the same mean "performance", but with lesser variance (tighter control).

While the mean calibration values are a natural and intuitive choice, they reflect only

¹By data set, we mean each of the data collections given in Table 5.1.

current performance. Since we are planning for an advanced ATC system, we would like to be able to estimate delays in a future system with not only more tightly controlled LTIs, but smaller ones on average. The question then becomes: what would these shorter LTIs be. We have argued in Chapter 3 that the 25th percentiles from the calibration model are strong candidates. These proposed values cannot be so unrealistically small that the controllers and pilots would be uncomfortable with them. At the same time, they should be appreciably smaller than the mean to be a meaningful improvement. The 25th percentiles satisfy these considerations: pilots and controllers go below this a quarter of the time anyway, and they are typically 20 seconds smaller than the mean values.

These two values—mean and 25th percentile—for the LTIs are based on empirical measurements of current operations. Nowhere have we incorporated the “standards”: what the LTIs would be if the category-based minimum separations were followed meticulously? Calculating the LTIs from separation standards would require, however, an estimate of aircraft velocities. Simpson, Odoni and Salas-Roche [29] offer the typical values of 150 knots for a Heavy aircraft, 120 for a Large and 90 knots for a Small aircraft. Using these values, and the standard separations, one can compute the LTIs. Because we use the “formula values” for the distance and velocity parameters, we term these the formulary LTIs. Table 5.2 gives the value of the formulary LTIs as calculated in [29]. Table 5.3 reproduces the mean and 25th percentile values from the calibration exercise, for the sake of reference.

		Heavy	Large	Small
Weight Class of Lead Aircraft	Heavy	96	150	240
	Large	72	75	160
	Small	60	75	100

Table 5.2: Formulary LTIs (in seconds)

Some of the formulary LTIs are smaller than their counterparts from the mean calibration LTIs, and some are larger. For two of the sequence categories which are most

	Base-1	Base-2	H/H	H/L	H/S	Lj/S	Lp/S
Mean	94	106	102	121	130	94	75
25-perc	73	83	86	104	113	76	63

Table 5.3: Calibration LTIs (in seconds)

represented in the data (L/L and L/H), formulary LTI values are smaller than the mean values from calibration. For the third, H/L, the formulary value is larger. Further, the 25th percentile values are smaller than the formulary values for all the most represented sequence categories. We shall later study the effects of varying the LTI constraint on the sequencing algorithm.

For each aircraft i in a data set under study, let its assigned “optimal” landing time from the algorithm be t_i^S and its actual landing time from the data be t_i^D . Recall that the time window for this aircraft is (a_i, b_i) . Thus the cost of the sequence according to the algorithm is $\sum_i C_Z(t_i^S - a_i)$, and the cost of the actual sequence is $\sum_i C_Z(t_i^D - a_i)$ (with C_Z denoting the type of cost function used). The difference between these two costs, that from the algorithm and that from the data set, is denoted B ; this denotes the benefits (or disbenefits) due to the sequencing algorithm.

Hence B will be the primary measure of how well our algorithms perform. At the same time, however, as we will stress, it is not the only factor to use in judging the benefits of sequencing. Other factors include important human factor considerations, such as workload. Moreover, B has to be interpreted carefully. Some of B derives from good sequencing and some from the assumptions we make in our model. One important assumption that we make, for instance, is that the LTIs between aircraft are adhered to strictly, that they cannot be larger or smaller than the average values we compute from data. Such an assumption is quite valid given the context of TATCA where precise control of aircraft is the goal. On the other hand, in the actual sequences from the data sets, such precise control was not used, and B can be higher or lower depending on how controllers perform on a given day, as we shall see.

Thus we study the effects of sequencing under these different objective functions and

LTI constraints. We programmed the algorithms from Chapter 4 in the "C" language, on SUN workstations. We then evaluated the optimal sequence for each of the five data sets mentioned above. We did so for the three successive models for the ASP. We begin discussing these results in the next section, with the static model.

5.2 Sequencing Using the Static Model

The static model is the least restrictive depiction of terminal area operations. We assume, for each data set, that we know well in advance the entry times of each aircraft into the terminal area. From these entry times, we can compute lower bounds for the landing times. Further, we assume that all aircraft can be held in stacks for 60 minutes. This, in turn, gives us the upper bounds for the landing times.² Hence the algorithm plays the role of a master planner who knows well in advance when each aircraft is going to enter the terminal area, and has to decide the order in which to land them.

5.2.1 The Mean Calibration LTIs as Constraints

We begin by assuming the mean calibration values as the LTI constraints, and minimizing cumulative time delay as the objective. Note that this objective function treats all aircraft equally, regardless of size or number of passengers carried. Table 5.4 gives the results of the static algorithm applied to each data set. We present the total cost of the optimal sequence and the total cost of the actual landing sequence for each data set. The costs are in total minutes of delay. We also compute the ratio of the former to the latter.

The first, somewhat expected, result is that the sequence suggested by the static algorithm typically produces lesser delay than the actual sequence. The amount of reduction varies between 13% on 6-09-89a and as much as 32% and 35% on 2-17-89 and 5-31-89b, respectively. There is one exception, however: in 5-31-89a, the actual delay is 13% less than the delay from the algorithm. Before we explain this anomaly, we shall discuss the reasons for which the sequencing algorithm performs better than the controllers.

²Appendix C discusses how these time windows are computed precisely.

Data Set	Optimal Value	Actual Value	Ratio
2-17-89	43	63	0.68
5-31-89a	118	110	1.06
5-31-89b	39	60	0.65
6-09-89a	78	90	0.87
6-09-89b	73	112	0.65
6-15-89	142	164	0.86

Table 5.4: Cumulative Delay (minutes) Results for the Static Algorithm

Accurate Information on Lower Bounds

The first reason is that the algorithm has an accurate estimate of the lower bound of the landing time of the aircraft. If two aircraft entered the terminal area 50 seconds apart, one at SCUPP and the other at BRONC, it may be unclear to the controller as to which one has smaller lower bound for the landing time. The algorithm knows these values perfectly. Hence, though the controller could easily have modified his landing order to improve the objective, he may not have known what the correct order was. Of course, the controller may not even have shared our objective function. We are using one which measures delay as the difference between actual landing time and the lower bound of the landing time. Hence in so far as we are imposing this objective on the controllers, we stand to observe some improvement over the one(s) they use. They could have one to minimize workload as well, as we shall discuss.

Better Sequences Based on LTI Differentials

There are instances in the actual sequence when, for example, the controller could have chosen the sequence H-H-L-L, but chose instead H-L-H-L, which is worse. In such cases, the algorithm chooses the better sequence in terms of weight classes. We note that any incremental delay to a given aircraft affects those following it equally, delaying them by the same amount. This could happen until the excess is "absorbed" by one or many large gaps of locally light traffic.

Data Set	Mean Position Shift
2-17-89	0.19
5-31-89a	0.56
5-31-89b	1.0
6-09-89a	0.84
6-09-89b	0.90
6-15-89	0.95

Table 5.5: Mean Position Shift Between Actual and Assigned Static Sequences

We are also interested in knowing how much the actual sequence differs physically from the assigned one. Thus, in Table 5.5, we present the mean position shift value between these two sequences. It shows that considerable resequencing is being performed in each iteration. Typically, at least one in two planes has its position changed, and often more. The sequencing algorithm takes advantage of the fact that there are different weight classes of aircraft in the data set. The richer the mix of aircraft, the larger the variety of sequences possible. Hence we would expect there to be greater scope for benefitting from sequencing if there is a rich mix. And the actual sequence may differ greatly from the assigned one. Indeed the lowest value of the mean position shift, 0.19, occurs when two runways are used on 2-17-89, and the mix landing on the primary was quite homogeneous. There were 3 Heavy and 28 Large jets out of a total of 31 aircraft that landed.

Tighter Control of LTIs

The third reason for a difference between the actual objective function value and the one from the algorithm is in our assumption of perfect control over LTIs. We assume that the LTIs are controlled perfectly to equal the constraint values, if they are binding. Thus, when we use the mean calibration LTIs as constraints, the L/L LTI always equals 94 seconds (with a single runway in use). In the actual sequence, this is of course not necessarily the case. But the assumption of perfect control cuts both ways. On days when the actual mean LTIs are higher, the algorithm performs better simply by assuming smaller values. On days when the actual mean LTIs are lower, the algorithm assumes

higher values than are used by the controllers.

This assumption of tight control over LTIs provides an explanation for the algorithm producing higher delays than the actual sequence on 5-31-89a. On that day, the mean LTIs were lower than the mean calibration values used for the constraints. The mean Base LTI on 5-31-89a is 99.7 seconds compared to the average calibration value of 106 seconds with two runways in use. Although the mean Big LTI (ignoring its subdivisions) equals the average calibration value, the disproportionate number of Base LTIs ensures a significant difference in overall mean. Hence, even if the controllers are landing aircraft with a worse sequence than the algorithm uses, but with smaller average intervals, they can do better in terms of the objective we have formulated.

Data Set	Optimal Value	Actual Value	Value of Actual Sequence w/mean Cal. LTIs
2-17-89	43	63	46
5-31-89a	118	110	139
5-31-89b	39	60	67
6-09-89a	78	90	89
6-09-89b	73	112	96
6-15-89	142	164	227

Table 5.6: Delay of Actual Sequence with Mean LTIs

One way to gauge the effects of the assumption regarding LTI values on the objective function is to consider the cumulative delay if we were to use the actual landing sequence, except that we constrain the LTIs to equal the mean values from calibration.³ Table 5.6 presents this cumulative delay value for each data set. We compare it to the cost of the actual sequence with actual LTIs. On some days, the cost of the sequence with actual LTIs is less than its cost with the calibration mean values, and on some days it is greater. Essentially, the difference depends on whether the mean LTIs (at least for the most represented sequence categories) on that day are lesser or greater than the respective

³We assume this whenever the LTI constraint is binding; if it is not, then the LTI is the smallest that is physically possible.

mean values from calibration.

This table, the delay with the actual sequence but with calibration mean LTIs, allows us to study the results of better sequencing, without the “noise” of the LTIs. We notice again that in all instances, the optimal sequence produces less cumulative delay than the actual sequence with the mean calibration LTIs. This is true on 5-31-89a as well.

Hence there are three reasons for which the algorithm produces lower cumulative delay than the actual sequence. The effect of any one factor can vary across data sets. Indeed, on one day (5-31-89a) the lower LTIs used by controllers negated any benefit from the sequencing algorithm. Also, we must remember, different controllers may have been used on the different days. Some of them may be more experienced than others, and there is internal variation in their own performance. Hence on a day on which we perceived 35% improvement, relatively inexperienced controllers may have been at work.

This raises another question. In improving the landing sequences, are we using procedures and practices that controllers do not follow? This is a very important problem. It relates to the way in which controllers give priority to aircraft of different weight classes, or from different fixes. It also concerns the important problem of workload. Perhaps even if the controllers do know of some benefits of resequencing, they do not think it worth the extra effort and control required. These are important points, and are central to the practical usefulness of any sequencing algorithm. We shall keep returning to these questions through the chapter. For now, we examine if the sequences change the standard deviation of delay.

Clearly we have reduced the mean delay suffered by aircraft. In doing so, however, do some aircraft get delayed disproportionately? One way to see if the distribution of delay is affected by the algorithm is by examining the standard deviations in delay. In other words, does the standard deviation of the delay differ between the actual sequence (with actual LTIs) and the optimal sequence? Table 5.7 looks at this question.

Using data from Table 5.4, combined with information on the number of aircraft in each data set, we can compute the mean delay to each aircraft in the actual and optimal sequences. The mean delays in the actual sequence (with actual LTIs) range from 1.6

Data Set	Optimal Sequence	Actual Sequence
2-17-89	2.36	2.16
5-31-89a	1.63	1.68
5-31-89b	1.37	1.83
6-09-89a	3.14	2.25
6-09-89b	2.30	2.10
6-15-89	1.89	1.81

Table 5.7: Standard Deviation in Delay for the Static Algorithm

to 2.9 minutes per plane; in the optimal sequence, mean delays are between 1.5 and 2 minutes, except for one which is 2.4 minutes. Given these average values for delay, the standard deviations above seem somewhat high. One can explain them somewhat by noting that while some aircraft do not get delayed at all, delays to others of 4 to 5 minutes are not atypical. Nevertheless, the standard deviation in delays does not seem to differ substantially or systematically between the actual and optimal sequences. Hence the reduction in delays from the static algorithm does not appear to come at the expense of a wider distribution of delays among aircraft. The danger of large differences in the standard deviation is that some aircraft may be getting penalized unfairly. Indeed, we look at this in another context later, when different objective functions delay aircraft of different weight-classes differently. We shall also return to this issue, toward the end of this chapter, when we examine whether our algorithms affect the way in which aircraft from different fixes are handled.

Of course, one question remains: while there are tangible reductions in delay, just how much are they worth. At one level, reducing delay on 5-31-89 from 63 minutes to 43 minutes, a difference of 32%, appears high. However, it was obtained for 31 aircraft, over just under an hour. Hence the reduction of delay for each aircraft was approximately $\frac{2}{3}$ of a minute. This is approximately equal to reducing each LTI by 40 seconds, ignoring the periods of sparse traffic when LTIs are not binding. Table 5.8 presents the reduction in delay per aircraft for each data set. Ignoring 5-31-89a, when the algorithm performed worse than controllers, the reduction is somewhat uniformly distributed between 0.23

minutes (about 13 seconds) and approximately one minute.

Data Set	Optimal Value	Actual Value	# of Aircraft	Red'n per Aircraft
2-17-89	43	63	31	0.66
5-31-89a	118	110	75	-0.1
5-31-89b	39	60	28	0.75
6-09-89a	78	90	38	0.42
6-09-89b	73	112	40	0.98
6-15-89	142	164	95	0.23

Table 5.8: Cumulative Delay Reduction (mins.) per Aircraft for Static Algorithm

It is difficult to say how much these reductions are really “worth”. On the one hand, a reduction of a minute in flying time, over a flight of more than one hour typically is minuscule. On the other hand, operating costs per minute are rather high and can accumulate to large amounts over many days. Hence we have to keep this issue of the true worth of the reduced delays in mind when assessing the benefits of sequencing. Nevertheless, it does appear, at least for the static case, that despite the operating constraints of the terminal area, some of the potential benefits of sequencing do appear realizable.

5.2.2 The Effects of Varying the LTI Constraint

So far we have just looked at the case of the LTI constraint equaling in value the mean calibration LTI. What happens if we look at other values for the LTI constraint, such as the 25th percentile or the formulary LTI. Does the nature of the results change? Do the optimal sequences change? If so, why?

Table 5.9 presents the values of cumulative delay for the optimal sequence, for each data set, and for each LTI constraint. We note, at the outset, that the cumulative delay is always less than the actual delay if the formulary or 25th-percentile LTIs are used. This is not only due to better sequencing, effects of which we have already seen, but also to smaller LTIs on average than were actually used. The 25th percentile LTI constraint produces the smallest cumulative delays, followed by the formulary values, and then the

Data Set	Actual Delay with Actual LTIs	Minimum Cumulative Delay		
		Mean LTIs	25th-Perc LTIs	Formulary LTIs
2-17-89	63	43	32	35
5-31-89a	110	118	44	79
5-31-89b	60	39	19	27
6-09-89a	90	78	43	63
6-09-89b	112	73	23	56
6-15-89	164	142	55	107

Table 5.9: Optimal Cumulative Delays with Different LTI Constraints to the Static Problem

mean. This is in increasing order of LTI values (for some of the most represented sequence categories), and thus conforms to expectations.

Of course with different models for the LTI constraint, the optimal sequence of aircraft will be different as well. This is for two reasons. The first is that the LTI values are simply different. So with the formulary LTIs, the H/S LTI (240 sec.) is 90 seconds higher than the H/L LTI (150 sec.). On the other hand, the corresponding difference is 9 seconds for the mean LTIs, and 7 seconds for the 25th percentiles. Hence the H/S sequence category is much more to be avoided if using the formulary values, than the other two types of LTIs.

The second reason for a difference in sequence is that some sequences are no longer feasible if LTIs are on the average smaller. If previous aircraft have been landing at a higher rate, then a particular LTI constraint that was binding before may no longer be binding at all. This is because the trail aircraft in question is not able to land early enough to make the constraint binding. We illustrate this with an example from the 5-31-89a data set. Two runways were being used on that day. So the mean Base LTI is 106 seconds. Table 5.10 shows the optimal sequence for landing three aircraft, with each of the different LTI constraint values.

The system starts at time 0, without loss of generality. Some aircraft have landed, and the landing time of the previous aircraft (whose identity was the same for all sequences) is given. Because aircraft have been landing at a higher rate using the 25th percentile or

formulary LTIs, the landing time of the previous aircraft is lower in these instances.

Aircraft	Weight Class	Lower Bnd on Land. time	Actual Land. Time	Assigned Landing Times		
				Mean	Formulary	25th Perc.
Previous	-	-	926	963	899	877
GAA846	L	1028	1076	1069	1223	1064
TAP324	H	841	1163	1175	959	960
AAL94	H	1073	1261	1277	1073	1147

Table 5.10: Example of Effect of Different LTI Constraints on Optimal Sequence

Two of the three aircraft are heavy, and one is Large. It is always optimal to keep the two Heavy aircraft landing one after the other (in a cluster). Breaking up the cluster in any way by inserting another aircraft is suboptimal, since the H/H LTI is the smallest of those with a Heavy aircraft in the lead. But the optimal sequence with the LTI constraint being equal to the 25th percentile of calibration LTIs breaks the cluster: one Heavy lands at 960, followed by the Large at 1064, and the last Heavy at 1147. The reason is as follows.

If the Heavy aircraft were to be clustered, the Large would land before or after the cluster. If it were to land before, we would have to wait until 1128 (its lower bound) for it to land, and then the Heavy aircraft would land 83 seconds later and 86 later still, respectively. The last of them would land at 1297, which is much later than the value of 1147, from the optimal sequence. If the Large were to land after the cluster, the first Heavy (TAP324), which can land at 960, would be sequenced first. Then we would have to wait for the other Heavy, AAL94, to land at 1073 at the earliest, and then for the Large to land at $1073 + 104 = 1177$. So it is optimal to break the cluster, as the algorithm does, to have the last of the three aircraft land at 1147.

This example, taken from the data, shows how the differing LTI values can induce changes in the optimal sequences. Essentially, the shorter 25th percentile LTIs make some landing orders suboptimal because one would have to wait very long for some particular aircraft. In our example, the Large was arriving so much later that it was not worth clustering the Heavy aircraft.

5.2.3 The Effects of Varying the Objective Function

Finally, we wish to explore the effects on the optimal sequence of using different types of objective functions. So far, the objective has been to minimize cumulative delays. Now we shall consider the other two objectives that we have proposed: minimizing passenger delays and minimizing the monetary (dollar) cost of delays.

An algorithm minimizing cumulative delays treats delays to all aircraft equally: one minute of delay to a B747 is worth the same as a minute of delay to a Cessna. Weighting the delay by the number of passengers on board eliminates this, somewhat unrealistic, assumption. The objective function of minimizing the dollar cost of delays (termed dollar delay) not only weights delays to different aircraft differently, but also makes the cost function of delays convex. Hence it acts against particularly long delays to any one aircraft. These different objectives not only measure different aspects of delay (to aircraft as units, to passengers, and in dollar terms), but also represent different functional forms for delay.

We begin by comparing the sequence suggested by the algorithm with the actual performance of the controllers for each data set, for the two objective functions not looked at so far. We want to see how the different optimizing criteria affect the optimal sequences, vis-a-vis themselves and the actual sequence, for a given LTI constraint. Hence, we assume without loss of generality that the LTI constraints are the mean of the calibration values. We have already examined the value of the optimal sequence and actual one for each data set, where the objective was to minimize cumulative delay. In this instance, the sequence suggested by the algorithm produced less cumulative delay than that used by controllers, in all but one instance: 5-31-89a. We now repeat this analysis with the other two objective functions.

Our dollar delay objective function was described in Chapter 4. For passenger delays, we assume that a typical Heavy (like a DC10) has 300 passengers, a typical Large jet (like a B737) has 150 passengers, a typical Large prop (like an ATR42) 50 passengers, and a typical Small (like a Beechcraft) 10 passengers. These figures are illustrative and represent order of magnitude estimates. They may not literally reflect the most common aircraft of each type as is seen in Boston, nor average passenger loads.

Date	Passenger Delay			Dollar Delay		
	Optimal	Actual	Ratio	Optimal	Actual	Ratio
2-17-89	7,377	11,780	0.63	738	1,144	0.65
5-31-89a	13,506	20,597	0.65	2,009	2,072	0.97
5-31-89b	3,086	7,558	0.40	574	937	0.61
6-09-89a	7,069	11,051	0.61	1,083	1,304	0.83
6-09-89b	6,452	17,806	0.36	1,038	2,004	0.52
6-15-89	14,923	21,756	0.69	2,318	2,602	0.89

Table 5.11: Comparison of Optimal Static Sequences with Different Objective Functions

Table 5.11 gives the objective function values of the optimal and actual sequences for each data set. Passenger delay is measured in passenger-minutes. We notice that, with these objective functions, the optimal sequence produces less value of delay than the actual one used by controllers, on all data sets, including 5-31-89a. This suggests that even when the actual LTIs are below the mean calibration values, they do not compensate for the poor performance of the actual sequence on the objectives above. The percentage improvement of the algorithm is much higher for these objective functions than for the cumulative delay one. For passenger delay, the difference is the greatest, with the algorithm producing sequences typically worth 60% the actual delay, and sometimes as little 36% (6-09-89b) and 40% (5-31-89b). The dollar delay function produces less improvement on each data set, ranging from 3% on 5-31-89a to 52% on 5-31-89b. Certainly this wide difference in the performances may be an indication that air traffic controllers do not themselves weight different aircraft differently in their mental sequencing algorithm.

The cumulative delay objective, as we have discussed, treats delay to each aircraft equally. And the other two objectives value one minute of delay to aircraft higher in increasing order of size (by weight-class). To what extent are these differences manifested in the sequences suggested by the algorithm? Table 5.12 shows the differences in average delay (in minutes) overall, and by weight-class, for each objective function and for each data set. This table displays some interesting patterns: as the objective function is weighted more in favor of Heavy and Large jet aircraft, so does the mean delay suffered

Date	Obj. Fun. of Delay	Average Delay (minutes)				
		Overall	Heavy	Large (jet)	Large (prop)	Small
2-17-89	# Aircraft	31	4	27	-	-
	Cumulative	1.4	2.2	1.3	-	-
	Dollar	1.5	1.5	1.5	-	-
	Passenger	1.5	1.5	1.5	-	-
5-31-89a	# Aircraft	75	14	45	10	6
	Cumulative	1.7	2.8	1.5	0.9	1.3
	Dollar	2.11	0.4	1.7	2.6	8.7
	Passenger	2.5	0.4	1.2	4.8	12.8
5-31-89b	# Aircraft	28	1	16	9	2
	Cumulative	1.4	0.5	1.5	1.6	0.4
	Dollar	1.8	1.2	1.3	1.4	8.2
	Passenger	2.7	0	0.6	3.1	18.8
6-09-89a	# Aircraft	38	5	26	3	4
	Cumulative	2.1	2.2	2.3	1.9	0.7
	Dollar	2.72	0.7	1.6	4.8	11.0
	Passenger	3.9	0	1.3	7.6	22.7
6-09-89b	# Aircraft	40	8	22	6	4
	Cumulative	1.8	3.9	1.2	1.2	2.1
	Dollar	2.3	0.7	1.0	2.9	11.0
	Passenger	2.4	0.6	0.9	3.8	11.6
6-15-89	# Aircraft	92	12	50	22	8
	Cumulative	1.5	3.8	1.0	1.8	0.9
	Dollar	2.08	0.8	1.5	1.7	8.8
	Passenger	3.6	0.4	1.1	3.3	25

Table 5.12: Average Delay by Objective Function and Weight Class

by such aircraft decrease.

For each data set, the overall mean delay is smallest when the objective is to minimize cumulative delay—quite logical, since this is the same as minimizing mean delay. It is highest for the passenger delay function because this gives most disproportionate weight to Heavy aircraft (300 passengers, versus 10 for a Small). The dollar delay function produces the middle value of mean overall delay for each objective function. This is because the Heavy aircraft are proportionately weighted less (compared to Large or Small) than the passenger delay function. Further, the convex costs weigh against delaying an aircraft, even a Small, for an extended duration.

This differential weighting also explains the individual differences in mean delay by weight class, as we change objective function. In each data set, as the objective function assigns greater weight to delaying the larger aircraft, so too does the mean delay suffered by these aircraft decrease. Consider, for example, 5-31-89a, a typical case. For the cumulative delay objective, the Heavy aircraft had the highest mean delay (2.8 minutes). But this decreased dramatically for the other two objective functions, to 0.4 minutes each. In these cases, the delays suffered by the Heavy and Large jet aircraft are “transferred” to the Small and Large prop aircraft. The transfer is most dramatic for the passenger delay function because, as we have said, this weights delays to larger aircraft most disproportionately. For instance, the mean delay for Small aircraft in this data set rises from 1.3 minutes (cumulative delay minimization) to 8.7 minutes (dollar delay minimization) to 12.8 minutes (passenger delay minimization).

These comparisons of the performance of the algorithm for the dollar and passenger delay objective functions illustrate many points. First, the actual sequence produces so much more delay (valued according to the objective) compared to the cumulative delay function, that it appears air traffic controllers may follow the latter objective rather than any others. Second, while these other objectives seek to minimize the costs of important aspects of delay, they do so at the expense of smaller aircraft. From a policy perspective, this may not be equitable to the smaller aircraft. Hence this keeps the dramatic improvements of the algorithm in perspective. Given the sensitive issues related to delay-

ing Small aircraft disproportionately, the passenger delay function would probably be met with strong opposition in a practical setting. Hence a function like dollar delay, possibly with a different exponential increase in cost with delay, may strike a balance between the cumulative delay and passenger delay, each of which can be termed inequitable, in opposite ways.

In this section we have discussed the static algorithm, and examined the results of better sequences, using many objectives and different LTI constraints. The effects of varying the objective and constraint are present in the same manner for the dynamic problem as well. Hence, when studying the dynamic algorithms, and comparing the sequences suggested by them to the actual sequence and to the static algorithm, we consider a canonical objective and LTI constraint. The canonical objective is to minimize cumulative delay, the one controllers seem to follow. The canonical LTI constraint is the mean value from calibration. The aim here is to judge the performance of the dynamic algorithms if they were applied today, with current LTI values (tightly controlled). Naturally, as we shall discuss, we shall reduce delays (by any measure) if we were to use the 25th percentile of LTIs, simply because landing rates will be higher. We begin in the next section, by studying the problem DASP-1.

5.3 Sequencing Using the Dynamic Algorithm with Fixed Time Windows

In this section, we present the results of the dynamic algorithm with fixed time windows applied to the data sets. We aim to compare the performance of this more constrained model with the static algorithm and the actual sequences. We shall do so for the canonical objective and constraints. We do not repeat the detailed analyses of the effects on the algorithm costs and sequences of varying the LTI constraint or the objectives. This is because these results closely parallel those presented previously, and can be explained by the same reasons discussed then.

The dynamic algorithm differs from the static one in two ways. First, there is no hold-

ing, so there is a narrower time window within which each aircraft has to land. Secondly, we do not assume that we know in advance what the arrival times of the aircraft are. Instead we emulate a controller who sees each aircraft only when it enters the terminal area. Then he resequences all those that have not landed, with the exception of those within 300 seconds of landing (roughly 15 miles). Thus there may be optimal sequences that are “missed” because we did not know in advance of an aircraft whose arrival could have been exploited to improve the sequence.

Data Set	DASP-1	Static	Actual
2-17-89	43	43	63
5-31-89a	118	118	110
5-31-89b	39	39	60
6-09-89a	78	78	90
6-09-89b	73	73	112
6-15-89	142	142	164

Table 5.13: Comparison of Minimum Cumulative (mins.) From DASP-1 and Static Sequences

Table 5.13 presents the results of the dynamic algorithm (DASP-1) applied to each data set. We also present the result from the static algorithm, and the cumulative delay associated with the actual sequence and the actual landing times in each data set. The results are quite surprising. They indicate that the more constrained, DASP-1 problem, with narrower time windows, and decisions made on less information, produces exactly the same minimum delays as the static algorithm. This is true in all cases. The implication is that the constraints we have added to DASP-1 are not really binding, at least with respect to this particular objective function.

Let us consider these additional constraints in turn. DASP-1 specifies a narrower time window within which aircraft must land. Since the lower bounds of the windows are the same for the static algorithm and DASP-1, the narrower range will affect the objective value only if the upper bound of the time window is a binding constraint. An examination of the sequences and landing times of each aircraft across all data sets shows that in no

case was the upper bound of the time window a binding constraint. Then is it also true that there is no value to the additional information embodied in the static model, i.e. knowledge in advance of the arrival time of each aircraft into the terminal area?

This is a subtler question. First, we ask if the optimal dynamic and static sequences can differ. In theory, the myopic horizon of each iteration in the dynamic algorithm need not produce a sequence that is optimal in the long run. We illustrate this problem in the following way. Consider the following aircraft Lp-Lj-H-Lp-Lj. Now there are only two distinct LTIs possible here Base and H/L. It is important to sequence the Heavy last, to avoid the larger H/L LTI. Other than this caveat, however, there are many identical sequences, in terms of equivalent cumulative delay (assuming that all aircraft can land at any time within consideration). Thus all permutations with the Heavy aircraft last are identical. The static algorithm can choose from all of them, and the dynamic algorithm possibly only from some of them, but the restriction may not matter.

The oddity is that such restrictions, when they occur, seem not to matter in all the data sets that we study. Actually, what happens is that the optimal sequences for the dynamic algorithm are indeed different from those for the static algorithm. But the differences do not lead to disparities in the objective function value. The dynamic version, with the fixed leadtime preventing resequencing beyond a point, does indeed produce fewer feasible sequences. But the mix of aircraft is relatively poor (many more Large aircraft than Heavy or Small ones) and the Base category includes so many weight-class combinations, that often the myopic sequencing does not perform poorly with respect to the objective. There are so many good sequences that a restricted choice among them seems not to affect the objective.

Table 5.14 gives the delay for each aircraft by weight class. We note that although the sequences have different values, the average delay sometimes differs within each weight class. Hence if we consider another objective function which does not weight a given amount of delay to all aircraft equally, then the restrictions of the dynamic model may indeed affect the optimal objective function value. Although the average differences are small, on the order of 0.1 when they exist, cumulatively, and weighted heavily, they might

Date	Algorithm	Average Delay (minutes)				
		Overall	Heavy	Large (jet)	Large (prop)	Small
2-17-89	# Aircraft	31	4	27	-	-
	Static	1.4	2.2	1.3	-	-
	Dynamic	1.4	2.2	1.3	-	-
5-31-89a	# Aircraft	75	14	45	10	6
	Static	1.7	2.86	1.55	0.9	1.3
	Dynamic	1.7	2.90	1.50	1.1	1.3
5-31-89b	# Aircraft	28	1	16	9	2
	Static	1.4	0.5	1.5	1.6	0.4
	Dynamic	1.4	0.5	1.4	1.7	0.4
6-09-89a	# Aircraft	38	5	26	3	4
	Static	2.1	2.2	2.3	1.9	0.7
	Dynamic	2.1	2.2	2.2	1.9	0.7
6-09-89b	# Aircraft	40	8	22	6	4
	Static	1.8	3.9	1.2	1.2	2.1
	Dynamic	1.8	3.9	1.2	1.2	2.1
6-15-89	# Aircraft	92	12	50	22	8
	Static	1.5	3.8	1.0	1.8	0.9
	Dynamic	1.5	3.8	1.1	1.4	1.1

Table 5.14: Average Delay by Weight Class for Static and DASP-1

alter objective function values. Referring to our previous example, let us assume that delays to the Lp aircraft cost less than delays to Lj aircraft. Then, if the dynamic algorithm cannot choose for instance Lj-Lj-Lp-Lp-H, the optimal objective function value may well be affected.

Date	Passenger Delay			Dollar Delay		
	Static	DASP-1	Ratio	Static	DASP-1	Ratio
2-17-89	7,377	7,377	1.00	738	738	1.00
5-31-89a	13,506	17,765	0.76	2,009	2,096	0.96
5-31-89b	3,086	3,253	0.95	574	574	1.00
6-09-89a	7,069	9,211	0.77	1,083	1,173	0.92
6-09-89b	6,452	7,553	0.86	1,038	1,137	0.91
6-15-89	14,923	20,904	0.72	2,318	2,373	0.98

Table 5.15: Performance of Static and DASP-1 Algorithms for Dollar and Passenger Delay Objective Functions

Consequently, we compare the performance of the static and dynamic algorithms for the other two objective functions (passenger delay and dollar delay) for each data set. The results are given in Table 5.15. For the Passenger delay function there can be relatively large differences, of as much as 18% on 6-15-89 for example. For the Dollar delay function, the differences, when they exist, are smaller: no more than 9% (on 6-09-89b). The reason is that the myopic dynamic algorithm does not perform as well as the ‘global’ optimization of the static algorithm. As we have discussed, this is because delays to different types of aircraft are now weighted differently. We should also mention that the optimal sequences for the dynamic algorithm differ by objective function for a given data set in the same way that they did for the static algorithm and for the same reasons.

One final point to consider is whether, for the cumulative delay objective function, varying the LTI constraint ought to affect the relative performance of the dynamic algorithm with respect to the static one. In theory, we mentioned that the dynamic algorithm did not do much worse than the static one because the mix of aircraft is relatively poor and thus there are many sequences that have equal costs. Restricting the choice set may not rule out the possibility that the dynamic algorithm chooses one optimal sequence.

Further, the upper bound on the landing times is not a binding constraint. Neither of these factors may be affected by the fact that we may use smaller formulary or 25-th percentile LTI constraints instead of the larger calibration mean values. Of course, there may arise a situation which, with different LTI constraints, could lead to differences in the static and dynamic sequences, using the cumulative delay as objective function. But we would not be surprised if there is only a marginal difference or none at all. We computed the optimal dynamic sequence and cumulative delay for each data set, for these other two LTI constraints. In all cases, as before, the static and dynamic problems had the same objective function values.

To conclude, the dynamic formulation does indeed constrain some of the static sequences. The effect is only marginally seen, if at all, with the cumulative delay function for many reasons. First the upper bound of the time window is not binding. Second, there are so many optimal sequences to choose from that the restriction is often not meaningful. This is because of the relatively poor traffic mix and the fact that the cost function does not differentiate between delaying, for an equal amount of time, a B747 or a Cessna. If an objective function does indeed discriminate in such a way, the dynamic algorithm does perform more poorly than the static one.

Of course, we have mentioned that the dynamic algorithm, as we have it, is not very restrictive. Indeed the results seem somewhat to indicate that themselves. Now, therefore, consider the most restrictive model of terminal area operations, DASP-2.

5.4 The Dynamic Algorithm with Shrinking Time Windows

Finally, we implement the dynamic algorithm with shrinking time windows on the data sets. This is the most restricted model of terminal operations, and was suggested by air traffic controllers as more representative of current practice than DASP-1. As discussed in detail in Chapter 4 and in Appendix D, the model recognizes that as an aircraft nears the airport it cannot have the same opportunity for speeding up or delay that it once ahead.

Thus time windows of aircraft do not remain constant as they did with DASP-1, but shrink. Further, there is a strict decision point with 20 nm left to fly on the nominal path, by which time the aircraft's position in the final sequence ought to be fixed. For aircraft from the northern fixes (BRONC, SCUPP, KHRIS, EXALT, LOBBY), this means that the length of the downwind leg should be determined by this point as well. For aircraft from the southern fixes (PVD, FREDO, WOONS), there is very limited room for path stretching: the maximum amount being 180 seconds, and tenable within 90 seconds of entering the terminal area.

Hence this model is much more restrictive operationally than the previous ones. The point we now wish to explore is whether and how much it affects the benefits of optimal sequencing. Once again, we begin with the canonical case: the LTI constraints are the mean calibration values, and the objective is to minimize cumulative delay.

Data Set	DASP-2	DASP-1	Static	Actual
2-17-89	43	43	43	63
5-31-89a	130	118	118	110
5-31-89b	39	39	39	60
6-09-89a	84	78	78	90
6-09-89b	74	73	73	112
6-15-89	170	142	142	164

Table 5.16: Comparison of Minimum Cumulative Delay (mins.) for All Algorithms

Table 5.16 presents the results of the DASP-2 algorithm, comparing it to cumulative delay with the static and DASP-1 algorithms, as well as the actual sequence. The results are somewhat surprising. The optimal value with DASP-2 equals that with DASP-1 for 2 data sets. It is marginally higher for the others, between 1% on 6-09-89b and 7% on 6-09-89a, except for 6-15-89, when the difference is 20%. On 6-15-89, the sequence with DASP-2 in fact performs worse than the controllers, by about 4%. The results are somewhat surprising, because given the rather severe restrictions we have now enforced on terminal area operations, we might have expected considerable curtailment of our freedom to resequence, and consequently higher optimal values for DASP-2.

On the days when there is a perceptible difference, the explanation lies with the tighter constraints. Consider the case of two arrivals, one from SCUPP and one from Providence. An arrival from SCUPP nominally takes 855 seconds to travel to the runway, whereas the one from Providence takes only 555 seconds nominally. Hence, if they were expected at the runway at about the same time, the Providence aircraft would enter the terminal area (and be “seen” by our algorithm) 300 seconds later. Now consider the scenario as happened on 6-15-89, on more than one occasion: an aircraft from SCUPP was being expedited to land at the runway, and when it was just past the decision point, an arrival from Providence appeared. It would have been optimal to land the Providence arrival earlier, as there was a relatively large gap between the SCUPP arrival and its predecessor. But the commitment had been made to land the SCUPP aircraft (it was past the decision point) and the Providence aircraft was delayed. Further the delays accumulated through the successive aircraft. The DASP-1 algorithm in this circumstance would not have committed to land the aircraft from SCUPP and would have delayed it further. This is because it would have assumed that it could exploit the long downwind leg even at that point.

5.4.1 Differential Delays to Northern and Southern Aircraft

This method of delaying Providence aircraft is in fact observable in general. Table 5.17 compares the mean delay to aircraft from the northern and southern fixes for each data set, for the DASP-1 and DASP-2 algorithms, and the actual sequence. The mean values for the Static algorithm are very similar to those for DASP-1 and hence are not presented. First we examine the differences between DASP-1 and DASP-2. We see that for each data set, the mean delay given to aircraft from the southern fixes is higher for DASP-2 than for DASP-1. On the contrary, the delay to those from the northern fixes decreases.

Essentially, the way the algorithms (DASP-1 and DASP-2) work is as follows. When aircraft from the northern fixes enter the system, there are relatively large gaps between them and those in the terminal area that entered from the southern fixes (which are closer to the runway). The algorithms expedite the northern aircraft to land at the runway close to their lower bounds. Because the algorithms employ no look-ahead policy, they do

Date	Fixes	# Aircraft	Actual	DASP-1	DASP-2
2-17-89	Southern	18	1.14	0.57	0.82
	Northern	13	3.24	2.53	2.18
5-31-89a	Southern	45	1.23	1.4	1.9
	Northern	30	2.07	2.13	1.51
5-31-89b	Southern	17	1.9	1.38	1.54
	Northern	11	2.3	1.42	1.17
6-09-89a	Southern	24	2.18	2.02	2.34
	Northern	14	2.7	2.14	1.99
6-09-89b	Southern	19	2.5	1.26	2.37
	Northern	21	2.85	2.37	1.4
6-15-89	Southern	37	0.83	1.16	2.2
	Northern	55	2.4	1.8	1.61

Table 5.17: Average Delay for Aircraft from Northern and Southern Fixes

not foresee (as the controller probably does) that there are likely to be arrivals from the southern fixes which may have similar expected arrival times at the runway. When these arrivals from the southern fixes do materialize, the DASP-1 algorithm delays the northern aircraft if it is better to do so. The DASP-2 algorithm, with shrinking time windows and an earlier decision point, may not be able to delay the northern aircraft by very much, if in fact it can at all. Hence the no look-ahead aspect of the DASP-2 algorithm can lead to worse sequences. The DASP-1 algorithm also suffers from having no look-ahead, but has much greater flexibility to adapt to the dynamics of the new arrivals. Hence we see why the mean delay for arrivals from southern fixes is greater with DASP-2 than with DASP-1.

What do the controllers do? They state that they also let the southern arrivals influence the sequence. Because they do indeed know that there will be arrivals from the south, they do not expedite the northern arrivals as our algorithm does. Rather they leave gaps in which to sequence the southern arrivals. Unfortunately, it is difficult to discern this

from the figures on mean delay. For instance, from Table 5.17, the mean delay from actual and static sequences decreases for aircraft at both types of fix (except on 5-31-89a, and on 6-15-89 when the delay for aircraft from southern fixes increases under the algorithm). The differences between the actual sequence and the DASP-1 algorithm occur because of better sequencing, different control, etc., as we have discussed before. Further, the proportional change in mean delay for each type of fix does not follow a consistent pattern. For example, on 2-17-89 and 6-09-89b, the proportional decrease in delay to southern aircraft is very high for the DASP-1 algorithm. On other days, it goes the other way. What is quite clear, however, is that both in the algorithms and the actual sequences, the southern aircraft do get delayed on average more than the northern ones. Hence, in all cases, though to differing degrees, the southern aircraft influence the sequence strongly. Indeed, with reference to 6-15-89, the day on which the DASP-2 algorithm did worse than the controllers, the actual sequence had the smallest average delay to southern arrivals (0.83 minutes) whereas the DASP-2 algorithm, with limited look-ahead and restrictive constraints, had a very high average delay. Perhaps that day, IFR with heavy traffic and all landings on 4R, the controllers knew best that the southern arrivals had to be given priority.

5.4.2 Varying the Objective Functions and Constraints

Table 5.18 presents the optimal objective function values for the passenger and dollar delay functions, using DASP-2. We compare the performance with respect to the actual sequence. We assume that the LTI constraints are represented by the mean values from calibration. DASP-2 does not perform as well on these objectives as the static and DASP-1 algorithms (refer to Table 5.15), as is to be expected. For the passenger delay function, it performs much closer to the actual sequence than the other algorithms.

There are, however, the interesting cases of 6-09-89b and 6-15-89 when, using the passenger delay objective, the algorithm cannot get a feasible solution. We recall that average delay to aircraft is greatest with the passenger delay objective than with the others. Hence, aircraft are on average (and Small ones in particular) landing closer to

their upper bounds than with this objective compared to the others. Then, essentially what happens is that at some point a plane (from from the South in both cases) has a landing time that is much higher than it would be with other objectives. This is because its predecessors are landing later on average than with other objectives. This particular aircraft then has to be delayed more than the constraints as we have them permit, and the algorithm produces an infeasible solution: the upper bound of the time window is violated. Since we have formulated constraints based on the advice of controllers, this is another indication that the passenger delay objective might not be the one which they are using.

Date	Passenger Delay			Dollar Delay		
	Optimal	Actual	Ratio	Optimal	Actual	Ratio
2-17-89	7,330	11,780	0.63	738	1,144	0.65
5-31-89a	18,555	20,597	0.90	2,197	2,072	1.06
5-31-89b	3,904	7,558	0.90	1,133	937	1.21
6-09-89a	10,755	11,051	0.94	1,386	1,304	1.06
6-09-89b	Infeasible	17,806	-	1,173	2,004	0.59
6-15-89	Infeasible	21,756	-	2,558	2,602	0.98

Table 5.18: Comparison of Optimal DASP-2 Sequences with Different Objective Functions

We had noted earlier that for the cumulative delay objective, the static and dynamic algorithms performed equally well, regardless of the type of LTI constraint. With the DASP-2 case as well, the performance of the algorithm with respect to DASP-1 is the same for all LTI constraints.

5.5 Implementation of Sequencing

Our application of sequencing has indicated that the algorithms can perform well compared to the actual sequences used by controllers, for a variety of objective functions and constraints. In this final section of the chapter, we consider the implication of our results for advanced terminal area automation. In particular, we discuss the potential for the use of such algorithms to aid controllers in the future. At the outset, it is extremely important

to state that any such algorithm would function purely in an advisory capacity. It might suggest sequences, but the controller would in no way be obliged to follow the suggestions.

The algorithms that we have developed (even the most restricted one, DASP-2) show a reduction in cumulative delay compared to actual sequences, using the mean calibration LTI. Hence, the comparisons based on the terms most close to current operations do indeed raise the possibility of gains to be had from sequencing. The algorithms run very fast: even for the largest data set, the DASP-2 algorithm implemented on a SUN 3 workstation, takes no more than 10 seconds for the entire data set. So clearly there is no computational limit to its being used for real-time sequencing. Further, it makes use of information, such as the lower bounds on arrival times that are not available to the controller at present (though they conceivably could be). Finally, the implementation for DASP-2 takes into account the important operational constraints as expressed by controllers for Boston. Hence the model is relatively realistic. There are, however, other concerns that need to be addressed.

5.5.1 Workload and Overtakes

The most important concern is workload. The sequencing algorithms do allow overtakes to occur. Overtaking requires close monitoring and guidance from the part of the controller, especially if the aircraft are from within the same stream. Hence, even if the controller sees that a particular improvement is possible, he may not want to attempt it if it involves additional work in a working environment that is already stressful. Overtaking is also a concern in so far as first-come-first-serve policies are violated. Pilots are especially averse to being overtaken: a global optimum need not be a local optimum for each aircraft.

We have not addressed the issue of overtaking directly in our model. One could add some constraint to the algorithm of the form that overtaking is not allowed unless it results in a certain minimum improvement in the objective function. For a dynamic programming formulation, operating on a dynamic problem of varying size and no look-ahead, this is difficult to implement. One way would be to constrain certain orderings as fixed and then to re-solve the problem to see the change induced in the objective function.

We explore the issue of overtaking in another way. We examine the number of over-

takes, within stream and across stream, for the sequences suggested by DASP-1 and DASP-2, for each data set. Indeed, there is an interesting question here: if an aircraft from Providence just enters the system, and is sequenced to land ahead of an aircraft from SCUPP (for example) which could have landed first, is that an overtake? Arguably, it is, but we do not consider it as such. Given the fact that Providence arrivals are given priority, by controllers, we choose a more restricted definition of overtaking. We define an overtake as a reordering of the landing sequence of the aircraft already in the terminal area. An overtake is within stream if the aircraft are from the same fix, and across stream if they are from different fixes. This again is a somewhat simplistic division, since aircraft from different fixes do indeed merge into one stream even before the decision point. However, we need to make this simplifying assumption, given the limitations of our computer model: it does not have a "radar map" of the location in space of each aircraft at each instant. Note that an across stream overtake is effected by reducing the speed of one aircraft and increasing that of another. Hence it does not require as much supervision as a within stream overtake.

Date	DASP-1		DASP-2	
	Across	Within	Across	Within
2-17-89	0	0	0	0
5-31-89a	1	0	0	0
5-31-89b	2	0	1	0
6-09-89a	3	0	2	0
6-09-89b	12	3	9	1
6-15-89	14	3	14	3

Table 5.19: Overtaking with Sequencing Algorithms

Table 5.19 presents the number of within stream and across stream overtakes for DASP-1 and DASP-2 for each data set. Except for the last two data sets there is not much overtaking taking place. In fact, on 2-17-89, there is none. This does not mean that our sequence was identical to the actual one. Indeed, the actual one itself may have used some overtaking. Moreover, we may have placed entering aircraft from Providence

differently into the existing stream for landings and, as we discussed, we do not include these as overtakes. The DASP-2 algorithm typically has fewer overtakes than DASP-1 (as we would expect), and the same number on 6-15-89 (and 2-17-89).

The only days with a substantial number of overtakes are 6-09-89 and 6-15-89. The latter was a very large data set, and in the former, one particular aircraft (a Small) was having its position changed frequently over about 4 resequences. In a practical situation, such frequent resequencing would be avoided. The overtakes on 6-15-89 were spread relatively uniformly over the 3 hour data set. Further, the number of within stream overtakes is very small. No more than 3 in any data set. In fact 3 such overtakes in 3 hours on 6-15-89 is arguably not excessive. It may, however, have been excessive over 1.5 hours on 6-09-89b.

It is difficult to say how much of an increase in workload would be involved by using the sequences suggested by the algorithm. One important fact to keep in mind, however, as suggested by one controller⁴ is that the very existence of sequencing algorithms reduces the workload of controllers: they no longer have to devise a sequence in which to land aircraft. The fact remains, however, that the sequences suggested by our algorithms may not be preferable to them. However, one good way to study the acceptability of our sequences might be with a simulation involving air-traffic controllers. This would require, though, a far richer model of the terminal airspace than our implementation provides. Further, it would have to be connected with a simulated radar giving the position in space of every aircraft.

To conclude, the algorithms we have developed are very promising in that they indicate that improvements are realizable from better sequencing. These improvements are seen on a variety of data sets, under relatively realistic models of operating constraints, for a variety of objective functions and LTI constraints. These improvements accrue from better timing, control and physically better sequences.

There are two important issues that are still open. One is the important question of how these sequences would affect the workload of controllers is still open. The algorithms

⁴Mr. Walter Brown, at MIT Lincoln Laboratory

do not suggest many overtakes in their sequences (and very few within stream ones). Nevertheless, those that there are might themselves be difficult to implement for controllers. At the same time, we should mention, that these concerns are all assuming current operational practice. They may be less pressing if advanced radar and computers were brought to bear on terminal area operations. These might give controllers a degree of flexibility and accuracy that they may not currently possess, and hence make such resequencing easier.

The other important question is how much the benefits of sequencing are worth anyway. A reduction of an aircraft's flying time by one minute is not particularly tangible to its passengers. Nevertheless, in terms of annual operating costs, the savings can be large. Further, we must remember that this analysis was performed at Boston, a relatively uncongested airport compared to Chicago O'Hare. There it may be possible to accrue more benefits from sequencing. In heavily congested airports, to use an analogy from queueing theory, even a minor improvement in the service rate (ρ) from sequencing, can lead to significantly reduced delays. Finally, we must keep in mind that the aim of this study was to see if, despite the constraints of the terminal area traffic, gains from sequencing could be realized. While that does indeed appear to be the case, we must carefully evaluate the extent of the benefits, and consider the impacts on the operational preferences and procedures of controllers.

Chapter 6

Conclusion

This thesis examines two important issues relating to terminal area air traffic control operations: the time intervals between aircraft at landing, and the sequencing of aircraft in the near terminal area. The study is motivated by the goals of advanced terminal area air traffic control, which include tighter spacing of aircraft for landing and better sequencing for them. Hence we wanted to perform an empirical study of spacings between aircraft now, and developed a statistical model for the landing time intervals (LTIs) between aircraft, based on data from Boston. Further, we used this model in a study of the aircraft sequencing problem. In this Chapter, we review the main findings of these two main parts of the thesis. For each part (in Sections 6.1 and 6.2), we discuss extensions of the current work and, importantly, for a work of this nature, the applications of our results. We close with some final comments in Section 6.3.

6.1 The Model for Landing Time Intervals

6.1.1 Review of Results

Our model and analysis is based on a very extensive data collection for Logan Airport, Boston. We collected about 70 hours of data by radar and by hand, for a variety of weather conditions and runway configurations. We concentrated our data collection in the evening hours of heavy traffic. We used this data in a two stage approach to develop the LTI model. The first stage involved the calibration of the model and was based on 10

data sets. In the second stage, this model was validated on 8 completely different data sets.

Our analysis focussed on the mean LTI because statistical tests revealed that there was no significant variation in the standard deviation of the LTIs. Further, we only considered LTIs on the primary runway, when two were in use, because the traffic on the secondary is usually light and the time intervals rather large and not very meaningful. The aim of the calibration exercise thus was to determine what factors affect the mean LTI, and to quantify these effects.

We have identified two factors that have a significant effect on the mean LTI. These are the runway configuration (which subsumes the wind and weather variables) and the sequence category of the LTI. We have two very interesting findings with respect to these two factors. The first is that the mean LTI on the primary runway, when two of them are used for landings, is on average larger than it is when only one is used. Given that a single runway landing configuration is typically used in bad weather (IFR), and two-runway configurations are used in good weather (VFR), this finding was interesting. The explanation lies in the distribution of traffic. We showed that, when two landing runways are used, roughly 8-10% of overall traffic is diverted to the secondary runway, and the reduced strain on the primary leads to longer average LTIs there. By contrast, in bad weather, much more traffic lands on the primary, rendering the LTIs there smaller on average. Hence the larger mean LTI on the primary runway in VFR compared to IFR is due possibly to it being less stressed. It remains to be seen whether, if the primary is as stressed in VFR as IFR, the mean LTIs on it would be comparable for the two weather conditions.

The second interesting finding of the study is that the extra wake vortex separations based on weight-class are followed in VFR even when they are not mandatory. We discover that the mean LTI for sequence categories not requiring additional separation is significantly less than it is for those requiring it. Furthermore, this is true in general across all days. This affirmation of the safety consciousness of pilots has an important cautionary effect for the implementation of an advanced Air Traffic Control system. In a system of

the future, one cannot expect pilots to ignore in VFR the extra wake-vortex separations that they maintain in IFR.

We partitioned LTIs into 7 homogeneous types: groupings of sequence categories and runway configurations. These are Base-I, Base-II, H/H, H/L, H/S, Lj/S and Lp/S. The Base grouping comprises those sequence categories requiring only the 2.5 nm IFR separation. The mean Base LTI differs according to whether one or two landing runways are in use. The other LTIs are considered separately, either because they require extra minimum distance separations (H/L and H/S), or because there are different air traffic control procedures used for them (H/H, Lj/S and Lp/S).

The aim of the analysis was to study LTIs as they are now, as well as to develop some idea of what may be appropriate values for an advanced system to use. We have argued that the 25-th percentile of current LTIs might be one such value: it is below the mean and thus would increase throughput, and since pilots and controllers go below it 25% of the time, it may not be unacceptably low to them. With this in mind, we determined three statistics to be of importance and estimated them for each LTI type. The statistics are the mean, the standard deviation, and the 25-th percentile. We compute estimates of these statistics from the calibration data for each LTI type.

Given the fact that many assumptions went into the LTI model, and, had we proceeded differently, that it could have taken a different shape with different estimates of the statistics of interest, it was important to test the model. So we developed a procedure to validate the model. This procedure involved, first, a completely new data collection, somewhat smaller than the first, but large enough to permit statistical testing. We found that this new data lacked in Small aircraft, and had to test only 5 LTI types, omitting Lp/S and H/S. We tested the validation data with the calibration model for each LTI type for each statistic. We found no reason to reject the null hypothesis of equality for any of the respective statistics. Further, we performed a bootstrap analysis to get an overall "judgement" on the issue. Again, we found that there was no reason to reject a null hypothesis of no difference on all counts: mean, standard deviation and 25th percentile. Indeed, the chance of a more extreme difference than that which we observed, given the

null hypothesis, is 84%. This provided reassurance for the methods and results of the calibration model.

It is important to note again that the model could well have taken a different shape had we proceeded differently. And the calibration and validation data could suggest many, different, models. Hence our model is by no means correct in any absolute sense; rather, it is a plausible representation of LTIs at Boston, and is one supported by the data.

6.1.2 Applications and Future Work

The model that we have developed, particularly the numerical estimates, are, of course, valid only for Boston. But the methods that we have used to identify salient factors relating to LTIs and to quantify them could be applied elsewhere. Indeed, the statistical techniques have had to be tailored to the problem at hand, and the methodological questions that have arisen are of interest in themselves. For instance, we developed a statistic based on linear ranks for equality in the 25th percentile. It would be interesting to see if a nonparametric test could be developed that was independent of the assumption of equal ranks. Indeed, if this could be generalized for testing equality for any percentiles of a distribution.

While the study has been performed for Boston, in general one would assume that similar studies would have to be performed elsewhere, should advanced terminal air traffic control techniques become more commonly used. Hence, it would be of great interest to see if some of our findings can indeed be replicated elsewhere. The discovery that extra wake-vortex separations are maintained even in VFR, when they are not required, may arise from safety consciousness on the part of pilots, and hence may be true at other airports as well.

The second interesting finding, regarding the pattern of traffic distribution between primary and secondary runways, and the smaller Base LTIs on the primary in IFR, compared to VFR, would depend, in part, on operations at Boston. Nevertheless, in so far as these findings have run counter to the prevailing notion in Air Traffic Control, that VFR operations are "more efficient" than IFR ones, the issue deserves to be explored more fully

at other airports. For example, is it true in airports where there are parallel operations in IFR, that the LTIs in IFR are larger than those in VFR?

The importance of this question is in setting the Arrival Acceptance Rate (AAR) for the airport. If indeed IFR operations on the primary runway are as efficient as VFR operations, if not more, ought not the AAR to reflect that. At Boston, the IFR single runway operations constitute the bottleneck in airport operations and thus we cannot schedule more planes in VFR to take advantage of relatively underused capacity. Such a schedule would overwhelm the airport in IFR. But in other airports, if there are parallel operations in IFR, and the IFR LTIs are not significantly larger than VFR ones, the AAR should reflect the fact, if it currently does not do so.

Finally, we have identified the runway configuration as a salient variable in determining LTIs. It subsumes both the weather conditions and the wind. These two variables combine to specify the runway configuration. Hence in developing capacity models for an airport, it may be worthwhile to consider runway configurations as a salient factor, as opposed to weather conditions alone. The weather conditions are, of course, important, but along with the wind they determine a runway configuration, which has a significant effect on LTIs. Another attractive reason for using the runway configuration is that there are relatively fewer of them, and many of them have similar capacities. Thus one may not need a very fine grained model based on a variety of weather conditions to model the airport capacity.

6.2 The Aircraft Sequencing Problem

6.2.1 Review of Results

We developed three models for the sequencing problem: the static, the dynamic model with fixed time windows (DASP-1) and the dynamic model with shrinking time windows (DASP-2). The first model (static) assumes that the entry times of aircraft into the terminal area are known with certainty for all the aircraft under consideration. The entry time of an aircraft into the terminal area also gives a lower bound on its landing time at the runway (depending on the entry fix). The static model assumes that aircraft can be

held in stacks and that the upper bound for landing times is very large. We solve the static problem optimally using a Dynamic Programming algorithm.

The dynamic models both assume that the problem is dynamic and an aircraft's entry time is known only when it actually enters the terminal area. In addition, the aircraft can only be delayed by speed control or path length variation within the terminal area. Hence there is no holding at stacks. This specifies a smaller upper bound for the aircraft's landing time. The algorithms for the dynamic problem re-solve, upon the entry of each new aircraft, for the optimal sequence for those currently within the terminal area. The only restriction is that any aircraft that is within a certain distance of the runway has its landing position (and time) fixed. Hence the dynamic algorithm uses the static algorithm as a subroutine each time the sequencing problem is re-solved. Moreover, the dynamic algorithm does not necessarily find the optimal solution because it has no look-ahead capability.

The models DASP-1 and DASP-2 differ in two respects. The first is that the DASP-2 model presents a much more restrictive view of terminal area ATC operations. It limits the degree to which southern arrivals can have their paths stretched. It defines an earlier decision point (the point at which the position in sequence is fixed) for aircraft. The DASP-1 model assumes that the decision point is at 300 seconds before the assigned landing time, whereas the DASP-2 model assumes it is with 20 nm left to fly on the nominal path to the runway. This means that all decisions to fly an extended downwind leg have to be made relatively earlier. Another important distinction between the two models is with respect to the time windows. The DASP-1 model assumes that the bounds remain fixed as an aircraft transits through the terminal area. The DASP-2 model, however, recognizes that the time windows themselves shrink because lost opportunities for speed up or delay or path stretching cannot be regained. Hence the time windows are not fixed, and are termed to be shrinking.

We apply the three algorithms to 4 data sets from Boston. We use actual air traffic data, and compare the sequences suggested by the algorithm to those actually used by controllers. Our aim is to minimize delay to aircraft. Delay is defined as the difference

between an aircraft's earliest possible landing time and latest possible landing time. We develop three different objective functions for delay—minimizing cumulative delay, passenger delay and cost of delay—and three different types of LTI constraints, two of them (mean and 25th percentile) based on the model developed previously, and the third on an assumed strict adherence to separation minima.

We discover, for the variety of objective functions and constraints, there can be gains to be had from better sequencing. The gains accrue because of better timing, control and physically better sequences. One surprising result is that the static and DASP-1 algorithms perform equally well for the cumulative delay objective. One reason is that the upper bounds on landing are not binding. The other reason is that the relative paucity in the mix of aircraft and the equivalent LTIs for a variety of sequence categories, mean that the restrictions of the dynamic problem (less information on arrival times) do not affect the objective. If the objective function weights delays to different aircraft differently (as the passenger and dollar delay one do) the DASP-1 algorithm does perform worse than the static algorithm.

Equally surprising is that the DASP-2 algorithm does not perform substantially worse than the DASP-1 algorithm. The restrictions do affect traffic in that aircraft from the southern fixes get delayed more on average by DASP-2 than by DASP-1. This is because the algorithm, with no look-ahead, hastens the northern aircraft to land since it sees no southern aircraft competing for similar landing times. When the southern aircraft do enter, later, whereas DASP-1 may have delayed the northern arrivals if it were better to do so, DASP-2 with shrinking time windows and earlier decision points cannot. Nevertheless, even DASP-2 indicates that gains are to be had from better sequencing.

Of course, the controllers may have other concerns with the sequences suggested. A primary concern is workload. Even if there are gains to be had from better sequencing, the increase in workload to a controller in an already stressed environment may counter these gains. We examined this issue by looking at the number of overtakes that occur, and found that relatively few of them do take place. Moreover of those that occur, the within stream ones, which require more attention from the controller, are a small proportion of

the total. Nevertheless, if the algorithm were to be applied and, from the benefits we observe, there may be reason to do so, the issue of workload would be important. Hence it would be valuable to do a more realistic simulation with controllers to see what they may not like about the sequences suggested.

Finally, there is the important question of what a reduction of delay by some 30 seconds or a minute might mean to an aircraft on a flight of over one hour. In other words, are the reductions in delay we have noticed really minor? It can be argued both ways. The amount per flight is quite intangible, but nevertheless operational savings, especially when cumulated over a year, can be high. Further, we must remember that the aim of this study was to see if, despite the constraints of the terminal area traffic, gains from sequencing could be realized. That does indeed appear to be the case. Moreover, these gains may be more at facilities that are more congested than Boston, where even minor reductions in the service rate (ρ) of the airport, through sequencing, could lead to significantly reduced delays.

6.2.2 Future Work

Algorithmically, the sequencing method may benefit from having some limited look-ahead capability. We do not currently solve the problem to optimality in the dynamic cases. But the fact that the dynamic sequence equals the static one exactly when minimizing cumulative delay suggests that the marginal improvement may not be valuable for this problem. Another algorithmic extension is to the multiple runway case. Here we not only decide the landing sequence, but assign each aircraft to a particular runway. Psaraftis examined this for his problem with all aircraft on stack and no time windows. Given the more realistic model we have developed, though, the multiple runway algorithm would also have to incorporate time windows. This may indeed restrict the problem size, since not all assignments to individual runways would be feasible.

As far as testing for implementation goes, the algorithm is far from complete. The method has only been implemented for landings on 4R. It needs to be generalized at Boston for other runways as well. Further, the model has no inbuilt artificially intelligent

method of knowing if certain resequences would increase workload tremendously. This would ideally be incorporated on the basis of elaborate testing with controllers. But the algorithm and method are sufficiently general, and specific models easily developed for individual airports using the time windows concept, that sequencing can be studied at other locations as well. Each model would be specific to a given airport, but the algorithm would be general.

Lastly, there is the question of look-ahead and the sequencing horizon. We have assumed that all sequencing was taking place in the relatively restrictive confines of the terminal area. In an advanced system, sequencing could commence much earlier, before the aircraft was within the TRACON. This might result in better sequences, and more easily implementable sequencing changes. Also, this opens the possibility of using data on arrival times of aircraft from the Central Flow Control Facility for earlier sequencing.

6.3 Final Remarks

In conclusion, this thesis has sought to apply the techniques and methodology of Operations Research and Statistics to an important problem related to the future of air traffic control in the terminal area. We have built a model to describe the landing time intervals between aircraft at Boston, and studied the effects of better sequencing. The insights that we have gained, and the models that we have developed, should be of use in designing an advanced ATC system for Boston. Yet much field testing needs to be done before implementation, especially with respect to the sequencing algorithm, and determining the appropriate LTIs for an advanced system.

Appendix A

Summary of Data Sets Used for Calibration

We summarize below eight of the data sets collected for calibration (the other two are described in section 3.2). We give the number of aircraft landing on each runway, the AAR's for the primary and secondary runways, and weather and wind conditions. We also describe any peculiarities of the day, such as frequent runway configuration changes or unusual weather patterns. Thus, in addition to supplying quantitative information on the amount and type of data collected, we hope to convey a qualitative sense of the aviation operating condition under which we obtained them.

9-13-87:

The data was collected by hand in IFR conditions from 4:00 p.m. to 9:00 p.m. Visibility was between $\frac{1}{8}$ statute miles in the first hour, $\frac{3}{4}$ statute miles for the next hour, and $\frac{1}{2}$ statute miles after that. The ceiling was 500 feet. Thus the weather conditions were IFR2-3. The wind for the first $3\frac{3}{4}$ hours was from 110° at 5 knots. It then changed to 130° at 7 knots. The runway configuration for landing correspondingly changed from 4R to 15R. Fifty minutes later it changed to 22L. Since there was no change in wind, this latter change may have been to distribute noise equitably among surrounding communities—15R is a particularly unsavory runway for reasons of noise pollution: the landing path is over land largely. The AAR was 34 except when 15R was used, when it was 30. The actual number of landings was at or above the AAR. There were 176 landings in all, 136 on 4R, 28 on

15R and 12 on 22L.

Time (P.M.)	Total Number of Landings	Runway		
		4R	15R	22L
4:00 - 5:00	36	36	0	0
5:00 - 6:00	39	39	0	0
6:00 - 7:00	34	34	0	0
7:00 - 8:00	34	27	7	0
8:00 - 9:00	33	0	21	4

Table A.1: Number of Landings on 9-13-87 by Runway and Time

10-11-87:

The data was collected in IFR conditions from 4:00 p.m. to 9:00 p.m. The visibility was between 1.1 statute miles and 1.3 statute miles. The ceiling varied from 1,000 feet to 600 feet. Thus the conditions were IFR-1/IFR-2. The wind was from 20° at 11 knots to 360° at 13 knots. There were ILS approaches to 4R throughout this period. There were about about 8 transition landings to 4L in the first hour, when the ceiling was highest. Then there were none. The AAR was 34 planes per hour throughout. The actual number of arrivals fell below the AAR in the first hour, but was at or above it afterwards. There were 175 landings during the observation period.

Time (P.M.)	Total Number of Landings (Rwy 4R)
4:00 - 5:00	30
5:00 - 6:00	34
6:00 - 7:00	39
7:00 - 8:00	38
8:00 - 9:00	34

Table A.2: Number of Landings on 10-11-87 by Runway and Time

5-13-88:

The data was collected by radar, in excellent VFR conditions, from 4:00 p.m. to 8:00 p.m. The runway configuration used for landing was 22L/R (with AAR 34/10) for the first 3 hours and 20 minutes. Then on it was 27/22L (with AAR 34/26). The number

of landings on the primary was at or above the AAR for that runway, but the secondary (22R initially, and 22L after 7.20 p.m.) was not used to such a level in the last two hours.

There were 111 landings in all, with 121 on 22L, 28 on 22R and 22 on 27.

Time (P.M.)	Total Number of Landings	Runway		
		22L	22R	27
4:00 - 5:00	45	34	10	1
5:00 - 6:00	44	33	11	0
6:00 - 7:00	38	35	3	0
7:00 - 8:00	44	20	3	21

Table A.3: Number of Landings on 5-13-88 by Runway and Time

5-18-88:

The data was collected by hand in IFR conditions from 4:00 p.m. to 7:00 p.m. The ceiling was 600 feet overcast and visibility between 1.5 statute miles and 1 statute mile. This implies IFR-2 conditions. The wind was from 040° at 14 knots. Aircraft landed on 4R throughout the day. The AAR was 34 planes per hour, but more landed in each hour, 113 landing in all.

Time (P.M.)	Total Number of Landings (Rwy 4R)
4:00 - 5:00	35
5:00 - 6:00	42
6:00 - 7:00	36

Table A.4: Number of Landings on 5-18-88 by Runway and Time

3-30-89:

Data was collected by radar between 3:00 p.m. and 6:00 p.m. under very poor weather conditions. There were heavy thunderstorms with driving rain. A thunderstorm front extended 100 (statute) miles from SW to NE; moreover, it was stationary in the region of the airport. The cloud ceiling was 500 feet, thus weather conditions were IFR-2. For the first 25 minutes, 15R was used for landings. It accommodated 15 aircraft. In the remaining 2.5 hours, 4R was used for landings, and it accommodated only 64 of them.

There were periods as long as 20 minutes without a landing. On a few occasions, pilots refused to land, and had to be returned to hold at an entry fix after they were well within the TRACON area. This was the worst weather encountered in any data set.

Time (P.M.)	Total Number of Landings	
	15R	4L
3:00 - 4:00	15	12
4:00 - 5:00	0	24
5:00 - 6:00	0	28

Table A.5: Number of Landings on 3-30-89 by Runway and Time

5-31-89:

The data was collected in intermediate IFR-VFR conditions, with fluctuating runway configurations. The collection period went from 3:00 p.m. to 7:00 p.m. Visibility was 8 statute miles and the ceiling 2,000 feet. The wind was from 110° at 6-8 knots. This remained relatively unchanged throughout the collection period. The initial landing configuration was 4R/15Rc4L. At 5:48 p.m., however, a tanker carrying natural gas passed through a channel on the approach to 4R and bordering its threshold. The arrival configuration was thus changed to 15R. After the tanker passed through the channel, a half an hour later, the configuration changed again to 4R. But the 15Rc4L approach to the secondary runway was closed shortly because of a drop in the ceiling. There were 143 landings during the collection period. 114 were on 4R, 21 using the circling approach to 4L, and 16 on 15R. The AAR was 34/10 when 4R/15Rc4L was used and 34 when only 15R or 4R accepted landings. The AAR was not reduced to the customary 30 for 15R was used because it was only to be for a short while.

6-09-89:

The data was collected between 2:00 p.m. and 5:00 p.m. in poor IFR conditions. The visibility was 7 statute miles but the cloud ceiling only 1000 feet. There was intermittent rain and thunderstorm warnings were in effect. The wind was from 070° at 10 knots. The arrival runway was 4R and the AAR 34 planes/hour throughout. The airport was

Time (P.M.)	Total Number of Landings	Runway		
		4R	4L	15R
3:00 - 4:00	34	28	6	0
4:00 - 5:00	29	21	7	0
5:00 - 6:00	41	29	7	5
6:00 - 7:00	39	26	1	11

Table A.6: Number of Landings on 5-31-89 by Runway and Time

operating at approximately this level. There were 102 arrivals in in the collection period.

Time (P.M.)	Total Number of Landings (Rwy 4R)
2:00 - 3:00	36
3:00 - 4:00	33
4:00 - 5:00	33

Table A.7: Number of Landings on 6-09-89 by Runway and Time

6-15-89:

Data was collected from 5:00 p.m. to 8:00 p.m. The weather was very poor with rain and fog. The ceiling was 600 feet and the visibility was 1.5 miles. Thus conditions were IFR-2. The wind was at 14 knots from 110° . Runway 4R was used for landings with an AAR of 34 planes an hour. The number of arrivals was above this value at all times. There were 110 arrivals during the collection period.

Time (P.M.)	Total Number of Landings (Rwy 4R)
5:00 - 6:00	36
6:00 - 7:00	36
7:00 - 8:00	38

Table A.8: Number of Landings on 6-15-89 by Runway and Time

6-28-89:

This data set was collected in excellent VFR conditions with visibility over 10 miles and the cloud ceiling above 10,000 feet. The data was collected between 3:00 p.m. and

6:00 p.m. The wind was at 9 knots from 70° . The airport was operating at its highest capacity runway configuration, 27/22L, with AAR of 34/26. The system was never really stressed either on the primary runway or the secondary. There were 124 landings in all, 73 of them on 27 and the remaining 51 on 22L.

Time (P.M.)	Total Number of Landings	
	27	22L
3:00 - 4:00	21	14
4:00 - 5:00	31	16
5:00 - 6:00	21	21

Table A.9: Number of Landings on 6-28-89 by Runway and Time

Appendix B

Descriptions of Statistical Tests

In this Appendix, we give technical descriptions of the statistical procedures and tests used in this thesis. Each of the tests is listed in a different section.

B.1 Difference of Means Test

The test works as follows. Suppose we are given two data sets, denoted X and Y respectively, with n and m data points each. We wish to test whether $H_0 : \mu_X = \mu_Y$ is true or not. We estimate the true means μ_X, μ_Y by the sample means \bar{X} and \bar{Y} respectively. Likewise we estimate the true variances σ_X^2 and σ_Y^2 by the sample variances s_X^2 and s_Y^2 , respectively, where, for example,

$$s_X^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n - 1}.$$

We take the test statistic Z defined as

$$Z = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \tag{B.1}$$

We note by a Central Limit Theorem (CLT) based approximation that Z is distributed as a standard Normal. The CLT approximation is quite accurate when the combined sample exceeds 15 data points (see Hamburg (1983), [15]). Thus at a 5% level, one would reject the null hypothesis, using a two-tailed test if $\Phi(|Z|) \geq 0.975$.

B.2 The C-test

We wish to aggregate the results of difference of means tests on individual pairs of data sets, and make an aggregate statement on the overall data. The test procedure that we develop is, in general, as follows. Assume that we have N sets of data x_i (e.g. Base LTIs) and y_i (e.g. Big LTIs), where i denotes the data set, $i = 1, \dots, N$. Each data set x_i (or y_i) has n_{x_i} (or n_{y_i}) sample values. These sample sizes are not necessarily equal. We wish to test if the true means $\mu_{x_i} = \mu_{y_i}$ for all i . One easy way is to take the test statistic, C , of the sum of the differences in the individual sample means (\bar{x}_i and \bar{y}_i) over all data sets. We normalize each term by the sum of respective sample variances. Thus we have that

$$z_i = \frac{(\bar{x}_i - \bar{y}_i)}{\sqrt{\frac{\sigma_{x_i}^2}{n_{x_i}} + \frac{\sigma_{y_i}^2}{n_{y_i}}}} \quad (\text{B.2})$$

Then we have that

$$C = \frac{\sum_i^N z_i}{\sqrt{N}} \quad (\text{B.3})$$

We wish to test the null hypothesis that $C = 0$.

The distribution of each z_i should be a standard normal, by an approximation based on the central limit theorem, for large enough sample sizes. This is because it is the average of many values. As long as we have 12 -15 values in all ($n_{x_i} + n_{y_i}$), the approximation would be reasonable. C is the (normalized) sum of independent¹ standard normals, and thus is distributed as $N(0, 1)$. It is then a simple matter to see if $\Phi(|C|) > 0.975$, using a two-sided test at the 5% level. Note that if any x_i or y_i have no sample values,² then the i th row is ignored in the test.

B.3 Wilcoxon Rank-Sum with Blocked Treatments

The blocked treatment Wilcoxon procedure works as follows. Let there be b data sets (or blocks), each consisting of two types of observations, termed controls and treatments for

¹The data sets are taken on different days and are quite independent of each other.

²Actually, less than 2 sample values—in order to compute a variance.

ease of exposition. The total number of observations in data set i is N_i , with n_i treatment values and m_i control values. Thus:

$$N = N_1 + \cdots + N_b; n = n_1 + \cdots + n_b; m = m_1 + \cdots + m_b \quad (\text{B.4})$$

We denote the treatments ranks from the i th block by $S_{i1} < \cdots < S_{in_i}$. Let

$$W_s^i = S_{i1} + \cdots + S_{in_i} \quad (\text{B.5})$$

We use the test statistic

$$W_s = \sum \frac{W_s^i}{N_i + 1} \quad (\text{B.6})$$

This is the weighted sum of the rank-sums from each block. Even for a relatively small number of blocks, as we have, we can use a normal approximation for the distribution of W_s [17]. Using this approximation, we have that under the null hypothesis of no treatment effect, $E(\frac{W_s^i}{N_i+1}) = \frac{n_i}{2}$ and $Var(\frac{W_s^i}{N_i+1}) = \frac{m_i n_i}{12(N_i+1)}$. We can then easily compute $E(W_s)$ and $Var(W_s)$ using, for the latter, the fact that the results in each block are independent.

B.4 A Rank-based Test for Equality in 25th Percentiles

Consider the following problem: there are two data sets, X with m observations, and Y with n observations. We wish to test the null hypothesis H_0 : the 25th percentile of X = the 25th percentile of Y . A review of the statistical literature did not reveal any test that has been designed specifically for this purpose. Hence we need a test that is sensitive to differences specifically at the 25th percentile.

We now present a nonparametric test which we have designed to test for equality in the 25th percentile between two sets of data. The test, which we shall refer to as the Percentile test, is based on ranks, like the Wilcoxon Rank-Sum test. It works as follows.

Consider once more the two data sets, X and Y defined earlier. We assume that X and Y arise from continuous distributions $F(x)$ and $F(x - \Delta)$ respectively. The parameter Δ is known as the shift parameter. Thus if $\Delta > 0$, the Y 's are larger than the X 's. Under the null hypothesis, we assume that X and Y are independent and that $\Delta = 0$. We will

develop a test that is designed to capture the alternate hypothesis that $\Delta \neq 0$ and that is sensitive to differences in the 25th percentiles of X and Y . The intuition underlying the test that we use is simple. We combine the data, and compute the 25th percentile of the combined sample. We then compute the proportion of the Y (we could, equally, use X) data that is above this value. If this proportion is “too high” above, or “too low” below expected value of $\frac{3}{4}n$, then we can reject the null hypothesis.

We define the variable R^* as the rank vector of the variables $X_1, \dots, X_m, Y_1, \dots, Y_n$ with $N = n + m$. Thus R_1^*, \dots, R_m^* denote the joint ranks of X_1, \dots, X_m respectively, and R_{m+1}^*, \dots, R_N^* the joint ranks of Y_1, \dots, Y_n . We have assumed continuous distributions for X and Y and thus we suppose, without loss of generality, that there are no tied ranks. Under the null hypothesis, $\Delta = 0$, the joint ranks are a sampling from a single continuous distribution. This means that each of the $N!$ possible orderings of the ranks is equally likely. In other words, the vector R^* is distributed uniformly over \mathcal{R} , the set of permutations of the integers $1, \dots, N$.

We define z as the rank of the largest element at or below the combined 25th percentile.

$$z = \begin{cases} \frac{N+1}{4} & \text{if } (N+1) \bmod 4 = 0 \\ \lfloor \frac{N+1}{4} \rfloor & \text{otherwise} \end{cases} \quad (\text{B.7})$$

We use the following test statistic:

$$S = \sum_{i=1}^N c(i)a(R_i^*) \quad (\text{B.8})$$

where

$$c(i) = \begin{cases} 0 & i = 1, \dots, m \\ 1 & i = m + 1, \dots, N \end{cases} \quad (\text{B.9})$$

and

$$a(i) = \begin{cases} 0 & i = 1, \dots, z \\ 1 & i = z + 1, \dots, N \end{cases} \quad (\text{B.10})$$

Thus S is just the number of Y elements that have a joint rank above the 25th percentile of the combined sample. We require the precise notation above to account for cases when the 25th percentile does not fall precisely at a data point.

S as defined above is a linear rank statistic (for details, see Randles and Wolfe, [27]). We can derive $E(S)$ and $\sigma^2(S)$ from the theory of linear rank statistics. An easier way, however, is to note that, because each possible ordering of ranks is equally likely, the distribution of S is hypergeometric:

$$P(S = k) = \frac{\binom{n}{k} \binom{m}{(N-z)-k}}{\binom{N}{(N-z)}}, k = 0, \dots, \min(N-z, n) \quad (\text{B.11})$$

Further, we have (see Drake, 1967 [9]) that

$$E(S) = \frac{n}{N}(N-z) \quad (\text{B.12})$$

and

$$\sigma^2(S) = \frac{nmz(N-z)}{(N-1)N^2} \quad (\text{B.13})$$

We note that if $z \sim \frac{N}{4}$, then $E(S) = \frac{3}{4}n$, which makes intuitive sense under the null hypothesis.

Informally, we will reject the null hypothesis if the sample value of S is not “close enough” to its expected value. For the values of N that we shall be dealing with (greater than 25) we can invoke the Central Limit Theorem to approximate the hypergeometric with a normal of the same mean and variance. Thus, denoting \hat{S} as the sample value of S , and $p = \frac{\hat{S}-E(S)}{\sigma(S)}$, we will reject the null hypothesis at a 5% level if $\Phi(|p|) \geq 0.975$.

This test, which we refer to as the percentile test, can be used to test whether the 25th percentiles of each of the LTI types in calibration and validation are significantly different or not. Further, they can be combined (by adding the P values, and comparing to an appropriate normal distribution) to test the overall strength of the validation in this respect. This test is similar to Mood’s test for equality in medians (see Hettmansperger (1984), in which he compares the actual proportion of one data set which is above the combined median.

While the percentile test has some intuitive appeal, it is based on a rather strong assumption: that every possible ordering of the ranks is equally likely. This derives from

the basic assumption that the distribution of Y is just the distribution of X shifted over by some Δ , i.e. X and Y arise from continuous distributions $F(x)$ and $F(x - \Delta)$ respectively. Of course, this presupposes, at a fundamental level, that the shapes of the distributions of the random variables X and Y are identical, which in a realistic problem need not be the case. For example, if the range of Y is contained in the range of X , then items with extreme joint ranks are less likely to be from Y than from X . The question then is how badly does the percentile test perform in pathological instances, when the distributions of X and Y differ considerably.

We performed a series of tests on uniform distributions. Let us fix X to be $U(0, 100)$, and Y also to be $U(a, b)$ such that the 25th percentile of the distribution is 25 (i.e. $a + \frac{b-a}{4} = 25$). We wish to judge the performance of the Percentile test in comparing the 25th percentiles of X and Y (which we know are equal). But we want to do this for values of a and b , the upper and lower limits, respectively, of the Y distribution, that result in a markedly different distribution from that of X .

We perform the following simulation: we draw a random sample of size 100 each from X and Y , and perform a two-sided Percentile test at the 5% level. We repeat this procedure 1000 times and tally the value of R , the total number of times we reject the null hypothesis of no difference in 25th percentiles. We know that the null hypothesis is true, and that on any particular $P(\text{False Rejection} | H_0) = 0.05$. Thus, if the test works correctly, R is distributed binomially with parameters 1000 and 0.05, since each test is independent of the previous one, using different data. We can make a normal approximation to the Binomial (since we have 1000 tests, and the p parameter is not too low). Hence R is distributed normally with $E(R) = 1000 \times 0.05 = 50$ and $\sigma^2(R) = 1000 \times 0.05 \times 0.95 = 47.5$ (or $\sigma(R) = 6.89$). We wish to perform a right-tailed test on R , implying that H_0 is being rejected more frequently than we would prefer. So, at a 5% level, if the test is performing well for a given form of the distribution of Y , we should see less than $50 + (1.96 \times 6.89) \sim 64$ rejections.

Table B.1 gives the actual value for R for three simulations carried out with successively narrower (hence more skewed relative to X) distributions for Y . The results indicate that

<i>X</i> Dist.	<i>Y</i> Dist	R
$U(0, 100)$	$U(10, 70)$	40
$U(0, 100)$	$U(15, 55)$	62
$U(0, 100)$	$U(20, 40)$	70

Table B.1: Number of Rejections of H_0 in 1000 Tests

the percentile test performs quite well for quite skewed distributions of Y relative to X . The test performs acceptably (at a 5% level of acceptance) when $Y = U(10, 70)$ and when $Y = U(15, 55)$, with 40 and 62 rejections respectively. When $Y = U(20, 40)$, we reject the null hypothesis 70 times, indicating that the result is becoming sensitive to the assumption of all ranks being equally likely. Nevertheless, these distributions are more skewed than what we expect from the LTI data. Clearly the simulation results, are limited in scope—they compare two distributions of the same type, and do it only for uniform ones at that. Nevertheless, they provide some reassurance that the test does not fail miserably in circumstances that differ quite considerably from those which are assumed to hold. Thus, in comparing 25-th percentiles of the calibration and validation LTIs, we shall be using the Percentile test.

B.5 The Bootstrap Method

We illustrate the Bootstrap method [10] with the following problem. Given a random sample $X = X_1, \dots, X_n$ from some unknown probability distribution function, F , we wish to estimate the sampling distribution of some prespecified random variable $R(X, F)$ on the basis of the observed data. We let $X = X_1, \dots, X_n$ and $x = x_1, \dots, x_n$ denote the random sample and its observed realization respectively.

$$X_i = x_i, X_i \sim_{ind} F, i = 1, 2, 3, \dots, n. \quad (\text{B.14})$$

As an example, if μ_F is the mean of F , we may wish to estimate the bias in the data:

$$R(F, X) = \bar{x} - \mu_F \quad (\text{B.15})$$

However, we do not know what F is. So we estimate $R(F, X)$ using the Bootstrap method (see Efron (1979) [10]).

The Bootstrap method works as follows:

1. Construct a sample probability distribution \hat{F} , putting mass $1/n$ at each point x_1, \dots, x_n .
2. With \hat{F} fixed, draw a random sample of size n with replacement from \hat{F} . Call this sample distribution X^* with realization x^* .

Thus we have:

$$X_i^* = x_i^*, X_i^* \sim \text{ind}\hat{F}, i = 1, 2, 3 \dots, n. \quad (\text{B.16})$$

We call this the bootstrap sample.

3. Approximate the sampling distribution of $R(X, F)$ by the Bootstrap distribution of $R^* = R(\hat{F}, X^*)$.

In our example, $R(F, X)$ is as in (B.15), and

$$R^*(\hat{F}, X^*) = \bar{x}^* - \bar{x} \quad (\text{B.17})$$

The essence of the method is that the distribution of R^* , which can be calculated exactly if data X is observed, equals the desired distribution R , if $F = \hat{F}$. This is true because $\hat{F} \rightarrow F$ as $n \rightarrow \infty$. Hence, ideally, one should have large sample sizes.

Appendix C

Calculating Times-to-fly from Fix to Runway

C.1 Preliminaries

Jet traffic enters the Boston TRACON from three directions: over the fixes Providence, BRONC¹ and SCUPP. Propeller traffic enters the area at five entry fixes, EXALT, WOONS, FREDO, LOBBY, KIRIS. Figures C.1 and C.2 show the entry fixes and flight paths to runway 4R for jets and props respectively. We shall refer to these figures in detail later. The air traffic controller who does the sequencing for landings, and merging for final approach, is known as the final vector controller.² This controller “sees” the aircraft as they enter the TRACON. There is usually then a commitment to land the aircraft without returning it to the entry fix for holding.

If the traffic is light, then an aircraft may travel at a high speed along the shortest path to the designated runway for landing. Usually, however, aircraft can be delayed (given speed reductions) to give way in the landing stream to another aircraft from another fix. These delays follow prescribed conventions depending on the landing runway and the entry fixes of the aircraft. For example (from Figure C.1), for landings on 4R, aircraft from Providence are given priority for landings as they have a short, straight run from the fix

¹The traffic is from the west and enters over the VOR station at Gardener, but the entry fix is called BRONC.

²If holding stacks are in use, there is typically only one.

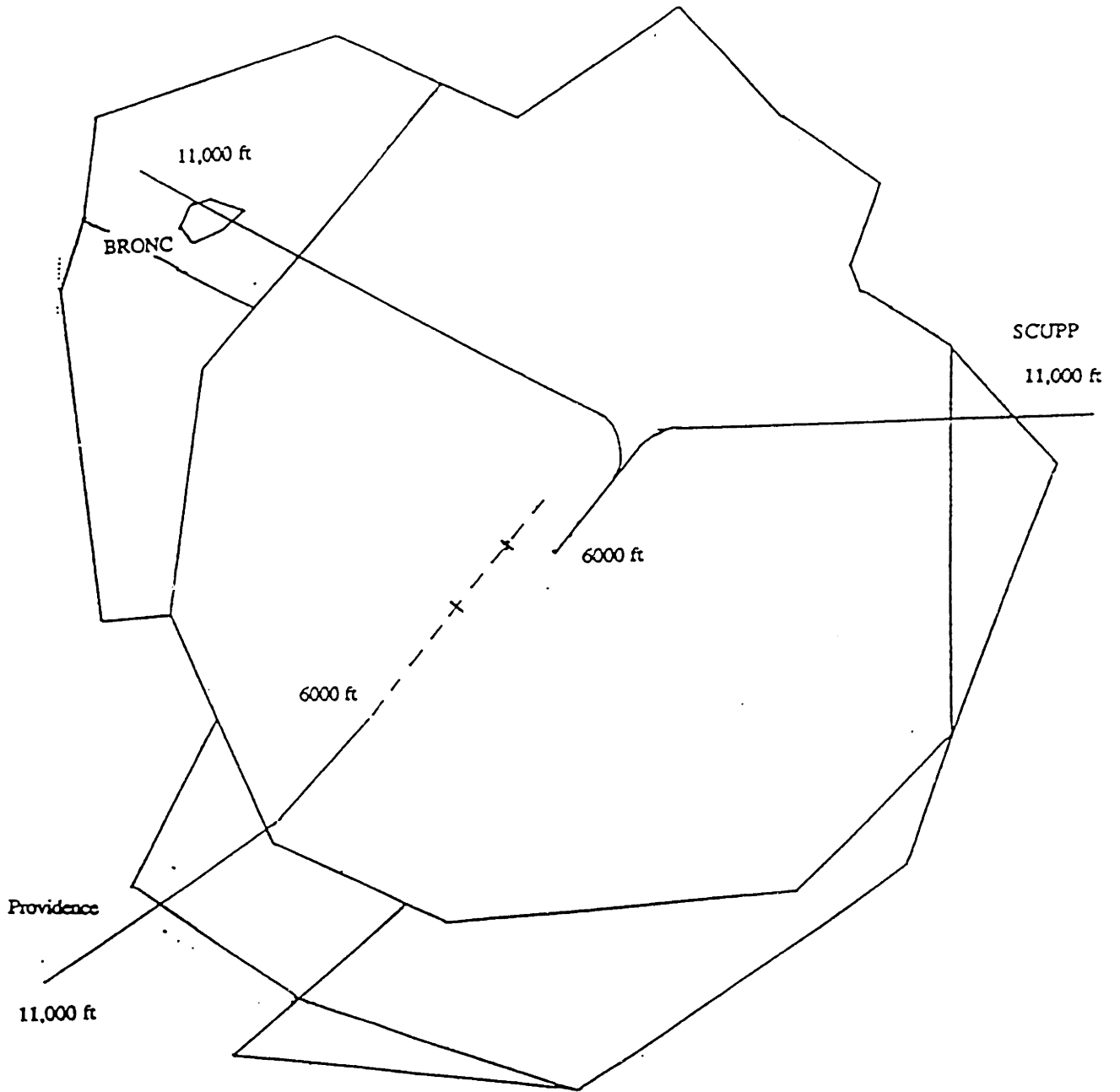


Figure C.1: Jet Approach Routes to Runway 4R at Boston

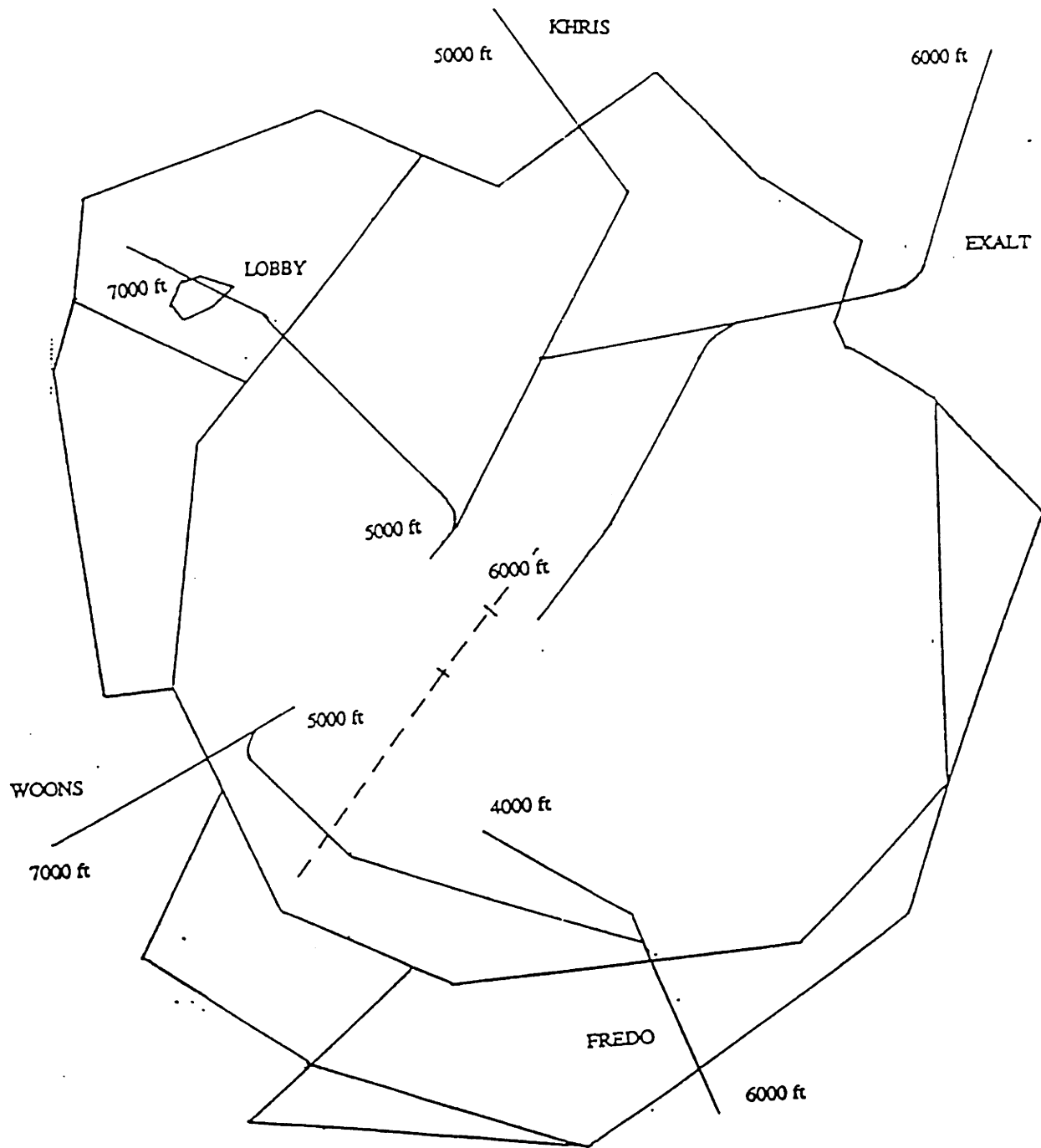


Figure C.2: Prop Approach Routes to Runway 4R at Boston

to the runway. Aircraft from Gardener and Scupp are merged into a single stream which is itself merged into the landing stream from Providence.

When congestion increases, the aircraft along each individual stream may be made to do any or all of go slower, travel more circuitous routes, and merge at a point further from the runway. The time taken for an aircraft to land given that it has just entered the terminal area thus has a lower bound and an upper bound. The lower bound corresponds to the fastest that the aircraft can transit the TRACON to landing; the upper bound corresponds to the most that it can be delayed. These bounds depend on the performance characteristics of an aircraft (typical cruise speed, for example) and also, as explained, on the runway in use and the entry fix. These time-to-fly bounds, when added to the entry time of an aircraft into the terminal area give the time windows for its landing time.

In this Appendix, we discuss, for traffic landing on runway 4R,³ how we obtain the time-to-fly (TTF) for aircraft from entry fixes to the runway. The TTF usually has a typical or nominal value, which can be increased or decreased as discussed above; the upper and lower bounds for the TTF help define the time windows within which aircraft must land, given their entry times. We wish to know the TTF and its bounds, for both jet and propeller traffic, from their respective entry fixes to the runway threshold of 4R. To be more specific, we are interested in the TTF from a circular marker (set up by GATEX) of 30 nm radius centered at the intersection of 4R and 9 at Logan airport, to the 4R threshold. The entry fix is important because it determines the flight path to the fix. Most of the fixes are just outside the 30 nm marker, but some propeller fixes (e.g. EXALT) are just within it.

Essentially, the TTF is determined by the typical flight path (in space) from the fix to the runway. The TTF is primarily based on knowing the length of various segments and at what speeds they are traversed. Methods developed at Lincoln Labs by Sturdy et. al. (1989) [31] can be used to fine tune this rough measure of TTF by considering the effect of aircraft weight, the relationship between actual ground speed and the Indicated Air Speed (IAS), the wind and the altitude, and also the geometry of the turn to final

³We explain our focus on this runway shortly.

approach for landing.

This methodology enabled Sorensen (1990) [30] to use the typical flight paths in space from the fixes to the runways, the typical speeds along various segments, and standard ways in which controllers delay or speed up aircraft, all to compute the nominal TTF and its upper and lower bounds for various types of aircraft and various fixes. Unfortunately, however, Sorensen computed these values only for a typical Boeing 727-200 Series type jet landing on 4R, from the fixes SCUPP and PVD (Providence). Since we require this information for the other jet fix (BRONC) and for prop arrivals from their respective fixes, we have to make estimates for them.

We also consider only 4R as Seagull provides enough basic information about the landings on this runway that we can make inferences about TTFs for arrivals to it from all fixes. For other runways, however, a paucity of information (relating specifically to flight paths and aircraft controllability) make such TTF estimation much more difficult. In Sections C.2 and C.3, we explain how we calculate the values of TTF, with upper and lower bounds, for SCUPP and Providence entrants respectively. Then we discuss (in Section C.4) how to estimate them for arrivals from the third jet fix, BRONC, and for prop arrivals.

C.2 The TTF From SCUPP to 4R

Figure C.3 shows the typical flight path from SCUPP to 4R. The numbers on the graph indicate altitude points. Also, on the graph are denoted the three last stages of the landing procedure: the downwind leg (against the direction of final landing, and hence with the wind), the base-leg (perpendicular to the runway landing direction) and the final approach for landing. On final approach the aircraft has intercepted the localizer for the runway in question and it descends along the glide slope (roughly 3°). This path in Figure C.3 is a typical manoeuvre for aircraft landing from fixes to the north of the airport. We notice that the aircraft enters the terminal area at about 11,000 ft. and at 250 knots.⁴ The

⁴This is the Indicated Airspeed (IAS), but we assume that there is no wind, and that the IAS is thus equal to the true groundspeed.

altitude decreases gradually along the flight path to about 5,000 ft. at which point the downwind leg commences. The speed of the aircraft also decreases to about 210 knots for the downwind leg. Aircraft from BRONC and prop aircraft from northern fixes will merge with this landing stream at about the point that the downwind leg commences. The props enter the terminal area at about 6,000 ft. and typically travel at 200 knots;⁵ thus they do not have to reduce much speed or altitude by the time they reach the merge point. We shall assume that the props behave identically to the jets from this point on. This is obviously a simplification, but it may not be too many inaccurate since aircraft weight and performance seem only to have a second order effect on the TTF, as we shall discuss later.

Given this description of a typical flight path, we can list the strategies which are mainly used for timing control. They are control of:

1. The points along the aircraft's projected flight path where it is commanded to decelerate.
2. Path stretching or corner stretching to create an extended downwind leg (this would apply to aircraft along the flight path outlined for SCUPP).
3. Path stretching for an aircraft flying nearly straight in or on a "dog-leg" approach (e.g. from Providence to 4R).

There are other methods for fine tuning, such as varying the rate at which speed is reduced, or the angle of turn to capture the localizer. But these do not yield much in terms of time differences, as we shall see.

Given these alternatives, Table C.1 shows the variation in TTF from SCUPP to 4R (from Sorensen (1990)). We assume no wind or variation in aircraft weight. From this, we see that the nominal TTF is 970 seconds, but it can be increased by 89 seconds by flying slowly (at 210 knots) after entering SCUPP, and by a further 580 seconds with a long downwind leg, giving an upper bound of $970 + 89 + 80 = 1639$ seconds. Also the nominal TTF can be decreased by 80 seconds by flying fast into the base and flying a base close to

⁵This is true of the "high-performance" props which are the most common of their type at Logan: ATR42, BE1900, DO28, DASH8 etc.

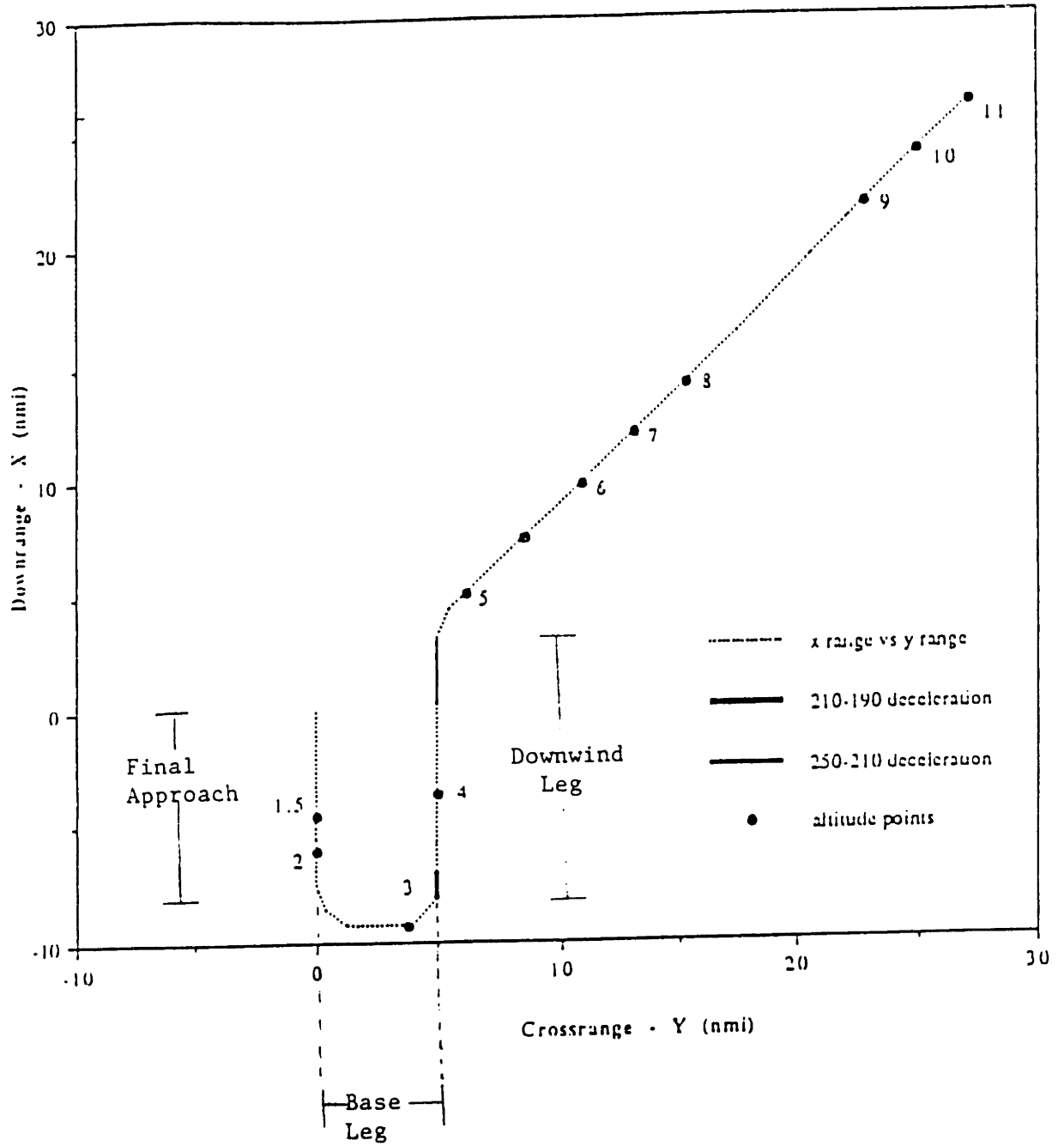


Figure C.3: Flight Path from SCUPP to 4R

Scenario	TTF (sec)	Δ TTF
1. Nominal - Base 10.8 nm from Rwy Threshold	970	0
2. Fast - 250-210 decel right before base turn	948	-22
3. Slow - 250-210 decel right after SCUPP	1059	89
4. Short Downwind Leg (DWL) - Base 9.3 nm from threshold	905	-65
5. Short DWL Fast	890	-80
6. Max Long DWL - Base 28.4 nm from threshold	1550	580

Table C.1: Variations in TTF for the SCUPP Approach to 4R

the threshold, giving a minimum of 890 seconds. Thus the bounds on the TTF are (890, 1639), or roughly between 15 min. and 27 min.⁶

The values in the above table, however, are from the entry fix to the runway, and not from the 30 nm circular marker (our point of interest) to the runway. The entry fix is important because it determines the flight path of the aircraft between the 30 nm marker and the runway. Thus what we also wish to find out is the time to fly from the entry fix to the 30 nm marker. This is obtained from the data in Table C.1 along with a navigation map for Boston. All distances and flight paths discussed in this Appendix have been deduced from Figures C.1 and C.2 and from a navigation map for the Boston area [36]. Naturally, a scope for error is in trying to relate paths from a map with no scales (e.g. Figure C.1 or C.2 to paths on a detailed one with them. But in doing this we are helped by Figure C.3 (and Figure C.4, its counterpart for Providence) which show the distance travelled from the fix. Further, controllers at Lincoln Laboratory have concurred with estimates that we have made.

The nominal TTF assumes a speed of 250 knots up to the start of the downwind leg. The distance from SCUPP to the circular marker is 8 nm (the flight path initially follows the radial to the airport). Thus the nominal flight time above has to be reduced by $\frac{8}{250} \times 3600 = 115$ seconds. Also, of the 89 seconds time gained by flying at 210 knots between SCUPP and the runway, 22 seconds is obtained outside the 30 nm marker. So a net amount of 67 seconds is gained by speed control outside the decision point. Hence we have a nominal TTF of 855 seconds and the bounds are (775, 1502).

⁶We shall present bounds for the TTF as an ordered pair (a, b) , where a is the lower bound of the TTF and b is the upper bound.

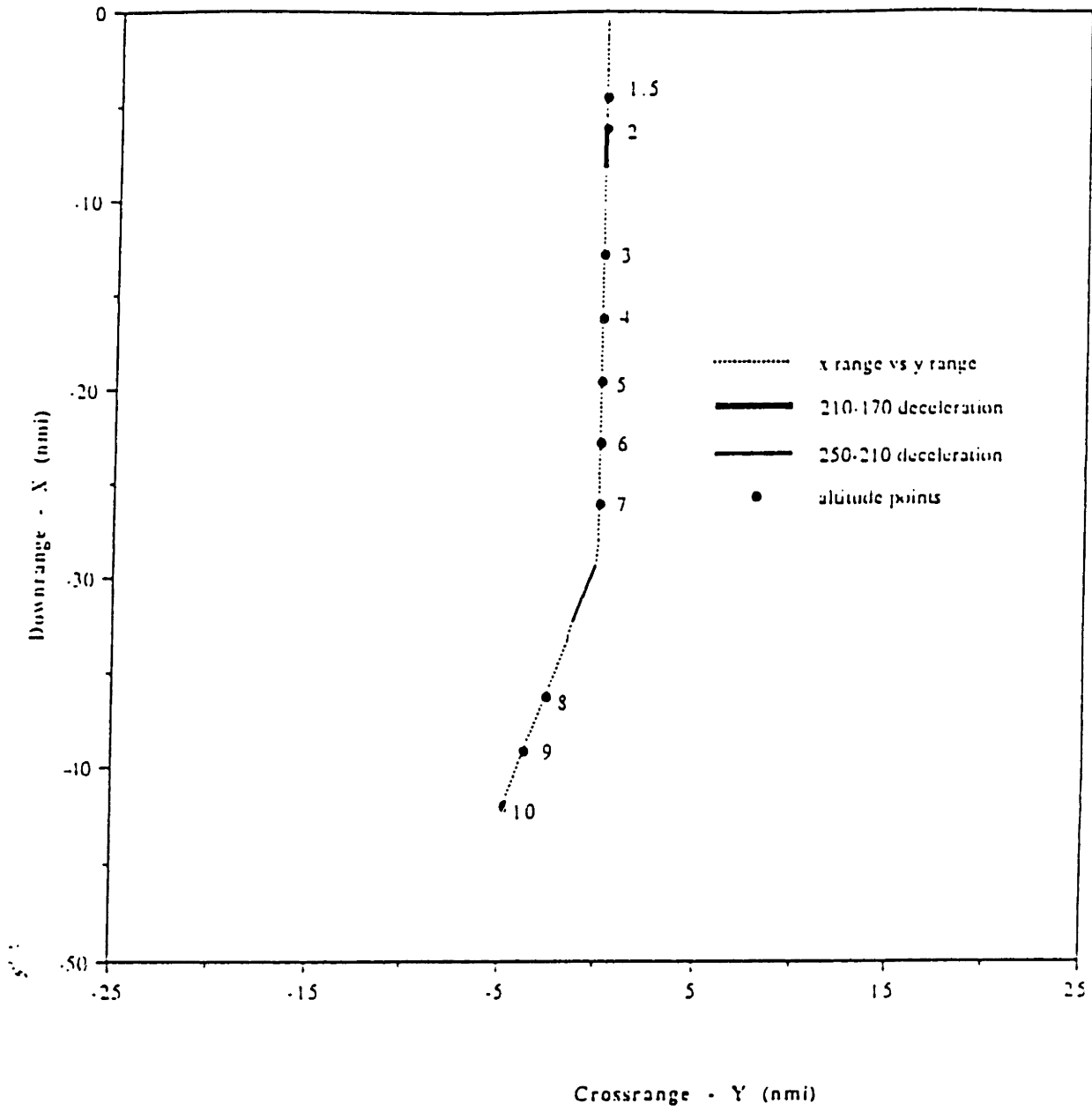


Figure C.4: Flight Path from Providence to 4R

Besides the speed and distance control parameters mentioned earlier, other timing control measures have a marginal effect. Deciding where on the downwind leg to place the 250-210 deceleration decreases TTF by at most 15 seconds. Deciding how gradually to decrease the velocity over a large segment (250-210 knots) increases TTF by at most 7 seconds and decreases it by 22 seconds. Varying the geometry for capture of the base leg yields between -14 and 8 seconds. These facts overall capture a time window of (-51,15) seconds.

It shall be convenient for us (in Appendix C) to partition the differences in the time-to-fly into that arising from speed control and that from path length variation. For arrivals from SCUPP, the path length variation can produce a difference in the nominal TTF of (-80, 580), and the speed control a difference of (-51,82). Given a nominal TTF of 855 seconds, the total TTF is thus bounded by (724,1517). Thus the flying time is roughly between 12 min. and 25 min.

The question arises: how sensitive are the TTF values to the assumption that the aircraft in question is a Boeing 727-200. Another aircraft will have a different weight, deceleration profile, and landing speed. Sorensen (1990) suggests that the weight has a marginal effect on TTF: an aircraft weighing 20,000 lb. more than the one being modeled in the nominal case would have decreased the TTF from SCUPP to 4R by only 10 seconds. Further, varying the descent and deceleration profile (not where you decelerate but how you do so given that you are going to do so) can change the TTF by 22 seconds at most. Hence we do not think that approximating other jets by the Boeing 727-200 (and later approximating high-performance props by the same) will cause great error in the TTF computed.

C.3 The TTF from Providence to 4R

In this section, we discuss the TTF bounds for the approach from PVD to 4R. The nominal approach time from PVD to 4R is 728 seconds. Figure C.4 gives the approach path. The time windows are not increased very much by speed control in this path. By flying fast

one can decrease the TTF by 20 seconds, by flying slow one can increase it by 54 seconds. So speed control gives us a differences in the nominal TTF of (-20,54).

As for arrivals from SCUPP, the other main means for controlling the TTF is by path stretching. Sorensen does not address this question directly. A controller at Lincoln has said that path stretching is not often used for arrivals from Providence. Since it is a short leg from Providence to 4R, arrivals from there are given priority over ones from BRONC and SCUPP. Nevertheless, for our purposes we are interested in the maximum and minimum possible time windows, regardless of how much they are exploited in practice.⁷ According to the controller, if need be, path stretching for flights from Providence takes the form of a detour to the north-east of at most 15 nm and back. During this maximal detour of 30 nm, the aircraft usually fly at 210 knots. Hence the total time that an aircraft can be delayed in this way is $\frac{30}{210} \times 3600 = 514$ seconds. Path stretching hence gives us differences in the nominal TTF of (0,514).

Again, we need to adjust the nominal TTF for the fact that PVD is outside the 30 nm circular marker. In fact it is about 42 nm from the runway. The extra distance of 12 nm—again the flight path for this segment from the fix to the circular marker is along a radial from the airport—is flown mostly at 250 knots, but deceleration commences about 2 nm from the marker. The difference is relatively insignificant, and we assume that the entire 12 nm is flown at 250 knots; this takes time 173 seconds. So the nominal TTF is reduced from 728 sec to 555 seconds. And the lower and upper bounds are (535, 1123), adjusting for what differences can arise from speed control and path stretching, This is roughly between 9 minutes and 19 minutes.

Now we need to approximate the nominal TTFs and their upper and lower bounds for jet traffic from BRONC as well as for prop traffic from their respective fixes.

⁷We shall look more closely at this practical consideration later.

C.4 Estimating the Other TTFs

C.4.1 The TTF for Jets from BRONC

As mentioned earlier (and as can be verified from Figure C.1), the flight path from BRONC intercepts that from SCUPP at the point where the turn to the downwind leg commences. This point is 31.5 nm and 26 nm from their respective crossings of the 30 mile marker. From the merge point on, their flight paths coincide, as do all the adjustments (speed changes and path stretching) that can be employed. The time for a SCUPP arrival to this intercept (from the 30 nm marker), is 374 seconds, at the nominal velocity of 250 knots. Likewise, for a BRONC arrival, it is 454 seconds. Thus the nominal TTF from BRONC to 4R is given by the nominal TTF from SCUPP to 4R plus the additional time from BRONC to the intercept point. This is thus equal to $970 + 454 - 374 = 1050$ seconds.

The time windows that are applicable for the flight path of a SCUPP entry are also appropriate for an entry from BRONC. The one difference is that the flight from SCUPP can be delayed by 82 seconds in the terminal area by speed control. The flight from BRONC travels 31.5 nm between entering the 30 nm marker and the intercept, compared to 26 nm for the corresponding SCUPP entry. Thus the BRONC arrival can be delayed by a further $3600 \times (31.5 - 26) \left(\frac{1}{210} - \frac{1}{250} \right) = 15$ seconds. Added to the speed control windows for SCUPP—which were (-51,82)—we have that their counterparts for BRONC are (-51,97). Path stretching can cause the same difference to the nominal path for BRONC as for SCUPP, i.e. (-80,580). Thus, including the nominal TTF of 1050 seconds, the overall TTF is for the BRONC entrant is $(1050 - 80 - 51, 1050 + 580 + 82) = (919, 1727)$, roughly between 15 and 29 minutes.

C.4.2 The TTFs for Props from the Northern Fixes: EXALT, KHRIS, LOBBY

Propeller aircraft enter the TRACON for 4R from a variety of lower altitude fixes. The entry points north of the airport are the fixes EXALT, KHRIS and LOBBY. The aircraft enter these fixes at between 5,000 and 7,000 feet (see Figure 2) and follow a similar

downwind leg pattern as the jets, starting at roughly the same point. Because they can make sharp or fast turns, however, their base legs may be a bit shorter. We do not assume so, however. The props may merge with the jet landing stream when the jets are on final approach or beginning it.

What makes computing the TTFs for props difficult is that the flight paths are not very clearly specified and have to be deduced from Figure C.2 and local navigation charts. Further, aerodynamic properties of these aircraft are different from those of jets. For instance, the amount of time saved by employing a sharp angle of turn may not be the same (the props should be able to make sharper turns). Hence we have to make certain simplifying assumptions.

We assume that the props in question are the high performance variety, which are the most common type of props landing at Boston. These include the ATR42, BE1900, DO28 etc. (often called prop-jets). These aircraft enter the 30 nm marker at about 200 knots and continue at this speed up to the point where they intercept the jet landing stream. We assume also that the time-to-fly and controllability thereof are the same as for jets from this point on. This assumption is quite a reasonable one, given the previous arguments that approximating all manner of jets by a B727 does not lead to significant errors. The reason is that the TTF has only a marginal dependence on an aircraft's descent profile and deceleration profile, as well as weight. For SCUPP, these factors make for a 30 second increase to the 970 second nominal TTF. The difference in the TTF between, for example a B727 and a B747 should be almost nonexistent in calm conditions, and a second order effect in windy conditions.⁸ There is a difference in the TTF only during descent and this is due primarily to difference in the drag-to-weight ratios for aircraft. What would be true, however, is that props should be capable of making much sharper turns onto final approach than jets. And hence the component of the time window relating to the geometry of turning to final may be the most affected by assuming that the prop performs like a B727.

Given the description of the flight path above, the TTF is composed of two parts:

⁸Jim Sturdy, Lincoln Laboratory, Personal Communication.

Entry Fix	TTF from Speed Control		TTF from Path Stretching		Overall TTF	
	Nominal	Bounds	Nominal	Bounds	Nominal	Bounds
KHRIS	540	(0,95)	596	(-65,580)	1136	(1071, 1811)
EXALT	504	(0,89)	596	(-65,580)	1100	(1035, 1769)
LOBBY	567	(0,100)	596	(-65,580)	1163	(1098, 1843)

Table C.2: Variations in TTF for Northern Prop Approaches to 4R

1. The TTF from the entry point to the commencement of the downwind leg. The nominal speed for this portion is 200 knots, but can be as slow as 170 knots.
2. The TTF from the start of the downwind leg to the runway: typically a 20 nm distance. From here on, these aircraft typically fly the nominal speed of jets (which have slowed to 200 at this point). The speed cannot be increased, but the prop, like the jet, can fly a short or a long downwind leg.

Again we see that these two segments partition the effects of speed control and path stretching respectively on the TTF. The TTF bounds for the prop are obtained from the TTF bounds for the two segments. The TTF for Segment 1 can be increased by flying at 170 knots over the distance from the entry point to the merge point; the TTF cannot be reduced, however.⁹ The TTF for Segment 2 can be increased by 580 seconds by flying a long downwind leg (see Table C.1) or decreased by 65 seconds by flying a short one. This gives overall controllability of (-65,580). Given these values, Table C.2 gives the TTF and bounds for props from the northern fixes to 4R.

C.4.3 TTFs For Props From the Southern Fixes: FREDO and WOONS

This follows the same pattern as the previous subsection. The difference is that traffic from these southern fixes merges with that from Providence at roughly 20 nm from the airport. We identify this merge point from the navigation map and from Figure C.1. The distance from the entry point to the merge point is 20 miles for a WOONS arrival and 10 for an arrival from FREDO. Because these are relatively short legs, speed control does not greatly affect the bounds for the TTF. As in the previous section, there is not much

⁹Jets can be brought in at 250 knots up to the point where they would normally have to reduce from 210 knots to 190 knots. Hence one can gain time over the (relatively) short stretch which would otherwise have been flown at a speed of 210 knots.

Entry Fix	Nominal TTF (sec)	Bounds on TTF
Providence	555	(535,1123)
BRONC	1050	(919,1727)
SCUPP	855	(724, 1517)

Table C.3: Nominal TTF and Bounds for Jet Arrivals

Entry Fix	Nominal TTF (sec)	Bounds on TTF
EXALT	1100	(1035,1769)
FREDO	720	(720, 1144)
KHRIS	1136	(1071,1811)
LOBBY	1163	(1098, 1843)
WOONS	522	(522, 914)

Table C.4: Nominal TTF and Bounds for Prop Arrivals

scope for Props flying faster than 200 knots, they can primarily only be slowed down.

As before, we can derive the nominal TTF values as well as the upper and lower bounds. The bounds from speed control derive from flying at 170 knots from the 30 nm. marker to the point where the prop traffic merges with the jet traffic from Providence. These values are (0,64) for traffic from FREDO and (0,32) for traffic from WOONS. Path stretching for traffic from the southern prop fixes is somewhat different than that for the jets from Providence. The props, instead of flying north-east, are made to fly north-west on a circular path of at most 20 nm. The props then merge with the jet traffic about 10 nm. from the threshold, as before. This path stretching can increase the TTF by at most 360 sec.¹⁰ Thus the path stretching gives changes in the TTF of (0,360).

We can, once more, get the overall bounds, by combining those from path stretching with those from speed control. The nominal TTF for arrivals from FREDO is 720 sec. and the bounds are (720,1144). Thus the window for the TTF is roughly between 12 minutes and 19 minutes. The nominal TTF for arrivals from WOONS is 522 seconds with bounds of (522, 914). Thus the window for the TTF is approximately between 9 min. and 16 min.

Hence we have outlined the method used by Sorensen to calculate the TTF nominal

¹⁰The props fly this stretched path usually at 200 knots.

values and bounds for jet aircraft from SCUPP and Providence to 4R. We have developed approximate bounds for jet entries from BRONC, as well as for prop aircraft coming in from their respective fixes. The values obtained for the latter involve making some major assumptions, but should hopefully give reasonably accurate answers. We summarize the nominal TTF and time windows for jet aircraft from their respective fixes in Table C.3. Likewise, Table C.4 summarizes such details for prop arrivals.

Appendix D

Computing Shrinking Time Windows

D.1 Introduction

We have developed two models of dynamic sequencing in the terminal area. The first, referred to as DASP-1, assumes that the upper bound of the landing time window for a given aircraft remains fixed as the aircraft transits the terminal area. The second, DASP-2, considers shrinking time windows, as we explain in Chapter 4. In Appendix B, we discussed how to compute the maximum and minimum times-to-fly (and hence maximal time windows) for aircraft entering at the various fixes. In this Appendix, we develop a method to compute shrinking time windows: the margin for expedition or delay available to a particular aircraft at a particular time in its path in the terminal area. As before, we assume that all landings are on Runway 4R.

Consider an aircraft that enters the TRACON at time 0. At any time $t > 0$, let the position of the aircraft be given by $p(t)$. We want to know $(a(t), b(t))$, the time windows of the aircraft at this time. Clearly, both $a(t)$ and $b(t)$ are functions of $p(t)$. The question is what these functions are. Of course, at the time of entry, $t = 0$, $(a(0), b(0))$ are just the original (“maximal”) time windows computed in Appendix B. Thus if the aircraft were entering at SCUPP, $(a(0), b(0)) = (724, 1517)$.

We wish to compute the time windows as they shrink with time. First, we will explain

the restrictions on manoeuvre and delay, as they have been outlined to us by air traffic controllers for Boston. Then we shall try to develop a model to approximate $(a(t), b(t))$ as the plane moves from its particular entry fix to the runway. The restrictions outlined by the controllers are as follows:

1. All sequencing decisions must be made with 20 nm left to fly on the nominal flight path. We shall refer to this point as the “decision point.” For jets from Providence, this is roughly 13 miles from entry; for the props from WOONS, it is 10 miles and for those from FREDO about 20 miles. For jets from BRONC or SCUPP and for props from the northern fixes, this decision point is toward the beginning of the nominal downwind leg. For such arrivals, the controllers stress, one may decide to send a plane on an extended downwind leg, but any such decision must be made by the decision point.
2. For jets from Providence, the path stretching manoeuvre of 30 nm is used only in emergencies. Under normal circumstances, an aircraft from Providence cannot have its path stretched by more than 180 seconds, and this decision is made within 5 miles of entry.¹ The constraints for props from the southern fixes resemble those for jets from Providence.

Trying to model these restrictions, and computing the shrinking time windows, involves a major conceptual leap over DASP-1. There, we were concerned only with the bounds on TTF: how much could an aircraft be delayed or how much could it be expedited. We obviously based these bounds on how controllers actually delay or expedite aircraft. But we did not really model how in some particular case a given amount of delay or expedition was achieved. In trying to model the restrictions above, however, we need to make some assumptions on how delays are effected. This is because we need to know if an aircraft is past the decision point or not. Further, the air traffic controllers would prefer to use speed control to path stretching as it is less disruptive, and requires less effort in merging aircraft. Hence if we wish to delay or speed up an aircraft by a certain amount, we assume that it is done by speed control first and then by path stretching. This assumption is based on a rule of thumb used by controllers. It may not be followed always, but it is a necessary simplifying assumption. We cannot, at the present level of detail of the algorithm, second guess every controller decision. As we discuss in Chapter 5, an important topic for the

¹Walter Brown and Charles Crone, personal communication.

future, if the algorithm were to undergo testing, would be to include some way to decide how particular time delays are effected.

D.2 Shrinking Windows for Aircraft from the Northern Fixes

Let us consider aircraft from the northern fixes: jets from SCUPP or BRONC, and props from LOBBY, EXALT and KHRIS. We shall focus now on what we call incremental time-to-fly differences from the nominal path. So, if for SCUPP the nominal TTF is 855 seconds, and the upper and lower bounds are (724,1517), then the incremental window is $(724 - 855, 1517 - 855) = (-131, 662)$. This incremental window gives the limits to which an aircraft can be delayed or speeded up with respect to the nominal TTF. We refer to the windows like (724, 1517) mentioned earlier, as the “absolute time windows”.

As mentioned earlier, the time windows arise from path stretching and from speed control. Thus we can partition any overall incremental time window (l_O, u_O) into portions (l_S, u_S) and (l_P, u_P) , arising from speed control and path stretching respectively. Hence we have

$$(l_O, u_O) = (l_P + l_S, u_P + u_S). \quad (\text{D.1})$$

For instance, referring to SCUPP again, we have $(l_O, u_O) = (-131, 662)$, $(l_S, u_S) = (-51, 82)$, and $(l_P, u_P) = (-80, 580)$.

The incremental windows for the northern fixes have a nice property: the portion from speed control arises before the decision point. The portion from path stretching (or shortening) arises after the decision point, and is due to a long downwind leg, or a short one. Given this special structure, we can compute the time windows $(a(t), b(t))$ for any aircraft at some time t rather easily.

First at any given time t , let us denote the incremental time windows left (overall) by $(l_O(t), u_O(t))$. This shows how much the aircraft can be expedited or delayed, with respect to the nominal path from the current point to the runway. Hence, we denote T_{nom} as the nominal landing time at entry and $T_{nom}(t)$ as the nominal landing time given the

current position of the aircraft at time t . Clearly $T_{nom} = T_{nom}(t)$, only if the aircraft has been following the nominal path from entry to time t . Otherwise, if it has been delayed or expedited in the interim, flying the nominal path from this point on would not make it land at the same time as if it had followed the nominal path from entry to runway. Since $(l_O(t), u_O(t))$ represent the incremental time windows applied to the nominal path from the current time to the runway, we have that

$$(a(t), b(t)) = (T_{nom}(t) + l_O(t), T_{nom}(t) + u_O(t)). \quad (\text{D.2})$$

If the aircraft is past the decision point at time t , then its position in the landing sequence is fixed. On the other hand, if it is before the decision point, then the entire window (l_P, u_P) is available, along with some portion of (l_S, u_S) , depending on how close to the decision point the aircraft is.

To compute the proportion of the speed control window that is left, we make another assumption. We assume that the incremental windows from speed control diminish linearly with the path length travelled. Thus if we are some proportion q in time along the path to the decision point, the incremental window left is $((1 - q) * l_S, (1 - q) * u_S)$. This assumption again is not too inaccurate, because it assumes that all the speed control is done uniformly over the distance travelled. This is indeed the case, except for the fact that some time increase or decrease may be had by the rate of at which speed is reduced from 250 knots to 210 knots.

The important problem, then, is to find out whether the aircraft is past the decision point or not. The nominal flying time for any aircraft from the decision point to the runway is 375 seconds, (see Sorensen (1990) [30]). Consider a given aircraft entering the 30 mile marker at time τ . Let it have an assigned landing time of T_{ass} . Let it have a nominal arrival time at the runway of T_{nom} . Then let Z denote the amount it is either expedited or delayed in the terminal area. So $Z = T_{ass} - T_{nom}$. Given Z , and the fact that the controller first uses speed control before short or long downwind legs, to control the landing time, we can easily compute the time T_{dp} at which an aircraft would cross the decision point. If $t > T_{dp}$, the position of the aircraft in sequence is fixed. If not, we know that we have all the path stretching window and some of the speed control windows. In

fact the absolute time windows are given by

$$(a(t), b(t)) = (T_{nom}(t) + l_P + (1 - \frac{t - \tau}{T_{dp} - \tau})l_S, T_{nom}(t) + u_P + (1 - \frac{t - \tau}{T_{dp} - \tau})u_S) \quad (D.3)$$

We illustrate now how we compute time windows for an aircraft from SCUPP. We have mentioned already that, for such an aircraft, $(l_O, u_O) = (-131, 662)$, $(l_S, u_S) = (-51, 82)$, and $(l_P, u_P) = (-80, 580)$. So, we use the following rules to compute the time T_{dp} at which the aircraft crosses the decision point:

1. The controllers want all delay or expedition to be by speed control. Thus we have that $-51 < Z < 82$. So the path from the decision point to the runway will be flown in the nominal manner. In this case, if the aircraft entered the terminal area at time τ , then the time T_{dp} at which the decision point is crossed is given by $T_{dp} = T_{ass} - 375$.
2. The controllers are going to use expedite the aircraft by using a short downwind leg, in addition to higher speeds. Then we have that $-140 \leq Z \leq -51$. Hence the nominal path from decision point to runway will be travelled in less than 375 seconds; more precisely it will be travelled $375 + Z - (-51)$ seconds. So we get $T_{dp} = T_{ass} - (375 + Z + 51)$.
3. The controllers are going to delay the aircraft by using a long downwind leg, in addition to slower speeds. Then we have that $82 \leq Z \leq 662$. Hence we have that the nominal path from decision point to runway will be travelled in $375 + Z - 82$ seconds. Thus $T_{dp} = T_{ass} - (375 + Z - 82)$.

Thus we can compute the value of T_{dp} for any aircraft from SCUPP. And then computing time windows is easy. The values of T_{dp} for the other northern fixes are computed by generalizing the above calculations. For any northern fix, we have that if $l_S < Z < u_S$, all delay or speed reduction is by speed control and $T_{dp} = \tau + T_{ass} - 375$. On the other hand, if $l_O \leq Z \leq l_S$ then $T_{dp} = T_{ass} - (375 + Z - l_S)$; or, if $u_S \leq Z \leq u_O$, $T_{dp} = T_{ass} - (375 + Z - u_S)$.

The question still remains: how do we compute $T_{nom}(t)$. We need to do this only if the aircraft is before the decision point. We know then, that the nominal path from the decision point to the runway is 375 seconds long. We assume that the time windows for speed control shrink linearly with time. Hence, this means essentially, that we are assuming that the aircraft travels at a constant speed during speed control (though the exact value of this speed can vary). Hence, at time t , the aircraft has covered the proportion $\frac{t - \tau}{T_{dp} - \tau}$

of its path from entry to the decision point. If it were to travel the rest at the nominal speed, then the time to reach the decision point would be given by the length of time that journey takes on the nominal path. This is given by the remaining proportion of the nominal time from entry to decision point:

$$\left(1 - \frac{t - \tau}{T_{dp} - \tau}\right)(T_{nom} - 375 - \tau) \quad (\text{D.4})$$

The landing time at the runway is 375 seconds later, and thus we have that

$$T_{nom}(t) = t + \left(1 - \frac{t - \tau}{T_{dp} - \tau}\right)(T_{nom} - 375 - \tau) + 375 \quad (\text{D.5})$$

.

D.3 Shrinking Windows for Aircraft from the Southern Fixes

The manoeuvrability of aircraft from the southern fixes is similar in one respect to that of aircraft from the northern fixes, and it is different in one respect. The similarity is that the speed control is done gradually up to the decision point. The difference is that the path stretching takes place at the very start of the flight in the TRACON. Again, however, we can assume that controllers prefer to expedite or to delay by speed control than by path stretching. Now what is important is not only to find out whether the decision point has been reached (so that the sequence is frozen), but also whether we are beyond the point of path stretching or not.

We mentioned earlier, that the point of path stretching is within 5 miles of entering the TRACON (or within 90 seconds)² and that one can delay an aircraft by at most 180 seconds by path stretching. Further, the amount of path stretching delay is decided by the 90 second point. We assume that the arrivals from WOONS and FREDO are similar to those from Providence in terms of how the time windows shrink. Hence, in addition to T_{dp} , we want to know when we cross the point of path stretching which we term T_{ps} . The

²This may be slightly more, if we begin speed control upon immediate entry, but we ignore the difference.

latter is easy: $T_{ps} = \tau + 90$. For Providence, there is some minor amount to be gained at flying a higher speed (250 knots instead of 210 knots), but for jet aircraft, which enter and fly into the TRACON at 200 knots, no such opportunities exist. Further, there are no downwind legs which can be shortened to expedite aircraft: path variation can only be used to produce delays.

If $t < T_{ps}$, then the entire path stretching time window is available and some (linearly shrinking) portion of the speed control window. If $T_{ps} \leq t \leq T_{dp}$ then only the (linearly shrinking) portion of the time window is available. Again, we have that the nominal time from the decision point to the base is 375 seconds. Since we can have no delay or expedition after the decision point, we have that $T_{dp} = T_{ass} - 375$. Finally, we can compute the linearly shrinking time windows as we did before.

If $t < T_{ps}$,

$$(a(t), b(t)) = (T_{nom}(t) + l_P + (1 - \frac{t - \tau}{T_{dp} - \tau})l_S, T_{nom}(t) + u_P + (1 - \frac{t - \tau}{T_{dp} - \tau})u_S) \quad (D.6)$$

where $T_{nom}(t)$ is given by (D.5).

Now we assume that $T_{ps} \leq t \leq T_{dp}$. Thus the path stretching time windows are no longer applicable, and the proportion of speed control left depends on the length of time travelled not including path stretching. First we compute $Z = T_{ass} - T_{nom}$. If $l_S \leq Z \leq u_S$, then no path stretching is used and the incremental windows are merely a function of the proportion of the path travelled:

$$(a(t), b(t)) = (T_{nom}(t) + (1 - \frac{t - \tau}{T_{dp} - \tau})l_S, T_{nom}(t) + (1 - \frac{t - \tau}{T_{dp} - \tau})u_S) \quad (D.7)$$

Again $T_{nom}(t)$ is given by (D.5).

On the other hand, if $Z > u_S$, there is path stretching. To be precise, we spend time $Z - u_S$ on the extended path. So if $t - T_{ps} \leq Z - u_S$, then we are still on the extended path, and the time window left is the portion after path stretching:

$$(a(t), b(t)) = (T_{nom}(t) + (1 - \frac{T_{ps} - \tau}{T_{dp} - \tau})l_S, T_{nom}(t) + (1 - \frac{T_{ps} - \tau}{T_{dp} - \tau})u_S) \quad (D.8)$$

where

$$T_{nom}(t) = T_{ps} + (Z - u_S) + (1 - \frac{T_{ps} - \tau}{T_{dp} - \tau})(T_{nom} - 375 - \tau) + 375 \quad (D.9)$$

Lastly, if $t - T_{ps} > Z - u_S$, we are beyond the path stretching and, adjusting for this, the time windows left are

$$(a(t), b(t)) = (T_{nom}(t) + (1 - \frac{t - \tau - (Z - u_S)}{T_{dp} - \tau - (Z - u_S)})l_S, T_{nom}(t) + (1 - \frac{t - \tau - (Z - u_S)}{T_{dp} - \tau - (Z - u_S)})u_S) \quad (D.10)$$

Again, we have to adjust for the time spent on the stretched path, and

$$T_{nom}(t) = t + (1 - \frac{t - \tau - (Z - u_S)}{T_{dp} - \tau - (Z - u_S)})(T_{nom} - 375 - \tau) + 375 \quad (D.11)$$

One final point to mention is that fixing the position in sequence of one aircraft also fixes the position of all those that would land before it. This is an important for the following reason. Consider an aircraft from SCUPP that is just passed the decision point, but is being sent on an extended downwind leg. Thus it may still have 30 miles left to fly to the runway. Hence it may be sequenced to land after an aircraft from Providence that is still 25 miles left to fly from the airport, and before its own decision point. But the position in final sequence of the Providence arrival is now fixed *de facto*, because the position of its successor is.

D.4 Modifying the “Nearest-Neighbor” Technique

There is one other change to the algorithm that is necessitated by the shrinking time windows constraints. This is to alter the nearest neighbor technique by which we chose the best aircraft to land from each weight-class category at any given time. The problem is that the aircraft with the smallest preferred landing time is not necessarily the best one to sequence. For instance, it may be the case that one aircraft, from SCUPP, has a smaller lower preferred landing time than another from Providence. It may also happen, however, that the latter has a very tight upper bound and has to be landed before the former. Hence, the algorithm now has to identify such cases, and revise the choice of “best” aircraft from each category. We do so as follows.

Consider the weight-class category L , with aircraft S_L , which have time windows S_L^W . Initially, we had ordered all the aircraft by increasing order of lower bounds. Suppose

that we are considering sequencing L_1 , the first aircraft in L , at time T_{L_1} . Let the second aircraft L_2 have time windows (a_{L_2}, b_{L_2}) . Let us define LTI between two aircraft from L as $t_{L,L}$. If $T_{L_1} + t_{L,L} > b_{L_2}$, we clearly have an infeasible sequence, and the upper bound constraint of aircraft L_2 is violated. In such circumstances, we reorder the aircraft L_1 and L_2 , making the latter aircraft the “best” from this category to sequence. We would never have to search beyond one adjacent aircraft, because there is enough spacing between aircraft entering the terminal area to guarantee that any one aircraft’s landing would violate the upper bounds of two other aircraft which have higher lower bounds for landing.

Bibliography

- [1] Abundo, S., "An Approach for Estimating Delays at a Busy Airport", Unpublished S.M. Thesis, Operations Research Center, M.I.T., June, 1990
- [2] Andreatta, G. and G. Romanin-Jacur, "Aircraft Flow Management under Congestion", *Transportation Science*, Volume 21, No. 4.
- [3] Andrews, J. and J. Welch, "The Challenge of Terminal Air Traffic Control Automation", Proceedings of the 34th Annual Air Traffic Control Association Conference, Boston, Fall, 1989.
- [4] Blackman, S., *Multiple-target Tracking with Radar Applications*, Artech House, Dedham, MA., 1986.
- [5] Chambers, J., W. Cleveland, B. Kleiner and P. Tukey, *Graphical Methods for Data Analysis*, Duxbury Press, 1983.
- [6] "Construction to Start in 1989 on New Denver Super-hub", *Airport Forum*, December, 1988.
- [7] Crone, C. and M. Moore, "Modes of Operation for Runway Configuration 4RR/L and 9 at Logan International Airport", Memorandum, Lincoln Laboratory, Lexington, MA., November, 1989.
- [8] Dear, R., "The Dynamic Scheduling of Aircraft in the Near Terminal Area", Technical Report R76-9, Flight Transportation Laboratory, M.I.T., September, 1976
- [9] Drake, A.W., *Fundamentals of Applied Probability Theory*, McGraw-Hill Book Company, 1967.
- [10] Efron, B., "Bootstrap Methods: another look at the jackknife", *The Annals of Statistics*, Vol. 7, No. 1, January, 1979.
- [11] "Enroute Metering to Boston-Logan Airport", Order ZBW 7210-482A, Federal Aviation Administration, U.S. Department of Transportation, Boston, February, 1988.
- [12] *FAA Aviation Forecasts Fiscal Years 1988-1999*, Report FAA-APO 88-1, Federal Aviation Administration,

- [13] *Handbook of Mathematical Tables for Probability and Statistics*, 2nd Edition, The Chemical Rubber Company, Inc., 1968.
- [14] Hettmansperger, T.P., *Statistical Inference Based on Ranks*, John Wiley & Sons, 1984.
- [15] Homburg, M., *Statistical Analysis for Decision Making*, 3rd Edition, Harcourt Brace Jovanovich, 1983.
- [16] Larsen, R.J., and Marx, M.L., *An Introduction to Mathematical Statistics and its Applications*, Prentice-Hall, 1981.
- [17] Lehmann, E., *Nonparametric Statistics*, Holden-Day Inc., 1978.
- [18] Magill, S.A.N., A.C.E. Tyler, E.T. Williamson and W. Finch, *Computer-Based Tools for Assisting Air Traffic Controllers with Arrivals Flow Management*, Report 88001, Royal Signals and Radar Establishment, HMSO, London, 1988. February, 1988.
- [19] Miller, R.G., *Beyond ANOVA, Basics of Applied Statistics*, John Wiley & Sons, 1986.
- [20] Moore, M., "Description of ATC and Facilities at Boston TRACON", Project Memorandum 42PM-TATCA-0004, Lincoln Laboratory, Lexington, MA., 1989.
- [21] Mundra, A., *A Description of Air Traffic Control in the Current Terminal Airspace Environment*, MITRE Report MTR-88W 00167, McLean, VA., 1989.
- [22] Newell, G., "Airport Capacity and Delays", *Transportation Science*, Vol. 13, No. 3, August, 1979.
- [23] Odoni, A., "The Flow Management Problem in Air Traffic Control", in *Flow Control of Congested Networks*, Editors: A. Odoni, L. Bianco and G. Szego, Springer-Verlag, Berlin, 1987.
- [24] Pratt, J., and J. Gibbons, *Concepts of Nonparametric Theory*, Springer-Verlag, New York, 1981.
- [25] Psaraftis, H., "A Dynamic Programming Approach to the Aircraft Sequencing Problem", Technical Report R78-4, Flight Transportation Laboratory, M.I.T., October, 1978.
- [26] Psaraftis, H., "A Dynamic Programming Approach for Sequencing Groups of Identical Jobs", *Operations Research*, Volume 28, Number 6, 1980, pp. 1347-1359.
- [27] Randles, R.H., and Wolfe, D.A., *Introduction to The Theory of Nonparametric Statistics*, John Wiley & Sons, 1979.
- [28] *Research Plan for Terminal ATC Automation*, Memorandum, Lincoln Laboratory, Lexington, MA., 1988.

- [29] Simpson, R., and A.R. Odoni and F. Salas-Roche, "Potential Impacts of Advanced Technologies on the ATC Capacity of High-Density Terminal Areas", Technical Report, Flight Transportation Laboratory, M.I.T., 1986.
- [30] Sorensen, J.A., "Flight Timing Control Sensitivity Analysis for TRACON Approaches to Boston Runway 4R", Technical Memorandum 90112-01, Seagull Technology Inc., March, 1990.
- [31] Sturdy, J., Andrews, J., and Welch, J., "Aircraft Trajectory Prediction for Terminal Automation", AIAA Guidance, Navigation and Control Conference, Boston, 1989
- [32] *TATCA Concept and Development Plans*, Memorandum, Lincoln Laboratory, Lexington, MA., 1988.
- [33] Terrab, M., "Ground Holding Strategies for Air Traffic Control", Technical Report 196, Operations Research Center, M.I.T., February, 1990.
- [34] Thompson, S.D., "Aircraft Landing Sequence Optimization Under Constraints", Project Memorandum 41PM-TATCA-0001, Lincoln Laboratory, Lexington, MA., 1990.
- [35] Trivizas, D., "Parallel Parametric Combinatorial Search—Application to Runway Scheduling", Technical Report 87-4, Flight Transportation Laboratory, M.I.T., February, 1987.
- [36] United States Government Flight Information Publication Series, "Enroute Low Altitude—U.S. Chart L-26", National Ocean Service, 1989.